Three Essays in Experimental and Behavioral Economics

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Vasudha Chopra

December 2022

This work is dedicated to my maternal uncle, the one who was taken away before his time. The one whom I have never met yet feel extremely close to. The one who I hope is proud of my accomplishments - Captain Dr. Ajay K. Mahindru

Acknowledgements

I would like to thank my academic advisor, Dr. Christian A. Vossler, without whose constant support this work would not have been possible. I am beyond grateful to have an exceptionally brilliant, supportive, extremely critical and massively detail-oriented person overlooking my research. I have seen a huge change in the the way I approach research since the time I joined the Ph.D. program, and I know it is for the better. His guidance carried me through various ups and downs in my dissertation writing process for which I shall remain indebted. I want to thank him for all his efforts, but most of all for his time commitment towards my research. I am proud to be his advisee and I hope to make him proud with my future accomplishments, each and everyone one of which he shall always remain a part of. As my teacher and mentor, he taught me more than I could ever give him credit for. He has shown me, by example, what a great researcher looks like.

I consider myself very lucky to have some of the brilliant minds on my committee - Dr. Scott Gilpatric, Dr. J. Scott Holladay and Dr. Philip Brookins. I would like to thank each of them for their insightful comments and suggestions that greatly improved the quality of my research. Every conversation with them has been productive and has helped me find my way through unexpected hurdles. I would also like to thank my entire committee not just for being supportive and understanding throughout the journey but also letting my defense be both challenging and enjoyable at the same time.

The entire faculty and staff at the Department of Economics deserves a special mention not just for the academic and logistical support they provided but also for being a large part of my personal growth during the past five years. Dr. Jean Gauger has not only been my teacher and mentor but also a wonderful friend with whom I have shared some amazing times. It is rightly said that those who love you stick by you through all your ups and downs. When I say this I think of three people in particular - Sherri Pinkston, Erin Turman and Beth Weissmueller. Thank you for everything you have done and continue doing for students. It keeps students sane, to say the least.

Last but definitely not the least, I would like to thank my family without whose constant love and support I would not be here today. They continue to inspire me. My mother has always been the one to push me to pursue my dreams and my father has been the source of all of my strength. My younger brother always tends to say the right thing at the right time and fiercely protects as well as guides me. I could not have imagined living so far apart from the three of them but they never let me feel alone. My gratitude also goes to my loving friends back home and in Knoxville who made this entire journey worthwhile. A Ph.D. in itself is one of the most challenging tasks I have ever attempted to accomplish but moving to an entirely different country and finding such love and support, practically a home far away from home, has been the most wonderful surprise. From long-distance friendships spanning multiple time-zones to having found lovely friends in Knoxville, I consider myself to be extremely lucky. This work is a testament to the support and encouragement received from Aanchal Singhania, Ayushi Choudhary, Tanvi Khurana, Srishti Goyal, Abhinav Srivastava, Disha Mendiratta, Divya Pahuja, Sonal Dhawan, Bokseong Jeong, Himadri Palikhe, Hieu Nguyen, Richard Beem, Ge Wu, Ewa Zawojska, Indrani Singh, Beant Kapoor, Swati Mishra, Kirtan Davda, Esha Dutta, Keertana Tallapragada, Debshikha Banerjee, Krutika Desai, Hemant Bonde, Ajay Dwivedi, Prabuddha Prakash, Ruhani Sagar, Soumendu Sarkar, Kaushik Roy Chowdhury and many others.

"Success is the sum of small efforts, repeated day in and day out"

– Robert Collier

Abstract

This dissertation presents three essays that use experimental economics methods. The first essay examines how behavior in inter-group contests is altered when players have incomplete information on their opponent. The game is a Tullock contest with heterogeneous groups (differences in cost-of-effort, prize value, and group-size), and players only know the probability their opponent is a particular type. For cost and value treatments, incomplete information increases effort in uneven contests but has no effect in even contests. Group-level effort is higher in group-size treatments, but incomplete information does not systematically alter effort. Overall, group-level effort is much higher than what standard theory predicts; however, extending the theory to consider behavioral motives (altruism, utility of winning) helps to reconcile theory and data.

The second essay examines the effects of using hypothetical bias reduction procedures in stated preference surveys designed to elicit demand for potential public policies. These procedures are commonly used due to a concern that people treat surveys as hypothetical choice settings, leading to elicitation bias. However, accumulated evidence indicates that most respondents perceive that their decisions are consequential, which questions the use of these procedures. I test popular bias reduction procedures in consequential settings: cheap-talk, solemn oath, and certainty adjustment. The oath increases willingness to pay (WTP), whereas certainty adjustment leads to large decreases in WTP estimates. Cheap-talk does not alter

mean WTP. These results have important implications for practitioners, and provide a new vantage point from which to evaluate the appropriateness of these procedures.

The third essay examines the effects of using non-binding, exogenous team goals on worker effort in a weakest-link team production game. The experimental design varies the team goal (and whether a goal is present) and task complexity level (simple or complex), lending itself to identify a causal effect of complexity on goal effectiveness. Preliminary results indicate that easy goals may have a detrimental effect while difficult goals may increase team production. Interestingly, as complexity increases, while physical effort may decrease, cognitive effort increases. Further, the effects of varying goal-difficulty on production may not be monotonic. Study outcomes are expected to have important managerial implications.

Table of Contents

List of Tables

List of Figures

1

Who are we up against? Heterogeneous group contests with incomplete information

Abstract

This study examines how behavior in inter-group contests is altered when players have incomplete information on their opponent. We model a Tullock contest where there are two possible types of groups that are heterogeneous in the incentives they face, and players only know the probability their opponent is a particular group type. In the theory and complementary experiment, we compare three sources of heterogeneity – differences in cost-of-effort, prize value, and group size. For the cost and value treatments, we find that incomplete information increases effort in uneven contests but has no effect on average in even contests. Group-level effort is higher in group size treatments, but incomplete information does not systematically alter effort. Overall, group-level effort is much higher than standard theory predicts, and we consider behavioral theories to help reconcile theory and experiment. Based on observed patterns in the data, an extended model with in-group altruism (and possibly a non-monetary utility of winning) is supported. Moreover, bounded rationality, free-riding and time trends may be important in explaining some of the effects of incomplete information.

Keywords: inter-group competition; heterogeneous contests; Tullock contests; incomplete information; public goods; group size paradox; experiments; utility of winning; in-group altruism

1.1 Introduction

In many settings, people are engaged in inter-group competitions where opponents may be at an advantage or disadvantage. Advantages in these “uneven” contests may arise from various sources such as talent differentials across groups that can be thought of as decreasing the relative (effort) cost associated with attaining a level of output. R&D firms or legal teams may face opponents that have fewer researchers or employees, thus leading to a potential advantage due to the increased opportunity to put forth productive effort. Some public or private organizations have a larger resource base or otherwise can better incentivize desirable actions from their members through bonuses, thus increasing the marginal returns from effort. While competing teams are rarely on equal footing in all important dimensions, also endemic to group contests is that the agents in one group have uncertainty over the extent of their disadvantage or advantage over the competing team. An academic research lab applying for a grant may know who their main competition is but is unlikely to precisely know another lab’s talent allocation, number of researchers (which can decrease the time to complete the project), or the motivation (such as from salary raises tied to external funding) of another lab to obtain the grant. This study uses theory and experiments to examine how behavior in potentially uneven group contests is altered when players have incomplete information on the incentives facing their opponent.

We study a Tullock rent-seeking contest between two groups that are potentially heterogeneous in the terms of the incentives they face, and while players know the type of their own group, they only know the probability that the opponent is of a particular group type. Players within a group are identical, group-level effort is an additive function of individual efforts (perfect substitutes), individual efforts are not directly observable to teammates, and the winning team receives a prize that is evenly split. We consider three potential sources of advantage: cost-of-effort, prize value, and group size. The experimental design varies as treatments whether teams have complete or incomplete information on their opponent's type, along with the potential source of advantage.

In doing so, we make three contributions to the literature. This is the first experiment to study the effect of incomplete information in a heterogeneous inter-group contest. Second, we test whether the source of the potential advantage matters. While these sources of advantage have been studied in group contest settings to a limited degree, prior work has only examined them in isolation. Third, the experimental design and theoretical framework allow us to provide new insight on possible behavioral motives in inter-group contests. For instance, in contrast to the case where teams differ with respect to the prize value or cost-of-effort, the standard theory model (i.e., each person maximizes their own expected payoff) predicts that being in a relatively large group is not an advantage in the sense that it does not alter group-level effort or the probability of winning the contest. This and other theoretical predictions that differ across contest types help us to identify underlying behavioral motives for expending costly effort in group contests.

Our theoretical framework builds on Tullock's canonical model Tullock (1980), Katz et al. (1990) who model a group Tullock contest setting with inter-group heterogeneity, and Malueg and Yates (2004) who study individual contests with incomplete information over the prize value. To our knowledge, the only prior theoretical treatment of group contests with incomplete information is Eliaz and Wu (2018), who study uneven all-pay contests where two teams may differ in size and have incomplete information on the value of the other team's prize.

A stylized fact in the group contest literature is that effort far exceeds what is predicted by the self-interest model (e.g., Cason et al. (2012); Leibbrandt and Sääksvuori (2012)). Moreover, the literature highlights that there may be one or more behavioral motivations for overbidding and perhaps they might even be inter-dependent (Sheremeta (2010); Mago et al. (2016)). Motivated by this evidence, we consider three possible behavioral motives in the theory: nonmonetary utility of winning, in-group altruism, and out-group hostility. The second and third motives may be attributed to parochial altruism or social identity. While Sheremeta (2018) highlights these possible motivations in his review of the group contest literature, evidence in support of these motivations comes from contests between individuals and non-contest experiments.[1]

While experimentally examining incomplete information in a group contest setting is novel, we note that prior studies have used experiments to examine the effects of uncertainty in lottery contests and all-pay auctions involving competition between individuals (see Dechenaux et al. (2015)). In most of these studies, players are ex ante symmetric, but a parameter value (e.g., corresponding to per-unit effort cost) for an individual is determined by taking an independent draw from a common uniform distribution. For example, Brookins and Ryvkin (2014), for a contest among four players, find theoretically and experimentally that incomplete information increases effort among advantaged players, and decreases effort among disadvantaged players. Also relevant is Boosey et al. (2017), who introduce uncertainty in the number of players in a contest by varying the (independent) probability a player enters the contest, along with the maximum number of possible participants. They find that when the participation probability is low, an increase in the maximum possible number of competitors increases effort while the opposite is true when the participation probability is high. Fallucchi et al. (2013) analyze the impact of information feedback on individual effort in an ex-ante homogeneous Tullock contest. Although their contest structure differs from ours, in general they find information feedback about other members in the team (i.e., their opponents) raises individual and group expenditures.

In his survey of the group contest literature, Sheremeta (2018) notes that few studies incorporate heterogeneity between groups. As exceptions, Heap et al. (2015) test the effects of providing teams with unequal endowments, which allows

[1] Sheremeta (2018) also discusses relative payoff maximization and cognitive limitations as possible explanations for over-expenditure of effort. Relative payoff maximization is captured by a model of parochial altruism in this group contest setting. While we do not formally model bounded rationality, we consider the effects of cognitive ability through a proxy measure in the data analysis.

the advantaged team to contribute more towards winning the competition, and Bhattacharya (2016) examines contests between groups that differ in either the probability of winning (when groups expend equal effort) or their effort cost. These investigations, along with related ones involving symmetric groups with intra-group heterogeneity, generally find that advantaged players contribute relatively more effort.[2]

A few experiments examine heterogeneity in group contests in the form of different group sizes. Rapoport and Bornstein (1989) and Kugler et al. (2010) examine contests between three- and five-player groups and find that even in cases where theory predicts the smaller team should expend more collective effort, larger groups are instead more likely to win. One distinction in our design is that, similar to Abbink et al. (2010) and Ahn et al. (2011), who study contests between an individual and a group of four, we fix the value of the winning prize for an individual. In doing so, the difference in group size is the only potential source of advantage across competing groups.

As predicted, incomplete information does increase effort in uneven contests with either cost or value heterogeneity. But the effect of incomplete information in even contests is a null effect, on average, regardless of the potential source of heterogeneity. Consistent with standard theory, we find that group-level (and contest-level) effort is higher for advantaged groups, when the source of the advantage is either a lower cost-of-effort or a higher prize value. Nevertheless, in contrast to theory which predicts a null effect, effort increases with group size. In fact, effort levels are highest in our group size treatments. The high levels of effort we observe in all contests, along with patterns in the data, suggest that behavioral motives have a potentially important role. A theory that includes non-monetary utility of winning and in-group altruism helps to reconcile data with theory. Besides, we also find some support that deviations from theory are driven by boundedly rational participants.

[2]The same result is evident from experiments involving heterogeneous contests between individuals (see Dechenaux et al. (2015)).

1.2 Theory

We first develop models of a heterogenous group Tullock contest under complete and incomplete information, while assuming that players are solely driven by self-interest. These models build on Tullock's canonical model of a rent-seeking contest Tullock (1980), Katz et al. (1990) who model a group Tullock contest setting with inter-group heterogeneity, Malueg and Yates (2004) who study individual contests with incomplete information over the prize value. As overbidding (i.e., excessive effort expenditures) relative to the predictions of standard models is ubiquitous in the group contest literature, we then consider models that incorporate behavioral motives hypothesized to explain overbidding.

Consider a contest between two groups. Group g consists of N_g risk-neutral players, and groups compete to win a prize. Regardless of their individual actions, the value of the prize to each player on the winning team is v_g, and there is no prize for the losing team. This may, for instance, characterize a setting where the group prize is a (local) public good that is non-excludable and non-rival in consumption. All players in both groups simultaneously and independently expend effort x_{ig} at a (constant) per-unit cost of c_g. Player efforts within a group are perfect-substitutes, such that group-level effort, X_g, is simply the sum of player efforts; i.e., $X_g = \sum_{i=1}^{N_g} x_{ig}$. The probability of winning, p_g, depends on the relative effort of the competing teams. In particular, we use the contest success function (CSF) of Tullock (1980) for the standard lottery case:

$$p_g = \frac{X_g}{X_g + X_{-g}} \tag{1}$$

When both teams expend the same collective effort, each has a win probability of $\frac{1}{2}$. Otherwise, with this CSF, the team that exerts more effort has a higher probability of winning. Throughout the analysis we assume all players within a group are identical with respect to cost and value parameters, but there may be heterogeneity in either the cost, value, or group size across competing groups. We limit the analysis here to settings where there is at most one source of heterogeneity.[3] We consider two information conditions. In the complete information condition, each player has perfect knowledge of the incentives (i.e., parameters) facing their team as well as their opponent. In the incomplete information condition, players do not know with certainty one of the parameters that the opposing team faces.

1.2.1 Complete Information

When all players know the parameters (cost, value, and group size) characterizing their own group as well as their opponent, the expected payoff of player i in group g is:

$$\pi_{ig} = p_g v_g - c_g x_{ig} = \frac{X_g}{X_g + X_{-g}} v_g - c_g x_{ig} \tag{2}$$

A self-interested, risk-neutral individual maximizes [2] by choosing effort x_{ig}, yielding the first-order condition:

$$\frac{X_g}{(X_g + X_{-g})^2} v_g - c_g = 0 \tag{3}$$

The maximization problem for a representative player from team $-g$ is of course symmetric, giving rise to the first-order condition:

$$\frac{X_g}{(X_g + X_{-g})^2} v_{-g} - c_{-g} = 0 \tag{4}$$

The first-order conditions include only group-level efforts, and the theory is silent about individual effort. Assuming an interior solution, the Nash equilibrium is attained by solving [3] and [4] simultaneously for X_g and X_{-g}. The equilibrium is:

$$X_g^* = \frac{c_{-g} v_g^2 v_{-g}}{(c_g v_{-g} + c_{-g} v_g)^2}; X_{-g}^* = \frac{c_g v_{-g}^2 v_g}{(c_{-g} v_g + c_g v_{-g})^2} \tag{5}$$

[3]The theory logistically extends to settings where a team has an advantage in multiple dimensions. Ambiguity in the various comparisons of course arises if a group holds an advantage in one dimension, but a disadvantage in another.

In terms of group-level efforts, this equilibrium is unique. In terms of individual effort, there is one symmetric equilibrium and multiple asymmetric equilibria in which the sum of the individual efforts equals the group-level equilibrium Baik (1993). This is the result of assuming constant marginal effort costs and the same prize value to each team member. Importantly, the equilibrium is not a function of group size, and the theory therefore predicts two groups that differ in size will exert identical group-level effort.

Using [5] we can obtain predictions for homogeneous and heterogeneous contests. To facilitate this, let $g(A, D)$, where A denotes that a group is of the type "advantaged" and D denotes the group type "disadvantaged". These labels are meant to identify group type and to be clear, when a group is "advantaged" (or "disadvantaged") this does not necessarily mean that this group has an advantage (or disadvantage) relative to their opponent. Contests where both groups are of the same type – either both advantaged or both disadvantaged – are referred to as "even" contests; otherwise, contests between types are "uneven". In contests with potential cost heterogeneity, the advantaged group is one where the per-unit cost of effort is lower, such that $c_A < c_D$. In a similar vein, advantaged and disadvantaged groups in the context of value and group size heterogeneity are defined by $v_A > v_D$ and $N_A > N_D$, respectively.

Equilibria for even and uneven contests under complete information are provided in Table 1.1.[4] For clarity, in the table we drop the subscripts on parameters held fixed across groups within a comparison set. When both teams are advantaged based on cost or prize value, group effort is strictly higher when compared to an even contest between disadvantaged teams. Moreover, in an uneven contest, the advantaged team exerts relatively more effort than their opponent. Interestingly, the effort of a disadvantaged team is lower when in an uneven contest relative to an even contest. This can be labelled a discouragement effect as the fact that they are playing against an advantaged opponent lowers the chance they win, which serves

[4]All tables and figures for this chapter and the following chapters are located in the Appendix.

to disincentivize effort. Following the previous discussion, in the special case where players are only self-interested, group size differences do not matter, i.e., $X_A^* = X_D^*$ for both even and uneven contests.

1.2.2 Incomplete Information

In the incomplete information setting, players do not know with certainty the parameters that the opposing team faces. Instead, they know the opponent will either be of the advantaged or disadvantaged type with some probability. Let $0 < r < 1$ denote the probability that any contest opponent is advantaged, which is common knowledge. A player on team g then forms an expectation based on the probability r the opponent is advantaged and the probability $1-r$ that the opponent is disadvantaged. The maximization problem is:

$$\max_{x_{ig}} U_{ig} = \{r(p_g|A) + (1-r)(p_g|D)\}v_g - c_g x_{ig} = \left(r\frac{X_g}{X_g + X_A} + (1-r)\frac{X_g}{X_g + X_D}\right)v_g - c_g x_{ig} \tag{6}$$

The associated first-order condition is:

$$\left(r\frac{X_A}{(X_g + X_A)^2} + (1-r)\frac{X_D}{(X_g + X_D)^2}\right)v_g - c_g = 0 \tag{7}$$

This gives rise to two equations based on whether $g = A$ or $g = D$. This in turn leads to a system of two equations and two unknowns. In the general case where $0 < r < 1$, the solution is intricate as the problem can only be reduced to a cubic equation. In the special case of $r = 1/2$, which coincides with the experiment, the algebra is simpler. In this case, and focusing on the case of cost heterogeneity, the (pure strategy) symmetric Bayesian-Nash equilibrium is

$$X_A^* = \left(\frac{v_A}{c_A}\frac{4\frac{c_A v_D}{c_D v_A} + (1 + \frac{c_A v_D}{c_D v_A})^2}{8(1 + \frac{c_A v_D}{c_D v_A})^2}\right); X_D^* = \left(\frac{v_D}{c_D}\frac{4\frac{c_A v_D}{c_D v_A} + (1 + \frac{c_A v_D}{c_D v_A})^2}{8(1 + \frac{c_A v_D}{c_D v_A})^2}\right) \tag{8}$$

As in the complete information game, group-level effort is independent of group size. Table 1.2 presents the equilibria for the incomplete information condition for $r = 1/2$, and as before, we focus on cases where there is inter-group heterogeneity across types with respect to effort cost, prize value, or group size.[5] Given the expectation about opponent's strategies, individuals in either an advantaged or disadvantaged group have unique effort levels that do not depend on the type of group they are actually competing against, which is of course unknown. With $r = 1/2$, effort in the incomplete information setting is the average effort (for the same group type) across even and uneven contests with complete information.

As in the complete information case, for both cost and value heterogeneity, the advantaged team is predicted to exert more effort relative to the disadvantaged team which, in turn, increases their probability of winning an uneven contest. With group size heterogeneity, equilibrium effort is $v/4c$ regardless of group size or r.

Next, we summarize the effects of incomplete information on effort through three propositions. In the appendix, we provide proofs for the $r = 1/2$ case. We further use numerical calculations to illustrate that the first two propositions hold for any $0 < r < 1$, and to illustrate how expected contest-level effort varies with r.

Proposition 1: When players are solely self-interested, for group contests with cost or prize value heterogeneity, incomplete information increases contest-level effort in ex post uneven contests. For contests with group size heterogeneity, incomplete information has no effect.

Proposition 2: When players are solely self-interested, for group contests with cost or prize value heterogeneity, incomplete information decreases contest-level effort in ex post even contests. For contests with group size heterogeneity, incomplete information has no effect.

[5] For the standard model, the group contest equilibria are identical to those based instead on a contest among individuals. It follows that similar results can be found in Malueg and Yates (2004) and Fey (2008) for the $r = 1/2$ case.

Proposition 3: When players are solely self-interested, for group contests with cost or prize value or group size heterogeneity, as compared to the benchmark complete information condition, the average contest-level effort is equal under both information conditions if $r = 1/2$.

Theory predicts that incomplete information alters effort expenditures in contests where groups differ according to effort cost or prize value. For uneven contests, an advantaged team increases effort under incomplete information. This is because the team does not know for sure that the other team is disadvantaged and increases effort accordingly. The effect of incomplete information is increasing in r and the extent of the advantage. For the special case of $r = 1/2$, the effort from a disadvantaged team also increases with incomplete information. As a result, group-level (i.e., aggregate effort of each team) and contest-level effort (i.e., the sum of effort across the disadvantaged and advantaged team) is strictly higher under incomplete information when compared to complete information.

For even contests, that incomplete information decreases group-level effort is intuitive. Note first that both teams exert equal effort in this case, and effort is higher when competing teams are advantaged rather than disadvantaged. With incomplete information, a team does not know the opposing team's type. Either team suspects that they are playing against an advantaged opponent with a probability less than 1, and so this lowers effort relative to the case where they know for sure the opponent type. This is due to a discouragement effect (Fonseca (2009), Kimbrough et al. (2014)). Put simply, the disadvantaged team strategically lower effort when competing with an advantaged opponent. When considering the average contest-level effort, unconditional on contest type, the differential effects of incomplete information across uneven and even contests of course will counteract. When $r = 1/2$, the effects completely offset.

As the standard model predicts that group size does not affect group-level effort, it follows that incomplete information has no effect. While Proposition 3 technically holds for heterogeneous group size contests, it of course does so trivially. In the section below, we consider extended models that incorporate behavioral motives.

1.2.3 Behavioral models

The literature hypothesizes that behavioral motivations are important in group contests, given the systematic experimental evidence that effort expenditures far exceed what is predicted by the standard self-interest model. Here, we consider as possible motives a non-monetary utility of winning, altruism towards one's group members, and hostility towards members of the competing group.[6] Here, we discuss these motives separately, and focus on their implications for comparisons between complete and incomplete information contests.

Non-monetary utility of winning

To allow for the possibility that players may have a 'joy of winning', we can specify that the "prize" associated with winning the contest has both monetary and subjective, non-monetary components. Denote this "overall" prize value as $w(v_g)$, and we assume that $w(v_g) > v_g$ and is (weakly) increasing in v_g. For notational convenience let $p_g(r)$ denote the probability of winning, conditional on the chance the opponent is an advantaged type, which is of course either 0 or 1 for the complete information case. The expected utility for a player can then be expressed as:

$$U_{ig} = p_g(r)w(v_g) - c_g x_{ig} \tag{9}$$

The equilibria for complete and incomplete contests can be obtained by substituting $w(v_A)$ and $w(v_D)$ for v_A and v_D, respectively, in the formulas presented in Table 1.1 and 1.2. It is straightforward to see that incorporating a non-monetary utility of winning increases predicted effort for both advantaged and disadvantaged groups, but

[6] Prior to conducting the experiment, we only considered a model with in-group altruism. In response to feedback on a prior version of this paper (Chopra et al., 2020), we were motivated to consider other leading hypotheses for the over-expenditure of effort in group contests.

does not alter the (directional) effects of incomplete information for any contest type. Propositions 1 to 3 continue to hold. The utility of winning may be an increasing function of the group size, N. While we do not formally model this possibility, the in-group altruism model discussed below captures a similar phenomenon.

In-group altruism

A natural extension is to assume that players derive utility based on the payoffs of other players within the group. Let $\alpha > 0$ denote the weight placed on these payoffs. The expected utility for a representative player is then (assuming players are symmetric):

$$U_{ig} = p_g(r)v_g - c_g x_{ig} + \alpha \sum_{j \neq i} \{p_g(r)v_g - c_g x_{jg}\} = \{(1+\alpha(N_g - 1))[p_g(r)v_g - c_g x_{ig}]\} \quad (10)$$

As for the case of a non-monetary utility of winning, in-group altruism serves to increase effort as now the marginal return from effort is higher. For the cost and prize value contests, regardless of information condition, altruism increases equilibrium effort by a factor of $1 + \alpha(N_g - 1)$. The Propositions therefore continue to hold.

Of interest is that the model with in-group altruism predicts that group size matters, as utility is now a function of the gains and losses to other group members. Further, an increase in N_g increases the marginal utility from effort, and it follows that group-level effort is higher for an advantaged group. In fact, the predicted effects of incomplete information for the cost and prize value contests, as summarized by the Propositions, now also hold for group size contests. As evident from Table 1 and Table 2, for contests with possible cost or value heterogeneity, the ratio of advantaged and disadvantaged team effort is equal to the ratio of the cost (c_D/c_A) or value (v_A/v_D) parameters, respectively. With in-group altruism, the ratio of advantaged and disadvantaged group effort is equal to the ratio $[1 + \alpha(N_A - 1)]/[1 + \alpha(N_D - 1)]$.

Out-group hostility

Last, we consider the possibility that a player derives disutility from the payoffs of the competing group. This could arise out of hostility towards the competitors or instead reflect preferences for relative payoff maximization. Let $\beta > 0$ denote a weight placed on the payoffs of the other group. Focusing first on the complete information case, the expected utility for a representative player is then:

$$U_{ig} = p_g(r)v_g - c_g x_{ig} - \beta\{(1 - p_g(r))N_{-g}v_g - c_g X_{-g}\} \tag{11}$$

The effect of out-group hostility varies in interesting ways across the heterogeneous value and cost contests. For the latter, the marginal effect of increasing effort is independent of the cost parameter for the other team. Equilibrium effort is scaled by βN relative to the standard model, and thus has an effect that parallels in-group altruism. For value contests, the marginal effect of effort is a function of the other team's prize value. Therefore, the effect of out-group hostility is enhanced in an uneven contest.

As in the case of in-group altruism, when motivated by out-group hostility, group size matters and can lead to differences in inter-group contests where opponents only differ in size. Here, this effect arises because an increase in effort reduces the probability that the other team wins and, in turn, decreases the expected payoff by each member of the competing group. The overall effect of this decrease in the other team's win probability intensifies as the size of the other group increases, as in the case of a heterogeneous value contest.

Incomplete information introduces complexity for the group size and value contests. To see this, the expected utility for a representative player is:

$$\begin{gathered} U_{ig} = p_g(r)v_g - c_g x_{ig} - \beta\{r(1 - (p_g(r)|A))N_A v_A + \\ (1 - r)(1 - (p_g(r)|D))N_D v_D - r c_A X_A - (1 - r)c_D X_D\} \end{gathered} \tag{12}$$

and increasing effort alters the win probabilities, and in turn the expected payouts to the other team, conditional on the (latent) type of the other group. While advantaged groups continue to exert more effort than disadvantaged groups, the effects of incomplete information are in general ambiguous. When β and/or the extent of the advantage is relatively small, Propositions 1 and 2 continue to hold for prize value contests, and the same directional effects arise for contests that differ with respect to group size. The predictions are otherwise in the opposite directions. In the case of cost heterogeneity, the mathematics simplify, and equilibrium effort is scaled by relative to the standard model. Propositions 1 to 3 continue to hold.

1.3 Experimental Design

In an experiment session, participants are randomly placed into groups, and then paired with a competing group. Players are randomly rematched into groups prior to each of 20 independent decision rounds. The total number of rounds is not disclosed to minimize possible end-of-game effects. In a round, the task of each player is to decide how many points to contribute to a "group project". Contributing points ("effort" in the theory) comes at a constant per-unit cost, and participants can select any integer amount between 0 and 50 points (inclusive). To avoid negative earnings, in each round a participant receives a "fixed income" sufficient to cover any effort costs. After all choices are made, the points contributed are added up for both groups, and the probability a group wins is given by equation [1]. Each member of the winning group receives a prize, the value of which is the same for all, regardless of how many points they contributed.

We employ a 2 x 3 between-subjects design that varies the information condition (complete or incomplete information) and the potential source of heterogeneity (cost, value, or group size). Experiment parameters are summarized in Table 1.3. As in the theory, we characterize a group as either advantaged or disadvantaged. Regardless of treatment, a disadvantaged group has three players (i.e., $N_D = 3$), and players therein can contribute at a cost of 1 lab dollar per point ($c_D = 1$) in attempt to win a prize that yields a payoff of 50 lab dollars per player ($v_D = 50$). To construct advantaged groups, we varied the relevant parameter by a factor of three. For a group with a cost advantage, effort cost is 1/3 per point; i.e., $c_A = 1/3$. For a group with a prize value advantage, $v_A = 150$, and a group with a size advantage has nine members ($N_A = 9$).

Players engage in a mix of even and uneven contests. In each round, a team has a 50% chance of being assigned the parameters for an advantaged team, and determinations are made independently for each team. For example, in a cost treatment, each group has a 50% chance of facing a 1 lab dollar effort cost and a 50% chance of a 1/3 lab dollar cost. Overall, this means that there is a 25% chance that both teams are disadvantaged, a 25% chance that both are advantaged, and a 50% chance of an uneven contest between advantaged and disadvantaged teams.[7]

In the complete information condition, players have full knowledge of the incentives (cost, value, and group size) facing members of their team as well as the competing team. In the incomplete information condition, players have full information on the parameters facing their own team only. They do know that the other team has a 50% chance (i.e., $r = 1/2$) of being advantaged, and that the status of each group is determined independently. In other words, when a team is advantaged, this provides no additional information on the state of the competing group. Consistent with the theory, under the incomplete information condition players are uncertain about a single parameter facing their opponent but know the values for the other two parameters.

[7] The group size treatment presents a logistical challenge given the number of participants in a session is fixed. To overcome this, we included 18 participants in a session, which allowed for them to play in a mix of even and uneven contests. The experiment software is programmed in such a way that gives rise to a participant playing in an uneven contest 50% of the time, an even small group contest 25% of the time, and an even large group contest 25% of the time, as stated in the experiment instructions. Important for the incomplete information treatment, a participant's knowledge of the size of her group provides no information on the size of the competing group.

1.3.1 Theoretical predictions and testable hypotheses

The experimental design lends itself to testing group contest theories in several ways. The main hypotheses to be tested based on group-level effort are summarized below, and are deliberately written as "null" hypotheses:

H1. In an uneven contest, incomplete information has no impact on group effort.
H2. In an even contest, incomplete information has no impact on group effort.
H3. Unconditional on contest type, incomplete information has no effect on effort.
H4. Group effort does not vary according to the source of the advantage.
H5. Group effort is equal for advantaged and disadvantaged groups.

The first three hypotheses relate to the three Propositions, and tests of the first four hypotheses are unique to this study. While Hypothesis 5 has been previously tested under conditions of complete information, our experimental design provides an additional test under incomplete information as well as across multiple sources of advantage.

Column (1) in Tables 1.4 and 1.5 present theory point predictions from the standard theory model for the complete and incomplete information conditions, respectively. Given the three-fold difference in heterogeneous parameters (e.g., $\frac{c_A}{c_D} = 3$), when engaged in an uneven contest, theory predicts an advantaged team exerts three times more effort than a disadvantaged team when teams differ in terms of cost-of-effort or prize value. Theory predicts that effort outcomes from cost and value heterogeneity are identical across all contest types and so are the information effects. In contrast, when teams differ in their respective group size, the self-interest model predicts that there is no advantage for the larger team and hence, there are no information effects.

Models that incorporate in-group altruism or out-group hostility predict that larger groups exert a higher group-level effort, although the effect of information depends on the specification and its parameters. In the special case of the in-group altruism model with $\alpha = 1$, the point predictions for the cost, prize value, and group size contests are identical. Further, unlike the case of the standard model, behavioral motives can lead to differences when comparing prize value and cost treatments. The variation in the experimental design therefore provides a platform from which to identify the importance of alternative behavioral theories.

1.3.2 Pilot experiment and power analysis

To help inform the experimental design, a pilot experiment was conducted using the cost treatment with incomplete information. Participants were drawn from the same population and experimental procedures followed the final protocols described later.[8] Based on the estimated group-level variances from the pilot (within and across periods), we settled on a plan to run three sessions of 18 participants for each of the cost and value treatments, and four sessions of 18 participants each for the two group size treatments.[9] The additional group size sessions help adjust for the fact that fewer group-level observations are generated from these treatments.

Based on the econometric methods employed and the planned sample sizes, power calculations suggest that we can detect a minimum treatment effect size of 9.4 units of effort based on 80% power and a 5% significance level (two-sided test) when testing Hypothesis 1 for the group size treatments, and an effect size of 8.4 when testing the same hypothesis based on either the cost or value treatments. For tests of Hypothesis 2, these figures are 8.9 and 8.4, respectively. Tests of Hypothesis 3 are powered

[8] For this reason, we include data from the pilot in the analysis.

[9] Sample sizes, and target minimum detectable effect sizes, are based on predictions from the in-group altruism model with $\alpha = 1$. For heterogeneous prize value and cost contests, this model predicts group-level efforts are equal to the standard theory prediction multiplied by group size ($N = 3$). Further, in this special case, predictions for all three sources of advantage are identical. In the Abbink et al. (2010) experiment, the four-person teams in no-punishment treatments expend 1,035 points on average, which is 4.1 times the prediction of standard theory (250). Bhattacharya (2016), with three-person groups, finds that group-level effort averages 689 in even contests, and effort for advantaged and disadvantaged teams in uneven contests average 951 and 555, respectively. These effort levels are 4.3 to 5.5 times standard theory predictions. Chen and Li (2009) assume in their theory that individuals place the same weight on their group members' payoffs as their own, which is consistent with the assumption $\alpha = 1$.

to detect somewhat smaller differences, given that data from all even and uneven contests are pooled (7.7 for group size treatments, 7.0 for value and cost treatments). For Hypothesis 4 and Hypothesis 5, minimum detectable effect sizes range from 7.0 to 10.5, and from 9.5 to 11.0, respectively.

Power calculations are of course only approximations as the true underlying outcome distributions are unknown. We expect lower variation in group-level effort for the complete information cases, and to the extent this is true, the calculations above are under-estimates of the minimum detectable effect sizes. Moreover, controlling for other factors, such as participant characteristics, in the econometric models is expected to increase power as these factors should be uncorrelated with treatment assignment.

1.3.3 Experimental procedures

A typical experimental session proceeds as follows. Participants are randomly assigned an ID number and a computer station in the laboratory. The same moderator reads instructions aloud and follows several protocols that are clearly mentioned in the consent form as well as in written instructions provided to participants. All decisions are made on the computer. The experiment was programmed and conducted using the software z-Tree (Fischbacher, 2007).

Prior to the group contest experiment participants complete a (paid) risk elicitation task of the sort popularized by Holt and Laury (2002). Following standard procedures, the outcome of this task is not revealed until the end of the session. After reading instructions for the group contest experiment, participants take a quiz designed to test and educate participants on earnings calculations and the incentives they face. Participants are paid for correct answers and provided detailed answers to the questions posed. Participants then proceed through one unpaid training round and any questions are answered by the moderator prior to the 20 paid rounds.

For complete information treatments, the computer decision screen displays all three parameters (cost, value, group size) in effect for the participant's group as well as the opponent group. For incomplete information treatments, identical information is displayed except for the one parameter for the opponent group for which there is uncertainty. In this case, the two possible values are displayed. After all decisions are entered, a result screen reveals which team won, total effort for the participant's group, and earnings for the round. Participants do not see the individual efforts of their team members, nor do they receive any direct information on the choices of their opponent. Participants earn money based on the outcome in each of the 20 paid decision rounds. The experiment concludes with a demographic questionnaire. Representative instructions and the questionnaire are provided in the appendix.

1.3.4 Participants

Eighteen experiment sessions were conducted during the summer and fall of 2019 as well as fall of 2020. Including the pilot, we have data from 360 participants.[10] All sessions were conducted in a designated experimental economics laboratory at a major public research university. Undergraduate students were recruited from a large existing database that had previously registered to receive invitations for economics experiments. People were not allowed to attend more than one session of the experiment. Earnings were dominated in "lab dollars" and exchanged for U.S. dollars at an announced exchange rate. As theory predicted earnings in the value treatments to be considerably higher, we used an exchange rate of 120-to-1 for the value treatments, and 90-to-1 for the remaining treatments. The experiment lasted approximately 50 minutes and on average participants earned $18 for the session.

Table 1.6 describes the experiment data. Overall, 42% of participants are female, 56% had participated in a prior economics experiment, and 47% can be characterized

[10]Due to variations in participant show-up rates, there are 42 participants in the complete information cost treatment and 66 participants in the incomplete information cost treatment. Revising our power calculations based on these realized sample sizes has only a negligible effect. We met exactly our sample size targets for the other treatments.

as risk averse based on the incentivized risk elicitation task. The average score on our instructions quiz is about 86%. Sixty percent of participants answered all quiz questions correctly, 27.5% answered three correctly, and the remainder answered 2 or fewer questions correctly. Responses from the post-experiment questionnaire suggest that the vast majority (88%) felt they were sufficiently compensated. In response to a Likert-scale question that ranged from "1" ("poorly understood") to "5" ("well understood"), the vast majority (89%) selected a 4 or 5, indicating a strong self-assessment of how well instructions were understood.

1.4 Results

Column (3) in Tables 1.4 and 1.5 present the observed group-level efforts by group type, source of heterogeneity, contest type, and information condition. Consistent with prior experiments, actual effort far exceeds what is predicted by standard theory. For cost-of-effort and prize value contests, effort is two to four times higher, depending on the comparison, and is three times higher on average. For group-size contests, effort for advantaged groups exceeds theory predictions by an order of magnitude, and effort is roughly three times higher relative to disadvantaged groups; in contrast, standard theory predicts no differences due to group size. For the cost and value treatments, in uneven contests, the observed effort for disadvantaged groups is much higher (about double) under incomplete relative to complete information; in contrast, the opposite is true in group size treatments. For group size treatments relative to the cost and value treatments, effort for advantaged groups is about 60% and 40% higher, respectively, in complete and incomplete information contests.

We begin the analysis with linear regression models of group-level outcomes: effort and the probability of winning. We cluster standard errors by period within a session, which allows for heteroskedasticity and contemporaneous correlation across groups. Recalling that participants are randomly re-sorted into groups every period, nearly all groups will be unique. For models that include participant characteristics, we use group-specific averages. We also provide some exploratory analysis by including source of heterogeneity by round effects to control for trends observed in the raw group-level effort data. Then, we explore individual behavior by analyzing variation in free-riding and a measure of within-group variation. In these regressions, we cluster standard errors by individuals, which allows for heteroskedasticity as well as within-subject serial correlation. Last, we draw conclusions about behavioral motives, and how to best reconcile the data with the theory models we consider.

1.4.1 Group-level effort

Tables 1.7 and 1.8 present regressions that allow for tests of information effects for uneven and even contests, respectively. When interpreting results, it is important to keep in mind that with incomplete information participants do not know whether they are engaged in an even or uneven contest, and as a result theory predicts differences across information conditions. Further, when testing for information effects this allows us to include all the data from the incomplete information treatments regardless of the contest type.[11] Specification (1) estimates the effect of incomplete information, averaged across all potential sources of heterogeneity. Specification (2) includes interactions to allow for tests of information by advantage source. Specification (3) adds control variables, defined in Table 1.6, to the interactions model.

In uneven contests, when averaged across heterogeneity sources, incomplete information significantly increases group-level effort by 11 points on average. This positive and significant effect is largely driven by the cost and value treatments, where the effects are 20 and 19 points, respectively. For group size treatments, the estimate of the information effect is negative and significant in specification (2),

[11] Tables A.3 and A.4 in the appendix present results that restrict the data used from the incomplete information treatments to only include observations from what are in actuality uneven and even contests, respectively. As participants in the incomplete information treatments do not know whether they are playing in an even or uneven contest, restricting the data has the expected effects – treatment effects are very similar but less precisely estimated.

although this effect is statistically zero when control variables are included.[12] Thus, we reject Hypothesis 1 for the value and cost treatments and in turn find support for Proposition 1. The self-interest model does not predict an information effect for the group size treatments which is reflective of the results, although based on other evidence this model overall inadequately explains behavior. Based on specification (3), effort decreases as the experiment progresses. Groups with a higher proportion of players with prior experience in economics experiments, and groups with a higher average GPA, put forth less effort on average. Groups with more players that identify as female select a higher effort choice.

For even contests, we find no effect of information when averaging across all sources of heterogeneity. Thus, we fail to reject Hypothesis 2. This is contrary to Proposition 2 where we expect incomplete information to lower effort for cost and value heterogeneity. In fact, based just on the value treatments, there is a *positive* and marginally significant effect of incomplete information.[13] The effects of control variables are similar to what we find for uneven contests, with the exception that here risk aversion decreases effort.

Table 1.9 pools data across the two contest types. Theoretically, and as implied by the point predictions in Table 1.4 and 1.5, effort for an advantaged (disadvantaged) group under incomplete information is equal to the effort, averaged across uneven and even contests, for an advantaged (disadvantaged) group under complete information. Thus, theory predicts the expected effort to be the same regardless of the information condition; however, empirically there are significant differences. We find, pooling over sources of heterogeneity, incomplete information increases group-level effort by approximately 6 points. By allowing the information effects to vary with the source

[12]We randomized treatments across sessions prior to implementation, but by chance have a slight imbalance of characteristics across the two group size treatments, with the complete information treatment utilizing a higher proportion of females, and a lower proportion of experienced and risk averse participants.

[13]We present regressions restricted to advantaged groups in Table A.5 in the appendix, which reveal large and positive effects of incomplete information for the cost and value treatments, although this effect is only statistically significant for the latter.

of advantage, we find that incomplete information increases effort in the value and cost treatments. In contests with potential group size heterogeneity, this effect is insignificant when we include control variables. Thus, we fail to reject Hypothesis 3 for the group size treatments but reject this hypothesis for the value and cost treatments.

Result 1. In uneven contests, incomplete information increases group-level effort when one team has either a cost or value advantage but has no effect when one team has a group size advantage.

Result 2. In even contests, there is marginal evidence that incomplete information increases effort in the value treatment. Incomplete information has no effect in the cost and group size treatments.

Result 3. When the data is pooled across the two contest types, incomplete information increases group-level effort for the value and cost treatments but has no effect for the group size treatment.

The regressions reported in Table 1.9 allow us to test Hypothesis 4 across a few dimensions. From specification (2) and (3), we deduce that effort for the group size treatment with complete information is statistically different and higher when compared with either the value or cost treatment. With incomplete information, we find no difference based on any pairwise comparison of treatments.[14] Comparisons based on the results in Tables 1.7 and 1.8 demonstrate that effort is equivalent across the value and cost treatments regardless of contest type. Further, these results reveal that the overall differences observed between the group size and other

[14]Based on specification (3) in Table 1.9, additional test results are as follows: value versus group size, complete information (F=26.85, p=0.00); value versus cost, incomplete information (F=0.04, p=0.84); value versus group size, incomplete information (F=0.06, p=0.80); cost versus group size, incomplete information (F=0.01, p=0.92).

treatments under complete information are driven by behavior in both uneven and even contests.[15,16]

Result 4. Based on a large set of comparisons, group effort is similar across cost and value treatments. On the other hand, many differences arise when comparing group size treatments with either the cost or value treatments.

Result 5. Group effort is higher for advantaged groups, regardless of the source of advantage or information condition.

1.4.2 Group-level effort and time trends

From results presented earlier, we found that group-level effort decreases as the experiment progresses in both uneven and even contests (see Table 1.7 and 1.8). This finding is both intuitive as well as empirically supported by the literature. This means that the information effect may also vary by round. To explore the same, we include an empirical model in Appendix A that allows the effects of the three sources of heterogeneity (cost, value and group size) on group-level effort to vary by round. This helps control for the possible trends observed in the group-level effort in Figures 1.1-1.4. Tables A.6 and A.7 replicate Specification (1) and (2) from Tables 1.7 and 1.8 for uneven and even contests respectively. Specification (3) allows the source-specific information effect to vary by round.

A few comparisons are worth noting. In uneven contests, while the directional impacts of incomplete information, as predicted by theory, hold for cost and value treatments, the magnitudes are not similar anymore. An important reason for that is

[15] Based on specification (3) in Table 1.7, additional test results are as follows: value versus group size, complete information (F=42.43, p=0.00); value versus cost, incomplete information (F=0.05, p=0.82); value versus group size, incomplete information (F=0.09, p=0.76); cost versus group size, incomplete information (F=0.02, p=0.88).

[16] Based on specification (3) in Table 1.8, additional test results are as follows: value versus group size, complete information (F=3.60, p=0.05); value versus cost, incomplete information (F=0.08, p=0.78); value versus group size, incomplete information (F=0.21, p=0.64); cost versus group size, incomplete information (F=0.08, p=0.78).

when we account for the time trends, the effects are significant for value and group-size treatments in the complete information condition but not for cost treatments. In the incomplete information condition, the opposite is true. Given this empirical specification, group-effort in the complete information treatments (for value and group-size) declines as rounds progress while in the incomplete information, it declines for the cost treatments. Therefore, the estimated information effect is also expected to change. Figures 1.1-1.4 depict a convergent behavior towards the later rounds (more so for group-size treatments) so it makes sense to investigate how the treatment effect changes.

Recall that in *uneven* contests, incomplete information increases group-level effort (empirically true for cost and value treatments). The estimated information effect at the average round (round=10) is 15.78 ($p = 0.00$), 8.65 ($p = 0.10$) and -27 ($p = 0.00$) while it is 8.91 ($p = 0.07$), 5.18 ($p = 0.37$) and -39 ($p = 0.00$) at the onset of the last 5 rounds (round=15) for cost, value and group-size respectively. The effect shrinks for cost and value treatments as rounds progress as expected but it becomes more pronounced for the group-size treatments which is due to a rather large time trend.

In *even* contests, there are null effects of incomplete information. The only time trend that is statistically significant is for the group-size treatments in an incomplete information. What this means for the average information effect is that it is null for cost and value treatments at the average round as with other specifications but interestingly for the cost treatments, the effect is in the direction that theory predicts for the last 5 rounds and magnitudes range from 13 points to 19 points. For group-size treatments, this effect is null at the average round but for the last 8 rounds, the information effect is in line with theory and magnitudes range from 16 to 32 points.

Summarizing, as rounds progress, the information effect in uneven contests shrinks but still remains in the expected direction for cost and value treatments. On the other hand, the effect of information in even contests is visible and more pronounced in the later rounds for cost and group-size treatments. This provides a strong evidence that the information effect has important time trends that need to be accounted for in the analysis. Finally, it is worth noting that one of our results that highlights an equivalence between cost and value treatments is robust to inclusion of time trends but only in the case of complete information.

1.4.3 Probability of winning

Table 1.11 presents regressions that estimate differences in the chances of winning between advantaged and disadvantaged groups in uneven contests for each source-of-advantage.[17] Separate regressions are run for the two information conditions. The probability of winning is endogenous and determined by the relative effort of the competing groups. Standard theory predicts that, in uneven contests where one team has either a cost or prize value advantage, that the advantaged team has a higher chance of winning. In particular, the three-fold advantage we implement gives rise to the advantaged group being three times more likely to win; in other words, theory predicts the advantaged team has a 75% chance of winning. For the group size treatments, as group-level effort is invariant to group size, theory predicts there will be no differences in the probability of winning across advantaged and disadvantaged teams.

With complete information, win probabilities are consistent with a roughly three-fold advantage for all treatments. For the group size treatment, from specification (1), the advantaged and disadvantaged groups have 75% and 25% chances to win, respectively. The advantaged team has an 80% and 74% chance of winning, respectively, in the cost and value treatments. Under incomplete information, the win percentages remain very close to the 75%/25% split for the group size treatment. However, advantaged groups in the value and cost treatments have a significantly lower chance of winning than what theory predicts: 62% and 67%, respectively, based on specification (3). This is largely driven by the fact that, as illustrated

[17] A parallel analysis for even contests is uninformative given that both advantaged and disadvantaged teams will have a 50% win probability by construction. It is also for this reason that the data from the incomplete information treatments are restricted to uneven contests, regardless of the fact that participants did not know the contest type.

in Table 5, the actual ratio of advantaged to disadvantaged group effort is noticeably less than 3-to-1. There are significant differences in the chances of winning (about 5 percentage points) for a particular group type across both information conditions when comparing cost and value treatments.

1.4.4 Within-group heterogeneity

Last, we briefly investigate heterogeneous behavior within groups by estimating regressions based on individual-level effort choices. As in other social dilemma games, the possibility arises for players to free-ride off the effort expenditures of other players. About 21% of individual-level effort expenditures (1506 of 7200 observations) are zero, so it makes sense to take a closer look at the extent of free-riding. Model (1) in Table 1.12 presents a linear regression where the dependent variable is an indicator that equals 1 in cases where the participant contributed 0 effort.[18] Being in an uneven contest increases free-riding by 8 percentage points. Competing on an advantaged team decreases free riding by 13 percentage points. Incomplete information has no effect. There is significantly more free riding in the group size treatments relative to either the cost or value treatments, and the estimated difference is approximately 12 percentage points. When effects are considered in tandem, the highest rate of free riding comes from players on a small team that are, with or without their full knowledge, competing against a larger team. Of course, regardless of what may be true in theory, the optics for those on a small team are bleak. Players with prior participation in economics experiments and those classified as risk averse are more likely to free ride, whereas females are less likely to free ride. On average, free riding is 14 percentage points more likely in the last round of the experiment relative to the first round.

We analyze as a second measure of within-group heterogeneity the squared deviation of a player's effort from the group mean; i.e., $(x_{ig} - x_g)^2$. Given random

[18]While we continue to use linear regression because of its robustness properties, similar results arise if we instead estimate a probit model.

re-sorting into groups, x_g is specific to the particular group one is in for a specific decision round. In the extreme case where each group member makes the same effort choice, the measure equals zero. Analysis of this outcome variable is presented as Model (2) in Table 1.12. Participant characteristics are strongly correlated with this variance measure. Within-group variation decreases with risk aversion, as well as experience in prior economics experiments. The latter is suggestive of a learning effect. The contribution variance, however, does not vary as the experiment progress. Overall, most of the variation in the experimental design does not appear to impact within-group variation. The main exception is that there are larger disparities among members of an advantaged team. This is somewhat unexpected, given that free riding is less likely for advantaged team members. As a possible explanation, some players may feel a stronger frustration of losing when on an advantaged team and, in turn, over-expend. If players hold expectations that their team members will behave this way, a logical response is to then contribute significantly less.

1.4.5 Consideration of behavioral motives

The stark differences in effort levels between the experiment data and the predictions of standard theory, along with discrepancies between theory and observation with respect to the effect of incomplete information, suggest the potential importance of behavioral motives. One robust finding in the various comparisons is that effort is quite similar across cost and prize value treatments. This contrasts with the predictions of the out-group hostility model. Further, it implies that, if a non-monetary utility of winning is important, this utility is proportional to prize value. In contrast, if non-monetary utility is additive, this would also lead to differences between these two treatments. The high effort levels in all treatments, including the group size treatments, suggests that in-group altruism is an important driver. A model that only allows for a non-monetary utility of winning continues to predict that group-level effort is invariant to group size.

Based on these observations, we fully developed the theory to incorporate both a non-monetary utility of winning, with $w(v_g) = v_g(1+\gamma)$, and in-group altruism.[19] With this specification, incomplete information increases effort in uneven contests, decreases effort in even contests, and has no effect on the average contest. That is, Propositions 1 to 3 continue to hold for the cost and value treatments, and these directional effects now also apply to heterogeneous group size contests. Derivations and proofs are provided in the appendix.

To gain an understanding of the underlying structural parameters of the extended theory model, we considered estimating these parameters by selecting values that minimize the sum of squared deviations between the observed values presented in Table 1.4 and Table 1.5 and the theoretical predictions (i.e., the equilibria based on α and γ). An issue that arises is that, at least for the cost and value treatments, α and γ are perfectly linearly dependent and one cannot identify them separately. As a compromise, we set $\gamma = 0$ and estimated α separately for each of the six treatments. These estimates vary from 0.66 to 1.29.[20] Column (2) in Table 4 and 5 present theory predictions based on these estimates. If we instead restrict the parameter to be equal across treatments, the estimate is $\alpha = 0.99$.

While the extended theory does a better job than the standard theory in justifying the high effort levels in all contests we study, it fails to explain why we observe no effect of incomplete information in even contests, and no effect of incomplete information in uneven contests for the group size treatments. We thus turn to other possible explanations.

Sheremeta (2013) provides evidence that bounded rationality explains overbidding relative to standard theory in individual contests. To explore whether cognitive limitations may explain patterns in the data, we include as specification (4) in Tables

[19] We assume that in-group altruism is tied to both monetary and non-monetary values of winning. Excluding the latter from the altruism term leads to theoretical differences across cost and prize value treatments.

[20] We obtain the following estimates: cost-of-effort (complete) = 0.72; prize value (complete) = 0.66; group size (complete) = 1.29; cost-of-effort (incomplete) = 0.92; prize value (incomplete) = 0.90; group size (incomplete) = 1.26.

1.7 to 1.9, regressions that allow for an interaction effect between GPA and the indicator for incomplete information. To the extent that GPA serves as a proxy for cognitive ability this interaction measures differences in the estimated incomplete information effect due to cognitive ability. For both uneven and even contests, GPA has a statistically insignificant effect for the complete information treatment, but there is a large and negative coefficient on the interaction term, indicating that the incomplete information effect decreases with an increase in cognitive ability. The incomplete information treatments are more complex, and this buttresses the notion that GPA is a proxy for cognitive ability.

For even contests, if we calculate the effects of incomplete information conditional on high GPAs, the null effect becomes a negative effect.[21] Therefore, bounded rationality is a potential explanation for why we observe a null effect for incomplete information, which is in contrast to Proposition 2. These negative effects of GPA on effort in the uneven contests are not large enough to negate the positive effect of incomplete information on effort in cost and value contests, i.e., Proposition 1 continues to hold for high GPA groups.

Bounded rationality also has the potential to explain the effect of incomplete information in uneven contests where teams differ in terms of group size. The group size treatments were conducted with sessions of 18 participants, and the lab space is small. As this participant count was not announced, in the incomplete information setting some participants may have thought that a contest between two groups of nine players was not possible logistically and instead it was more plausible that the other team would be of a different size. Of course, such a phenomenon would be unique to the group size treatments. From Tables 1.4 and 1.5 the observed group-level efforts for an uneven contest with complete information, and an uneven contest under incomplete information, are virtually identical, which is suggestive evidence of this phenomenon.

[21] Negative and significant treatment effects emerge when GPA is at least 0.25 grade points higher than the sample average for the cost and group size treatments, and at least 0.50 grade point higher for the value treatment.

Further, there are important time trends in the data which when accounted for provide evidence that at least for cost and group-size treatments, the effect of information is in line with theory in the later rounds. Theory depicts a one-shot game while the experiment may have learning to some extent. In even contests, if participants initially had different subjective probabilities regarding facing a certain type of team (advantaged or disadvantaged) but eventually revised their probabilistic beliefs to be closer to true objective probabilities in the experimental design, then it provides one potential explanation of why the expected negative impact only shows up in the later rounds. Although speculative in nature, this provides another channel for bounded rationality concerns and learning effects to cause deviations from theory.

As with other games characterized by social dilemma, there is the potential in group contests to free ride off the efforts of others. Free-riding incentives provide another reason why incomplete information does not increase effort in uneven contests where teams differ by size. We find evidence of higher free-riding as rounds progress especially for the group size treatments. Moreover, our data analysis reveals that the highest amount of free-riding occurs in disadvantaged (i.e., small) groups under incomplete information while free-riding is comparable for big and small groups under complete information. If we alter the extended theory to account for such behavior, e.g., by decreasing N_D while holding N_A constant, this lowers equilibrium efforts for both teams under incomplete information. This behavior has virtually no effect on effort under complete information while it lowers effort for both teams under incomplete information which then serves to counteract the anticipated information effect for group size treatments in uneven contests.

1.5 Conclusion

This study uses theory and experiments to study the impact of incomplete information on group-level effort in a heterogeneous inter-group competition. Importantly, we investigate three different potential sources of advantage, which provides a platform from which to evaluate a standard theory based only on self-interest as well as potential behavioral extensions to this theory. Specifically, we study contests where teams differ in terms of the cost of effort, the prize value, or group size. Standard theory predicts group-level effort is invariant to group size, and in turn that incomplete information has no effect. Based on the selected experiment parameters, standard theory predicts equivalent effort across cost-of-effort and prize value contests, but alternative theories predict differences in some cases.

The experiment reveals that group-level effort is (1) much higher than predicted by standard theory for potential each source of advantage, (2) increases with group size, and (3) similar across cost-of-effort and prize value treatments. This evidence supports an extended, in-group altruism model that assumes people derive utility from payoffs that accrue to other group members. In addition to in-group altruism, players may derive a non-monetary utility of winning and, if so, the data suggests that utility is proportional to the monetary prize value. The experimental evidence does not lend support to a model of out-group hostility.

While the extended theory helps to rationalize the effects of altering group size and the high effort levels relative to standard theory predictions, the theory fails to explain all the observed effects of incomplete information on effort. We offer as potential explanations bounded rationality and the propensity to free ride as group size increases. Ultimately, additional experiments are needed to substantiate our claims as well as to provide alternative tests of the theories we consider. Future investigations would benefit from measures of cognitive ability, altruism, non-monetary utility of winning, subjective beliefs, and other potential behavioral drivers, determined through questionnaires or other experiments. Using a 'partners' rather than a 'strangers' experimental design, or using an intervention to promote group identity, are possible ways to exogenously vary the effects of altruism through the experimental design. While our conclusions are open to alternative interpretations, we note that our study is the first in the group contest literature to investigate alternative theories.

Some of our results serve to reinforce and extend prior findings. In an uneven group contest with cost-of-effort heterogeneity, Bhattacharya (2016) finds that advantaged teams contribute significantly more effort than disadvantaged teams. This is consistent with our results, and we further demonstrate that this advantage effect holds for different sources of advantage as well as with incomplete information. In individual lottery contests involving four players, incomplete information over marginal cost-of-effort causes players with a low (high) marginal cost to submit higher (lower) bids over complete information (Brookins and Ryvkin, 2014). In a group contest setting, we find that incomplete information also alters effort both in theory and in practice. The theory also suggests that, while incomplete information increases effort for an advantaged team in an uneven contest, its effect on a disadvantaged team depends on parameter values. Last, consistent with the group contest literature but inconsistent with a standard theory model of self-interest, increasing group size does lead to higher group-level effort. Prior group contest experiments use groups with five or fewer players and thus we demonstrate that the stylized fact continues to hold with nine-member groups. We find that being in a larger group is in practice a more significant advantage relative to a cost-of-effort or prize value advantage.

On a final note, our findings have relevance for contest design. When competing groups are asymmetric (uneven contests), according to theory incomplete information causes only a wasteful increase in efforts (lower efficiency) as there are no significant changes in the probability of winning for advantaged and disadvantaged groups. Thus, if the organizer cares about efficiency, they can engender this by promoting transparency (e.g., by providing details on the competitors). On the other hand, our experimental results suggest that if the goal is to reduce free-riding behavior, then the contest designer may prefer less transparency as we observe less free-riding under incomplete information.

Bibliography A

Abbink, K., J. Brandts, B. Herrmann, and H. Orzen (2010). Intergroup conflict and intra-group punishment in an experimental contest game. *American Economic Review 100*(1), 420–47.

Ahn, T.-K., R. M. Isaac, and T. C. Salmon (2011). Rent seeking in groups. *International Journal of Industrial Organization 29*(1), 116–125.

Baik, K. H. (1993). Effort levels in contests: The public-good prize case. *Economics Letters 41*(4), 363–367.

Bhattacharya, P. (2016). Inter-team contests with power differential. *Journal of Economic Behavior & Organization 132*, 157–175.

Boosey, L., P. Brookins, and D. Ryvkin (2017). Contests with group size uncertainty: Experimental evidence. *Games and Economic Behavior 105*, 212–229.

Brookins, P. and D. Ryvkin (2014). An experimental study of bidding in contests of incomplete information. *Experimental Economics 17*(2), 245–261.

Cason, T. N., R. M. Sheremeta, and J. Zhang (2012). Communication and efficiency in competitive coordination games. *Games and Economic Behavior 76*(1), 26–43.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–57.

Chopra, V., H. M. Nguyen, C. A. Vossler, et al. (2020). Heterogeneous group contests with incomplete information. Technical report.

Dechenaux, E., D. Kovenock, and R. M. Sheremeta (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics 18*(4), 609–669.

Eliaz, K. and Q. Wu (2018). A simple model of competition between teams. *Journal of Economic Theory 176*, 372–392.

Fallucchi, F., E. Renner, and M. Sefton (2013). Information feedback and contest structure in rent-seeking games. *European Economic Review 64*, 223–240.

Fey, M. (2008). Rent-seeking contests with incomplete information. *Public Choice 135*(3), 225–236.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Fonseca, M. A. (2009). An experimental investigation of asymmetric contests. *International Journal of Industrial Organization 27*(5), 582–591.

Heap, S. P. H., A. Ramalingam, S. Ramalingam, and B. V. Stoddard (2015). 'doggedness' or 'disengagement'? an experiment on the effect of inequality in endowment on behaviour in team competitions. *Journal of Economic Behavior & Organization 120*, 80–93.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Katz, E., S. Nitzan, and J. Rosenberg (1990). Rent-seeking for pure public goods. *Public Choice 65*(1), 49–60.

Kimbrough, E. O., R. M. Sheremeta, and T. W. Shields (2014). When parity promotes peace: Resolving conflict between asymmetric agents. *Journal of Economic Behavior & Organization 99*, 96–108.

Kugler, T., A. Rapoport, and A. Pazy (2010). Public good provision in inter-team conflicts: effects of asymmetry and profit-sharing rule. *Journal of Behavioral Decision Making 23*(4), 421–438.

Leibbrandt, A. and L. Sääksvuori (2012). Communication in intergroup conflicts. *European Economic Review 56*(6), 1136–1147.

Mago, S. D., A. C. Samak, and R. M. Sheremeta (2016). Facing your opponents: Social identification and information feedback in contests. *Journal of Conflict Resolution 60*(3), 459–481.

Malueg, D. A. and A. J. Yates (2004). Sent seeking with private values. *Public Choice 119*(1), 161–178.

Rapoport, A. and G. Bornstein (1989). Solving public good problems in competition between equal and unequal size groups. *Journal of Conflict Resolution 33*(3), 460–479.

Sheremeta, R. M. (2010). Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior 68*(2), 731–747.

Sheremeta, R. M. (2013). Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys 27*(3), 491–514.

Sheremeta, R. M. (2018). Behavior in group contests: A review of experimental research. *Journal of Economic Surveys 32*(3), 683–704.

Tullock, G. (1980). Efficient rent seeking, towards a theory of the rent-seeking society, edited by buchanan, j., tollison, r., and tullock, g.

2

Are we doing more harm than good? Hypothetical bias reduction techniques in potentially consequential survey settings

Abstract

Proponents and critics of using stated preference surveys to quantify the monetary benefits of public goods have traditionally been concerned that the method suffers from an upward "hypothetical bias". This bias is purported to arise because respondents view their survey choices as hypothetical, and there is a large experimental literature that documents persistent and often large discrepancies between hypothetical and revealed preference measures of demand. More recently accumulated evidence suggests that most field survey respondents instead perceive that their decisions have consequences (i.e., impact their future welfare), and are not hypothetical, which raises questions about the efficacy of the various hypothetical bias reduction techniques now in use. To address this question, we test popular bias reduction procedures in both hypothetical choice and incentive compatible, real payment settings: cheap talk, solemn oath, and certainty corrections. While on one hand we find that the oath reduces willingness to pay (WTP) in a hypothetical setting, on the other, it increases demand in an incentive compatible setting. Applying the common rules for ex post adjustment of choices based on stated response certainty leads to significant and large decreases in WTP estimates for both hypothetical and incentive compatible settings. Cheap-talk does not affect WTP per se but has a rather large effect on the variance of the WTP distribution. Our results suggest that survey researchers should make use of screening questions to determine how to best target hypothetical bias reduction techniques to those prone to bias.

Keywords: hypothetical bias, consequentiality, stated preferences, experiments, solemn oath, cheap talk, certainty corrections

2.1 Introduction

What are the benefits, in money terms, from improving local public schools, developing a new car safety device, or reducing air pollution? The leading technique for estimating benefits for non-market goods, as well as goods not yet brought to market, involves the use of carefully constructed sample surveys. These stated preference survey methods have the ability to estimate benefits for situations far beyond the scope of existing data, and further to estimate benefits not directly tied to market transactions, including values not associated with direct use (e.g., value of preserving ecosystem services). Despite the widespread use of these methods in government cost-benefit analyses as well as environmental litigation, their validity is subject to an ongoing debate (Johnston et al., 2017). At the center of the debate is the methods' ability to elicit behavioral intentions in a survey that reflect choices people would actually make in a related setting where there are direct financial consequences, such as when voting in an actual public referendum.

Much of the literature from laboratory experiments documents that people tend to over-value goods and services in a hypothetical choice setting, relative to a parallel setting where choices for the same goods have financial consequences (Loomis, 2011). In other words, when the choice is hypothetical, some people state they would pay for a good whereas in similar situations when there are consequences, they would end up not paying. This is referred to as "hypothetical bias". Given the method's widespread use in public policy, hypothetical bias can distort value estimates. When these methods were being developed, it was a common belief among practitioners that people view such surveys as completely hypothetical. Based on this traditional view, the literature devised techniques designed to correct for the bias. Laboratory experiments have demonstrated that techniques such as "cheap-talk", the "oath" and certainty adjustment approaches can be effective in obtaining value estimates that are comparable to consequential settings. This evidence, in turn, has motivated many practitioners to adopt these techniques in field surveys.

Vossler and Evans (2009) highlight that there is a potential disconnect between laboratory research on hypothetical bias and realities of the field: for surveys that are designed to inform public decision-making, responses have the potential to influence a government action. As stated preference surveys are often of high quality, funded by government agencies, and relate to important public policy issues, it is plausible that this information ultimately informs policy. In turn, it seems only logical that a subset of respondents may perceive their choices as consequential and may be motivated to respond truthfully (Carson and Groves, 2007). Supporting this claim, recent evidence suggests that most – but not all – survey respondents in fact view their choices to have consequences (Herriges et al. (2010); Zawojska et al. (2019)). Despite the recent take on stated preference surveys, the use of hypothetical bias reduction techniques has continued. Moreover, many practitioners include reduction techniques in field surveys, regardless of whether they are interested in testing its effects. This brings up an important question about the intended (or unintended) effects of reduction procedures on value estimates in a consequential setting. This is the question this study seeks to address, and we do so through an experiment that examines the effects of using popular hypothetical bias reduction techniques in an incentive compatible, consequential setting. Loomis (2014) highlights several *ex-ante* and *ex-post* techniques that have explored in the literature. *Ex-ante* approaches are administered prior to asking the valuation question and popular examples include a consequentiality design, honesty (e.g., solemn oath), cheap talk, social desirability and cognitive dissonance minimization approaches. More recent ex-ante methods also include the Bayesian Truth Serum (Barrage and Lee, 2010).

Consequentiality pertains to the idea that participants view a hypothetical survey as real when they believe that there is a positive probability that the survey will result into a policy action (Vossler et al., 2012). Therefore, a consequentiality design must ensure that the survey has a potential impact on the respondent's future utility.[1]

[1]Recently Vossler et al. (2020) show that consequentiality is one of the many necessary conditions besides respondent beliefs and belief consistency for a stated preference survey to be incentive compatible.

The“oath” primarily involves participants voluntarily making a promise just as the one made in real world courts. It works as a commitment device to reveal truthful preferences by asking participants to swear upon their honor to provide honest answers (Jacquemet et al., 2013). The oath has been shown to reduce hypothetical bias in a voting referendum (Jacquemet et al., 2017). Aside from its use in non-market valuation, the oath has also proven its effectiveness in tax evasion games by increasing compliance rates (Jacquemet et al., 2020). “Cheap-talk”, introduced by Cummings and Taylor (1999), is best characterized as an information nudge designed to motivate survey responses that better reflect actual behavior. Since then, many studies have explored cheap-talk (Murphy et al. (2005); Aadland and Caplan (2006); Morrison and Brown (2009); Barrage and Lee (2010)). Social desirability designs as well as cognitive dissosance approaches focus on minimizing positive responses arising out of a tendency to project a positive self or social image. Some examples include asking respondents their views on how other respondents would behave which serves to highlight whether people care about social acceptability. Cognitive dissonance specifically offers more bid choices in a dichotomous choice format. More elaborate payment card formats can help lower the need for such methods to an extent, such as the one we use in this study.

Ex-post techniques are administered after the valuation question has been posed to the respondents and primarily includes certainty follow-up questions. The “certainty adjustment” approach asks respondents to gauge how certain they are about their stated choices (e.g., their vote in the advisory referendum), and uses this information to re-code responses of those who state they are uncertain. There are two main formats for certainty adjustment, one includes a 10-point follow up scale while the other uses a qualitative scale with the former being more popular. The method involves recoding a set of positive responses to negative based on a certain threshold and therefore relies on respondent uncertainty as a dominant mechanism in reducing hypothetical bias. In a meta-analysis, Little and Berrens (2004) highlight the effectiveness of such methods in making hypothetical and real responses comparable.

In this paper, we use an experiment to study the effects of the solemn oath, a quasi-cheap talk script, and certainty corrections, when combined with an incentive compatible elicitation mechanism. Our research contributes to the existing literature in two ways. First, it identifies the impact of using reduction techniques (specifically the oath, cheap-talk and certainty adjustment) in a consequential setting. While prior laboratory studies have tested these approaches for public and private goods with non-market attributes, they have done so only in hypothetical choice settings. To our knowledge, there is only one study by Jacquemet et al. (2017) that explores the oath in a real payment setting. They do not find any evidence of the oath statistically altering votes in the real treatment; however, we do note that their design is based on a single bid amount. We depart from previous studies by using an incentive compatible payment card elicitation. This provides a more efficient way to identify the underlying willingness-to-pay (WTP) distribution and overcomes a limitation of prior studies that only elicit a yes/no response to a single price (which precludes welfare estimation). It provides us with the benefit of testing for the effects of the reduction procedures at various points in the WTP distribution. Second, the design of this study incorporates features from field experiments to enhance lab-to-field generalizability.[2]

It is important to explain why we focus on a subset of ex-ante and ex-post reduction methods. The methods such as consequentiality scripts have been shown to improve validity in field settings, however, both consequentiality scripts and post-experiment consequentiality questions are troublesome to use in an incentive compatible setting as they would cast doubt on the incentive compatibility of the mechanism so to speak. Further, social desirability and cognitive dissonance approaches are rarely employed in the field. Therefore, it stands to reason to focus on the commonly used methods, especially ones where effects on WTP and value estimates are not clear.

[2]The treatments in this study are thus intended to capture behavior in a field setting where stated preference survey respondents perceive that their responses will be considered by the authorities.

It is also vital to mention why we make use of a quasi-cheap talk script instead of an actual cheap talk script. The primary element of a cheap talk script is to explain the phenomenon of hypothetical bias. Specifically, it states that people behave differently in hypothetical versus real choice settings, explains potential reason for the bias and provides the direction of the bias. When administered in a purely consequential, incentive compatible setting, an actual cheap talk script presents a challenge by compromising the incentive compatibility and validity of the mechanism. This primarily works through altering respondent beliefs and, in part, creating confusion about whether the survey responses have direct consequences or not. For this reason, we devise a quasi-cheap talk script that highlights existence of a positive bias in an experimental market setting as opposed to other non-experimental settings. The script states that this bias potentially arises due to a couple of reasons - "house money" effects i.e., respondents are more likely to vote in favor of a good because they view their budget constraint differently with the easily obtainable house money in the laboratory versus in a parallel realistic setting; social pressure due to the experimenter or other participants. Our quasi-cheap talk script shares the elements of a standard cheap talk script in that it discusses that there exists a bias, highlights potential reasons for the bias and asks the participants to consider voting as they would in an incentive compatible, consequential setting.[3]

Based on the more recent view of stated preference surveys, for hypothetical bias reduction techniques to be useful in the field, they should be able to correct the bias of those who view the elicitation to be hypothetical (and are thus prone to bias) while not distorting the preferences of those who view the elicitation to be consequential. While we know how these measures are expected to alter WTP in hypothetical choice settings, their effects in a consequential choice setting are unclear. Although we expect

[3]The quasi-cheap talk script does not explicitly frame the bias as "hypothetical" because it would be problematic in an incentive compatible setting. This means that we do not gain much by showing its effectiveness in a hypothetical setting as it is slightly different from the standard cheap-talk by Cummings and Taylor (1999). Therefore, we only include it in the incentive compatible treatment as discussed in the experimental design section.

the reduction approaches to not alter WTP meaningfully in a consequential setting, it is possible that these techniques distort value estimation in which case it is important to understand when and where to apply the bias reduction approaches such that they do not alter true preferences.

Our data collection is still in progress. The present findings, as follows, help throw light on the questions posed above. Importantly, comparison of WTP estimates across the baseline hypothetical treatment and an incentive compatible treatment confirms existence of a positive and significant hypothetical bias as is well documented in numerous studies and meta-analyses (Murphy et al. (2005); Loomis (2011); Penn and Hu (2019)). Further, the solemn-oath and certainty adjustment approaches reduce hypothetical bias in a purely hypothetical treatment. In contrast, the pseudo cheap-talk script does not meaningfully after WTP distributions in an incentive compatible setting but does lower variance in extreme responses. What is more interesting is that an exploratory analysis of the oath leads us to find results that are completely novel to the literature, to the best of our knowledge. In an incentive compatible treatment, we find that there is marginal evidence that the oath increases WTP. The more sophisticated empirical tests of the oath reveal that it meaningfully alters WTP in an incentive compatible setting thereby increasing demand.

Coming to the certainty adjustment approach, we find that the common technique that uses certainty thresholds to recode "yes" votes to "no" votes lowers WTP estimates thereby inducing an unintended downward bias. This is primarily because we find no differences in stated certainty levels for "yes" votes across the hypothetical and consequential treatments; there are some differences in "no" votes. This is a strikingly unique finding and is in contrast to the common belief that "yes" votes in a hypothetical treatments are less certain. A certainty adjustment approach that adjusts both "yes" and "no" votes also ends up altering WTP estimates, however, the bias is relatively smaller as compared to the common technique. Finally, using data from the post-experiment questionnaire, we also explore and test behavioral mechanisms underlying the reduction techniques.

2.2 Experimental Design

2.2.1 Valuation

This study elicits preferences for an environmental public good, in particular participants are asked to vote on proposals to fund a tree-planting project that involves planting and maintaining 160 trees in the Appalachian Mountains. As two of the treatments need to provide an actual mechanism through which participants can fund the project, we collaborate with the organization One Tree Planted. Participants are given information including broad benefits of tree planting and reforestation, and specific information about increased water storage, carbon capture and avoided nutrient runoff. One benefit of using this good is to enhance the lab-to-field generalizability given the values elicited in this study and in many field surveys designed to inform public decisions are tied to passive or indirect use.

Participants are asked to vote yes or no on a set of proposals to fund the tree-planting project. The only variation across proposals is the cost to the participant, and participants are asked to make choices simultaneously (i.e., all proposals are presented at once, rather than in a sequence). The amounts considered in the referenda are: \$0, \$1, \$2, \$3, \$4, \$5, \$6, \$8, \$10, \$12, and \$15. To eliminate strategic considerations, and to make the elicitation incentive compatible for the IC treatments, after all votes are made, only one vote per participant is randomly selected and counted. A simple majority-vote implementation rule is used to determine whether the vote passes. Prior research demonstrates that this version of a "payment card" is approximately demand revealing in an induced-value setting (Vossler and McKee, 2006), and further that elicited values in a home-grown value setting are not statistically different from those elicited using the more common approach of asking each respondent to vote at a single cost (Vossler and Zawojska, 2020).

2.2.2 Treatments

We employ a between-subjects experimental design with five distinct treatments. In the two baseline treatments, we vary the elicitation to be either incentive compatible (IC) or hypothetical (HYP). The IC treatment serves as a benchmark to measure hypothetical bias as well as compare effects of reduction techniques. Procedures are the same across IC and HYP except that the HYP treatment highlights that the choices will not have direct consequences i.e., no money will be deducted from earnings and no trees will be planted. Further, we include as treatments specific reduction techniques i.e., the "oath" and a "quasi-cheap-talk" script. The oath script is administered and kept identical across HYP+Oath and IC+Oath treatment (see appendix for the wording of the oath). The last treatment is IC+CheapTalk which includes a quasi-cheap talk script as discussed earlier.

As the certainty adjustment technique relies on re-interpreting stated preferences based on answers to a follow-up certainty question, there are no additional treatments needed to study this technique. While prior studies only elicit response certainty in hypothetical settings, we do so in an incentive compatible setting and for both "yes" and "no" votes. A certainty question is included in all treatments, and, as the question is encountered after participants are asked their preferences for a public good, it neither contaminates elicited preferences nor response certainty. Our complete experimental design is a total of seven treatments, two of which are run with the baseline HYP and IC treatments i.e., use the data from the same respondents (HYP+Cert and IC+Cert).[4]

2.2.3 Testable hypotheses and power analysis

Based on our experimental design, we are primarily interested in testing the following null hypotheses:

[4] While we note that by including a certainty question in all treatments, we can speak to interaction effects of reduction techniques, it certainly is not the primary focus in this study.

H1: WTP is equal across HYP and IC treatments.

H2: The Solemn oath has no effect on WTP in a HYP treatment.

H3: The solemn oath has no effect on WTP in an IC treatment.

H4: Certainty adjustment has no effect on WTP in a HYP treatment.

H5: Certainty adjustment has no effect on WTP in an IC treatment.

H6: The quasi-cheap talk script has no effect on WTP in an IC treatment.

We base the power considerations on a pilot session conducted with 20 participants using the HYP+Oath treatment. All procedures closely follow that of the actual experiment and therefore we include the data from the pilot session in the analysis. We use a Monte Carlo simulation approach that assumes the data will be analyzed using an interval regression model where the null hypothesis states the equality of means across treatments and is evaluated based on a chi-squared test at a 5% level of significance. Results from the HYP+Oath treatment reveal a standard deviation of 5.05. The incentive compatible treatment from Vossler and Zawojska (2020) reveal a standard deviation of 3.87. As mentioned earlier, they use a similar good and rely on similar procedures in the incentive compatible treatment, so this provides as a relevant benchmark for the standard deviation. With 80% power, our study can identify a minimum detectable effect of $1.60 when comparing across the hypothetical treatments (HYP v. HYP+Oath), $1.25 when comparing across the two incentive compatible treatments (IC v. IC+Oath and $1.50 across HYP and IC treatments.[5]

Vossler and Zawojska (2020) find that mean WTP in an incentive compatible, payment card treatment is approximately $4. While the hypothetical bias ratio may vary based on the valuation context, typically studies find WTP in a hypothetical setting is twice that of WTP in a consequential setting (Murphy et al. (2005); Loomis (2011); Penn and Hu (2019)). A rule of thumb for an ex-post calibration factor suggested by the NOAA panel is also 2. Considering these numbers, the study is

[5]For the incentive compatible treatment, by using estimates of standard deviation from Vossler and Zawojska (2020), we are able to base our power considerations on a larger sample size as opposed to just basing it on our pilot sample of 20 participants.

well-powered to detect difference between HYP and IC treatments. With respect to the effects of reduction procedures, the design is capable of detecting treatment effects in the range of 20 to 30% relative to the two baseline HYP and IC treatments. Finally, based on a 40% treatment effect identified by Penn and Hu (2019) for the cheap-talk script, we expect the quasi-cheap talk script to be able to identify a difference of \$1.6 when comparing between IC and IC+CheapTalk. Overall, the analysis suggests sample sizes of 150 per treatment (HYP, HYP+Oath, IC, IC+Oath) and 70 for the IC+CheapTalk treatment should be sufficient to identify said differences.

2.2.4 Experimental Procedures

A typical experimental session proceeds as follows. Participants are randomly assigned an ID number tied to their order of entry into the online experiment via Zoom. The experiment instructions are displayed using Zoom's screenshare feature, and the same moderator reads instructions aloud while participants follow along. In addition, the moderator follows several lab protocols mentioned in the consent form before getting the experiment started. Participants are informed that instructions contain only true information, and their decisions will be kept anonymous. All decisions are made on the participants' personal computers. The moderator encourages questions, which are asked through online chat, and exchanges are private to the inquiring participant and the moderator. The experiment is programmed and facilitated using the software z-Tree (Fischbacher, 2007) as well as z-Tree unleashed (Duch et al., 2020).

There are four stages to the experiment. In stage one, participants complete a (paid) risk elicitation task of the sort popularized by Holt and Laury (2002). Participants make a series of binary choices between playing a lottery or receiving a fixed amount of money. After the choices have been made, one of the scenarios is randomly selected to be binding and participants are paid based on the selected outcome. Following standard procedures, the outcome of this task is not revealed until the end of the session. In stage two, participants go through a voluntary team contribution game to earn money that will be available for the next stage i.e., the voting experiment. This task helps obtain homegrown values for the good while also overcoming possible "house money" effects. Participants are placed in groups of either two or three members. In each group, participants simultaneously and independently make decisions about how many of the experimental tokens to contribute to the team project and how many to keep for themselves. The game is repeated for 20 consecutive rounds and participants are randomly re-matched into groups at the start of each round. For each decision round, player's earnings are a sum of their income from the team project and the tokens they kept for themselves. After each decision round, participants receive feedback about the total team project contribution from all group members and their own payoff.

In stage three, participants go through the voting experiment where they are provided with detailed information on the tree planting project (see Appendix B). This information and other implementation procedures are kept constant across all treatments to rule out potential confounds. For treatments with the oath, the script, which is programmed in Qualtrics, and includes a signature field is shown before revealing the context of the voting experiment. It is made clear to the participants that their decision of signing the oath will be kept anonymous, is completely voluntary and does not affect the outcome or earnings from the experiment. Following the project description, participants are asked to vote on a set of proposals, presented simultaneously, that vary only in the cost of the tree-planting project. In the IC treatment with the quasi-cheap talk script, participants are shown the script prior the voting decisions. Once the voting decisions have been made, participants are presented with a certainty follow-up question immediately before revealing the voting outcome. After the certainty decisions have been submitted, participants are shown the randomly selected proposal being considered (i.e., the cost), the percentage of yes/no votes from all participants, and whether the referendum passed. For IC treatments only, if the referendum passes, participants' individual cost is subtracted from their earnings (in stage one), and the money is set aside for the tree-planting project. Procedures are in place to ensure that payment is actually sent to One Tree Planted. Further, once payment is received by One Tree Planted, participants are sent a written verification of this as well as a certificate that acknowledges that students at the university have funded 160 trees in the Appalachian Mountains.

In the final stage of the experiment, participants fill out a questionnaire to collect their socio-demographic characteristics, as well as elicit information which will be used to test for multiple behavioral mechanisms identified by prior literature.[6] Loomis (2011) points out that the underlying causes for hypothetical bias are not fully understood and apart from these often-used reduction approaches, the social psychology literature offers alternative behavioral explanations that may explain the bias. Some of these include costs of lying, guilt aversion, and happiness.

2.2.5 Participants

Twenty seven experiment sessions were conducted during spring, summer and fall of 2021 as well as spring and summer of 2022. Including the pilot, we have data for 522 participants. All sessions were conducted online and facilitated via Zoom. The experimenters made use of the designated experimental economics laboratory at The University of Tennessee to run the experiment while the participants joined via Zoom.[7] Participants were recruited from a large existing database of undergraduate students that had previously registered to receive invitations for economics experiments at University of Tennessee and the Appalachian State University. Participants were not allowed to attend more than one session of the experiment. Earnings were dominated in “lab dollars” in the second stage of the experiment and exchanged for U.S. dollars at a common announced exchange rate. For the voting stage, however, all values were

[6]The complete instructions and questionnaire are mentioned in detail in the appendix.

[7]To be able to conduct the entire experiment in the same format and minimize confounding effects we chose to do the experiment online and thus, it was in part due to the uncertainty faced during the COVID-19 pandemic.

denominated in U.S. dollars. The experiment lasted approximately 80 minutes and on average participants earned approximately $27 for the session.

2.3 Results

We begin the analysis by reporting descriptive statistics for the sample (Table 2.1). The average age of participants is approximately 21 years, 57% of participants are female, about 58% state they are currently employed (in a part- or full-time job). Sixty-one percent of the participants have previously participated in a laboratory experiment and the average GPA in the sample is 3.39.

Table 2.2 reports the estimated empirical survival functions by treatment. The percentage of "yes" votes decreases in all treatments as the bid amount increases. In the HYP treatment, the percentage of "yes" votes is lower with oath while it is slightly higher in the IC treatment with oath relative to respective controls (HYP and IC). To test the distribution of "yes" votes more formally, we use the Kolmogorov-Smirnov test where the null hypothesis tests the equality of distributions across treatments. There are considerable differences between HYP and IC votes, with deviations between 30 to 45 percentage points ($p < 0.01$) for costs between $3 to $10 (largest at $5). This suggests that WTP is lower in the IC treatment i.e., there is presence of a significant hypothetical bias. This result is not surprising as many meta-analyses report existence of hypothetical bias based on a substantial amount of lab experiments. This evidence not only helps to substantiate our other results but also highlights that the stylized results from the literature hold with respect to hypothetical bias and effectiveness of techniques in reducing the bias. Comparisons across the two hypothetical treatments (HYP and HYP+Oath) and the two incentive compatible treatments (IC and IC+Oath) do not depict statistical differences based on the Kolmogorov-Smirnov test statistic.[8]

[8]Given the differences across these two comparisons are smaller relative to the one between HYP and IC, we need some more data for the Kolmogorov-Smirnov test to be powered to test differences. Based on the estimates, we expect to find statistical differences as we collect more data.

When comparing IC treatments with and without the oath, we surprisingly see that the IC+Oath dominates the distribution of the baseline IC treatment along the entire distribution. We do note that based on a Fisher exact test of differences, the number of "yes" votes is statistically higher for IC+Oath relative to IC shows at \$4 and \$5 ($p < 0.05$) for the aggregate sample suggesting oath may drive an increase in "yes" responses in an incentive compatible setting. Results get more interesting on slicing the data by participant pool and gender. For the UTK sample, in addition to differences between IC and IC+Oath as noted above, there are significant differences between HYP and HYP+Oath as well for cost amounts between \$10-15. Slicing the data by gender, we find that for males, oath decreases "yes" responses in a HYP treatment (for costs between \$6-15) while it increases "yes" responses in an IC treatment (at \$3). On the other hand, for females, oath decreases "yes" responses in a HYP treatment (at \$8) while it increases "yes" responses in an IC treatment (for costs between \$3-5).

The largest difference between HYP and HYP+Oath occurs at \$10 (although it is about the same at \$12 and \$15) i.e., about 10% decrease in "yes" votes as a result of the oath script. This percentage is consistent with prior studies that find that the oath lowers the number of "yes" votes and WTP in hypothetical scenarios (Carlsson et al. (2013); Jacquemet et al. (2013), Jacquemet et al. (2017)). Interestingly, the largest statistically significant difference between IC and IC+Oath occurs at \$5 i.e., about 15% increase in "yes" votes as a result of the oath script.

Results from the IC v.s. IC+Oath comparison are surprising because ideally, we would expect the oath to not alter responses in an incentive compatible treatment assuming it already the mechanism already provides incentives to be truthful. Moreover, these results slightly contradict those found by Jacquemet et al. (2017) i.e., the oath does not statistically alter votes in the IC treatment. However, we do note that their design is based on a single bid amount and there exists a possibility that the oath may have differing impacts at different points in the WTP distribution. Finally, the comparison between IC and IC+Oath is rather novel when it comes to the literature and warrants further investigation as to why, if at all, the oath increases the number of "yes" votes in a consequential setting. We discuss some of the possible mechanisms in section 3.4 below.

Comparisons between IC and IC+Cheap-Talk may sound strange at first because it is really a pseudo cheap-talk script that we are analyzing here because of the issues with the actual cheap-talk script mentioned early on. Based on Table 2.2, cheap-talk seems to slightly increase percentage of "yes" votes at lower end of the cost distribution while there is an opposite effect at the higher end. Overall, the two distributions are not statistically different as per the Kolmogorov-Smirnov test statistic.

Turning to the collected data on response certainty, based on Fisher exact tests conducted separately for each cost amount, we find no significant differences in the certainty levels for "yes" votes across HYP and IC treatments for majority of the cost amounts except \$2 and \$3. We do, however, find that certainty levels of respondents who vote "no" are statistically different ($p < 0.05$) at almost all cost amounts besides \$0. Looking at the mean difference in stated certainty levels across HYP and IC treatments, we find that respondents in an incentivized setting tend to be more certain about their "no" votes relative to a hypothetical setting. This evidence contrasts what is usually conjectured in the literature i.e., respondents are expected to less certain of "yes" votes in a hypothetical setting. Certainty for "no" votes is only selectively explored in prior literature but the argument can be made that respondents should be more certain of their "no" votes in a hypothetical treatment relative to an incentive compatible treatment.

We use a pooled probit model to analyze data from the experiment and estimate WTP distributions. With a payment card elicitation format, the data is recorded as a yes/no vote for each of the stated cost amounts therefore, it can be analyzed using a limited dependent variable model. Willingness to pay (WTP) is a latent dependent variable as we cannot observe the true WTP value, instead we only observe a yes or no vote at separate cost amounts. We assume that one's valuation is a linear function of covariates such that $V_i^* = x_i\beta + u_i$, where x_i is a vector of covariates, β is a vector of unknown parameters, and u_i is a mean-zero error term. To facilitate estimation, we assume a normal distribution for the errors, with $u_i \sim Normal(0, \sigma_i^2)$. This gives rise to the log-likelihood function:

$$\mathcal{L} = \sum_{i=1}^{N} \Big\{ y_i \ln \Phi(x_i\beta) + (1 - y_i) \ln(1 - \Phi(x_i\beta)) \Big\}, \tag{2.1}$$

where Φ denotes the normal CDF. As illustrated by Haab et al. (1999), conclusions drawn from hypothetical bias experiments can be sensitive to the assumption of a common error variance across treatments. We therefore will allow the error variance to differ across treatments by specifying a standard deviation function $\sigma_i = z_i\gamma$, where the z_i denote treatment-specific indicators.

Table 2.3 presents results from the analysis. In Specification (1), we include the treatment indicators and obtain estimated treatment effects. The base category for all regression tables is the IC treatment i.e., the model intercept is the estimate of WTP for the IC treatment and all coefficients are denoted as differences in WTP estimates and may be interpreted relative to the IC baseline. This specification also allows for the standard deviation to vary by treatment. Specification (2) adds the individual-level controls to Specification (1).

2.3.1 WTP distribution and the hypothetical bias

In this sub-section, we examine the extent of hypothetical bias. Based on Specification (1) in Table 2.3, WTP in the HYP treatment is \$5.6 higher relative to the IC treatment. This means that there exists a positive hypothetical bias such that WTP in the HYP treatment is statistically higher than that of the IC treatment. The estimated WTP amounts by treatment are as follows: \$4.17 (IC), \$4.64 (IC+Cheap-Talk), \$5.23 (IC+Oath), \$8.63 (HYP+Oath) and \$9.8 (HYP). Hypothetical bias, which is simply a ratio of WTP in the HYP to that in the IC treatment, is approximately 2.35. This number does not change significantly when the effects of controls are included. This suggests that If the nature of the survey is hypothetical, then WTP is more than double as compared to estimates in a setting with binding consequences. Prior literature and various meta-analyses find that the ratio generally lies between the range of 1.3 to 3 which somewhat validates our preliminary findings.[9] Although we use a conservative estimate of 1.3 while computing expected sample sizes from the pilot data, a higher ratio increases the likelihood that our analysis will be well powered across various between-group comparisons.

2.3.2 Solemn-oath and willingness to pay

Results from Table 2.3 show that when comparing the WTP estimates across HYP and HYP+Oath, the "oath" technique does not statistically lower WTP for the aggregate sample. The results are robust to inclusion of controls.[10,11] Participant characteristics include age, money earned, GPA, gender, earnings as well as indicators for comprehension of experiment instructions, employment, and prior experiment experience. We find similar results from an interval regression model such that the oath is ineffective in altering probability of "yes" votes in a hypothetical treatment. However, the magnitude of differences in WTP with and without the oath are large and have been fairly consistent. Over the course of data collection, results have changed with respect to the oath but given large magnitudes, it more likely than not that we will find a statistically significant effect of the oath reducing WTP. It has been shown that the oath help reduce WTP in a hypothetical scenario based on the findings from prior literature (Jacquemet et al. (2011), Jacquemet et al. (2013),

[9]Note that hypothetical bias ratio may vary based on the valuation context as well as definition of variables used in meta-analyses, however, a rule of thumb as per the NOAA panel is considered as a ratio of about 2 which lies within the extreme bounds as suggested by prior literature.

[10]Based on Specification (1), additional test statistics by treatment are as follows: IC versus IC+Oath (chi2=2.36, p=0.12); HYP versus HYP+Oath (chi2=1.81, p=0.17) and IC versus HYP (chi2=28.72, p=0.00).

[11]Note that the observations are slightly lower when control variables are included in Table 2.3 due to a technical error in the software that led to loss of the questionnaire data.

Jacquemet et al. (2017)). The expected results of our study are very likely to be similar to what some of the prior lab experiments find in a HYP setting.

The effects of the oath in a consequential scenario are largely unexplored when it comes to the non-market valuation literature. We use a chi-squared test of differences to compare WTP in the IC treatment (with and without oath). In contrast to the results in a HYP treatment and consistent with Fisher exact test results reported earlier the oath marginally ($p < 0.10$) increases WTP in the incentive compatible elicitation setting by about a dollar and 6 cents i.e., 25% increase (see Specification (1) in Table 2.3). Jacquemet et al. (2017) find somewhat contrasting results in an incentive compatible setting, however, as mentioned earlier, their elicitation method relies on a single bid amount. Certainly, for some cost amounts we find similar results as Jacquemet et al. (2017), but, we also find evidence of marginal positive effects at some of the cost amounts ($4 and $5). Our study adds to theirs by testing effects of the oath on the WTP distribution in an incentive compatible, consequential setting.

Assuming this result holds up or even intensifies after additional data collection, it suggests caution in using this approach in a field survey setting where there is presumably a mix of respondents with and without consequentiality beliefs. This result, however, is not robust to inclusion of controls, but this in part can be explained by the fact that, due to the technical issue mentioned above, fewer observations are available from which to measure WTP for the IC+Oath treatment.

Summarizing our results for the oath technique, it helps reduce hypothetical bias and lowers WTP in a hypothetical setting. As such the oath serves as a promising ex-ante technique in reducing the degree of hypothetical bias. The mechanism through which the oath operates is honesty and may be explained by commitment theory. It posits that an individual is more likely to reveal the truth if they have made a prior commitment of honesty (Joule and Beauvois, 1998). On the other hand, it may also have unintended effects on WTP in a consequential setting and therefore casts doubt on its usefulness in field settings where a subset of respondents may believe the survey is consequential.

2.3.3 Pseudo Cheap-talk and willingness to pay

This is the first experiment to test the effects of a Cheap-talk script *look-alike* in an incentivized setting. Based on the script's observed directional impacts in a hypothetical setting, it is natural to expect that the script should lower WTP responses. Results from Table 2.3 show that when comparing the WTP estimates across IC and IC+Cheap-Talk, the "cheap-talk" script does not statistically lower WTP. Instead, there is a slight increase in WTP (about 47 cents) but this difference is statistically insignificant. These results do not change with inclusion of controls. However, there is a decrease in the standard deviation of the estimated WTP distribution across IC and IC+Cheap-Talk. This evidence suggests that in an incentivized setting, Cheap-talk reduces the more extreme "yes" responses which generally do contribute to hypothetical bias in a hypothetical treatment. So, based on the preliminary data, while the pseudo cheap-talk script does not meaningfully after WTP distributions, it does lead to lower noise in an incentive compatible setting. Further, such a script could also alter a respondent's belief that the survey may not be consequential which in turn alters the WTP estimates and this could be an unintended consequence of using such scripts in field settings.

2.3.4 Certainty adjustment and willingness to pay

Utilizing data from the certainty adjustment, we start by testing for differences between stated certainty levels for "yes" and "no" votes separately. Table 2.4 shows the differences between stated certainty levels across IC and HYP based on fischer's exact test (top panel). The bottom panel shows point differences in the mean certainty response between IC and HYP. Findings show that there are only a few differences in stated certainty levels pertaining to "yes" votes across HYP and IC while there are a lot of differences in stated certainty levels for "no" votes. Moreover, the findings show that participants are less certain for "yes" votes at low cost levels in an IC setting relative to HYP, although these differences are not significant suggesting similar certainty levels for "yes" votes. The opposite is true for "no" votes and these differences are statistically significant at cost $\geq$ \$4. The immediate consequence of this result is that if "yes" votes do not differ in certainty across HYP and IC treatments, then using an *ex-post* certainty adjustment method will reduce "yes" responses in an IC treatment and induce a downward bias in WTP.

Next, we analyze the effects of "correcting" WTP in the HYP treatment using various certainty thresholds on the estimated WTP. The common method from prior literature is to simply recode "yes" responses to "no" based on a certainty cut-off which may be context dependent (Morrison and Brown, 2009). That said, we use multiple thresholds to be able to find the one that minimizes the extent of hypothetical bias specific to our context.

Below we present results from the analysis of the certainty adjustment technique using multiple threshold levels estimated with a pooled probit regression (Table 2.5). Responses in the HYP treatment are adjusted for certainty levels i.e., denotes estimates for HYP+Cert treatment. In Table 2.5, we report the WTP estimates pertaining to four different certainty thresholds from four separate pooled probit regressions. In addition, we report the gap between the adjusted WTP i.e., HYP+Cert and the benchmark IC treatment (\$4.17; from specification (1) in Table 2.3). This gap measures a remainder of the hypothetical bias after adjusting for uncertainty in responses. Covariates are excluded from the regressions, although this has no important effect on the findings. The first column in the top panel uses a lower (i.e., weaker) threshold of 7 i.e., all "yes" votes for which respondents state a certainty level strictly less than 7 (Cert$\geq$7) are recoded as "no". The next three columns follow the same method with certainty thresholds of 8, 9 and 10.

We find that as the threshold increases i.e., as we move from left to right across in Table 2.5, the WTP decreases suggesting that the certainty adjustment technique helps reduce hypothetical bias but there are still significant differences between the adjusted WTP and the benchmark IC for thresholds of 7, 8 and 9. Increasing the threshold to 10 completely eliminates hypothetical bias. Most of the participants report certainty levels as high as 9 or 10, therefore it is not surprising that a certainty threshold of 10 makes WTP from adjusted HYP and IC comparable relative to lower thresholds. The bottom panel of Table 2.5 presents certainty-adjusted WTP estimates using the IC treatment, and the same cutoff rules as above. As a point of comparison, reported are differences between the adjusted and unadjusted estimates. Of interest is that all recoding schemes reported in the table result in lower, and statistically different WTP estimates. This type of recoding procedure is designed to lower WTP and so it is not surprising that as thresholds increase from 7 through 10, the absolute value of the WTP gap between the adjusted IC and the benchmark IC treatment increases. Therefore utilizing the common certainty recoding technique, this method decreases WTP in a consequential setting by a considerably large magnitude (\$2.08 with Cert$\geq$10).

We further consider two adjustment rules that work by translating the certainty levels into probabilities, which is arguably less ad hoc. In a recent work by Penn and Hu, they refer to the two approaches we consider as the "asymmetric uncertainty model" (ASUM) and the "symmetric uncertainty model" (SUM).[12] For the ASUM, "yes" votes are translated into probabilities using the formula:

$$\Pr(\text{yes} \mid \text{vote=“yes”}) = \frac{4}{9} + \frac{5}{90} * \text{certainty level}$$

which converts a "yes" vote to a probability between 50% and 100%. With this conversion, someone who votes "yes" but indicates the lowest level of certainty (a "1") is interpreted as being indifferent between "yes" and "no". At the other extreme, someone who votes "yes" and indicates the highest level of certainty ("10") is assumed to have a probability of 100%. With the ASUM, "no" votes are unaltered.

[12] Please note that this recent work by Penn and Hu is currently under review for publication. Given that this work is not available in the public domain, in lieu of a working paper, we obtained permission from the authors to use their unpublished manuscript in order to give due credit. Once published, we will make sure to provide a proper reference to this article.

For the SUM, both "yes" and "no" votes are given a probabilistic interpretation. The "yes" votes are converted using the formula for ASUM, whereas the "no" votes are converted to probabilities using the formula below:

$$\Pr(\text{yes} \mid \text{vote=``no''}) = \frac{5}{9} - \frac{5}{90} * \text{certainty level}$$

which converts a "no" vote into a probability of voting yes that is between 0% (certainty level of 10) and 50% (certainty level of 1). As we asked participants to indicate a certainty level (10-point scale) for each of the eleven referenda they voted on, we convert the data into a panel and analyze the set of voting choices using a pooled probit model. We account for correlation in within-person responses using cluster-robust standard errors. By including the cost of the referendum as a covariate in the model, we can readily recover WTP estimates (Cameron and James, 1987). For both specifications, estimation is carried out with a pooled fractional probit model.

Table 2.6 presents results based on ASUM and SUM applied separately to the HYP and IC treatment data. The SUM adjustment procedure does very little to WTP for either the HYP or IC data. This is likely attributable to the fact that most "yes" and "no" voters expressed a high level of certainty, with no meaningful asymmetries in certainty levels across "yes" and "no" votes. While ASUM noticeably lowers both WTP estimates, hypothetical bias is still substantial. The adjustment when applied to IC data also shows statistically lower WTP estimates for both SUM and ASUM. Summarizing the results from certainty adjustment and assuming these results will generalize, a conclusion that could be drawn from this exercise is that both SUM and ASUM potentially help reduce hypothetical bias in a HYP treatment, however, SUM relatively induces a smaller bias in an IC treatment. It is surprising, however, that SUM actually increases WTP in an IC treatment as opposed to ASUM but the effect is small.

2.3.5 Exploratory analysis

A deeper dive into the analysis of "oath" reveal interesting results. There is evidence that effectiveness of the oath may depend on factors such as cultural influences (Carlsson et al., 2013). Given that laboratory exploration of the oath is relatively limited as compared to other reduction methods, it makes sense to examine any other relevant factors that may be driving differences in how the oath affects WTP. The most obvious starting point relevant to our study is to slice the data by the participant pool as well as gender. Using the same econometric specification (as in Table 2.3), Table B.1 (in Appendix B) presents the results by allowing the treatment effect to vary by participant pool and by gender in Specification (3) and (4) respectively.

As highlighted in the beginning of the results section, we do find differences in how the oath affects WTP in the two samples (at UTK and ASU). Restricting the analysis to participants from UTK, we find strong evidence that the "oath" reduces WTP in a HYP treatment as anticipated. There are also interesting gender effects of the oath i.e., we find that while the "oath" decreases WTP for males, it has a null effect for females in a HYP setting. Looking at the comparison between IC and IC+Oath in Table B.1 (specification 4), it seems that the oath does increase WTP by a large amount - \$1.5 for females but has a null effect for males. These results suggest that the oath increases demand in an incentive compatible setting for females. Although still speculatory in nature, it seems that there might be a shift in preferences arising out of the "oath" script. Many studies have documented gender effects to a large extent and shown that it is more likely that females display pro-environmental preferences as compared to males (Hunter et al. (2004); Torgler et al. (2008)). Based on this evidence, it may prove to be important for future studies of the "oath" to control for potential gender effects when computing value estimates especially in a field setting. There is no evidence of any heterogeneous effects of the pseudo Cheap-talk script on WTP.

2.3.6 Insights from the post-experiment questionnaire

The mechanisms underlying the oath technique largely include commitment to "honesty" as explored by Jacquemet et al. (2017). The honesty explanation relies on the assumption that participants are more likely to tell the truth once a promise has been made or in this case, the oath has been voluntarily signed. From the analysis of the post-experiment questionnaire, we find that self-reported honesty levels are statistically identical across treatments with and without the oath and therefore does not serve lend support to honesty as a behavioral mechanism for the present sample. Although, we do find some differences in the self-reported happiness levels across treatments i.e., participants report being more happy in IC relative to IC+Oath and in HYP relative to HYP+Oath.

The broader literature on truth-telling suggests that the oath may be associated with other competing motivations such as lying costs, guilt aversion or for that matter even cognitive effort that may affect people's willingness to tell the truth. With respective to cognitive effort or attention so to speak, Jacquemet et al. (2017) analyze the time taken in individual responses when the oath is used. They conclude that the oath in no way increases the salience of the referendum or the good in question based on the finding that participants spend a similar amount of time responding to the referendum regardless of whether the oath technique was used. In our study, we use the post-experimental questionnaire to gain insight about whether an individual is willing to lie when there is more to gain from the lie as well as their preferences about guilt aversion. We do so by including measures of guilt aversion and the respondent's willingness to lie. Based on the responses, we find that participants are statistically less prone to lying and more prone to being guilt averse with the oath than without. This difference is only statistically significant for the participants in an incentive compatible treatment. This provides suggestive evidence that one or more mechanisms besides honesty may be in play when the oath script is administered and could be driving differences in WTP observed across the IC and IC+Oath treatments.

As people are more likely to be guilt averse after the oath is administered, it is possible that this "guilt" is tied to selfish behavior, in particular not being supportive (voting "yes") for the tree-planting project. Funding tree plantings is a publicly spirited thing to do, and the oath may nudge some to behave inline with their "true" or virtuous self or otherwise behave pro-socially to avoid a higher disutility from guilt. A recent lab experiment demonstrates that participants give considerably more in a standard voluntary contributions mechanism (VCM) game after the solemn oath is administered (Hergueux et al., 2022). In that paper, the authors argue that the oath provides "the intrinsic motivation necessary for players to behave according to their underlying social preferences". An additional result supports the plausible nudge theory behind the oath. When analyzing data from the personality questions, we construct a measure of *agreeableness* and find that in general participants state they are more agreeable in the IC+Oath treatment.

With respect to the certainty adjustment technique, recall that the underlying mechanism is straightforward and largely stems from "respondent uncertainty". If respondents are more uncertain in a hypothetical setting, by using that information to adjust their responses from a "yes" vote to a "no" vote, we can attempt to lower the extent of the hypothetical bias. However, if this is not true to begin with, then applying such a method in the field where there some participants may believe the survey to be *consequential* will result in a downward bias in WTP estimates for that sub-population. Based on our data, participants state statistically identical certainty levels across hypothetical and incentive compatible treatments and therefore, it is clear that using this method to recode the positive responses in such cases may be counterproductive.

2.4 Conclusion

Researchers often use techniques designed to reduce hypothetical bias and such techniques are commonly used in field survey research designed to inform public decision-making. The effectiveness of these procedures has been established through lab experiments, which allow for comparisons between hypothetical (with reduction procedures) and incentive compatible settings. Crucial to the reduction procedures and their effects may be recent evidence that most respondents to field studies perceive their choices will influence a policy decision, and as such a "hypothetical choice" scenario may be a poor characterization of the field. In turn, our research poses a question of whether these techniques still remain relevant in the face of new evidence and whether *ex-ante* and *ex-post* techniques widely used in the field have "unintended" effects on value estimates.

We find that the solemn-oath helps reduce hypothetical bias, and based on the preliminary data, there is evidence to suggest that it does meaningfully alter WTP in an incentive compatible setting. While the literature has studied the "oath" in a hypothetical setting using well-designed controlled laboratory experiments, it is yet to be widely tested with different student populations. Moreover, comparisons in an incentive compatible setting are rather novel to the literature and based on observed heterogeneous effects of the oath, it could potentially suggest underlying differences in how individual characteristics may be one of the factors affecting how the oath is perceived.

Of interest is also the evidence on mechanisms underlying the oath. While we find the oath to help lower the number of "yes" votes in a hypothetical referendum if the sample is restricted to UTK or males, the explanation does not seem to be honesty. We find that there are no significant differences in honesty levels across treatments with and without the oath and so as per present data, honesty does not explain why the oath may reduce WTP. Exploration of other mechanisms reveal that participants are less prone to lying and more prone to being guilt averse with the oath in an incentive compatible referendum. There is marginal evidence that the oath influences WTP upward. If we were to speculate, it seems there may be other mechanisms at work with the oath than previously explored and those warrant further investigation. We do suspect that the oath may influence preferences and therefore the WTP but refrain from making any inferences until more of the data is collected. In contrast to the results of the oath, the pseudo cheap-talk script does not meaningfully after WTP distributions in an incentive compatible setting, however it does alter variance of the distribution.

In the case of certainty adjustment, the general perception is that those who indicate they are uncertain about their yes vote would have instead voted no in a consequential setting. The common procedure is to then recode uncertain yes votes as no. Of course, in a consequential setting, people can still be uncertain, and the common method of certainty adjustment would then bias WTP estimates downward. To move this literature further, we assess the merits of different recoding procedures that account for uncertainty in both yes and no votes. First, based on the common technique, we find that participants in our study state high certainty levels for "yes" votes and therefore, certainty thresholds of 9 and 10 lead to identical WTP distributions between the hypothetical and the benchmark incentive compatible referendum. Interestingly, we find that there are no differences in participants' stated certainty levels for "yes" votes across the hypothetical and incentive compatible treatments. This means that the adjusted "yes" responses based on stated certainty bias WTP downward. Based on an alternate technique, we find that adjusting both "yes" and "no" votes and instead converting data into probabilities helps lower hypothetical bias but it can explicitly alter WTP in an incentive compatible setting.

Considering the evidence, it makes sense to talk through some of the implications of the results. The traditional view on stated preference surveys suggests that respondents view such surveys as purely *hypothetical* and therefore their responses are prone to hypothetical bias. To reduce hypothetical bias, popular *ex-ante* and *ex-post* reduction techniques have been employed by researchers in the field. A more recent take on stated preference surveys suggests that most respondents may view the survey as *consequential* and so their WTP is likely to stem from a belief about consequentiality. Evidence from our study suggests that blindly applying reduction procedures to *consequential* settings may end up unintentionally distorting value estimates and doing more harm than good. Our results help identify a need for meaningfully segregating participant population and placing them into bins based on their *consequentiality beliefs*. Further, recent field evidence supports the notion that controlling for consequentiality enhances external validity (Vossler et al. (2012); Vossler and Watson (2013)).

Lloyd-Smith et al. (2019) first proposed the possibility of asking about consequentiality beliefs beforehand and find that doing so increases the fraction holding this belief. A problem with this approach may be that it could perhaps comprise the incentive compatibility of the mechanism itself and so it is important to ensure that the consequentiality beliefs are elicited in a rigorous manner. For instance, one could think of designing a sub-section of the post-experimental questionnaire devoted to beliefs by asking a set of screening questions and better identifying respondents who view the survey as hypothetical and those who view it as consequential. Based on this information, researchers are better able to apply *ex-ante* or *ex-post* reduction methods only to the sub-sample who view the survey as hypothetical.

Bibliography B

Aadland, D. and A. J. Caplan (2006). Cheap talk reconsidered: New evidence from cvm. *Journal of Economic Behavior & Organization 60*(4), 562–578.

Barrage, L. and M. S. Lee (2010). A penny for your thoughts: Inducing truth-telling in stated preference elicitation. *Economics letters 106*(2), 140–142.

Cameron, T. A. and M. D. James (1987). Efficient estimation methods for" closed-ended" contingent valuation surveys. *The review of economics and statistics*, 269–276.

Carlsson, F., M. Kataria, A. Krupnick, E. Lampi, Å. Löfgren, P. Qin, and T. Sterner (2013). The truth, the whole truth, and nothing but the truth—a multiple country test of an oath script. *Journal of Economic Behavior & Organization 89*, 105–121.

Carson, R. T. and T. Groves (2007). Incentive and informational properties of preference questions. *Environmental and resource economics 37*(1), 181–210.

Cummings, R. G. and L. O. Taylor (1999). Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American economic review 89*(3), 649–665.

Duch, M. L., M. R. Grossmann, and T. Lauer (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance 28*, 100400.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Haab, T. C., J.-C. Huang, and J. C. Whitehead (1999). Are hypothetical referenda incentive compatible? a comment. *Journal of Political Economy 107*(1), 186–196.

Hergueux, J., N. Jacquemet, S. Luchini, and J. F. Shogren (2022). Leveraging the honor code: Public goods contributions under oath. *Environmental and Resource Economics*, 1–26.

Herriges, J., C. Kling, C.-C. Liu, and J. Tobias (2010). What are the consequences of consequentiality? *Journal of environmental economics and management 59*(1), 67–81.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Hunter, L. M., A. Hatch, and A. Johnson (2004). Cross-national gender variation in environmental behaviors. *Social science quarterly 85*(3), 677–694.

Jacquemet, N., A. James, S. Luchini, and J. F. Shogren (2017). Referenda under oath. *Environmental and resource economics 67*(3), 479–504.

Jacquemet, N., A. G. James, S. Luchini, and J. F. Shogren (2011). Social psychology and environmental economics: a new look at ex ante corrections of biased preference evaluation. *Environmental and resource economics 48*(3), 413–433.

Jacquemet, N., R.-V. Joule, S. Luchini, and J. F. Shogren (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management 65*(1), 110–132.

Jacquemet, N., S. Luchini, A. Malezieux, and J. F. Shogren (2020). Who'll stop lying under oath? empirical evidence from tax evasion games. *European Economic Review 124*, 103369.

Johnston, R. J., K. J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T. A. Cameron, W. M. Hanemann, N. Hanley, M. Ryan, R. Scarpa, et al. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists 4*(2), 319–405.

Joule, R.-V. and J.-L. Beauvois (1998). Se soumettre... en toute liberté. *Sciences humaines* (86), 38–41.

Little, J. and R. Berrens (2004). Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis. *Economics Bulletin 3*(6), 1–13.

Lloyd-Smith, P., W. Adamowicz, and D. Dupont (2019). Incorporating stated consequentiality questions in stated preference research. *Land Economics 95*(3), 293–306.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys 25*(2), 363–370.

Loomis, J. B. (2014). 2013 waea keynote address: Strategies for overcoming hypothetical bias in stated preference surveys. *Journal of Agricultural and Resource Economics*, 34–46.

Morrison, M. and T. C. Brown (2009). Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. *Environmental and Resource Economics 44*(3), 307–326.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics 30*(3), 313–325.

Murphy, J. J., T. Stevens, and D. Weatherhead (2005). Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism? *Environmental and Resource economics 30*(3), 327–343.

Penn, J. and W. Hu (2019). Cheap talk efficacy under potential and actual hypothetical bias: A meta-analysis. *Journal of Environmental Economics and Management 96*, 22–35.

Torgler, B., M. A. Garcia-Valiñas, and A. Macintyre (2008). Differences in preferences towards the environment: The impact of a gender, age and parental effect.

Vossler, C. A., S. Bergeron, M. Doyon, D. Rondeau, et al. (2020). Revisiting the gap between the willingness-to-pay and willingness-to-accept for public goods. Technical report, CIRANO.

Vossler, C. A., M. Doyon, and D. Rondeau (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics 4*(4), 145–71.

Vossler, C. A. and M. F. Evans (2009). Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management 58*(3), 338–345.

Vossler, C. A. and M. McKee (2006). Induced-value tests of contingent valuation elicitation mechanisms. *Environmental and Resource Economics 35*(2), 137–168.

Vossler, C. A. and S. B. Watson (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization 86*, 137–147.

Vossler, C. A. and E. Zawojska (2020). Behavioral drivers or economic incentives? toward a better understanding of elicitation effects in stated preference studies. *Journal of the Association of Environmental and Resource Economists 7*(2), 279–303.

Zawojska, E., A. Bartczak, and M. Czajkowski (2019). Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management 93*, 63–84.

3

Incentives, goals and task complexity: Studying the effects of non-monetary incentives on team performance

Abstract

This study examines the effects of using non-binding, exogenous team goals on worker effort in a weakest-link team production game. The experimental design varies the team goal (and whether a goal is present) and task complexity level (simple or complex), lending itself to identify a causal effect of complexity on goal effectiveness. Further, the design also varies goal difficulty (easy, moderate and difficult). Preliminary findings suggest that using team goals can alter production, but relationships between goal difficulty and production are not monotonic. While an easier goal reduces individual production, a rather challenging, difficult goal has no impact relative to no goal. At the same time, only a difficult goal seems to improve team production relative to no goal. A non-binding goal also helps minimize wasted performance thereby enhancing within-group coordination. There is evidence that easy goals may discourage individual production but at the same time this is not true for team production, i.e., effects differ when it comes to the weakest-link worker. Interestingly, when complexity increases i.e., higher cognitive costs are placed on individuals, the magnitude by which difficult goals increase team production is relatively smaller as goal difficulty increases. Further, as task complexity increases, while physical effort decreases, cognitive effort increases. Outcomes from the study are expected to highlight the types of goals managers should set based on the amount of cognitive load a task places on individuals in a team, and therefore this research has important managerial implications.

Keywords: complexity; non-binding goals; cognitive effort; team production; weakest-link; non-monetary incentives; real-effort task

3.1 Introduction

In team production settings, problems of performance management and coordination failure among employees often necessitate the use of managerial interventions (Zehnder et al., 2017). Economic theory has traditionally focused on the use of monetary rewards to incentivize employees. However, evidence from behavioral economics (Frey and Jegen, 2001) and managerial economics (Gómez-Miñambres, 2012; Corgnet et al., 2015) suggests that providing non-monetary incentives, such as through a non-binding, wage irrelevant goal (e.g., a production or sales expectation), can also help foster performance. A non-binding goal is an attractive mechanism for managers as goals are presumably costless, and managers may not have direct control over monetary resources to help incentivize effort.

Examples of non-binding goals are evident in workplaces; a sales manager may suggest a recommended sales target or an earlier deadline for completion of other workplace tasks. Such non-binding goals may be thought of as behavioral "nudges". There is a large literature validating the use of behavioral nudges in a variety of contexts, e.g., nutrition and health (Samek, 2019; Vecchio and Cavallo, 2019), tax compliance (Fonseca and Grimshaw, 2017), and workplaces (Bulte et al., 2020; Wu and Paluck, 2021) among others.

An important and related issue is how to motivate team performance when tasks are complex, i.e., are objectively cognitively challenging. Based on an extensive review of the leadership literature in economics, Zehnder et al. (2017) highlight that as tasks become more complex, performance decreases and so does the effectiveness of monetary incentives. While studying task complexity in team settings is novel, predictions from studies focused on individual rather than group incentive mechanisms suggest that an increase in complexity leads to a decrease in the effect of monetary incentives on performance, with the exception of individuals that have high skill and a strong belief that they can accomplish the task (Bonner and Sprinkle, 2002). Laboratory evidence from accounting research supports this prediction and indicates that the probability of monetary incentives affecting an individual's performance positively decreases as task complexity increases (Bonner and Sprinkle, 2002). For this reason, Zehnder et al. (2017) recommend that researchers explore the use of both monetary and non-monetary incentives in complex task settings.

This paper studies task complexity in a strategic team setting and investigates the effects of non-binding goals on team performance. To the best of my knowledege, the only study that analyzes non-binding goals in a team setting is that of Fan and Gómez-Miñambres (2020). Using a laboratory experiment, they find that non-binding goals are effective at not only increasing performance but also in improving team coordination by reducing wasted performance. In their study, Fan and Gómez-Miñambres (2020) allow a subject who acts as a manager in order to assign a non-binding goal. While in theory the goal is allowed to differ in its difficulty level, in practice it turns out that about 50% of the time, managers set unreasonable goals i.e., goals that are too challenging for the weakest-link member. Further, what remains unclear is how task complexity interacts with non-binding goals. The experimental design in the present study, however, explicitly varies the goal which provides a relatively cleaner way of identifying effects of goal difficulty on performance respectively for simple and complex tasks.

The contributions of this study to the literature are as follows. First, I study task complexity in a strategic team production game and empirically explore a causal relationship between complexity and team performance.[1] Second, in a setting with monetary rewards present, I study the effects of introducing and varying a non-binding team goal. The specific behavioral mechanisms by which non-monetary incentives interact with task complexity and team performance are largely unexplored in the economics literature. To that end, a third contribution is to help identify behavioral mechanisms with the aid of a theoretical framework.

[1]It is important to mention the type of complexity that this paper speaks to. Campbell (1999) identifies several sources from which objective complexity may arise. In this paper, I specifically study complexity arising out of uncertainty, high information and cognitive load and unknown consequences of action.

Preliminary findings suggest that using team goals can alter production, but relationships between goal difficulty and production are not monotonic. While an easier goal reduces individual production, a rather challenging, difficult goal has no impact relative to no goal. At the same time, only a difficult goal seems to improve team production relative to no goal. A non-binding goal also helps minimize wasted performance thereby enhancing within-group coordination. There is evidence that easy goals may discourage individual production but at the same time this is not true for team production, i.e., effects differ when it comes to the weakest-link worker. Interestingly, when complexity increases i.e., higher cognitive costs are placed on individuals, the magnitude by which difficult goals increase team production is relatively smaller as goal difficulty increases. Further, as task complexity increases, while physical effort decreases, cognitive effort increases. Data collection is still in progress and these findings should be considered preliminary. Nevertheless, I expect the qualitative findings to persist with an increase in sample size.

3.2 Related Literature

Prior research has extensively studied the effects of group incentives as well as the effects of task complexity on team performance, often in isolation from each other. Some key takeaways from the various strands of literature are as follows. First, as task complexity increases, such that a higher cognitive load is placed on individuals, the probability of success on tasks and task performance itself tends to decrease (Campbell, 1988; Zehnder et al., 2017). While monetary rewards have been explored in order to incentivize performance on complex tasks, studies find that the probability that monetary rewards directly improve performance decreases as tasks become more complex (Bonner and Sprinkle, 2002; Zehnder et al., 2017). Considering this evidence, the literature encourages exploration of monetary as well as non-monetary incentives such as goals, in tandem (Zehnder et al., 2017; Locke and Latham, 2019). The intuition behind this is simple. There is a less than direct relationship between effort and performance in a complex task relative to one that may be cognitively less challenging. Motivated by this guidance, I attempt to explore the interactions between monetary and non-monetary incentives in a complex team production game.

Second, non-binding goals have been shown to improve team performance as well as coordination (Fan and Gómez-Miñambres, 2020). For decades, the literature on goal-setting has largely been concentrated to the study of individuals, however, recent studies have investigated goal setting in teams. Further, studies on goal-setting primarily span the management, leadership and empirical psychology literature; however, recently goals have also explored in the economic literature (Gómez-Miñambres, 2012; Corgnet et al., 2015). The literature does not directly explore the effects of non-binding goals on team performance when tasks are complex. However, research focused on individual decision-making settings finds that while goals are effective at increasing performance, their effectiveness depends on task complexity (Wood et al., 1987). In a team setting, it is natural to expect that an incentive combination of non-binding goals and monetary incentives may be effective in improving performance. Fan and Gómez-Miñambres (2020) study a weak-link team production game and use theory and experiments to explore the effects of non-binding team goals on performance. The present study extends this study in two important ways. First, I use a complex task, and vary the complexity level. Second, Fan and Gómez-Miñambres (2020) assign a manager to each team who sets a goal at the start of each decision round. This potentially clouds identification, and fundamentally changes the game into one where both the manager can influence workers (through the goal choice) and workers can influence the manager (through their behavior). Instead, I exogenously impose goals that are objectively easy, moderately challenging, or difficult for most participants.

Experiments on goal-setting in teams do not vary task complexity and therefore the question of how effective such goals are on performance in complex tasks remains open. One exception to this is the study by Nahrgang et al. (2013). The authors focus on different types of binding goals, specifically, learning and performance goals and analyze their impact on team performance while varying the level of task complexity. Given that their design is focused on testing "binding" goals without varying goal difficulty, the present study differs with respect to theirs in important dimensions. One, the setting I study is a weak-link team production game. The types of goals I study are non-binding and therefore serve as a non-monetary incentive. Also, the present study distinguishes goals with respect to their difficulty level rather than their content as in Nahrgang et al. (2013). Finally, I study a combination of monetary and non-monetary incentives (goals) which may have a differential effect on team performance as complexity level of the task changes.

Studies at the individual-level that explicitly vary goal difficulty find that goals are more effective when they are specific as opposed to vague and difficult as opposed to easy (Locke and Latham, 1990). As goals become more difficult, they become effective motivators, however, there are some exceptions to this finding. As goal difficulty increases, there may be some ambiguous effects on performance depending on task complexity, an individual's self-efficacy (their belief of reaching the goal), and goal commitment (Latham et al., 2002). What is unclear from the literature so far is the effect of goal difficulty in conjunction with task complexity on team performance. This is also where the present study's design can contribute to the literature by identifying effects of goal difficulty on team performance. Further, the idea here is to add a dimension of task complexity thereby contributing to the literature studying task complexity and its interplay with incentives.

The literature studying task complexity has separately explored the impact of monetary incentives and goals on team performance using real-effort tasks (Allison et al., 1993; Fan and Gruenfeld, 1998). Findings suggest that group incentives may be more effective with complex tasks and require team members to coordinate with each other. Overall, the literature is inconclusive about how the effectiveness of incentive schemes is altered by changes in task complexity primarily because either task complexity is not explicitly discussed in the study or not varied in the experimental design (Fan and Gruenfeld, 1998; Allison et al., 1993). Prior experiments only study task type in isolation, some of which are low-powered comparisons as well as do not directly vary the complexity level of the task within the experimental design (van Vijfeijken et al., 2002). Therefore, it is hard to make a clean comparison across studies of how changes in complexity of the task affect the incentive schemes and their effectiveness.

The idea of combining monetary and non-monetary incentives has been supported by prior research. The importance of wage-irrelevant goals has also been shown in principal-agent models (Corgnet et al., 2018). However, research on the interaction of monetary and non-monetary incentives is limited and more so when it comes to complex task environments. van Vijfeijken et al. (2002) propose that a combination of the two performance management methods i.e., incentives and goals may enhance performance depending on the task characteristics.[2] Brandts and Cooper (2007) analyze the effects of financial incentives and communication on coordination in a weakest-link game. While the study does not have an element of task complexity, it is useful to point out that combinations of monetary and non-monetary incentives have been explored when studying team performance. Participants assigned the role of manager choose the bonus rate for their assigned team, and the communication allowed varies across treatments (no communication, one-way communication, and two-way communication). The overarching result is that effective communication between managers and employees about benefits of high effort is a much more effective tool than increasing financial incentives (Brandts and Cooper, 2007).

Finally, an important goal of the present study is to highlight the underlying behavioral mechanisms through which team goals operate. In its simplest sense, a team goal acts as a coordination device. Social psychology, however, has suggested the importance of investigating alternative mechanisms tied to effectiveness of goals. The most popular idea in the goal-setting literature is to think of goals as a reference point (Corgnet et al., 2015; Fan and Gómez-Miñambres, 2020). The goal-setting

[2]The task characteristics specifically refer to the task complexity that determines the cognitive load placed on an individual.

literature suggests that difficult goals increase performance by motivating individuals to put forth more effort through an increase in the intrinsic rewards from goal achievement. On the flip side, expectancy theory and self-efficacy theory from social psychology suggest that difficult goals have two competing effects on performance. On one hand, a difficult goal decreases the likelihood of attaining the goal which reduces motivation thereby decreasing effort. On the other hand, the intrinsic reward from goal attainment increases when a difficult goal is achieved which increases effort (Meyer et al., 1988). As such, only when high intrinsic rewards from a difficult goal outweigh the low probability of attainment, both theories can be reconciled in their predictions. From the task complexity literature, it is not clear how performance responds to a change in the difficulty level of the goal. Therefore, the present study complements prior literature by identifying the relationship between goal difficulty and task complexity. Moreover, by highlighting behavioral motives triggered by non-binding goals, I attempt to integrate the study of leadership in economics with social psychology to understand how intrinsic incentives influence behavior.[3,4]

3.3 Theory

The theoretical framework builds on the seminal model of a coordination game by Van Huyck et al. (1990). The game involves a team of n players. Each player simultaneously exerts effort e_i and is paid an amount A for each unit of team production. Assume that an individual's "production" is a nonlinear function of effort, i.e., $y_i = q(e_i, \epsilon_i)$, where ϵ_i is a random shock to output, uncorrelated with e_i. Team production is determined by a weakest-link production function that imposes extreme strategic complementarity. In particular, let team production be denoted

[3]In this study, leader behavior is captured by the non-binding goal condition but there is no physical leader per se.

[4]Table 3.1 in Appendix C summarizes the experimental literature on exogenous goal-setting and compares the key elements to that of the present study.

by $M(\vec{y}) = min(y_1, y_2, ...y_n)$ such that team production is determined by the lowest individual production among all team members.

Let $C(\cdot)$ denote the cost-of-effort function which depends on the level of effort exerted by an individual, e_i, their ability parameter θ_i and complexity cost of the task, ζ. As ζ increases, cost of complexity increases, i.e., the task becomes more complex, all else equal. The cost function is continuous, twice-differentiable and strictly convex in effort, i.e., $C_e(\cdot) > 0$ and $C_{ee}(\cdot) > 0$. Players are asymmetric depending on their ability level and therefore face asymmetric cost-of-effort functions. For simplicity, I assume that players have complete information about the ability parameter of each team member.

In the absence of a non-binding goal, the worker's maximization problem is:

$$\max_{e_i} \Pi_i^w = A \cdot M(\vec{y}(e_i)) - C(e_i; \theta_i, \zeta) \tag{1}$$

The associated first-order necessary condition is:

$$C_e(e_i; \theta_i, \zeta) \leq A \cdot y_e(\cdot) \tag{2}$$

Note that equation (2) holds with equality for the weak-link worker for any $e_i \in [0, e^*]$, where e^* is the solution to (2). Given the relationship between output, y and effort e, I can obtain $y^* = q(e^*)$. The cost function is also increasing in the level of task complexity and therefore team production (dependent on output of the weak-link worker) will be decreasing in the level of task complexity. Due to the nature of the production function, there are multiple equilibria. Any combination of e_i that leads to equal individual production (y_i) for all workers constitutes a pure-strategy Nash equilibrium.

3.3.1 Non-binding goals and behavioral theories

A non-binding goal is a type of managerial incentive that has been shown to enhance productivity in the workplace. Non-binding goals are particularly attractive because a manager with limited monetary resources is able to provide a costless incentive to motivate his/her employees. In this section, I consider behavioral extensions to the theory based on prior work in management and psychology. The most popular idea in this literature is to think about a goal as a reference point (Heath et al., 1999). Fan and Gómez-Miñambres (2020) extend the basic model by Van Huyck et al. (1990) to include a reference-dependent utility from a non-binding goal.[5] Using their setup, I allow the task to differ in its complexity costs and the team goal to differ in its difficulty level.

Other popular mechanisms highlighted in the psychology literature rely on the self-efficacy and expectancy theory to explain observed effects of goals. Ideas from self-efficacy and expectancy theory together suggest that although goals are helpful motivators, their efficacy depends on the difficulty level of goals. This is based on the evidence that in some cases, goals may either be counterproductive or perhaps may not alter production at all. This theory suggests that there are both costs and benefits arising out of a goal (and its difficulty level) which suggests that the directional effect of goals on production will depend on whether its benefits outweigh the costs or vice versa. Finally, goals may establish a group norm for behavior such that an individual's utility is decreasing in deviations from the group goal. This theory suggests that goals may or may not raise production in comparison to the baseline model (without a goal) depending upon the difficulty level of the goal. Below I discuss these behavioral theories in detail.

[5]Note that the utility is non-monetary because there are no direct monetary gains/losses tied to the goal. If, however, the goal increases production, then of course the monetary gains associated with that production level would be higher.

3.3.2 Reference-dependent utility

In a model with monetary incentives and non-binding goals, a worker's payoff Π_i^w is a sum of his/her monetary gains from team production and non-monetary gains (losses) from reaching (not reaching) the goal less cost of effort:

$$\Pi_i^w(y(e_i), g, \zeta, A) = \begin{cases} A \cdot M(\vec{y}(e_i)) + v(y_i - g) - C(e_i; \theta_i, \zeta), & \text{if } y_i > g. \\ A \cdot M(\vec{y}(e_i)) + \lambda(v(y_i - g)) - C(e_i; \theta_i, \zeta), & \text{if } y_i \leq g. \end{cases} \tag{3}$$

Here, $v(\cdot)$ is the goal-dependent non-monetary utility function such that $v(\cdot) > 0$ for $y > g$, $v(\cdot) < 0$ for $y < g$, and $v(\cdot) = 0$ for $y = g$. $v(\cdot)$ satisfies the properties of prospect theory in non-monetary terms as shown by Heath et al. (1999) and Fan and Gómez-Miñambres (2020). $\lambda > 1$ represents the loss-aversion parameter and g is the non-binding team goal. The goal is quantified in terms of the team production level, and so a higher g corresponds to a more difficult goal.

The necessary first order condition associated with (3), with respect to a worker's effort level for a given complexity level ζ, is the following:

$$\Pi_e^w(.) = \begin{cases} C_e(e_i; \theta_i, \zeta) \leq A \cdot y_e(\cdot) + v'(y_i - g), & \text{if } y_i > g. \\ C_e(e_i; \theta_i, \zeta) \leq A \cdot y_e(\cdot) + \lambda(v'(y_i - g)), & \text{if } y_i \leq g. \end{cases} \tag{4}$$

Assuming there is a non-monetary utility associated with the non-binding team goal, production should be weakly higher than in the case with only monetary incentives (the baseline). If $y_i > g$, utility is higher than in the baseline. If, however, the worker fails to reach a goal i.e., $y_i \leq g$, utility is lower than the baseline. The model suggests that goals (whether they are easy or difficult) help by increasing effort thereby also enhancing production. Consider an example to understand the relationship between goals, effort and performance. Suppose, Betty's usual test-score is 70 on a scale of 100 points. Assume that she derives a non-monetary utility from a goal g set at 80 points. As per goal-setting theory, a goal of 80 would increase her motivation to study harder, i.e., this increases her effort and possibly her performance. If she is able to reach the goal and score 82 points, then besides an increase in her performance, there is also an increase in her total utility (because she derives a positive non-monetary utility from reaching the goal). If she is not able to reach the goal, and falls short by say 4 points then her score is 76 points. There is still an increase in performance but she derives a negative non-monetary utility from not reaching the goal. In both cases, performance responds positively to the goal but utility may or may not. All in all, when I compare the baseline model with the goal-setting model, a goal increases performance.[6]

3.3.3 Self-efficacy and expectancy theory

Self-efficacy is defined as the belief in one's own ability of completing a task (Bandura, 1997). According to this theory, self-efficacy is an important determinant of performance. In part, self-efficacy has to do with the confidence of an individual about how likely they are to complete a certain task. Put differently, self-efficacy is an individual's belief of attaining a certain goal. The relationship between one's own self-efficacy belief and performance is moderated by goal difficulty. So, as goals become more difficult, an individual's self-efficacy decreases. Further, there is evidence that self-efficacy decreases as task complexity increases (Wood et al., 2000). Typically, cost of effort in a simple real-effort task such as a slider task refers to physical effort while in complex real-effort tasks, it may involve both physical and cognitive or mental costs. For complex tasks, cognitive effort (e.g. search for strategies) probably matters more than just physical effort; low self-efficacy may inflict an additional cognitive cost besides the complexity cost on an individual. While this cost may be high or low depending upon many alternative factors, based on the self-efficacy theory, a reasonable specification is that the cost is an increasing function of goal difficulty.

[6]It is possible that some goals may be set too high and may seem unattainable to the individual. In such cases, typically workers reject goals or are not committed to the goals which explains similar behavior in the baseline v. goal-setting model.

A similar idea is proposed by expectancy theory from the psychology literature. According to expectancy theory, introducing a non-binding goal gives rise to two competing effects: (1) expectancy, i.e., subjective probability of goal attainment; and (2) valence, i.e., expected value of goal attainment (Meyer et al., 1988). The first effect is often referred to as task-specific confidence which decreases as goals become difficult. It captures the idea proposed by self-efficacy theory. The second effect refers to the utility gain from goal attainment. The social psychology literature argues that assigned goals could have negative, positive or no effects on performance depending upon which of the two competing effects (expectancy or valence) outweigh each other (Meyer et al., 1983; Meyer et al., 1988). Most empirical studies show that goals in general increase performance so it is likely that the valence effect is stronger; however, there are some cases where a negative or a null effect may be expected such as in cases of unattainable goals, i.e., difficult goals that lead to a decrease in an individual's self-efficacy.

Taken together, both theories highlight that there are competing positive and negative effects from goals and their interaction with task complexity. To capture the competing effects from expectancy theory and self-efficacy theory, I also consider a model where goals add to a cognitive cost in addition to a non-monetary benefit. The net effect on performance depends on the two competing channels. This model suggests that besides an increase in utility from reaching a non-binding goal, the goal may also impose an additional cost that increases in the value of the goal. This is distinct from the cost function considered in the previous discussion in that this cost is also sensitive to the value of the goal.

In this model, a worker's payoff Π_i^w is the sum of her monetary gains from team production and non-monetary gains from reaching the goal less total cost (physical and cognitive cost):

$$\Pi_i^w(y(e_i), g, \zeta, A) = A \cdot M(\vec{y}(e_i)) + F(y_i \geq g) \cdot f(g) - C(e_i; \theta_i, \zeta, g) \tag{5}$$

where $f(\cdot)$ is the utility from the goal and it increases in the value of the goal i.e., $f'(\cdot) > 0$. $F(\cdot)$ is the probability that an individual reaches the goal. The probability increases as individual performance, y_i, increases and it equals 1 when the individual reaches the goal. The idea here is to capture that an individual's non-monetary gain is positive if they reach the goal, 0 otherwise. The cost $C(\cdot)$ is increasing in both the task complexity level and the value of the goal.

The necessary first order condition associated to (5), with respect to a worker's effort level for a given complexity level ζ is the following:

$$\Pi_e^w(.) = \begin{cases} C_e(e_i; \theta_i, \zeta, g) \leq A \cdot y_e(\cdot) + f(g), & \text{if } y_i \geq g. \\ C_e(e_i; \theta_i, \zeta, g) \leq A \cdot y_e(\cdot), & \text{if } y_i < g. \end{cases} \tag{6}$$

The equilibrium output will be weakly higher with a goal than without one, similar to the reference dependence model. But, an important difference here is that a relatively easier goal may increase output by more than what a difficult goal could. With an easy goal, the probability of reaching a goal $F(\cdot)$ is higher than with a difficult goal. So, it is more likely that an individual reaches the easier goal.

The interaction of task complexity and goal difficulty is important to this theory simply because while a difficult goal may add to an individual's cognitive cost, in a simple task, it is more likely to reach the goal by increasing effort than in a more complex task. The cost is such that it depends on an interaction between complexity and goal difficulty i.e., $\frac{dC''(\cdot)}{d\zeta dg} > 0$. This captures that a difficult goal triggers the cognitive cost in a complex task relative to a simpler task. If this is true, a simple goal does not add to the cognitive cost but only generates a non-monetary gain i.e., $f(g)$. Therefore, with a simple goal in any task (simple or complex), equilibrium output is higher with a goal than without and utility is higher if the individual is able to reach the goal.

The effects of a difficult goal in a complex task are slightly complicated. A difficult goal adds to the cost and also provides higher utility $f'(g) > 0$ but, at the same time, the probability of reaching a difficult goal is smaller for an easier goal, i.e., $F(\cdot)$ is decreasing in g. So, in cases where an individual believes she is very unlikely to reach a difficult goal, the goal only adds to her cost (refer to the FOC when $y_i < g$). In this case, the equilibrium output is lower with an easier goal and weakly lower than the no goal case. In the case an individual rejects the goal, the output will be the same as the basic model, i.e., without a non-binding goal.

For a simple task, the model predicts that the dominating mechanism is the utility gain from goal attainment while in a complex task, utility gain may be partly or fully offset by the cognitive cost of a goal. This explains why goals (even difficult ones) have a positive effect on relatively simpler tasks as compared to complex ones. If the non-monetary utility from goal attainment outweighs the cognitive cost that the goal imposes, then the goal should increase effort and performance. The opposite is true if costs outweigh the utility from goal attainment.

3.3.4 Social norms

In a team environment, it is natural for one's actions to be influenced by their peers. When production technology is such that it imposes a strong complementarity between team members' actions, effects of peer influence may be non-trivial. Research in the social psychology literature suggests that individuals tend to conform to peer behavior (Schnuerch and Gibbons, 2014). Akerlof and Kranton (2000) propose an identity model where individuals conform to a norm established by their social category. Further, norm-based interventions have been shown to foster what are considered positive behaviors such as reduced alcohol use or energy consumption (Miller and Prentice, 2016). This suggests existence and influence of social norms.

A non-binding goal may be thought of as an exogenously imposed norm. As such, social norms may arise in team production settings where team members derive a utility in conforming to the norm or a disutility in deviations from the norm. Social norms have been explored in the economics literature, however, their theoretical exploration in the context of non-binding goals is fairly limited. Fischer and Huddart (2008) study the existence of personal and social norms in a contracting model. While they study endogenous social norms, in the present study, it is more appropriate to consider a non-binding goal as an exogenous norm. The central idea is to add a cost function associated with deviations from the norm.

In a model with social norms, a worker's payoff Π_i^w is such that:

$$\Pi_i^w = A \cdot M(\vec{y}(e_i)) - C(e_i; \theta_i, \zeta) - h(y_i - g), \tag{7}$$

where $h(\cdot)$ is the social norm function such that $h'(\cdot)$, $h''(\cdot) > 0$. This indicates that the cost to an individual increases with a larger deviation from the social norm or goal 'g'. The cost is decreasing in the value of the norm such that $h_{yg} < 0$. As the goal value increases, effort increases, i.e., difficult goals increase effort. When comparing across models with and without goals, the social norm model is expected to predict the following. When the optimal performance in a team is such that $y^* < g_E$, where g_E is the easy goal, both easy and difficult goals are expected to increase performance relative to the baseline. However, when $y^* > g_E$, an easy goal will decrease effort and performance whereas a difficult goal will increase performance. This means that the model predicts that there exists a set of goals below which goals have negative effects on performance.

3.3.5 Main Hypotheses

Based on the study design and research question of interest, I am primarily interested in testing the following hypotheses. Because the behavioral theories give rise to differences in the directional effects of goals, expected effects in Hypotheses 3-6 are stated in reference to these theories.

H1: Individual production decreases as complexity increases.

H2: Team production decreases as complexity increases.

H3: Individual production increases when a non-binding goal is present if behavior is explained by the model of reference-dependence utility, ambiguous otherwise.

H4: Team production increases when a non-binding goal is present if behavior is explained by the model of reference-dependence utility, ambiguous otherwise.

H5: Individual production increases as goal difficulty increases if behavior is explained by the model of reference-dependence utility or social norms, ambiguous otherwise.

H6: Team production increases as goal difficulty increases if behavior is explained by the model of reference-dependence utility or social norms, ambiguous otherwise (depends on self-efficacy).

3.4 Experimental Design

The experimental design varies the: (a) presence/absence of a team goal; (b) goal type (easy, moderate and difficult), when a goal is present; and (c) complexity level of the task (low or high).[7] There are four between-subject treatments (no goal, easy goal, moderate goal, and difficult goal). The complexity level is varied within sessions, and whether the low or high level is encountered first will be randomized to help control for order effects. The treatments are summarized by Table 3.1.

3.4.1 Real-effort task: Ball-catching

The task employed is the ball-catching task introduced by Gächter et al. (2016). It requires participants to catch balls that fall from the top of the task box by using a tray at the bottom (see Figure 3.1). Participants can move the tray by clicking their mouse towards the left or right. The unique thing about this real-effort task is that "induced" costs are attached to each mouse "click". This gives the researcher a control

[7]Please note that throughout the rest of the paper, a simple task may also be referred to as a low complexity one while a complex task refers to the high complexity condition.

over costs such that the cost of complexity can be varied as per the experimental design. Balls fall at random in four separate columns as can be seen in Figure 3.1 and therefore add the element of "uncertainty" usually associated with complex tasks. The uncertainty exists throughout and forces participants to update strategies given the random falling pattern.

It is worth noting that this particular task helps separate physical effort from cognitive effort although the latter is still unobservable to an extent. Physical effort is defined as the number of clicks an individual makes in order to catch the balls. Cognitive effort, on the other hand, measures the amount of effort used in planning the scarce clicks. In the results section that follows, I analyze cognitive effort and discuss its measure in detail.

In the low complexity condition, the cost per click is 5 tokens while it is 20 tokens in the high complexity condition. The reward for catching one ball is 30 tokens; however, note that the group earns 30 tokens per catch only for the weakest or lowest scoring member of the group. The cost-to-prize ratio is 1/6 for the low complexity condition while for the high complexity condition it is 4/6, i.e., cost in the complex condition is four times that of the simple condition. When the cost of a mouse click and the cost-to-prize ratio is small (1/6), the number of predicted clicks is large. When the cost-to-prize increases (4/6), it increases cognitive effort as participants are expected to think hard and plan the number of clicks more carefully. Therefore, the high complexity condition depicts a scenario where physical effort and cognitive effort both matter for profit maximization. This is especially true given the weakest-link production technology in the theory and related experiment. Further, the task fits in the definition of a complex task as defined by Campbell (1988).

3.4.2 Pilot experiment and power analysis

To help inform the experimental design, a pilot experiment was conducted using the no goal treatment. Participants were drawn from the same population and experimental procedures followed the final protocols described later.

In addition to the ball-catching task described above, I also considered the verbal rule task, as employed by Oprea (2020). In this task, physical and cognitive costs are entirely unobservable. The task involves showing participants a verbal rule to be implemented on a sequence of letters selected randomly and shown one at a time. The Participants' task is to correctly implement the rule. The benefit of using this task is that it objectively defines complexity and it is easy to alter in order to make the task as simple or complex as needed. Two pilot sessions were conducted to test the verbal-rule task (Oprea, 2020) and the ball-catching task (Gächter et al., 2016). The pilot session for the verbal-rule task revealed lower than usual variation in measures of individual and team production across the low and high complexity conditions, partly due to the scoring rule chosen. Further, there was uncertainty in whether participants were just "guessing" to get at the correct answer. Based on pilot testing, it was deemed appropriate to use the ball-catching task for this study. Data from the ball-catching task pilot is not included in the analysis at this point but given that procedures and parameters were very similar, it may be included in future analyses.

Based on the estimated individual-level variances from the pilot (no goal treatment), the sample sizes required for tests to be sufficiently powered are N=80 (NG, EG), N=65 (MG) and N=50 (DG). The analysis assumes 10 decision rounds for each of the two complexity levels (low or high) and allows for correlation across rounds. Using the suitable econometric methods and the planned sample sizes, power calculations suggest the following.

In the simple condition, the minimum detectable effect (MDE) size is 1.5 units when comparing between NG and EG treatments. A comparison between NG and MG gives an MDE of 2 units while it is 2.5 units between NG and DG. In the complex condition, the minimum detectable effect size for the several between-subject comparisons is - 1.8 units (NG and EG), 2 units (NG and MG) and 2.3 units (NG and DG). Goal-setting studies find that the impact of goals on production ranges from 10-30% (Fan and Gómez-Miñambres, 2020). This means the smallest treatment effect in a low complexity task would be 2.3 units while that in a high complexity task will be 1.9 units. Overall, the total sample size of 275 participants should be sufficient to identify said differences.

Power calculations are of course only approximations as the true underlying outcome distributions are unknown. I expect lower variation in individual and group production for the goal treatment, and to the extent this is true, the calculations above are under-estimates of the minimum detectable effect sizes. Moreover, controlling for other factors, such as participant characteristics, in the econometric models is expected to increase power as these factors should explain variation in outcomes but be uncorrelated with treatment assignment.

3.4.3 Non-binding Goals

For a weakest-link team production game, there are important considerations about whether to set the team goal based on individual performance or that of the team. A team goal with this production technology basically targets the weakest-link worker within a team. Note that the manager's objective is to maximize monetary payoffs from team production, and a team goal consistent with profit maximization should be based on the weakest-link worker's performance. If goals are set with this notion in mind, the team should ideally respond to reasonable goals; if not, unreasonable goals may not have any effect on team production (Fan and Gómez-Miñambres, 2020).

In order to determine where to set the goals, I rely on the distribution of team production outcomes from the no goal treatments (N=30 including data from the pilot). For the simple task, the goals are 16 (easy), 20 (moderate) and 24 (difficult) and as for the complex task, they are 11 (easy), 15 (moderate) and 19 (difficult).

These goals reflect the 20^{th} (easy), 50^{th} (moderate), and 90^{th} (difficult) percentiles of the respective team production outcome distributions from the pilot.

3.4.4 Experimental Procedures

A typical experimental session proceeds as follows. Participants are assigned an ID number tied to their order of entry into the online experiment via Zoom. The experiment instructions are displayed using Zoom's screenshare feature, and the same moderator reads instructions aloud while participants follow along. In addition, the moderator follows several lab protocols mentioned in the consent form before getting the experiment started. Participants are informed that instructions contain only true information, and their decisions will be kept anonymous. All decisions are made on the participants' personal computers. The moderator encourages questions, which are asked through online chat, and exchanges are private to the inquiring participant and the moderator. The experiment is programmed and facilitated using the software z-Tree (Fischbacher, 2007) as well as z-Tree unleashed (Duch et al., 2020).

In all treatments, participants are randomly placed in three-person groups at the start of each round and therefore members of every group randomly change throughout the experiment.[8] The design is such that there is an exogenous manager/leader that depicts the two broad types of incentive conditions i.e., a monetary incentive with no team goal versus a monetary incentive with a team goal recommendation.

The participants first go through a paid risk elicitation task using the design by Holt and Laury (2002) in order to elicit risk preferences. Following this, during the second stage, the participants go through a loss aversion task. The goal here is to gauge whether participants' preferences are consistent with loss aversion therefore the task only varies the loss amount in each lottery while keeping the gain amount fixed.

[8] This particular design choice is made to ensure that any endogenous goal formation from repeated interactions with the same people overtime do not act as a confound in identifying effects of exogenously set goals.

It is implemented in accordance with procedures described by Bibby and Ferguson (2011) and Gächter et al. (2022).

In the third stage, participants engage in the team production experiment. In each round, every group member is assigned the ball-catching task to complete within a minute (60 seconds). After the task has been completed, the group members see a result screen with the individual and group outcomes (individual score, group score, total individual cost and individual earnings) i.e., the design incorporates individual and group feedback. Particularly for the goal treatments, participants are also shown whether their team met the goal in a round. Following similar procedures, each group plays the game for 10 rounds with the low complexity level, and then another 10 rounds with the high complexity level.[9]

The monetary payoff of a player is:

$$\Pi_i = A \cdot \text{group score} - \text{cost} \cdot \#\text{ of clicks} \tag{8}$$

Participants earn $A = 30$ tokens for every ball caught by the team, i.e., by the lowest performing member. The payoff is determined by subtracting the participant's total cost of clicking (determined by their individual # of clicks and cost per click) from the reward. At the end of the experiment, participants are paid their earnings from two separate rounds selected at random (one from the low complexity condition and the other from the high complexity condition).

Note that the group reward is determined by the parameter $A = 30$ while the cost tied to clicks is specific to task-type - low complexity ($cost = 5$) or high complexity ($cost = 20$). While the cost of clicks is observable given the unique real-effort task, cognitive costs are still unobservable and therefore total cost may still be weakly higher in the high complexity condition.

[9]Note that, the order of simple or complex tasks will be varied in the experiment overall, but, given the preliminary data, all sessions use the order of simple task followed by complex. The reason is simply to avoid order effects from confounding potential treatment effects in a small sample.

3.4.5 Participants

Four experiment sessions were conducted in July 2022. In total, I have data from 60 participants (not including the pilot). All sessions were conducted online and facilitated via Zoom. Undergraduate students enrolled at the University of Tennessee were recruited from a large existing database that had previously registered to receive invitations for economics experiments. People were not allowed to attend more than one session of the experiment. Earnings in the ball-catching task were dominated in "tokens" and exchanged for U.S. dollars at an announced exchange rate. The experiment lasted approximately 75 minutes and on average participants earned $19 for the session.

Table 3.4 describes the experiment data. Overall, 68% of participants are female, and 84% had participated in a prior economics experiment. Forty-one percent can be characterized as risk averse based on the incentivized risk elicitation task while 76% may be characterized as loss averse. Responses from the post-experiment questionnaire suggest that a majority (69%) felt they were sufficiently compensated. In response to a Likert-scale question that ranged from "1" ("poorly understood") to "5" ("well understood"), the vast majority (84%) selected a 4 or 5, indicating a strong self-assessment of how well instructions were understood.

3.5 Results

3.5.1 Individual and team production

I begin the analysis with linear regression models of production (i.e., number of catches) at the individual and group level, where the latter is as defined as the production of the group member with the lowest individual production. In regressions with individual-level observations, I cluster standard errors by participant and by decision round (for participants within the same session). This allows for within-person serial correlation as well as contemporaneous correlation across participants within the same session. For regressions with group-level observations, errors are clustered to allow contemporaneous correlations across groups within the same session. I define Round from 0 through 9 for both complexity conditions (low and high) such that the variable resets to 0 when the task changes in complexity.

Table 3.4 presents a basic regression analysis of the effects of task complexity on individual production. Model (1) pools data from all treatments while models (2) through (5) are specific to the goal condition (i.e., no goal, easy, moderate and difficult goals, respectively). In all models, I reject the null hypothesis that individual-level production is equal between the low and high complexity conditions. Based on Model (1), production at the individual-level is approximately 3.8 points lower in the high complexity condition, but the point estimates vary slightly across goal conditions with the highest difference being when a difficult goal is assigned (approximately 5.6 points). This result is not surprising as the complex condition increases both physical as well as cognitive costs for participants and therefore output is expected to decrease. Table 3.5 adds controls to all specifications from Table 3.4. While the magnitudes decrease due to presence of strong time trends, the directional results are robust to inclusion of controls for all models.

Table 3.6 presents the team production analog to Table 3.4. Based on the coefficients, I can reject the null hypothesis that team production is statistically equal between the low and high complexity conditions across Models (1), (2), (4) and (5). Altogether, there is support for Hypothesis 1 and 2.

Table 3.7 presents regression specifications that can be used to test Hypotheses 3-6. Specification (1) pools data from all goal treatments, and the variable "Goal" is an indicator variable that equals 1 if a goal (easy, moderate or difficult) is assigned to the team. The interaction of task type with the "Goal" indicator helps identify effects of having a goal in the low or high complexity conditions relative to the no goal setting. This model reveals that a goal has no effect on individual production, on average, in either of the two settings. A popular result from the in the experimental psychology literature is that non-binding goals help improve an individual's performance (Locke and Latham, 2002). This result has been established with respect to tasks that do not necessarily place a cognitive load on individuals, however, note that prior studies do not exogenously vary the goal. Instead, a subject acts as a manager and assigns goals to teams. As mentioned earlier, Fan and Gómez-Miñambres (2020) find that about 50% of the time, managers set goals that are too challenging or "unrealistic" such that team production does not respond to goals. This is evidence that managers may not always set goals that maximize team production. Turning to effects of goals on team production, Table 3.8 is the team analog to 3.7 and it depicts how goals influence the weakest-link worker's production. Results suggest that in a pooled model, goals have no effect on team production in either a low or high complexity condition.

To identify effects of goal difficulty on production, specification (2) in Table 3.7 and 3.8 allows the effects of easy, moderate and difficult goals on individual and team production to differ by task type. Specification (3) adds controls to (2). Results suggest that easier goals tend to lower individual production in the low complexity condition (by 1.5 points approximately); however, moderate or difficult goals have no impact relative to the no goal treatment. Adding controls does not change the quantitative effects of goals but depicts presence of a strong time trend. Based on specification (3) in 3.7 and 3.8, easy goals decrease individual production relative to no goals, however, there production is not altered when goals are either moderately challenging or difficult. With respect to team production, while easy goals have no significant impact on production, difficult goals appear to increase production. A difficult goal has a relatively large effect (almost double) on production compared when task complexity is low. Goal effectiveness is defined as the magnitude (in %) by which a non-binding goal increases production relative to a no goal setting. Based on the results so far, difficult goals seem to be more effective when task complexity is low compared to the high complexity condition.

It is also worthwhile to compare how production at the individual and team level differs across the three goal types while holding complexity fixed. Specification (3) in 3.7 and 3.8 can be used to compare the coefficients on the interaction of task with goal type. In the low complexity condition, this comparison reveals that individual production is statistically higher when a a moderate or a difficult goal is assigned relative to an easy goal ($p < 0.10$; $p < 0.01$). For the high complexity condition, individual production is statistically higher when a moderate or a difficult goal is assigned relative to an easy goal ($p < 0.05$; $p < 0.01$). Differences do not arise between moderate and difficult goals. With respect to team production, differences arise between difficult v. easy and difficult v. moderate goals in the low complexity condition ($p < 0.01$; $p < 0.05$). For the high complexity condition, differences arise between easy v. difficult ($p < 0.01$) and moderate v. difficult ($p < 0.05$). There are no differences across easy and moderate goals.

Effects of the control variables are as follows. Individual and team production decreases as the experiment progresses. This is expected in general because the weakest-link production technology induces coordination in production outcomes, i.e., higher-performing individuals are likely to learn that they are wasting effort early in the game, leading to lower effort as the experiment progresses. Prior experience in economics experiments, a higher average GPA, being risk averse and loss averse also lower individual production. The results are summarized below.

Result 1. Individual and team production reduces as complexity costs increase.

Result 2. Individual production is statistically lower between the baseline (no goal) treatment and treatments with a non-binding goal. This result holds for easy goals.

Result 3. Team production is statistically higher between the baseline (no goal) treatment and treatments with a non-binding goal. This result holds for difficult goals.

Result 4. Easier goals tend to lower *individual* production by a slightly larger magnitude for the low complexity condition, relative to the high complexity condition.

Result 5. The magnitude of goal effectiveness on *team* production is lower for the low complexity condition, relative to the high complexity condition. This holds true for difficult goals.

3.5.2 Physical and cognitive effort

In this section, I analyze individual and team effort, precisely the number of clicks.[10] Table 3.9 presents the results from the exact empirical specifications as before. When a task is complex, it reduces effort (i.e., the number of clicks) by approximately 6.5 units.[11] This is evidence that high complexity costs may demotivate individuals and discourage effort relative to tasks with a lower cost. This further solidifies Result 1. Moreover, effort decreases when an easy goal is assigned in a high complexity condition. There are no significant effects of moderate or difficult goals on effort. Accounting for task-specific time trends, it seems that effort is relatively higher with moderate goals relative to easy or difficult goals.

Analysis of team effort (Table 3.10) indicates that effort is lower when complexity is high (by about 5 units). Further, it depicts that difficult goals increases effort but the magnitude of goal effectiveness is much higher in the low complexity condition. Typically a complex task would discourage physical effort given that there is a relatively lower control over the output. This is what results indicate in general. But, when analyzing the effects of goals in conjunction with task complexity, differences arise only for a few comparisons. A case could be made about higher effort as complexity increases if say an individual has higher than average ability. A similar ambiguity holds for cognitive effort, but typically higher amount of cognitive effort is required in high complexity relative to low complexity conditions.

As the self-efficacy theory suggests that goals in conjunction with complex tasks could impose higher cognitive costs, it may be worthwhile to analyze cognitive effort for several goal conditions. While cognitive effort remains unobservable for the most

[10] Gächter et al. (2016) derive theoretical predictions for effort under different cost-to-prize ratios. By using a similar approach as theirs I show that the estimates of the ball-catching production function in the present study, $q(\cdot)$, is not very different from theirs. Please see section C.2 in Appendix C for details.

[11] Note that as the cost of complexity, measured by cost per click, increases, physical effort becomes a rather incomplete measure of the overall effort exerted. With clicking becoming more expensive, individuals need to exert more cognitive effort as well in order to utilize the scarce clicks. For this reason, I also analyze a measure of cognitive effort later on in the analysis.

part, this particular task does provide a crude way of measuring it. Insights from data analysis and the questionnaire specific to cognitive effort are discussed below.

Cognitive effort may be defined as "catches per click" as when individuals spend time carefully planning their clicks so as to catch more balls per click. If complex tasks do place a high cognitive cost on individuals relative to simple tasks then, the cognitive effort is expected to increase under such conditions. Table 3.11 reports the results for individual production while Table 3.12 depicts results for team production.

Specification (1) shows that cognitive effort is statistically higher in the high complexity condition relative to the low complexity condition by approximately 2 points. Put simply, for every click that an individual makes, they catch 2 more balls in the high complexity condition relative to the low complexity condition. The post-experimental questionnaire asked individuals to state whether they had to think more carefully and plan every click when the cost was higher (i.e., the high complexity condition) and 93% stated "Yes". Finally, with respect to the effect of goals on cognitive effort, there are no significant differences except marginal evidence that a difficult goal reduces cognitive effort in both task types. This is somewhat suggestive of a cognitive cost placed by a challenging, difficult goal as suggested by self-efficacy theory since the magnitude is higher for complex tasks (see interactions of task type with difficult goals in Tables 3.11 and 3.12).

Result 6. Task complexity appears to decrease physical effort but increase cognitive effort.

3.5.3 Within-group coordination and wasted performance

In this subsection, I analyze the effects of goals on individuals' attempts at coordination in production outcomes. I start by analyzing wasted production which is defined as the difference between an individual's production (number of catches) and the team production (the lowest performing member's number of catches) per round. This measure captures wasted performance on a team. By definition, this value is zero for the weakest-link member.

Table 3.13 presents specification to help identify the effects of goals on wasted performance or production.[12] Specification (1) depicts the pooled model and shows that a non-binding goal reduces wasted production regardless of its difficulty level. However, this is only true, on average, for the low complexity task. Therefore, coordination may not depend directly on goal difficulty for the low complexity task. Goals, on average, do not alter wasted production when task complexity is high.

Specification (2) and (3) show that easier goals reduce wasted production for both the low and high complexity levels while difficult goals do so only when complexity is low. While goal difficulty may have some opposing directional effects on production as highlighted in the prior analysis, overall it seems that a goal reduces the variance of production thereby reducing wasted performance relative to a no goal treatment. Easier goals reduce wasted production with a relatively higher magnitude when it comes to a high complexity task. However, difficult goals tend to reduce production only when complexity is low. Fan and Gómez-Miñambres (2020) find in a team setting that goals minimize wasted performance thereby enhancing coordination in production outcomes among team members, which is indicative of the results in the present study as well. Across all three specifications, there are no differences between wasted production between the low and high complexity levels.

Result 7. A non-binding goal helps reduce wasted performance thereby enhancing coordination.

3.5.4 Behavioral mechanisms underlying non-binding goals

The questionnaire includes several items to help evaluate the behavioral effects of non-binding goals. More specifically, it includes rating questions that help compute

[12] I chose to rely on linear regression models, although another alternative empirical specification could be a poisson regression.

an individual's own-assessment of their self-efficacy. These questions ask individuals to rate how confident they are when solving difficult problems, accomplishing goals set by a superior or their own personal goals. Further, there are a few Likert-scale rating questions that provide evidence of whether individuals have a tendency to follow a group's social norm.

Based on the preliminary data, the average self-efficacy score on a scale of 0 to 600 is 478 and 90% of the present sample state high self-efficacy. Further, it appears that individuals are more confident or possess a high self-efficacy belief in the no goal (NG) treatments relative to the easy goal (EG) treatment which suggest that both effort and production is expected to be lower with an easy goal relative to the no goal case. This is in fact what results reveal. There are no statistical differences in the overall self-efficacy belief score for the two other pairwise comparisons (NG and MG; NG and DG). The Likert-rating on the more direct social norm questions do not appear to provide any evidence with respect to social norms. There is evidence to show that participants are more likely to follow group behavior when no goal is present. This could be a potential reason why some of the differences between the no goal treatment and goal treatments appear relatively small. Regardless, the fact that easy goals reduce both production and effort seem to point toward the directional effects expected under the social norm model. If individuals anchor to the goal assuming that the team collectively aims for goal attainment, then an easier goal may not motivate individuals to try harder once the goal is attained.

The questionnaire also included an item designed to identify and control for any personal goals that individuals may have set for themselves. This is done to validate that individual's effort and production respond to exogenous team goals rather than unobservable personal goals. Seventy-seven percent of the participants state that they did not set any personal goal and therefore it is highly unlikely to be a confounding factor. Further, the relatively small proportion of participants who do set personal goals is balanced between the three goal treatments.

The loss-aversion task conducted in the second stage of the experiment reveals interesting findings. In a low complexity task, being loss averse tends to decrease individual production in treatments with no goal and an easy goal while it increases individual production when the goal is difficult. As task complexity increases, being loss averse only tends to lower individual production in treatments with no goal while increases it when goals are either moderately challenging or difficult. This would appear to further increase the treatment effect for the comparison between difficult goal v. no goal case while lowering the treatment effect between easy goal v. no goal. However, based on the results for individual production, it seems quite the opposite is happening. Despite the shrinkage effect that loss aversion may have on the treatment effect (easy v. no goal), easier goals still appear to lower individual production. Based on the evidence so far, it seems that a theory of self-efficacy is more likely to explain the results. A higher self-efficacy score in the no goal case may be responsible for higher production levels relative to the treatment with easy goals. The directional effects do seem in line with a theory of social norms, but at the moment there isn't overwhelming evidence supporting it based on the questionnaire data.

3.6 Conclusion

The goal of this study is to test the effectiveness of non-monetary incentives in a weakest-link team production game where individuals are paid based on team production and tasks are of a complex nature. I do so by conducting an online experiment where teams are engaged in tasks that differ in their complexity, and the level of a non-binding team production goal (and whether there is a goal) is varied across teams. Further, by studying complexity in a strategic team environment, this study helps identify how task complexity affects the relationship between incentives and performance.

Prior literature provides suggestive evidence that the effectiveness of monetary incentives decreases as tasks become more complex. Therefore, the management literature highlights the importance of non-monetary incentives and transformational leadership behaviors (Zehnder et al., 2017). The effects of goals on performance may be ambiguous and may very well depend on conditions such as task complexity, goal difficulty, and their interactions. The present study provides a direct test of non-binding goals under these conditions.

Preliminary results appear to support some conjectures while refute others. First, I find that production and effort levels decrease as complexity increases. Second, while an easier goal reduces individual production, a rather challenging, difficult goal has no impact relative to no goal. At the same time, only a difficult goal seems to improve team production relative to no goal. Third, a non-binding goal decreases wasted performance thereby enhancing within-group coordination. There is evidence that easy goals may discourage individual production but at the same time this is not true for team production suggesting that the effects of goals differ when it comes to the weakest-link worker. This is somewhat surprising given that individual-level studies find that goals in general improve performance (Corgnet et al., 2015). Finally, when complexity increases, i.e., higher cognitive costs are placed on individuals, the magnitude by which difficult goals increase team production is relatively smaller as goal difficulty increases. Further, as task complexity increases, while physical effort decreases, cognitive effort increases.

This is the first study that exogenously varies the goal to identify effects of goal difficulty. The results provide some insight on the underlying behavioral motivations of people when a non-binding goal is present. As production is not monotonically increasing in goal difficulty, this suggests that a theory of reference-dependent utility inadequately explains behavior in the experiment. Instead, results seem to be more consistent with a theory of self-efficacy and expectancy. Data from the questionnaire provides an indication that self-efficacy beliefs do differ across the goal and no goal treatments. Presently, self-efficacy is higher in the no goal case relative to other treatments and therefore production is expected to be similar or even higher in the no goal treatment than the goal treatments. Despite that, difficult goals appear to improve team production which suggests that difficult goals motivate the weakest-link worker to work harder. Given the opposite directional effects of easier goals, I also suspect that a theory of social norms may be a potential explanation for when goals are too easy. At present, the evidence on behavioral mechanisms is mixed. However, as I collect more data the present study will be sufficiently powered to identify how exogenous goals affect production. It will help understand how goal difficulty affects production depending on task type. Moreover, the study will be able to provide insight into the psychological make up of employees by identifying behavioral mechanisms underlying a non-binding goal.

Overall, this study provides important insights for managers. First, it explores the importance of non-monetary incentives that may be used in organizations. More specifically, by investigating a relationship between goal difficulty and task complexity, the study helps to identify effective goal-incentive combinations.[13] Present findings indicate that setting goals that are difficult may be better if the manager's goal is to improve team production because easier goals may end up in the team coordinating on lower production levels.

The broader goal of the study is to contribute to a more comprehensive view and understanding of incentives in workplaces. This type of a team production game not only captures environments in organizations but also collaboration between researchers and so the results are applicable to a wide variety of economic settings. A comparison that has been of interest but remains unexplored is how teams perform when multiple goals exist at the same time. Although experimentally it is easy to add a condition with more than one goal, it may be challenging to address it in the theory setup considered here. Nevertheless, it remains an open question for future research. The expected results as well as the experimental design further serve as

[13] While varying both the group incentive and goals would have been ideal, it leads to a very large design that may not be feasible. Another reason I consider variance is goals rather than monetary incentives is that the task complexity literature highlights that changes in monetary incentives are not closely related to performance and therefore, I expect that this comparison would not lead to interesting effects.

a building block to incorporate and more formally investigate the impacts of non-monetary leadership tools such as transformational and charismatic leadership styles on team performance. A potential area for future research would be to induce a competitive context such as a Tullock contest (Eisenkopf, 2014; Eisenkopf, 2020).

Bibliography C

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The quarterly journal of economics 115*(3), 715–753.

Allison, D. B., J. M. Silverstein, and V. Galante (1993). Relative effectiveness and cost-effectiveness of cooperative, competitive, and independent monetary incentive systems. *Journal of Organizational Behavior Management 13*(1), 85–112.

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman.

Bibby, P. A. and E. Ferguson (2011). The ability to process emotional information predicts loss aversion. *Personality and Individual Differences 51*(3), 263–266.

Bonner, S. E. and G. B. Sprinkle (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, organizations and society 27*(4-5), 303–345.

Brandts, J. and D. J. Cooper (2007). It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association 5*(6), 1223–1268.

Bulte, E., J. A. List, and D. Van Soest (2020). Toward an understanding of the welfare effects of nudges: Evidence from a field experiment in the workplace. *The Economic Journal 130*(632), 2329–2353.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of management review 13*(1), 40–52.

Chen, X. and G. P. Latham (2014). The effect of priming learning vs. performance goals on a complex task. *Organizational Behavior and Human Decision Processes 125*(2), 88–97.

Corgnet, B., J. Gómez-Miñambres, and R. Hernán-Gonzalez (2015). Goal setting and monetary incentives: When large stakes are not enough. *Management Science 61*(12), 2926–2944.

Corgnet, B., J. Gómez-Miñambres, and R. Hernán-Gonzalez (2018). Goal setting in the principal–agent model: Weak incentives for strong performance. *Games and Economic Behavior 109*, 311–326.

Earley, P. C., T. Connolly, and G. Ekegren (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *Journal of applied psychology 74*(1), 24.

Eisenkopf, G. (2014). The impact of management incentives in intergroup contests. *European Economic Review 67*, 42–61.

Eisenkopf, G. (2020). Words and deeds–experimental evidence on leading-by-example. *The Leadership Quarterly 31*(4), 101383.

Fan, E. T. and D. H. Gruenfeld (1998). When needs outweigh desires: The effects of resource interdependence and reward interdependence on group problem solving. *Basic and Applied Social Psychology 20*(1), 45–56.

Fan, J. and J. Gómez-Miñambres (2020). Nonbinding goals in teams: A real effort coordination experiment. *Manufacturing & Service Operations Management 22*(5), 1026–1044.

Fischer, P. and S. Huddart (2008). Optimal contracting with endogenous social norms. *American Economic Review 98*(4), 1459–75.

Fonseca, M. A. and S. B. Grimshaw (2017). Do behavioral nudges in prepopulated tax forms affect compliance? experimental evidence with real taxpayers. *Journal of Public Policy & Marketing 36*(2), 213–226.

Frey, B. S. and R. Jegen (2001). Motivation crowding theory. *Journal of economic surveys 15*(5), 589–611.

Gächter, S., L. Huang, and M. Sefton (2016). Combining "real effort" with induced effort costs: the ball-catching task. *Experimental economics 19*(4), 687–712.

Gächter, S., E. J. Johnson, and A. Herrmann (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision 92*(3), 599–624.

Gómez-Miñambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology 33*(6), 1223–1239.

Heath, C., R. P. Larrick, and G. Wu (1999). Goals as reference points. *Cognitive psychology 38*(1), 79–109.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Latham, G. P., E. A. Locke, and N. E. Fassina (2002). The high performance cycle: Standing the test of time. *Psychological management of individual performance 5*(6), 201–228.

Locke, E. A. and G. P. Latham (1990). Work motivation and satisfaction: Light at the end of the tunnel. *Psychological science 1*(4), 240–246.

Locke, E. A. and G. P. Latham (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist 57*(9), 705.

Locke, E. A. and G. P. Latham (2019). The development of goal setting theory: A half century retrospective. *Motivation Science 5*(2), 93.

Meyer, J., E. Konar, and B. Schacht (1983). Goal setting effects on motivation: A vie theory explanation. In *annual meeting of the American Psychological Association, Anaheim, CA*.

Meyer, J. P., B. Schacht-Cole, and I. R. Gellatly (1988). An examination of the cognitive mechanisms by which assigned goals affect task performance and reactions to performance 1. *Journal of Applied Social Psychology 18*(5), 390–408.

Miller, D. T. and D. A. Prentice (2016). Changing norms to change behavior. *Annual review of psychology 67*, 339–361.

Nahrgang, J. D., D. S. DeRue, J. R. Hollenbeck, M. Spitzmuller, D. K. Jundt, and D. R. Ilgen (2013). Goal setting in teams: The impact of learning and performance goals on process and performance. *Organizational Behavior and Human Decision Processes 122*(1), 12–21.

Oprea, R. (2020). What makes a rule complex? *American economic review 110*(12), 3913–51.

Samek, A. (2019). Gifts and goals: Behavioral nudges to improve child food choice at school. *Journal of Economic Behavior & Organization 164*, 1–12.

Schnuerch, R. and H. Gibbons (2014). A review of neurocognitive mechanisms of social conformity. *Social Psychology*.

Smithers, S. (2015). Goals, motivation and gender. *Economics Letters 131*, 75–77.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review 80*(1), 234–248.

van Vijfeijken, H., A. Kleingeld, H. v. Tuijl, J. A. Algera, and H. Thierry (2002). Task complexity and task, goal, and reward interdependence in group performance management: A prescriptive model. *European Journal of work and organizational psychology 11*(3), 363–383.

Vecchio, R. and C. Cavallo (2019). Increasing healthy food choices through nudges: A systematic review. *Food Quality and Preference 78*, 103714.

Wood, R., P. Atkins, and C. Tabernero (2000). Self-efficacy and strategy on complex tasks. *Applied psychology 49*(3), 430–446.

Wood, R. E., A. J. Mento, and E. A. Locke (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of applied psychology 72*(3), 416.

Wu, S. J. and E. L. Paluck (2021). Designing nudges for the context: Golden coin decals nudge workplace behavior in china. *Organizational Behavior and Human Decision Processes 163*, 43–50.

Zehnder, C., H. Herz, and J.-P. Bonardi (2017). A productive clash of cultures: Injecting economics into leadership research. *The Leadership Quarterly 28*(1), 65–85.

Appendix

Appendix A

Appendix

A.1 Tables

Table 1.1: Group effort: *complete* information

Heterogeneity	Contest Type	Equilibrium effort
Cost-of-effort	Uneven	$(X_A^*, X_D^*) = \left(\frac{vc_D}{(c_A+c_D)^2}, \frac{vc_A}{(c_A+c_D)^2}\right)$
	Even	$(X_A^*, X_A^*) = \left(\frac{v}{4c_A}, \frac{v}{4c_A}\right)$; $(X_D^*, X_D^*) = \left(\frac{v}{4c_D}, \frac{v}{4c_D}\right)$
Prize value	Uneven	$(X_A^*, X_D^*) = \left(\frac{v_D v_A^2}{c(v_A+v_D)^2}, \frac{v_A v_D^2}{c(v_A+v_D)^2}\right)$
	Even	$(X_A^*, X_A^*) = \left(\frac{v_A}{4c}, \frac{v_A}{4c}\right)$; $(X_D^*, X_D^*) = \left(\frac{v_D}{4c}, \frac{v_D}{4c}\right)$
Group Size	Uneven	$(X_A^*, X_D^*) = \left(\frac{v}{4c}, \frac{v}{4c}\right)$
	Even	$(X_A^*, X_A^*) = \left(\frac{v}{4c}, \frac{v}{4c}\right)$; $(X_D^*, X_D^*) = \left(\frac{v}{4c}, \frac{v}{4c}\right)$

Notes: An "uneven" contest refers to a case where an advantaged (A) group plays a disadvantaged (D) group. The advantaged team has either a lower cost of effort (i.e., $c_A < c_D$), higher prize value ($v_A > v_D$), or larger group size (i.e., $N_A > N_D$) relative to the disadvantaged team. In an "even" contest, both groups are advantaged or disadvantaged.

Table 1.2: Group effort: *incomplete* information

Heterogeneity	Equilibrium effort
Cost-of-effort	$X_A^* = \left(\frac{v}{c_A} \frac{4\frac{c_A}{c_D}+(1+\frac{c_A}{c_D})^2}{8(1+\frac{c_A}{c_D})^2} \right); \, X_D^* = \left(\frac{v}{c_D} \frac{4\frac{c_A}{c_D}+(1+\frac{c_A}{c_D})^2}{8(1+\frac{c_A}{c_D})^2} \right)$
Prize value	$X_A^* = \left(\frac{v_A}{c} \frac{4\frac{v_D}{v_A}+(1+\frac{v_D}{v_A}{}^{2}}{8(1+\frac{v_D}{v_A})^2} \right); \, X_D^* = \left(\frac{v_D}{c} \frac{4\frac{v_D}{v_A}+(1+\frac{v_D}{v_A})^2}{8(1+\frac{v_D}{v_A})^2} \right)$
Group Size	$(X_A^*, X_D^*) = \left(\frac{v}{4c}, \frac{v}{4c}\right)$

Notes: The equilibrium effort of advantaged and disadvantaged teams are denoted by X_A^{**} and X_D^{**}, respectively. An advantaged team has either a lower cost of effort (i.e., $c_A < c_D$), higher prize value ($v_A > v_D$), or larger group size (i.e., $N_A > N_D$) relative to the disadvantaged team. Equilibria correspond with $r = 1/2$, i.e., that there is a 50% chance the opponent is an advantaged team.

Table 1.3: Experiment Parameters

Source of heterogeneity	Group Type	Cost	Value	Group
Cost-of-effort	Advantaged	$c_A = \frac{1}{3}$	$\nu = 50$	$N = 3$
	Disadvantaged	$c_D = 1$	$\nu = 50$	$N = 3$
Prize value	Advantaged	$c = 1$	$\nu_A = 150$	$N = 3$
	Disadvantaged	$c = 1$	$\nu_D = 50$	$N = 3$
Group Size	Advantaged	$c = 1$	$\nu = 50$	$N_A = 9$
	Disadvantaged	$c = 1$	$\nu = 50$	$N_D = 3$

Table 1.4: Theoretical predictions and observed group effort: *complete* information

		(1) **Standard**		(2) **Extended**		(3) **Observed**	
Heterogeneity	Contest Type	X_A^*	X_D^*	X_A^*	X_D^*	X_A	X_D
Cost-of-effort	Advantaged	28.13	9.38	68.50	22.83	75.04	20.10
	Disadvantaged	37.50	12.50	91.34	30.45	80.79	49.43
Prize value	Advantaged	28.13	9.38	65.31	21.77	70.69	27.04
	Disadvantaged	37.50	12.50	87.09	29.03	76.84	43.72
Group Size	Advantaged	12.50	12.50	103.62	32.75	119.57	41.14
	Disadvantaged	12.50	12.50	141.96	44.87	128.50	51.91

Notes: X_A and X_D refer to effort for advantaged and disadvantaged groups, respectively. The standard model predictions are calculated using the equilibria presented in Table 1. The extended model predictions are calculated using the formulas in Table A1, conditional on utility parameters estimated from the data.

Table 1.5: Theoretical predictions and observed group effort: *incomplete* information

		(1) **Standard**		(2) **Extended**		(3) **Observed**	
Heterogeneity	Contest Type	X_A^*	X_D^*	X_A^*	X_D^*	X_A	X_D
Cost-of-effort	Advantaged	32.81	10.94	93.34	31.11	89.48	42.70
Prize value	Advantaged	32.81	10.94	91.86	30.62	84.85	51.66
Group Size	Advantaged	12.50	12.50	119.64	38.03	119.20	39.80

Notes: X_A and X_D refer to effort for advantaged and disadvantaged groups, respectively. The standard model predictions are calculated using the equilibria presented in Table 2. The extended model predictions are calculated using the formulas in Table A2, conditional on utility parameters estimated from the data.

Table 1.6: Description of data

Variable Name	Description	Mean	S.D.
Dependent variables			
Group Effort	Total points contributed by all group members	74.90	45.77
Probability of Winning	Calculated as a function of own and opponent group effort, using equation [1]	52.42	22.84
Individual Effort	Points contributed by the participant, 0 to 50 points	18.03	16.06
Effort Variance	Squared deviation of a participant's contribution relative to the group's mean contribution	145.81	211.13
Zero Effort	= 1 if participant contributed zero points; 0 otherwise	0.21	0.41
Treatment variables			
Advantaged	= 1 for advantaged groups; 0 otherwise	0.56	0.50
Incomplete	= 1 for incomplete information treatments; 0 otherwise	0.53	0.50
Uneven	= 1 for uneven contests; 0 otherwise	0.47	0.50
Cost	= 1 for cost treatments; 0 otherwise	0.33	0.46
Value	= 1 for value treatments; 0 otherwise	0.33	0.46
Group	= 1 for group size treatments; 0 otherwise	0.40	0.49
Control variables			
Risk Averse	= 1 if participant selected safe option at least six times in Risk Elicitation task; 0 otherwise	0.47	0.50
Experience	=1 if the participant had partaken in a prior economics experiment; 0 otherwise	0.56	0.50
Female	= 1 if participant is female; 0 otherwise	0.42	0.49
Round	Decision round in the experiment, 1 to 20	10.50	5.76
GPA	Participant GPA, recorded as midpoint of chosen interval	3.29	0.48

Table 1.7: Analysis of information effects: *uneven* contests

	Dependent: Group-level effort			
	(1)	(2)	(3)	(4)
Constant	56.80***	47.57***	47.89***	48.86***
	(1.88)	(2.33)	(2.39)	(2.46)
Value		1.30	1.60	1.131
		(3.22)	(3.05)	(3.09)
Group		32.79***	27.79***	26.80***
		(4.37)	(4.24)	(4.29)
Cost x Incomplete		20.01***	20.36***	19.99***
		(3.01)	(2.82)	(2.88)
Value x Incomplete		19.12***	19.08***	18.64***
		(3.37)	(2.92)	(2.87)
Group x Incomplete		−11.19**	−5.87	−5.79
		(4.83)	(4.39)	(4.42)
Incomplete	11.33***			
	(2.34)			
Experience			−17.77***	−17.85***
			(3.25)	(3.28)
Risk Averse			−5.33	−3.67
			(3.40)	(3.41)
Female			11.45***	11.29***
			(3.67)	(3.67)
Round			−1.31***	−1.29***
			(0.16)	(0.16)
GPA			−11.26***	2.37
			(3.61)	(5.81)
GPA x Incomplete				−21.14***
				(7.22)
R-squared	0.016	0.050	0.111	0.115
Observations	1498	1498	1498	1498

Notes: Cluster-robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All control variables are demeaned.

Table 1.8: Analysis of information effects: *even* contests

	Dependent: Group-level effort			
	(1)	(2)	(3)	(4)
Constant	66.52***	63.77***	47.89***	48.86***
	(2.21)	(3.79)	(2.39)	(2.46)
Value		−2.27	−3.99	−3.64
		(4.73)	(4.60)	(4.61)
Group		11.12*	6.78	6.49
		(5.92)	(5.89)	(5.93)
Cost x Incomplete		3.81	3.42	3.43
		(4.24)	(3.96)	(4.00)
Value x Incomplete		6.49*	6.85*	6.66*
		(3.80)	(3.60)	(3.52)
Group x Incomplete		−5.72	−2.00	−1.62
		(5.51)	(5.36)	(5.39)
Incomplete	1.60			
	(2.61)			
Experience			−15.94***	−16.44***
			(3.03)	(3.08)
Risk Averse			−8.68***	−7.28**
			(3.15)	(3.13)
Female			11.69***	11.36***
			(3.43)	(3.42)
Round			−1.19***	−1.20***
			(0.19)	(0.19)
GPA			−11.35***	0.52
			(3.59)	(5.75)
GPA x Incomplete				−18.35**
				(7.10)
R-squared	0.000	0.008	0.111	0.115
Observations	1566	1566	1566	1566

Notes: Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
All control variables are demeaned.

Table 1.9: Analysis of information effects: pooled over contest types

	Dependent: Group-level effort			
	(1)	(2)	(3)	(4)
Constant	62.02***	55.67***	56.81***	57.16***
	(1.53)	(2.51)	(2.42)	(2.46)
Value		−0.13	−0.94	−0.92
		(3.21)	(3.00)	(3.03)
Group		21.47***	16.91***	16.39***
		(3.82)	(3.85)	(3.86)
Cost x Incomplete		11.91***	12.17***	11.71***
		(3.15)	(2.87)	(2.89)
Value x Incomplete		12.45***	12.64***	12.40***
		(3.23)	(2.87)	(2.77)
Group x Incomplete		−7.967*	−4.10	−3.81
		(4.27)	(3.98)	(3.98)
Incomplete	6.10***			
	(2.07)			
Experience			−17.92***	−18.32***
			(2.97)	(3.00)
Risk Averse			−6.00**	−4.07
			(3.02)	(3.01)
Female			9.24***	8.96***
			(3.10)	(3.10)
Round			−1.15***	−1.15***
			(0.16)	(0.15)
GPA			−6.97***	3.54
			(3.11)	(4.04)
GPA x Incomplete				−21.92**
				(5.87)
R-squared	0.006	0.033	0.083	0.088
Observations	1988	1988	1988	1988

Notes: Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All control variables are demeaned.

Table 1.10: Analysis of advantage effects

	Dependent: Group-level effort			
	Complete info.		Incomplete info.	
	(1)	(2)	(3)	(4)
Constant	35.37***	58.53***	42.49***	69.28***
	(2.28)	(3.92)	(2.07)	(3.85)
Value	0.16	−0.91	9.2**	6.68**
	(3.07)	(2.80)	(3.58)	(3.17)
Group	12.95***	8.20***	−2.23	−3.56
	(2.91)	(2.98)	(3.14)	(2.72)
Cost x Advantage	42.42***	41.70***	46.78***	45.20***
	(4.43)	(4.30)	(2.31)	(2.36)
Value x Advantage	38.52***	38.70***	33.16***	33.46***
	(2.61)	(2.48)	(3.33)	(3.21)
Group x Advantage	75.37***	76.29***	78.23***	78.62***
	(4.61)	(4.56)	(4.42)	(4.25)
Experience		−21.03***		−16.56***
		(3.86)		(4.18)
Risk Averse		−1.32		−9.18**
		(3.70)		(3.83)
Female		0.46		15.066***
		(3.44)		(3.48)
Round		−0.91***		−1.24***
		(0.16)		(0.14)
GPA		1.98		−21.81***
		(3.08)		(4.14)
R-squared	0.458	0.494	0.442	0.526
Observations	912	912	1076	1076

Notes: Cluster-robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All control variables are demeaned.

Table 1.11: Probability of winning in *uneven* contests

	Dependent: Group-level effort			
	Complete info.		Incomplete info.	
	(1)	(2)	(3)	(4)
Constant	19.76***	20.39***	32.86***	32.60***
	(1.76)	(1.88)	(1.51)	(1.44)
Value	5.86**	5.23**	4.77**	4.82**
	(3.07)	(2.80)	(3.58)	(3.17)
Group	4.80*	2.24	−8.43***	−7.23***
	(2.71)	(3.03)	(2.39)	(2.54)
Cost x Advantage	60.48***	60.15***	34.28***	35.04***
	(3.53)	(3.57)	(3.02)	(1.98)
Value x Advantage	48.77***	48.79***	24.73***	25.00***
	(3.02)	(2.90)	(3.55)	(2.29)
Group x Advantage	50.89***	51.42***	51.14***	50.99***
	(4.10)	(4.08)	(3.70)	(3.08)
Experience		−4.70		−1.45
		(3.22)		(2.44)
Risk Averse		−5.17*		−5.56**
		(3.05)		(2.60)
Female		7.11*		1.20
		(3.99)		(2.94)
Round		−0.0005		−0.001
		(0.03)		(0.12)
GPA		−2.89		−7.55***
		(2.98)		(2.89)
R-squared	0.735	0.743	0.602	0.616
Observations	422	422	482	482

Notes: Cluster-robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All control variables are demeaned.

Table 1.12: Free-riding behavior and intra-group variation in effort

	(1) Dep Var: Zero Effort	(2) Dep Var: Contri. Variance
Constant	0.213***	118.6***
	(0.028)	(11.01)
Value	0.006	8.780
	(0.030)	(11.29)
Group	0.119***	20.61
	(0.032)	(13.32)
Incomplete	-0.037	12.40
	(0.026)	(10.98)
Advantaged	-0.129***	18.23**
	(0.017)	(7.723)
Uneven	0.083***	-1.140
	(0.013)	(6.688)
Experience	0.078***	-33.44***
	(0.027)	(11.78)
Risk Averse	0.072**	-21.33*
	(0.028)	(11.26)
Female	-0.072**	5.721
	(0.025)	(11.05)
Round	0.007***	0.034
	(0.001)	(0.494)
GPA	0.017	-10.98
	(0.033)	(11.01)
R-squared	0.082	0.015
Observations	7200	7200

Notes: Cluster-robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
All control variables are demeaned.

A.2 Figures

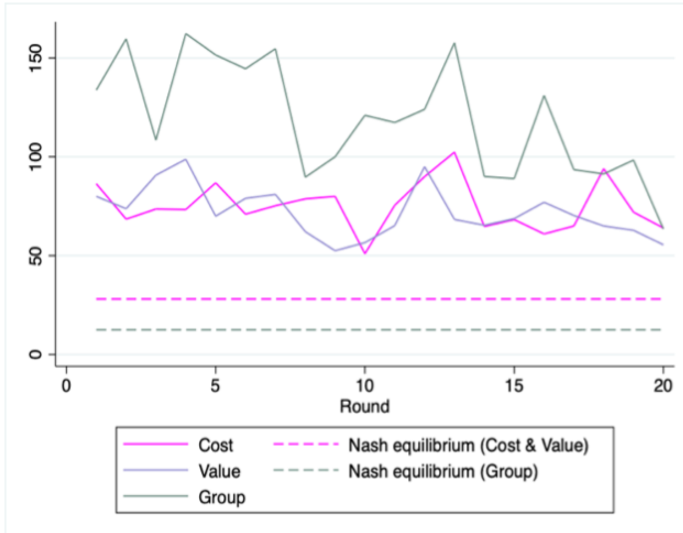

Figure 1.1: Complete Information, Advantaged

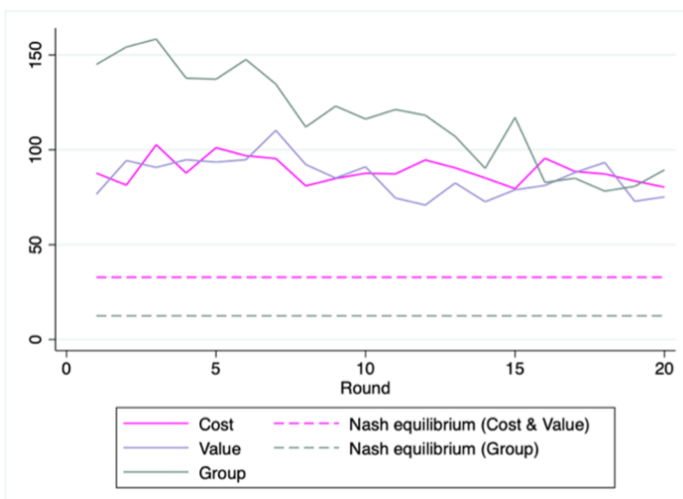

Figure 1.2: Incomplete Information, Advantaged

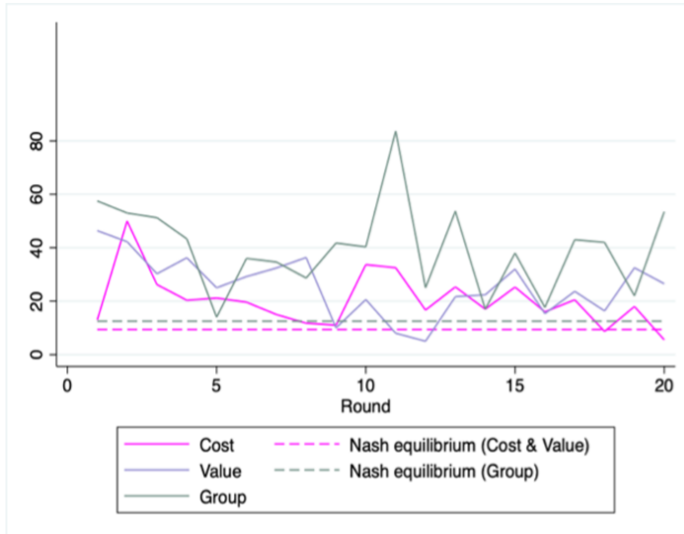

Figure 1.3: Complete Information, Disadvantaged

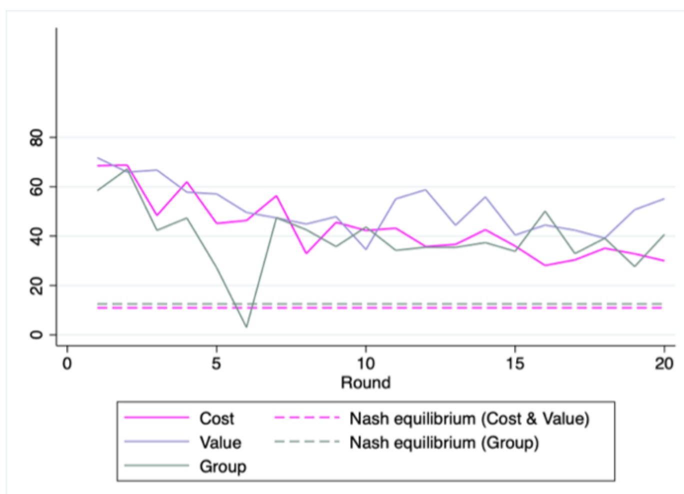

Figure 1.4: Incomplete Information, Disadvantaged

A.3 Theory

Extended Theory Model

The data lend support to a model that includes in-group altruism and, possibly, a non-monetary utility of winning that is proportional to the prize value. For the complete information case, the optimization problem is:

$$\max U_{ig} = [1 + \alpha(N_g - 1)] \frac{X_g}{X_g + X_{-g}} v_g \cdot (1 + \gamma) - c_g x_{i,g} - \alpha \sum_{j \neq i} c_g x_{j,g} \tag{A.1}$$

The first-order condition is:

$$[1 + \alpha(N_g - 1)] v_g \cdot (1 + \gamma) \frac{X_{-g}}{(X_g + X_{-g})^2} = c_g \tag{A.2}$$

The symmetric Nash equilibrium is:

$$X_g^* = \frac{[(1 + \alpha(N_{-g} - 1)](1 + \gamma) v_{-g}}{c_{-g} \left\{ 1 + \frac{v_{-g}[1+\alpha(N_{-g}-1)]c_g}{v_g[1+\alpha(N_g-1)]c_{-g}} \right\}^2}; X_g^* = \frac{[(1 + \alpha(N_{-g} - 1)](1 + \gamma) v_g}{c_g \left\{ 1 + \frac{v_g[1+\alpha(N_g-1)]c_{-g}}{v_{-g}[1+\alpha(N_{-g}-1)]c_g} \right\}^2} \tag{A.3}$$

Importantly, the equilibrium effort is higher for the advantaged group when the source of advantage is group size. Regardless of the source of advantage, relative to the standard model, group effort increases by a factor of $[1 + \alpha(N_g - 1)](1 + \gamma)$. Table A.1 presents the equilibria for each source of advantage.

In the incomplete information setting, the optimization problem is:

$$\begin{aligned} \max_{x_{ig}} U_{ig} = [1 + \alpha(N_g - 1)] \left\{ r \frac{X_g}{X_g + X_A} + (1 - r) \frac{X_g}{X_g + X_D} \right\} v_g \cdot (1 + \gamma) \\ - c_g x_{i,g} - \alpha \sum_{j \neq i} c_g x_{j,g} \end{aligned} \tag{A.4}$$

The first-order condition is:

$$[1+\alpha(N_g-1)]v_g(1+\gamma)\left(r\frac{X_A}{(X_g+X_A)^2}+(1-r)\frac{X_D}{(X_g+X_D)^2}\right)=c_g \tag{A.5}$$

The symmetric Bayesian-Nash equilibrium for $r=\frac{1}{2}$ is:

$$\begin{aligned}
X_A^{**} &= \frac{[1+\alpha(N_A-1)]v_A(1+\gamma)}{c_A}\left\{\frac{\frac{4c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)}+\left(1+\frac{c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)]}\right)^2}{\frac{8c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)]}\left(1+\frac{c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)]}\right)^2}\right\}\\
X_D^{**} &= \frac{[1+\alpha(N_D-1)]v_D(1+\gamma)}{c_D}\left\{\frac{\frac{4c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)}+\left(1+\frac{c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)]}\right)^2}{\frac{8c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)]}\left(1+\frac{c_Av_D[1+\alpha(N_D-1)]}{c_Dv_A[1+\alpha(N_A-1)]}\right)^2}\right\}
\end{aligned} \tag{A.6}$$

As in the case of complete information, relative to the standard model, effort is scaled by a factor of $[1+\alpha(N_g-1)](1+\gamma)$. Table A2 presents the equilibria for each source of advantage.

Table A.1: Group effort: *complete* information

Heterogeneity	Contest Type	Equilibrium effort
Cost-of-effort	Uneven	$(X_A^*, X_D^*) = \left(\frac{[1+\alpha(N-1)](1+\gamma)vc_D}{(c_A+c_D)^2}, \frac{[1+\alpha(N-1)](1+\gamma)vc_A}{(c_A+c_D)^2}\right)$
	Even	$(X_A^*, X_A^*) = \left(\frac{[1+\alpha(N-1)](1+\gamma)v}{4c_A}, \frac{[1+\alpha(N-1)](1+\gamma)v}{4c_A}\right)$; $(X_D^*, X_D^*) = \left(\frac{[1+\alpha(N-1)](1+\gamma)v}{4c_D}, \frac{[1+\alpha(N-1)](1+\gamma)v}{4c_D}\right)$
Prize value	Uneven	$(X_A^*, X_D^*) = \left(\frac{[1+\alpha(N-1)](1+\gamma)v_D v_A^2}{c(v_A+v_D)^2}, \frac{[1+\alpha(N-1)](1+\gamma)v_A v_D^2}{c(v_A+v_D)^2}\right)$
	Even	$(X_A^*, X_A^*) = \left(\frac{[1+\alpha(N-1)](1+\gamma)v_A}{4c}, \frac{[1+\alpha(N-1)](1+\gamma)v_A}{4c}\right)$; $(X_D^*, X_D^*) = \left(\frac{[1+\alpha(N-1)](1+\gamma)v_D}{4c}, \frac{[1+\alpha(N-1)](1+\gamma)v_D}{4c}\right)$
Group Size	Uneven	$X_A^* = \left(\frac{[1+\alpha(N_A-1)]^2[1+\alpha(N_D-1)](1+\gamma)v}{c([1+\alpha(N_A-1)]+[1+\alpha(N_D-1)])^2}\right)$ $X_D^* = \left(\frac{[1+\alpha(N_D-1)]^2[1+\alpha(N_A-1)](1+\gamma)v}{c([1+\alpha(N_A-1)]+[1+\alpha(N_D-1)])^2}\right)$
	Even	$(X_A^*, X_A^*) = \left(\frac{[1+\alpha(N_A-1)](1+\gamma)v}{4c}, \frac{[1+\alpha(N_A-1)](1+\gamma)v}{4c}\right)$; $(X_D^*, X_D^*) = \left(\frac{[1+\alpha(N_D-1)](1+\gamma)v}{4c}, \frac{[1+\alpha(N_D-1)](1+\gamma)v}{4c}\right)$

Notes: An "uneven" contest refers to a case where an advantaged (A) group plays a disadvantaged (D) group. The advantaged team has either a lower cost of effort (i.e., $c_A < c_D$), higher prize value ($v_A > v_D$), or larger group size (i.e., $N_A > N_D$) relative to the disadvantaged team. In an "even" contest, both groups are advantaged or disadvantaged.

Table A.2: Group effort: *incomplete* information

Heterogeneity	Equilibrium effort
Cost-of-effort	$X_A^{**} = \left(\frac{[1+\alpha(N-1)](1+\gamma)v}{c_A} \frac{4\frac{c_A}{c_D}+(1+\frac{c_A}{c_D})^2}{8(1+\frac{c_A}{c_D})^2} \right)$
	$X_D^{**} = \left(\frac{[1+\alpha(N-1)](1+\gamma)v}{c_D} \frac{4\frac{c_A}{c_D}+(1+\frac{c_A}{c_D})^2}{8(1+\frac{c_A}{c_D})^2} \right)$
Prize value	$X_A^{**} = \left(\frac{[1+\alpha(N-1)](1+\gamma)v_A}{c} \frac{4\frac{v_D}{v_A}+(1+\frac{v_D}{v_A}{}^2}{8(1+\frac{v_D}{v_A})^2} \right)$
	$X_D^{**} = \left(\frac{[1+\alpha(N-1)](1+\gamma)v_D}{c} \frac{4\frac{v_D}{v_A}+(1+\frac{v_D}{v_A})^2}{8(1+\frac{v_D}{v_A})^2} \right)$
Group Size	$X_A^{**} = \left(\frac{[1+\alpha(N_A-1)](1+\gamma)v}{c} \frac{4\frac{[1+\alpha(N_D-1)]}{[1+\alpha(N_A-1)]}+\left\{1+\frac{[1+\alpha(N_D-1)]}{[1+\alpha(N_A-1)]})^2\right\}}{8\left\{1+\frac{[1+\alpha(N_D-1)]}{[1+\alpha(N_A-1)]}\right\}^2} \right)$
	$X_D^{**} = \left(\frac{[1+\alpha(N_D-1)](1+\gamma)v}{c} \frac{4\frac{[1+\alpha(N_D-1)]}{[1+\alpha(N_A-1)]}+\left\{1+\frac{[1+\alpha(N_D-1)]}{[1+\alpha(N_A-1)]})^2\right\}}{8\left\{1+\frac{[1+\alpha(N_D-1)]}{[1+\alpha(N_A-1)]}\right\}^2} \right)$

Notes: The equilibrium effort of advantaged and disadvantaged teams are denoted by X_A^{**} and X_D^{**}, respectively. An advantaged team has either a lower cost of effort (i.e., $c_A < c_D$), higher prize value ($v_A > v_D$), or larger group size (i.e., $N_A > N_D$) relative to the disadvantaged team. Equilibria correspond with $r = 1/2$, i.e., that there is a 50% chance the opponent is an advantaged team.

Support of Propositions

Propositions 1 to 3 are based on a standard theory of self-interest. As demonstrated above, the extended theory equilibria are equal to the equilibria from the standard model multiplied by a scale factor that does not vary by information condition. For parsimony, here we prove the three Propositions for the extended model in the case of cost-of-effort heterogeneity when $r = \frac{1}{2}$. Parallel proofs for other sources of heterogeneity follow in a straightforward way. We then present the general solution for $0 < r < 1$, and the results of numerical calculations to provide further support of the Propositions. For convenience, throughout this appendix we define $\tilde{N} \equiv [1 + \alpha(N-1)] \cdot (1+\gamma)$. The standard theory arises when $\alpha = 0$ and $\gamma = 0$, in which case $\tilde{N} = 1$.

<u>Proof of Proposition 1:</u> We claim that contest-level effort in an uneven contest is higher with incomplete information. Using the solutions provided in Tables A.1 and A.2, we then need to show:

$$\frac{\tilde{N}v}{c_A}\left\{\frac{\frac{4c_A}{c_D} + \left(1+\frac{c_A}{c_D}\right)^2}{8\left(1+\frac{c_A}{c_D}\right)^2}\right\} + \frac{\tilde{N}v}{c_D}\left\{\frac{\frac{4c_A}{c_D} + \left(1+\frac{c_A}{c_D}\right)^2}{8\left(1+\frac{c_A}{c_D}\right)^2}\right\} > \frac{c_D v\tilde{N}}{(c_A+c_D)^2} + \frac{c_D v\tilde{N}}{(c_A+c_D)^2} \quad \text{(A.7)}$$

Combining terms, and dividing both sides by $v\tilde{N}$ yields:

$$\left\{\frac{\frac{4c_A}{c_D} + \left(1+\frac{c_A}{c_D}\right)^2}{8\frac{c_A}{c_D}}\right\}\frac{c_A+c_D}{(c_A+c_D)^2} > \frac{c_A+c_D}{(c_A+c_D)^2} \quad \text{(A.8)}$$

Dividing both sides by $\frac{c_A+c_D}{(c_A+c_D)^2}$ yields:

$$\left\{\frac{\frac{4c_A}{c_D} + \left(1+\frac{c_A}{c_D}\right)^2}{8\frac{c_A}{c_D}}\right\} > 1 \quad \text{(A.9)}$$

which simplifies to:

$$\frac{1}{2} + \frac{\frac{(c_A+c_D)^2}{c_D}}{\frac{8c_A}{c_D}} > 1 \tag{A.10}$$

Subtracting $\frac{1}{2}$ from both sides, and then multiplying both sides by $8c_Ac_D$ we obtain:

$$(c_A + c_D)^2 > 4c_Ac_D \tag{A.11}$$

Finally, this inequality simplifies to:

$$(c_A - c_D)^2 > 0 \tag{A.12}$$

which holds true for any $c_A < c_D$.

Proof of Proposition 2: We claim that incomplete information decreases effort in an even contest. Using the solutions provided in Tables A.1 and A.2, we then need to show:

$$\frac{\tilde{N}v}{4c_g} > \frac{\tilde{N}v}{c_g}\left\{\frac{\frac{4c_A}{c_D} + \left(1 + \frac{c_A}{c_D}\right)^2}{8\left(1 + \frac{c_A}{c_D}\right)^2}\right\} \text{ for } g = A, D \tag{A.13}$$

Cancelling terms on both sides, we are left with the following condition:

$$\frac{1}{4} > \left\{\frac{\frac{4c_A}{c_D} + \left(1 + \frac{c_A}{c_D}\right)^2}{8\left(1 + \frac{c_A}{c_D}\right)^2}\right\} \tag{A.14}$$

Expanding the r.h.s. of [A.8], and simplifying, we obtain:

$$\frac{1}{4} > \frac{c_Ac_D}{2(c_A + c_D)^2} + \frac{1}{8} \tag{A.15}$$

Subtracting $\frac{1}{8}$ from both sides, and then multiplying both sides by $8(c_A + c_D)^2$ yields:

$$(c_A + c_D)^2 > 4c_A c_D \tag{A.16}$$

As in the prior proof, this reduces to:

$$(c_A - c_D)^2 > 0 \tag{A.17}$$

which holds true for any $c_A < c_D$.

Proof of Proposition 3: We claim that average contest-level effort is the same under both information conditions when $r = \frac{1}{2}$. When $r = \frac{1}{2}$, there is a 50% chance of an uneven contest, a 25% chance of an even contest among disadvantaged teams, and a 25% chance of an even contest between advantaged teams. Using the equilibria presented in Table A.1, average contest-level effort under the complete information condition is:

$$\frac{1}{2}\left[\frac{c_D v\tilde{N}}{(c_A + c_D)^2} + \frac{c_A v\tilde{N}}{(c_A + c_D)^2}\right] + \frac{1}{4}\left(\frac{v\tilde{N}}{4c_A} + \frac{v\tilde{N}}{4c_A}\right) + \frac{1}{4}\left(\frac{v\tilde{N}}{4c_D} + \frac{v\tilde{n}}{4c_D}\right) \tag{A.18}$$

Rearranging terms,

$$\frac{1}{2}\left[\frac{c_D v\tilde{N}}{(c_A + c_D)^2} + \frac{v\tilde{N}}{4c_A}\right] + \frac{1}{2}\left[\frac{c_A v\tilde{N}}{(c_A + c_D)^2} + \frac{v\tilde{N}}{4c_D}\right] \tag{A.19}$$

Simplifying further and combining terms,

$$\frac{v\tilde{N}}{2}\left[\frac{4c_A c_D + (c_A + c_D)^2}{4c_A(c_A + c_D)^2}\right] + \frac{v\tilde{N}}{2}\left[\frac{4c_A c_D + (c_A + c_D)^2}{4c_D(c_A + c_D)^2}\right] \tag{A.20}$$

Last, multiplying the numerator and denominator of both bracketed terms by $\frac{1}{c_D^2}$, and simplifying, yields:

$$\frac{v\tilde{N}}{c_A}\left[\frac{\frac{4c_A}{c_D}+\left(1+\frac{c_A}{c_D}\right)^2}{8\left(1+\frac{c_A}{c_D}\right)^2}\right]+\frac{v\tilde{N}}{c_D}\left[\frac{\frac{4c_A}{c_D}+\left(1+\frac{c_A}{c_D}\right)^2}{8\left(1+\frac{c_A}{c_D}\right)^2}\right] \tag{A.21}$$

Under incomplete information, expected contest-level effort is simply $X_A^{**} + X_D^{**}$ as in expectation the contest includes one advantaged and one disadvantaged team. From Table A.2, one can easily verify that the first and second terms in [A.21] are the expected effort for advantaged and disadvantaged teams, respectively, for an incomplete information contest.

General solution for group contest with $\mathbf{0 < r < 1}$

Below we derive the closed-form solution for the case of cost-of-effort heterogeneity and $0 < r < 1$. Other cases follow in a similar fashion. First, beginning with the first order condition defined by equation [A.5], if $g = D, N_A = N_D = N$ and $v_A = v_D = v$, then:

$$\{(1-r)(X_D + X_A)^2 + 4rX_AX_D\}v\tilde{N} = 4X_Dc_D(X_A + X_D)^2 \tag{A.22}$$

In a similar vein, if $g = A, N_A = N_D = N$, and $v_A = v_D = v$, it follows that:

$$\{4(1-r)X_AX_D + r(X_D + X_A)^2\}v\tilde{N} = 4X_Ac_A(X_A + X_D)^2 \tag{A.23}$$

This gives us two equations and two unknowns. Dividing [A.22] by [A.23], and rearranging yields:

$$X_A = \frac{(1-r)}{r}\frac{c_D}{c_A}X_D - \frac{(1-2r)}{r}\frac{v\tilde{N}}{4c_A} \tag{A.24}$$

In the special case of $r = \frac{1}{2}$, the second term equals 0 and this yields the simple relationship $X_A = \frac{c_D}{c_A}X_D$. For convenience, let $\delta = \frac{1-r}{r}\cdot\frac{c_D}{c_A}$ and $\theta = \frac{2r-1}{r}\cdot\frac{1}{4c_A}$, in which case [A.24] can be written as:

$$X_A = \delta X_D + v\tilde{N}\theta \tag{A.25}$$

Now, substitute [A.25] into [A.22] to eliminate X_A:

$$\begin{aligned} &\{(1-r)(X_D + \delta X_D + v\tilde{N}\theta)^2 + 4r(\delta X_D + v\tilde{N}\theta)X_D\}v\tilde{N} \\ &= 4X_D c_D(X_D + \delta X_D + v\tilde{N})^2 \end{aligned} \tag{A.26}$$

Rearranging and combining terms in [A.26], we obtain the following cubic equation:

$$aX_D^3 + bX_D^2 + cX_D + d = 0, \tag{A.27}$$

where, $a = c_D(\delta+1)^2$, $b = v\tilde{N}(2c_D(\delta+1)\theta - r\delta - \frac{1}{4}\frac{1-r}{(\delta+1)^2})$, $c = (v\tilde{N})^2(c_D\theta^2 - r\theta - \frac{1}{2}(1-r)(\delta+1)\theta)$ and $d = -\frac{1}{4}(v\tilde{N})^3(1-r)\theta^2$. Last, dividing through by the coefficient a yields:

$$X_D^3 + a_1 X_D^2 + a_2 X_D + a_3 = 0, \tag{A.28}$$

where $a_1 = \frac{b}{a}$, $a_2 = \frac{c}{a}$ and $a_3 = \frac{d}{a}$. Applying established methods for solving a cubic equation (i.e., using a variant of Cardano's formula), the equation [A.28] has three real roots when $r \neq \frac{1}{2}$. The one root that satisfies the first-order condition of the maximization problem is:

$$X_D = 2\sqrt{-Q}\ \cos\ (\frac{1}{3}\nu) - \frac{1}{3}a_1 \text{ and } X_A = \delta\big(2\sqrt{-Q}\ \cos\ (\frac{1}{3}\nu) - \frac{1}{3}a_1\big) + v\tilde{N}\theta \tag{A.29}$$

where, $Q = \frac{(3a_2 - a_1^2)}{9}$, $R = \frac{9a_1a_2 - 27a_3 - 2a_1^3}{54}$, and $\nu = arccos(\frac{R}{\sqrt{-Q^3}})$. In the case of $r = \frac{1}{2}$, there are two real roots, but only one of them is non-zero. The solution in this case is:

$$X_D = 2R^{1/3} - \frac{1}{3}a_1 \text{ and } X_A = \delta(2R^{1/3} - \frac{1}{3}a_1) + v\tilde{N}\theta \tag{A.30}$$

Here, $R^{1/3} = -\frac{a_1}{3}$, and it follows that $X_D = -a_1$ which simplifies to the formulas presented in Table A.2.

Support of Propositions 1, 2 and 3 for $\mathbf{0 < r < 1}$

As mentioned earlier in the theory section, for uneven contests, incomplete information increases contest-level effort, and that the effect is increasing in r and extent of the advantage. Note that for an advantaged team, effort is increasing under incomplete information for any r. However, for the disadvantaged team, in general, the effect is ambiguous and depends upon the extent of the advantage together with the probability that the other team is advantaged. When the advantage is relatively small, the discouragement effect discussed previously is also small. Then only for very high r does incomplete information motivate lower effort. As the size of the advantage increases, however, the range of probabilities for which incomplete information discourages effort increases. Overall, the effect of incomplete information on the advantaged group unambiguously dominates its effect on the disadvantaged group and so more generally the contest-level effort is increasing under incomplete information.

For even contests, incomplete information decreases group-level effort. With incomplete information, a team does not know the opposing team's type. An advantaged team will only suspect they are playing another advantaged team with some probability less than 1, and as a result will be incentivized to put forth less effort relative to the case where the opponent is for sure advantaged. A disadvantaged team will suspect their opponent may be advantaged, and this also lowers effort relative to the case where they know for sure the opponent is disadvantaged. This is due to the discouragement effect.

When considering contest-level effort, unconditional on contest type, the differential effects of incomplete information across uneven and even contests of course will counteract. When the probability a team is advantaged is exactly 50%, there is no difference in expected effort between contests with complete and incomplete information. But, as the (negative) effect of incomplete effort in even contests between two disadvantaged teams is relatively small, for $r < \frac{1}{2}$ it is the case that expected effort is higher with incomplete information. This is because for $r < \frac{1}{2}$ the positive effect in uneven contests dominates the negative effect in even (disadvantaged) contests. The opposite is true when conditions make it more probable that the contest is between two advantaged teams, i.e., when $r > \frac{1}{2}$. Although the effects on expected effort (unconditional on contest type) are in general ambiguous, differences are relatively small.

As illustrated in Tables A.1 and A.2, under cost heterogeneity, the solutions for both the complete and incomplete information settings can be written as $X_g^{**} = v\tilde{N} \cdot f_g$ where the argument f_g is not a function of the altruism, non-monetary utility of winning, group size and prize value parameters. As a result, these parameters do not independently determine differences in effort across the information conditions. This remains true in the general case.[1] As such, any differences based on information condition depend on the extent of the cost advantage and r. Without loss of generality, we can normalize $c_D \equiv 1$ in which case $0 < c_A < 1$ and the size of the advantage is decreasing in c_A. It then suffices to show that the propositions hold for all possible combinations of c_A and r.

Presented as Figures A.1 to A.3 are surface plots, for specific contest types, of contest-level effort in the incomplete information contest minus the contest-level effort in the complete information contest for the case of cost-heterogeneity. These are based on $\tilde{N} = 3$ and $v = 50$. Figure A.1 corresponds to uneven contests, and is thus relevant for Proposition 1. The effort difference is always positive, and is strictly increasing in both the size of the cost advantage and the probability the opponent is an advantaged

[1]To see this, note that we can write $Q = (v\tilde{N})^2 \cdot f_1$, $a_1 = v\tilde{N} \cdot f_2$, and $R = (v\tilde{N})^3 \cdot f_3$, where f_1, f_2, and f_3 are functions that do not contain v or $\tilde{N}$. Then, [A.29] becomes $X_D = v\tilde{N}\left\{2\sqrt{-f_1}cos\left(\frac{1}{3}arccos\left(\frac{f_3}{\sqrt{-f_1^3}}\right)\right) - \frac{1}{3}f_2\right\}$.

team. Figures A.2 and A.3 correspond to even contests between advantaged and disadvantaged teams, respectively. Confirming Proposition 2, contest-level effort is strictly higher under complete information. For a contest between advantaged teams, this difference goes to zero as $r \to 1$, as expected, as in this limit the contest is a complete information contest between advantaged teams. The effect of information is maximal when both r and c_A approach zero. For a contest between disadvantaged teams, this difference goes to zero as $r \to 0$, as this converges to a certain contest between two disadvantaged teams. The effect of information is maximal when both r approaches 1 and c_A approaches zero. Effort differences are relatively larger for even contests that involve two advantaged teams.

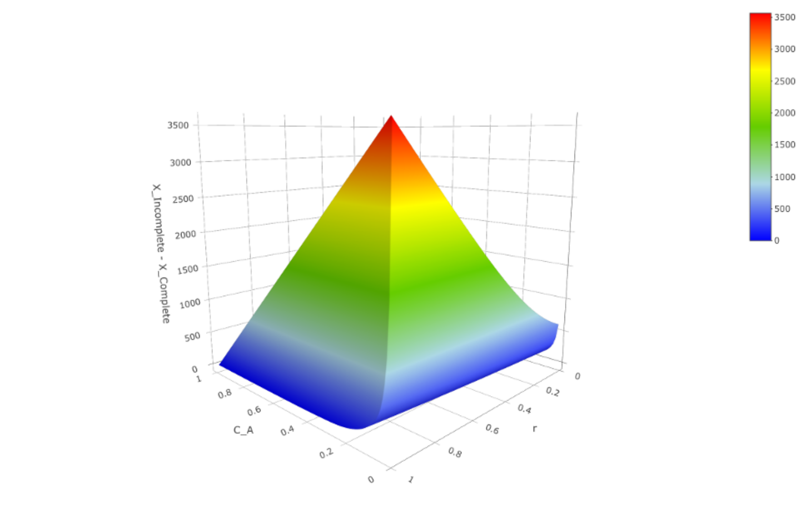

Figure A.1: Differences in contest-level efforts based on information condition: *uneven* contests

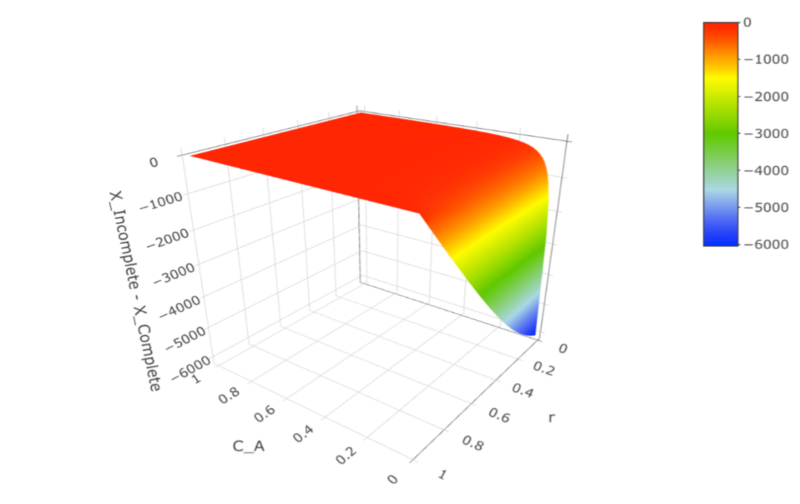

Figure A.2: Differences in contest-level efforts based on information condition: *even* contests between advantaged teams

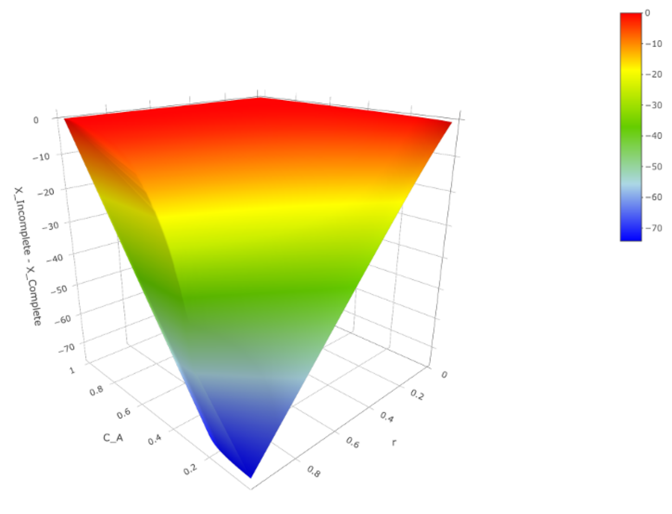

Figure A.3: Differences in contest-level efforts based on information condition: *even* contests between disadvantaged teams

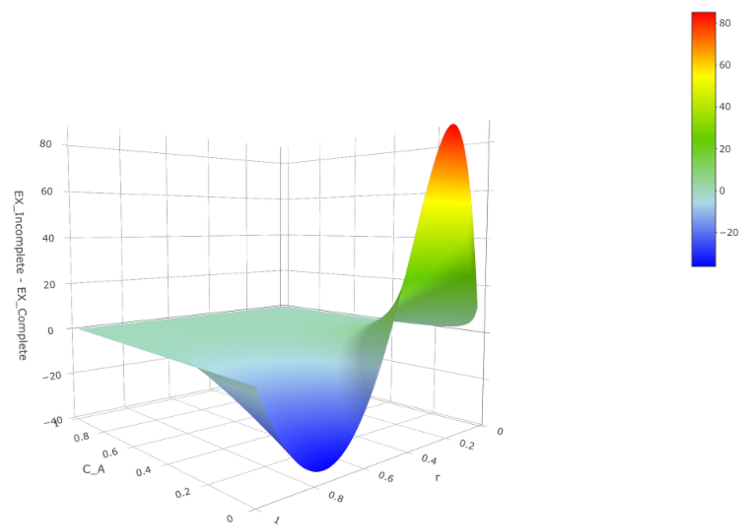

Figure A.4: Differences in expected contest-level efforts between incomplete and complete information conditions

Figure A.4 depicts differences in expected contest-level effort between the two information conditions. To be clear, this differs from the information provided in Figures A.1 to A.3 as effort is unconditional on contest type (e.g., even or uneven). When $r = \frac{1}{2}$, there is no difference in contest-level effort as proven analytically. As r deviates from this value, differences in expected effort arise due to information conditions but in general these differences are small when compared with the differences that arise from uneven contests and even contests between advantaged teams. The largest differences occur when $c_A \to 0$.

Deviating from $r = \frac{1}{2}$ in either direction increases the probability of an even contest, and from the prior results specific to contest types this would suggest expected contest-level effort would be higher with complete information. However, there turns out to be an asymmetry which is largely due to the fact that effort in an even contest between advantaged teams is considerably higher with complete information (see Figure A.2), but the information effect is relatively small for an even contest between disadvantaged teams (see Figure A.3). As a result, when $r > \frac{1}{2}$ and it becomes more likely that an even contest between advantaged teams will occur, overall effort is higher with complete information. On the other hand, when $r < \frac{1}{2}$ and it becomes more likely that an even contest between two disadvantaged teams will occur, expected effort is higher with incomplete information. Holding c_A fixed, the largest differences do not necessarily occur as r approaches 1 or 0 as there are competing effects. For instance, with $r > \frac{1}{2}$, while increasing r does increase the chance of an even contest between advantaged teams, as a countervailing effect the difference in effort for an uneven contest under incomplete versus complete information is also increasing with r.

A.4 Additional econometric analysis

Table A.3: Analysis of information effects: *uneven* contests, restricted sample

	Dependent: Group-level effort		
	(1)	(2)	(3)
Constant	56.80***	47.57***	48.00***
	(1.88)	(2.34)	(4.90)
Value		1.30	1.57
		(3.22)	(3.09)
Group		32.79***	27.78***
		(4.38)	(4.34)
Cost x Incomplete		20.99***	20.93***
		(3.24)	(3.09)
Value x Incomplete		19.35***	19.46***
		(3.87)	(3.41)
Group x Incomplete		−4.17	1.32
		(5.22)	(4.55)
Incomplete	13.48***		
	(2.55)		
Experience			−17.82***
			(4.55)
Risk Averse			−4.25
			(4.53)
Female			10.65**
			(4.92)
Round			−1.34***
			(0.18)
GPA			−8.96*
			(4.90)
R-squared	0.025	0.081	0.136
Observations	904	904	904

Notes: Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All control variables are demeaned. Estimation sample excludes observations from incomplete information treatments associated with even contests.

Table A.4: Analysis of information effects: *even* contests, restricted sample

	Dependent: Group-level effort		
	(1)	(2)	(3)
Constant	66.52***	63.77***	65.27***
	(2.21)	(3.79)	(3.61)
Value		−2.27	−3.71
		(4.74)	(4.54)
Group		11.12*	7.02
		(5.93)	(5.90)
Cost x Incomplete		3.05	3.55
		(4.75)	(4.43)
Value x Incomplete		6.27	6.47*
		(4.07)	(3.89)
Group x Incomplete		−10.81	−7.95
		(6.57)	(6.52)
Incomplete	−0.14		
	(2.96)		
Experience			−17.22***
			(3.81)
Risk Averse			−7.07*
			(3.94)
Female			8.34**
			(3.86)
Round			−1.01***
			(0.24)
GPA			−7.20*
			(4.10)
R-squared	0.000	0.011	0.058
Observations	1084	1084	1084

Notes: Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All control variables are demeaned. Estimation sample excludes observations from incomplete information treatments associated with uneven contests.

Table A.5: Analysis of information effects: *even* contests, advantaged groups only

	Dependent: Group-level effort		
	(1)	(2)	(3)
Constant	89.61***	80.80***	82.95***
	(3.35)	(5.88)	(5.50)
Value		−3.95	−5.15
		(6.52)	(6.16)
Group		47.70***	43.15***
		(8.27)	(8.15)
Cost x Incomplete		8.47	7.40
		(6.09)	(5.70)
Value x Incomplete		8.00**	8.28**
		(4.04)	(3.63)
Group x Incomplete		−10.01	−6.77
		(7.05)	(6.79)
Incomplete	3.93		
	(3.76)		
Experience			−22.36***
			(4.09)
Risk Averse			−5.396
			(4.54)
Female			15.26**
			(4.61)
Round			−1.20***
			(0.21)
GPA			−10.80**
			(4.44)
R-squared	0.002	0.187	0.264
Observations	729	729	729

Notes: Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All control variables are demeaned. Estimation sample excludes observations from incomplete information treatments associated with uneven contests.

Table A.6: Information effects with time trend: *uneven* contests

	Dependent: Group-level effort		
	(1)	(2)	(3)
Constant	56.80***	47.57***	52.48***
	(1.88)	(2.33)	(4.31)
Value		1.30	7.19
		(3.22)	(6.41)
Group		32.79***	45.65***
		(4.37)	(8.43)
Cost x Incomplete		20.01***	29.53***
		(3.01)	(5.22)
Value x Incomplete		19.12***	15.60**
		(3.37)	(6.71)
Group x Incomplete		−11.19**	−4.14
		(4.83)	(9.34)
Incomplete	11.33***		
	(2.34)		
Round x Cost			−0.49
			(0.37)
Round x Value			−1.01***
			(0.36)
Round x Group			−1.79***
			(0.55)
Round x Cost x Incomplete			−0.89*
			(0.47)
Round x Value x Incomplete			−0.31
			(0.55)
Round x Group x Incomplete			−0.54
			(0.70)
R-squared	0.016	0.050	0.091
Observations	1498	1498	1498

Notes: Cluster-robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.7: Information effects with time trend: *even* contests

	Dependent: Group-level effort		
	(1)	(2)	(3)
Constant	66.52***	63.77***	73.67***
	(2.21)	(3.79)	(7.75)
Value		−2.27	−5.58
		(4.73)	(10.19)
Group		11.12*	6.69
		(5.92)	(12.10)
Cost x Incomplete		3.81	8.34
		(4.24)	(8.30)
Value x Incomplete		6.49*	7.19
		(3.80)	(8.13)
Group x Incomplete		−5.72	13.62
		(5.51)	(11.00)
Incomplete	1.60		
	(2.61)		
Round x Cost			−0.91
			(0.62)
Round x Value			−0.64
			(0.53)
Round x Group			−0.52
			(0.84)
Round x Cost x Incomplete			−0.47
			(0.68)
Round x Value x Incomplete			−0.06
			(0.68)
Round x Group x Incomplete			−1.81*
			(0.96)
R-squared	0.000	0.008	0.046
Observations	1566	1566	1566

Notes: Cluster-robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

A.5 Experiment instructions and post-experiment questionnaire

Experiment Instructions for cost treatment with incomplete information

Thank you for participating in today's study. Please follow the instructions carefully. At any time, please feel free to raise your hand if you have a question.

You have been randomly assigned an ID number for this session. You will make decisions using a computer. You will never be asked to reveal your identity to anyone. Your name will never be associated with any of your decisions. In order to keep your decisions private, please do not reveal your choices or otherwise communicate with any other participant. Importantly, please refrain from verbally reacting to events that occur.

Today's session has three parts: Experiment 1, Experiment 2, and a short questionnaire. You will have the opportunity to earn money in both experiments based on your decisions. You will be paid your earnings privately, and in cash, at the end of the experiment session. We will proceed through the written materials together. Please do not enter any decisions on the computer until instructed to do so.

Are there any questions before we begin?

Please go ahead and click "Continue" to enter the experiment.

Experiment 1

Please click "Continue" and refer to your computer screen while we read the instructions.

We would like you to make a decision for each of 10 scenarios. Each scenario involves a choice between playing a lottery that pays $4 or $0 according to specified chances (Option A) or receiving $2 for sure (Option B).

You will notice that the only differences across scenarios are the chances of receiving the high or low prize for the lottery. At the end of the today's session, ONE of the 10 scenarios will be selected at random and you will be paid according to your decision for this selected scenario ONLY. Each scenario has an equal chance of being selected.

Please consider your choice for each scenario carefully. Since you do not know which scenario will be played out, it is in your best interest to treat each scenario as if it will be the one used to determine your earnings.

Before making decisions, are there any questions?

Once you are ready to submit your decisions, please click the "Submit" button.

Experiment 2

In this experiment, all money amounts are denominated in lab dollars, and will be exchanged at a rate of 90 lab dollars to 1 US dollar at the end of the experiment.

There will be many decision rounds in the experiment. You will not know the number of rounds until the experiment has been completed. Each decision round is separate from the other rounds, in the sense that the decisions you make in one round will not affect the outcome or earnings of any other round.

In each round, participants will be randomly placed into three-person groups.

In each decision round, your group will compete with one other group to determine which group wins a prize of 150 lab dollars. This prize will be evenly divided among all group members. If your group wins the prize, you will personally receive 150/3 or 50 lab dollars.

Your task in each decision round is to decide how many points to contribute towards a group project. Which group wins the prize depends upon the total contributions from your group relative to the total contributions of the opponent group. The chance your group wins the prize is determined by the following formula:

$$\boxed{\text{Chance of winning} = \frac{\text{Total contributions (Own)}}{\text{Total contributions (Own)} + \text{Total contributions (Opponent)}} \cdot 100\%}$$

Using this formula:

- If the total contributions from both groups are equal, then both groups have an equal chance of winning the prize; i.e., the chance each group wins the prize is 50%.
- If your group contributes more than your opponent, then your group has a higher chance of winning the prize. For example, if your group contributes twice as much, the chance your group wins the prize is 2 in 3 or 66.7%.
- If your group contributes less than your opponent, then your group has a lower chance of winning the prize. For example, if the opponent group contributes four times as much as your group, your group has a 1 in 5 or 20% chance to win.

You can contribute anywhere from 0 to 50 points (only in integer amounts) towards the group project.

While increasing contributions will increase the chance your team wins the prize, contributing points costs money. In particular, each point you contribute is associated with a per-point contribution cost.

The per-point contribution cost can have two values: either 1/3 of a lab dollar or 1 lab dollar. You will know the contribution cost when deciding.

In each round, you will receive 50 lab dollars in fixed income. This amount does not depend on your decision or whether your group wins this prize. Your earnings for the decision round will be calculated as follows:

IF your group wins... $\boxed{\text{Your Earnings} = 100 - (\text{points YOU contributed} * \text{cost})}$

IF your group does not win... $\boxed{\text{Your Earnings} = 50 - (\text{points YOU contributed} * \text{cost})}$

Before we continue, are there any questions?

Instructions quiz

At this time, we would like you to answer a few questions to help you understand how the experiment works. The good news is that you will be paid for correct answers. You may wish to first answer these using pen and paper. When you are ready, please read the instructions on your computer carefully, and click "I understand, Continue to Quiz" to submit your answers on the computer. If you have a question when working through the quiz, please raise your hand and your question will be answered privately.

1. Suppose the contribution cost is 1/3 of a lab dollar per point. You contribute 18 points. Your group wins the prize. How much money would you earn for this decision round (in lab dollars)?

a. 27 b. 44 c. 70 d. 94

2. If your group contributes a total of 60 points and the opponent group contributes a total of 100 points, what is the chance your group wins the prize?

a. 62.5% b. 37.5% c. 0% d. 50%

3. Suppose the contribution cost is 1 lab dollar per point. You contribute 40 points, and the total contributions from your group (including your own) are 50 points. Your group does not win the prize. How much money would you earn for this decision round (in lab dollars)?

a. 30 b. –40 c. 10 d. 0

4. Suppose the other two members of your group contribute a total of 20 points. The opponent group contributes 20 points. Therefore, if you contribute nothing your

group has a 50% chance of winning. By how much would you increase the chance your group wins if you contribute 10 points instead of contribute nothing?

a. 0% b. 5% c. 60% d. 10%

Proceeding through the experiment

At the start of each round, you will be randomly matched into a group of three players. Your group will then be randomly matched with another group. This means that both the members of your own group as well as the members of the opponent group will vary from one round to the next.

At the start of each round, the computer will randomly determine the contribution cost for each group. Both groups will each have a 50% chance of facing the low or high contribution cost. This random determination is done independently for each group, which means that in some rounds your contribution cost will be the same as your opponent, and in other rounds it will be different. In particular:

- There will be a 25% chance that both your group and the group you are competing with have a low contribution cost (1/3 of a lab dollar);
- There will be a 25% chance that both groups have a high contribution cost (1 lab dollar); and,
- There will be a 50% chance that one group will have a low cost while the other has a high cost.

You will always know the contribution cost for your group. Throughout the experiment, however, you will not know the contribution cost for the opponent group.

Note: In the corresponding complete information treatment, the above two sentences are replaced with: "You will always know the contribution cost for your group and the opponent group."

Your decision screen will include relevant information for both your own group and the opponent group. Know that the prize value and group size will never change during the experiment.

At the end of each decision round you will be shown a result screen with the contest result, the total points contributed by all your group members, and your earnings.

We will begin with a training round to help you understand the procedures.

Aside from decisions in this training round, you will be paid based on the outcome of each decision round. This means that it is very important to consider each decision prior to making it.

Before we continue, do you have any questions?

Post-experiment questionnaire (computerized)

Part 1: About the Experiment

We would now like for you to complete a short questionnaire. Please know that all responses will be treated as strictly confidential and will be used for statistical purposes only. The first questions relate to your experience in today's experiment.

1. Have you previously participated in a paid study that took place in an experimental economics laboratory?

a. Yes b. No

2. Please indicate your level of agreement with the following statement: "I understood well the instructions for Experiment 2."

1 - Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 - Strongly Agree

3. Please indicate your level of agreement with the following statement: "I was well compensated for my participation in this study."

1 - Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 - Strongly Agree

4. In the past twelve months, approximately how much money (cash, check, credit card, etc.) did you donate to a charity or non-profit organization?

5. In the past twelve months, what is the approximate fair market value of non-cash property (clothing, appliances, etc.) you donated to a charity or non-profit organization?

6. In the past twelve months, approximately how many hours did you spend doing volunteer work for a charity or non-profit organization?

7. Many classes at the University of Tennessee require students to work on assignments in groups. In these settings, do you usually contribute less, about the same, or more than other people in your group?

a. Less b. About the same c. More

Please use the following space to write any comments (positive or negative) you may have about the experiment.

Part 2: Demographics

The next questions tell us something about you.

1. What is your age?

2. How do you describe yourself?

a. Male b. Female c. Transgender d. Do not identify myself as female, male, or transgender

3. What is your academic major?

4. What is your current student classification?

a. Freshman b. Sophomore c. Junior d. Senior e. Master's Student f. Law Student g. Doctoral Student h. Other

5. What was your student status for the Spring 2019 semester?

a. Full-time student b. Part-time student c. Not a student

6. In what range is your cumulative GPA?

a. 0 to 2.0; b. 2.1 to 2.5; c. 2.6 to 3.0; d. 3.1 to 3.5; e. 3.6 to 4.0

7. How many economics courses have you completed at the university level?

8. How would you best describe your current employment status?

a. Employed Full-Time; b. Employed Part-Time; c. Self-Employed Full-Time; d. Self-Employed Part-Time; e. Unemployed

Part 3: Personality

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other. All questions below are to be rated from 1-7. 1 represents strongly disagree and 7 represents strongly agree.

I see myself as:

a. Extroverted, enthusiastic
b. Critical, quarrelsome
c. Dependable, self-disciplined
d. Anxious, easily upset
e. Open to new experiences, complex

f. Reserved, quiet

g. Sympathetic, warm

h. Disorganized, careless

i. Calm, emotionally stable

j. Conventional, uncreative

Appendix B

Appendix

B.1 Tables

Table 2.1: Description of data

Variable Name	Description	Mean	S.D.
Treatment variables			
IC	= 1 when elicitation is real; 0 otherwise;	0.44	0.50
HYP	= 1 when elicitation is hypothetical; 0 otherwise	0.44	0.50
Oath	= 1 when the oath script is employed; 0 otherwise	0.45	0.50
Control variables			
Risk Averse	= 1 if participant selected safe option at least six times in Risk Elicitation task; 0 otherwise	0.48	0.50
Experience	=1 if the participant had partaken in a prior economics experiment; 0 otherwise	0.61	0.49
Female	= 1 if participant is female; 0 otherwise	0.57	0.50
Employed	= 1 if participant is partly or fully employed, 0 otherwise	0.58	0.49
ASU	= 1 if participant is from ASU pool, 0 otherwise	0.20	0.40
Age	Recorded age of the participant	20.90	2.61
GPA	Participant GPA, recorded as midpoint of chosen interval	3.39	0.46
Earnings	Participants earnings from the experiment in $	27.14	4.34
Comprehension	Rating of instruction comprehension, scale 1 to 5	4.43	0.99

Table 2.2: Percentage of "yes" votes

	Treatment				
	HYP	HYP + Oath	IC	IC + Oath	IC + CT
\$0	94.78	95.72	93.91	93.16	94.82
\$1	94.78	91.45	80.00	82.05	86.20
\$2	92.17	88.03	66.95	74.35	75.86
\$3	87.82	87.17	53.91	64.10	67.24
\$4	86.08	81.19	44.34	58.11	53.44
\$5	76.52	75.21	33.04	47.86	44.82
\$6	62.60	55.55	26.08	32.47	25.86
\$8	47.82	48.71	18.26	22.22	15.51
\$10	41.73	31.62	12.17	15.38	10.34
\$12	35.65	26.49	10.43	11.96	3.44
\$15	33.91	24.78	7.82	11.96	1.72
Overall	68.53	64.18	40.63	46.69	43.57

Table 2.3: Willingness to pay regressions

Dependent Variable: Latent willingness to pay (WTP)		
	Pooled Probit Regression	
	(1)	(2)
Conditional mean function (σ) :		
Intercept (IC)	4.17***	4.30***
	(0.451)	(0.489)
IC + Cheap-Talk	0.47	0.14
	(0.663)	(0.720)
IC + Oath	1.06*	0.77
	(0.662)	(0.690)
HYP	5.63***	5.57**
	(0.798)	(0.845)
HYP + Oath	4.46***	4.41**
	(0.719)	(0.747)
Standard deviation function (σ) :		
IC + Cheap-Talk	−0.30***	−1.69**
	(0.112)	(0.108)
IC + Oath	0.04	0.01
	(0.140)	(0.141)
HYP	0.24	0.218
	(0.158)	(0.162)
HYP + Oath	0.13	0.05
	(0.147)	(0.143)
Controls	-	✓
Observations	5742	5610

Notes: All specifications allow for unequal variances by treatment. All control variables are demeaned so that estimated coefficients have a consistent interpretation across specifications. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.4: Follow-up certainty question responses (IC versus HYP)

Cost	**$0**	**$1**	**$2**	**$3**	**$4**	**$5**	**$6**	**$8**	**$10**	**$12**	**$15**
						Fisher exact test (p-value)					
"yes"	0.815	0.197	**0.098**	**0.012**	0.256	0.273	0.853	0.829	0.400	0.432	0.441
"no"	0.441	**0.070**	**0.030**	**0.033**	**0.010**	**0.009**	**0.072**	**0.001**	**0.015**	**0.038**	**0.029**
						Mean difference in certainty response (IC-HYP)					
"yes"	0.15	-0.05	-0.12	-0.30	-0.19	-0.38	-0.01	0.42	0.03	0.48	-0.51
"no"	2.67	1.92	1.43	1.47	1.35*	1.78**	1.44**	1.41**	1.17**	0.75*	1.02**

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Certainty-adjusted WTP estimates, recoding uncertain "yes" as "no"

Certainty-adjusted *hypothetical* WTP estimates				
	Cert.≥7	Cert.≥8	Cert.≥9	Cert.≥10
HYP treatment, adjusted	7.75***	6.89***	5.74***	4.46***
	(0.575)	(0.574)	(0.566)	(0.608)
HYP (adjusted) - IC	3.59***	2.73**	1.58**	0.29
	(0.731)	(0.730)	(0.724)	(0.757)
Certainty-adjusted *incentive-compatible* WTP estimates				
IC treatment, adjusted	3.35***	3.05***	2.54***	2.09***
	(0.387)	(0.362)	(0.360)	(0.359)
IC (adjusted) - IC	−0.82***	−1.11***	−1.62***	−2.08***
	(0.206)	(0.225)	(0.360)	(0.299)

Notes: Cluster-robust standard errors in parentheses. Column headings refer to the rule used to recode "yes" votes into "no" votes, e.g., "Cert≥7" means that all "yes" votes are recoded as "no" except for those associated with a certainty level of 7 or higher. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.6: Certainty-adjusted WTP estimates, using certainty levels as probabilities

Certainty-adjusted *hypothetical* WTP estimates		
	ASUM	SUM
HYP treatment, adjusted	8.79***	10.14***
	(0.287)	(0.352)
HYP (adjusted) - IC	4.63***	5.97***
	(0.747)	(0.755)
Certainty-adjusted *incentive-compatible* WTP estimates		
IC treatment, adjusted	3.78***	4.69***
	(0.213)	(0.242)
IC (adjusted) - IC	−0.38***	0.53***
	(0.070)	(0.137)

Notes: Cluster-robust standard errors in parentheses. ASUM refers to "asymmetric uncertainty model" and SUM refers to "symmetric uncertainty model". See text for details. *** p<0.01, ** p<0.05, * p<0.1.

B.2 Exploratory Analysis

Table B.1: Willingness to pay regressions (split by participant pool and gender)

Dependent Variable: Latent willingness to pay (WTP)				
	Pooled Probit Regression			
	(1)	(2)	(3)	(4)
Conditional mean function (σ) :				
Intercept (IC)	4.17***	4.30***	3.97***	3.97***
	(0.451)	(0.489)	(0.501)	(0.735)
IC + Cheap-Talk	0.47	0.14	0.63	0.11
	(0.663)	(0.720)	(0.802)	(0.999)
IC + Oath	1.06*	0.77	1.10	0.31
	(0.662)	(0.690)	(0.746)	(1.053)
HYP	5.63***	5.57**	5.67***	5.41***
	(0.798)	(0.845)	(0.850)	(1.245)
HYP + Oath	4.46***	4.41**	3.91***	2.71***
	(0.719)	(0.747)	(0.771)	(1.003)
ASU × IC			5.19***	
			(0.960)	
ASU × IC + Cheap-Talk			−0.52	
			(1.230)	
ASU × IC + Oath			0.51	
			(1.395)	
ASU × HYP			5.78***	
			(1.943)	
ASU × HYP + Oath			6.24***	
			(1.530)	
Female × IC				4.42***
				(0.653)
Female × IC + Cheap-Talk				0.66
				(0.940)
Female × IC + Oath				1.50*
				(0.910)
Female × HYP				5.66***
				(1.030)
Female × HYP + Oath				5.76***
				(0.994)
Other controls	-	✓	-	-
Observations	5742	5610	5742	5742

Notes: All specifications allow for unequal variances by treatment. Some coefficients are suppressed for brevity. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

B.3 Oath Script

We will present you with a form to sign. Please know that you are free to sign the form or not, and that your participation and earnings do not depend on this choice. The other participants will not be told whether you signed the form.

I undersigned, ______________ , swear upon my honor that, during the whole experiment I will:

Tell the truth and always provide honest answers.

Please sign below:

____

B.4 Pseudo-cheap talk script

Research suggests that people in experiments are more inclined to contribute money towards public goods. In a voting setting like ours, this means that more people would vote "yes" to pay for the tree plantings, at any given cost amount, than would if this were not an experiment. The question is then how we can get people to think about their votes in this experiment as they would if voting outside an experiment?

Let me tell you why I think that we see differences in behavior. I think that when we hear about a referendum that involves something that is basically good – helping people in need, improving environmental quality, or anything else – our basic reaction is to think: sure, I want to do this. I really want to vote "yes" to spend some of the money I have earned from the experiment.

But when making similar choices outside an experiment, we think differently about how we spend our own money to pay for something. We basically still would like to see good things happen, but when we are faced with the possibility of having to spend money that we didn't earn from an experiment, we think about our options differently: if I spend money on this, that's money I don't have to spend on other things. In other words, we are more likely to vote in a way that takes into account the limited amount of money we have. It is also possible that some people vote "yes" in an experiment because of social pressure, for example they think the experimenter or other participants want them to vote "yes". That is just my opinion, of course, but it's what I think may be going on in experiments like this.

So, if I were in your shoes, I would ask myself: if I were voting outside the experiment, and I had to pay money if the referendum passed, would I really want to spend my money this way? If I really did, I would vote "yes"; if I didn't, I would vote "no". In any case, I ask you to vote just exactly as you would if you were making decisions outside of an experiment. Please keep this in mind when voting in our referendum.

B.5 Experiment instructions

Instructions for incentive compatible treatments

In this experiment, you will be asked to vote in a referendum on whether all participants in the room will collectively fund an actual tree-planting project. If the referendum passes, you and the other participants will pay for the tree plantings using some of the money you have earned in the prior experiments. If the referendum does not pass, no money will be collected from you and no trees will be planted.

About the project . . .

The project involves planting and maintaining 160 trees in the Appalachian Mountains, a region that stretches from southern New York state to northern Alabama and Georgia. To carry out the tree planting, we will partner with the non-profit organization One Tree Planted. This organization plants trees to return formerly unproductive mining, logging, and agricultural land to a natural state.

The benefits of planting trees include . . .

Improved water quality. The intricate root systems of trees act like filters, removing pollutants and slowing down the water's absorption into the soil. This natural water filtration can lower costs associated with drinking water treatment.

Improved air quality. Trees help to clean the air we breathe by absorbing harmful pollutants. Healthy, strong trees act as carbon sinks, reducing our carbon footprint and reducing the effects of climate change.

Flood control. Trees play a key role in capturing rainwater and reducing the risk of natural disasters like floods and landslides.

Soil Stabilization. Trees reduce the effects of erosion caused by water and wind.

Wildlife habitat. Large populations of wildlife rely on forests for food, shelter, and water.

Payment procedures

If the referendum passes, we will subtract a specified amount from your prior earnings in today's session and set this aside. We will use this money to purchase the tree plantings while you are completing the post-experiment questionnaire. Since the price of the tree plantings is more than the amount we would collect from you, we will use money from a research grant to pay the difference.

We will forward the confirmation email we receive from One Tree Planted. Attached to this email will be a certificate. The certificate will acknowledge students at the University of Tennessee for funding 160 trees.

If the referendum does not pass, no money will be subtracted from your earnings. No money will be given to One Tree Planted and the tree planting project will not be funded.

Budget reminder: Please keep your budget in mind when voting and think about whether funding the project is worth it to you, and other things you can spend your money on.

Any questions about payment procedures?

The voting process

In this experiment we will ask you to vote YES or NO separately for several possible cost amounts, which will range from \$0 to \$15.

To determine the cost to you, the computer has been programmed to randomly select one of the stated cost amounts.

Your YES or NO vote to the randomly selected cost will be used to determine whether the referendum passes.

The referendum passes if a majority, more than half of the votes, are YES votes. Otherwise, the referendum does not pass.

If the referendum passes, each participant will pay the randomly selected cost and the tree planting project will be funded. If the referendum does not pass, no money will be collected, and the tree planting project will not be funded.

Any questions about the voting process?

Please go ahead and make your voting decisions.

Instructions for hypothetical treatments

In this experiment, you will be asked to vote in a hypothetical referendum on whether all participants in the room would collectively fund a tree-planting project. This referendum is hypothetical in the sense that, regardless of how everyone votes, no money will be subtracted from your earnings, and no trees will be planted. To be as clear as possible, this is not a real referendum, but we want you to imagine how you would vote if given the opportunity to fund a tree-planting project.

About the project . . .

The project would involve planting and maintaining 160 trees in the Appalachian Mountains, a region that stretches from southern New York state to northern Alabama and Georgia. To carry out the tree planting, we would partner with the non-profit organization One Tree Planted. This organization plants trees to return formerly unproductive mining, logging, and agricultural land to a natural state.

The benefits of planting trees include . . .

Improved water quality. The intricate root systems of trees act like filters, removing pollutants and slowing down the water's absorption into the soil. This natural water filtration can lower costs associated with drinking water treatment.

Improved air quality. Trees help to clean the air we breathe by absorbing harmful pollutants. Healthy, strong trees act as carbon sinks, reducing our carbon footprint and reducing the effects of climate change.

Flood control. Trees play a key role in capturing rainwater and reducing the risk of natural disasters like floods and landslides.

Soil Stabilization. Trees reduce the effects of erosion caused by water and wind.

Wildlife habitat. Large populations of wildlife rely on forests for food, shelter, and water.

Payment procedures

If this were a real referendum, and it passed, we would have subtracted a specified amount from your prior earnings in today's session and set this aside. We would have used this money to purchase the tree plantings while you were completing the post-experiment questionnaire. Since the price of the tree plantings would have been more than the amount we would have collected from you, we would have used money from a research grant to pay the difference.

We would have forwarded the confirmation email we would have received from One Tree Planted. Attached to this email would have been a certificate. The certificate would have acknowledged students at the University of Tennessee for funding 160 trees.

If this were a real referendum, and it did not pass, no money would have been subtracted from your earnings. No money would have been given to One Tree Planted and the tree planting project would not have been funded.

Budget reminder: Please keep your budget in mind when voting and think about whether funding the project would have been worth it to you, and other things you could have spent your money on.

The voting process

In this experiment we will ask you to vote YES or NO separately for several possible cost amounts, which will range from $0 to $15.

To determine the cost to you, the computer has been programmed to randomly select one of the stated cost amounts.

Your YES or NO vote to the randomly selected cost will be used to determine whether the referendum passes.

The referendum passes if a majority, more than half of the votes, are YES votes. Otherwise, the referendum does not pass.

Please keep in mind that this referendum is hypothetical. Regardless of whether the referendum passes, no money will be subtracted from your earnings and no trees will be planted. To be as clear as possible, this is not a real referendum, but we want you to imagine how you would vote if given the opportunity to fund a tree-planting project.

Before we proceed to the hypothetical referendum, we will ask you to answer a few questions to make sure you understand the procedures. The good news is that we will pay you 50 cents for answering each question correctly.

Appendix C

Appendix

C.1 Tables

Table 3.1: Summary of exogenous goal-setting experiments

Study	Goal-difficulty	Complexity	Goal Type	Setup	Design	Behavioral Theories
Earley et al. (1989)	Yes	Yes	Not mentioned	Individuals	Classroom	–
Nahrgang et al. (2013)	–	Yes	Binding	Teams	Lab	–
Chen and Latham (2014)	–	Yes	No incentives	Individuals	Lab	Automaticity
Smithers (2015)	Yes	–	Non-binding	Individuals	Lab	–
Corgnet et al. (2015)	Yes*	–	Non-binding	Individuals	Lab	Reference-dependence
Fan and Gómez-Miñambres (2020)	Yes*	–	Non-binding	Teams	Lab	Reference-dependence
This paper	Yes	Yes	Non-binding	Teams	Online/lab	Reference-dependence; Self-efficacy; Social Norms

Notes: * This study assigns a manager who sets non-binding goals that may or may not differ every period and it is not necessarily the case that goals increase in difficulty monotonically. Such a design means low variation in goals and that goals vary within a session. Goal-difficulty refers to whether the study varies the difficulty level of the goal explicitly. Complexity refers to whether the study varies the complexity level of the task. The last column titled "Behavioral theories" identifies whether the study highlights underlying mechanisms for goal-effectiveness.

Table 3.2: Summary of treatments

	Complexity	
Incentives	Low ($c = 5$)	High ($c = 20$)
Monetary (without goal)	NG	NG
Monetary (with easy goal)	EG	EG
Monetary (with medium goal)	MG	MG
Monetary (with difficult goal)	DG	DG

Table 3.3: Description of data

Variable Name	Description	Mean	S.D.
Treatment variables			
Low	= 1 when the task has a low complexity cost (cost=5); 0 otherwise	0.50	0.50
High	= 1 when the task has a high complexity cost (cost=20); 0 otherwise	0.50	0.50
Goal	= 1 when a goal (easy, moderate or difficult) is assigned; 0 otherwise	0.75	0.43
EG (Easy)	= 1 when an easy goal is assigned; 0 otherwise	0.25	0.43
MG (Moderate)	= 1 when a moderate goal is assigned; 0 otherwise	0.25	0.43
DG (Difficult)	= 1 when a difficult goal is assigned; 0 otherwise	0.25	0.43
Control variables			
Risk Averse	= 1 if participant selected safe option at least six times in Risk Elicitation task; 0 otherwise	0.41	0.49
Experience	= 1 if the participant had partaken in a prior economics experiment; 0 otherwise	0.84	0.36
Female	= 1 if participant is female; 0 otherwise	0.68	0.46
Employed	= 1 if participant is partly or fully employed, 0 otherwise	0.71	0.45
Loss Averse	= 1 if participant's loss aversion parameter $\lambda > 1$	0.76	0.42
Age	Recorded age of the participant	21.70	2.73
GPA	Participant GPA, recorded as midpoint of chosen interval	3.48	0.36
Earnings	Participants earnings from the experiment in $	18.76	3.25
Comprehension	Rating of instruction comprehension, scale 1 to 5	4.31	0.96
Round	Decision round in the experiment, 0 to 9 for each task (low and high complexity)	4.50	2.87

Table 3.4: Analysis of individual production

	Dep. Var.: Individual-level production				
	(1) Pooled	(2) No goal	(3) Easy	(4) Moderate	(5) Difficult
Constant	23.82***	24.21***	21.69***	23.40***	25.97***
	(0.349)	(0.713)	(0.637)	(0.705)	(0.622)
High	−3.82***	−3.21***	−2.59***	−3.82***	−5.66***
	(0.445)	(0.903)	(0.923)	(0.902)	(0.811)
Low × Round	−0.13**	−0.16	0.08	−0.01	−0.42***
	(0.061)	(0.127)	(0.116)	(0.126)	(0.105)
High × Round	−0.21***	−0.33**	−0.26**	−0.09	−0.150
	(0.070)	(0.152)	(0.131)	(0.147)	(0.131)
R-squared	0.179	0.154	0.207	0.166	0.252
Observations	1200	300	300	300	300

Notes: Cluster-robust standard errors in parentheses; clustered by round and subject. No controls. Task-specific trend is included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.5: Analysis of individual production (with controls)

	Dep. Var.: Individual-level production				
	(1) Pooled	(2) No goal	(3) Easy	(4) Moderate	(5) Difficult
Constant	30.20***	40.16***	30.66***	30.16***	38.01***
	(1.624)	(2.913)	(2.439)	(3.513)	(3.345)
High	−3.76***	−3.21***	−2.85***	−3.24***	−5.66***
	(0.452)	(0.911)	(0.991)	(0.903)	(0.818)
Low × Round	−0.13**	−0.16	0.08	0.01	−0.42***
	(0.061)	(0.123)	(0.120)	(0.109)	(0.109)
High × Round	−0.21***	−0.32**	−0.23*	−0.11	−0.15
	(0.069)	(0.137)	(0.121)	(0.137)	(0.117)
Experience	−1.83***	−0.14	−2.92***	−1.99***	−0.32
	(0.356)	(0.874)	(0.763)	(0.762)	(0.881)
Female	−0.51	2.60***	−1.62***	−1.87**	1.38*
	(0.321)	(0.623)	(0.485)	(0.772)	(0.765)
Risk Averse	−0.66**	3.38***	0.73	−4.73***	0.23
	(0.313)	(0.839)	(0.682)	(0.587)	(0.619)
GPA	−1.04**	−4.60***	−0.73	−0.99	−3.90***
	(0.449)	(0.801)	(0.722)	(1.035)	(1.002)
Loss Averse	−0.74**	−4.22***	−3.59***	1.19	1.20**
	(0.332)	(0.814)	(0.591)	(0.739)	(0.573)
R-squared	0.215	0.264	0.339	0.375	0.346
Observations	1160	300	280	280	300

Notes: Cluster-robust standard errors in parentheses; clustered by round and subject. Task-specific trend is included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.6: Analysis of team production

	Dep. Var.: Team production				
	(1) Pooled	(2) No goal	(3) Easy	(4) Moderate	(5) Difficult
Constant	19.92***	19.80***	17.99***	19.21***	22.67***
	(0.622)	(1.058)	(1.002)	(0.311)	(0.774)
High	−3.94***	−4.11**	−1.91	−3.92***	−5.82***
	(0.645)	(1.411)	(1.326)	(0.331)	(0.894)
Low × Round	−0.02	−0.05	0.21	0.13*	−0.35**
	(0.097)	(0.152)	(0.176)	(0.069)	(0.114)
High × Round	−0.22***	−0.17	−0.35***	−0.12	−0.26***
	(0.056)	(0.126)	(0.102)	(0.099)	(0.045)
R-squared	0.362	0.282	0.392	0.389	0.503
Observations	400	100	100	100	100

Notes: Cluster-robust standard errors in parentheses; clustered by round. Task-specific trend is included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.7: Analysis of individual-level production: goal effects

	Dep. Var.: Individual Production		
	(1)	(2)	(3)
Constant	23.49***	23.49***	30.57***
	(0.364)	(0.364)	(1.570)
High	−3.95***	−3.95***	−3.60***
	(0.579)	(0.580)	(0.621)
High × Easy Goal		−1.60***	−1.70***
		(0.584)	(0.594)
High × Moderate Goal		−0.38	−0.41
		(0.622)	(0.636)
High × Difficult Goal		0.11	−0.28
		(0.589)	(0.592)
Low × Easy Goal		−1.45***	−1.40***
		(0.480)	(0.489)
Low × Moderate Goal		−0.13	−0.57
		(0.500)	(0.497)
Low × Difficult Goal		0.59	0.20
		(0.478)	(0.491)
Low × Round			−0.13**
			(0.060)
High × Round			−0.21***
			(0.069)
Experience			−1.79***
			(0.361)
Female			−0.43
			(0.332)
Risk Averse			−0.81***
			(0.311)
GPA			−1.07**
			(0.442)
Loss Averse			−0.592*
			(0.341)
Low × Goal	−0.331		
	(0.410)		
High × Goal	−0.622		
	(0.505)		
R-squared	0.172	0.190	0.230
Observations	1200	1200	1160

Notes: Cluster-robust standard errors in parentheses; clustered by round and subject. Task-specific trend is included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.8: Analysis of team production: goal effects

	Dep. Var.: Team Production		
	(1)	(2)	(3)
Constant	19.60***	19.60***	19.66***
	(0.510)	(0.513)	(0.697)
High	−4.66***	−4.66***	−3.73***
	(0.595)	(0.598)	(0.884)
High × Easy Goal		−0.42	−0.42
		(0.540)	(0.429)
High × Moderate Goal		−0.18	−0.18
		(0.423)	(0.397)
High × Difficult Goal		0.74*	0.74**
		(0.435)	(0.336)
Low × Easy Goal		−0.68	−0.68
		(0.696)	(0.705)
Low × Moderate Goal		0.18	0.18
		(0.596)	(0.603)
Low × Difficult Goal		1.50**	1.50**
		(0.679)	(0.677)
Low × Round			−0.02
			(0.081)
High × Round			−0.22***
			(0.051)
Low × Goal	0.333		
	(0.587)		
High × Goal	0.0467		
	(0.379)		
R-squared	0.351	0.374	0.381
Observations	400	400	400

Notes: Cluster-robust standard errors in parentheses; clustered by round. Task-specific trend is included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.9: Analysis of individual effort

	Dep. Var.: # of clicks		
	(1)	(2)	(3)
Constant	13.67***	13.67***	15.20***
	(0.577)	(0.578)	(0.820)
High	-6.57***	-6.57***	-6.99***
	(0.759)	(0.76)	(0.748)
High × Easy Goal		-1.35**	-1.35**
		(0.624)	(0.617)
High × Moderate Goal		-0.08	-0.08
		(0.680)	(0.675)
High × Difficult Goal		0.89	0.89
		(0.747)	(0.743)
Low × Easy Goal		0.28	0.28
		(0.928)	(0.923)
Low × Moderate Goal		0.25	0.25
		(0.844)	(0.838)
Low × Difficult Goal		2.37***	2.37***
		(0.851)	(0.847)
Low × Round			-0.34***
			(0.119)
High × Round			-0.25***
			(0.080)
Low × Goal	0.96		
	(0.692)		
High × Goal	-0.18		
	(0.566)		
R-squared	0.224	0.235	0.246
Observations	1200	1200	1200

Notes: Cluster-robust standard errors in parentheses; clustered by round and subject. Task-specific time trend is included.
*** p<0.01, ** p<0.05, * p<0.1.

Table 3.10: Analysis of team effort

	Dep. Var.: (# of clicks)		
	(1)	(2)	(3)
Constant	7.760***	7.760***	8.23***
	(0.436)	(0.438)	(0.482)
High	-5.040***	-5.040***	-4.61***
	(0.566)	(0.569)	(0.499)
High × Easy Goal		-0.50	-0.50
		(0.450)	(0.367)
High × Moderate Goal		0.20	0.20
		(0.501)	(0.418)
High × Difficult Goal		0.58	0.58*
		(0.409)	(0.345)
Low × Easy Goal		-0.44	−0.44
		(0.514)	(0.528)
Low × Moderate Goal		-0.06	−0.06
		(0.565)	(0.584)
Low × Difficult Goal		2.54***	2.54***
		(0.644)	(0.609)
Low × Round			-0.10
			(0.069)
High × Round			-0.20***
			(0.028)
Low × Goal	0.68		
	(0.544)		
High × Goal	0.09		
	(0.402)		
R-squared	0.432	0.473	0.485
Observations	400	400	400

Notes: Cluster-robust standard errors in parentheses; clustered by round. Task-specific time trend is included.
*** p<0.01, ** p<0.05, * p<0.1.

Table 3.11: Analysis of individual-level cognitive effort

	Dep. Var.: catches/clicks		
	(1)	(2)	(3)
Constant	2.22***	2.22***	2.07***
	(0.117)	(0.117)	(0.146)
High	2.35***	2.35***	1.60***
	(0.331)	(0.332)	(0.350)
High × Easy Goal		-0.13	-0.11
		(0.416)	(0.411)
High × Moderate Goal		-0.43	-0.42***
		(0.430)	(0.423)
High × Difficult Goal		-0.72*	-0.71*
		(0.398)	(0.396)
Low × Easy Goal		0.16	0.16
		(0.208)	(0.208)
Low × Moderate Goal		-0.13	-0.13
		(0.144)	(0.143)
Low × Difficult Goal		-0.42***	-0.42***
		(0.133)	(0.133)
Low × Round			0.03
			(0.020)
High × Round			0.20***
			(0.053)
Low × Goal	-0.13		
	(0.135)		
High × Goal	-0.43		
	(0.348)		
R-squared	0.149	0.155	0.175
Observations	1162	1162	1162

Notes: Cluster-robust standard errors in parentheses; clustered by round and subject. Task-specific time trend included.
*** p<0.01, ** p<0.05, * p<0.1.

Table 3.12: Analysis of team-level cognitive effort

	Dep. Var.: catches/clicks		
	(1)	(2)	(3)
Constant	3.21***	3.21***	2.98***
	(0.365)	(0.367)	(0.402)
High	3.94***	3.94***	2.21***
	(0.812)	(0.816)	(0.690)
High x Easy Goal		-0.60	-0.40
		(0.932)	(0.798)
High × Moderate Goal		-0.39	-0.39
		(1.087)	(0.804)
High × Difficult Goal		-1.59*	-1.52**
		(0.803)	(0.749)
Low × Easy Goal		0.48	0.48
		(0.511)	(0.510)
Low × Moderate Goal		-0.36	-0.35
		(0.394)	(0.392)
Low × Difficult Goal		-0.92**	-0.92**
		(0.379)	(0.375)
Low × Round			0.05
			(0.047)
High × Round			0.43***
			(0.084)
Low × Goal	-0.27		
	(0.403)		
High × Goal	-0.89		
	(0.805)		
R-squared	0.242	0.261	0.314
Observations	363	363	363

Notes: Cluster-robust standard errors in parentheses; clustered by round. Task-specific time trend is included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.13: Analysis of wasted performance: individual-level

	Dep. Var.: Wasted Production		
	(1)	(2)	(3)
Constant	3.88***	3.88***	4.40***
	(0.342)	(0.342)	(0.436)
High	0.71	0.71	0.13
	(0.520)	(0.521)	(0.652)
High × Easy Goal		-1.18**	-1.18**
		(0.540)	(0.540)
High × Moderate Goal		-0.20	-0.20
		(0.590)	(0.590)
High × Difficult Goal		-0.63	-0.63
		(0.581)	(0.581)
Low × Easy Goal		-0.77*	-0.77*
		(0.450)	(0.450)
Low × Moderate Goal		-0.31	-0.31
		(0.481)	(0.481)
Low × Difficult Goal		-0.91**	-0.91**
		(0.447)	(0.447)
Low × Goal	-0.66*		
	(0.385)		
High × Goal	-0.66		
	(0.480)		
R-squared	0.015	0.015	0.018
Observations	1200	1200	1200

Notes: Cluster-robust standard errors in parentheses; clustered by round and subject. *** p<0.01, ** p<0.05, * p<0.1.

C.2 Predictions for effort: Ball-catching task

In this section, I present the theoretical predictions of individual effort arising from the ball-catching task. Gächter et al. (2016) derive the predicted number of clicks, however, their parameters are slightly different from the one used in this study. For this reason, I use the data in this study to estimate the individual production function i.e., relationship between catches and clicks and then use this estimate to predict number of clicks for the low and high complexity conditions respectively. In order to estimate the production function, I rely on the empirical strategy from Gächter et al. (2016). The functional form specification that fits the data is presented below and estimated using a random coefficients panel regression.

$$Catches_{it} = \beta_0 + \beta_1 Clicks_{it}^{0.5} + \beta_2 Clicks_{it}^{2} + (\delta_t + \omega_i + \mu_{it}) Clicks_{it}^{0.5} \tag{8}$$

where $Catches_{it}$ and $Clicks_{it}$ denote the number of catches (output, y) and the number of clicks (effort, e) by subject i in period t. δ_t is the period dummy, $\omega_i \sim (0, \sigma_\omega^2)$ denotes the subject-specific random effect and $\mu_{it} \sim (0, \sigma_\mu^2)$ is the randomly distributed error term.

In order to estimate equation (8), it is transformed by dividing throughout with $Clicks_{it}^{0.5}$ and then estimated using a standard random effects approach.[1] Coefficient estimates from the panel data regressions are reported in Table C.1. Column (1) reports estimates from the full sample (i.e., pooling low and high complexity conditions) while (2) and (3) provide estimates computed separately for both conditions. The estimates are fairly stable across the three models except for the squared clicks term in (3). This could be due to a relatively smaller sample but point predictions do not significantly change if we were to ignore the squared clicks term in (3). Models (2) and (3) are used predict the number of catches and clicks depending on task type.

[1]Please refer to Section 3.3 in Gächter et al. (2016) for details.

Table C.2 compares the predicted number of clicks with the observed averages for several conditions.[2] Results suggest that the observed number of catches and clicks are slightly different from what is predicted. In most cases, magnitudes are small, however, the predicted number of clicks in the low complexity condition is about 9 points higher than the actual number of clicks. One reason why this is the case may be that the weakest-link production setting greatly reduces variation between the weakest-link member and other team members in an effort to minimize own costs. This would reduce average clicks for all individuals and not just for the weakest-link member. The other plausible explanation is that predictions include goal treatments as well which may have counteracting effects on the number of clicks depending on goal type. If we were to remove the goal treatments, or restrict the specification to the data from teams, estimates may not be very stable given the present sample size.

Comparing estimates from regressions in Table C.2 to that of Gächter et al. (2016) shows some promise as coefficients are similar in magnitude. Finally, note that these predictions only account for the material cost of complexity i.e., the cost induced through clicks, however, it does not capture the cognitive costs so these predictions are more likely to be upper bounds of production and effort.

[2]Note that given the weakest link team-production function, I derive the point predictions for the weakest member and this provides a lower bound for other individuals whose catches may be weakly greater than that of the weakest-member.

Table C.1: Empirical production function: Panel data regressions

	Dep. Var.: Number of Catches		
	(1) Full sample	(2) Low ($c = 5$)	(3) High ($c = 20$)
Intercept	10.34***	10.23***	10.93***
	(0.304)	(0.499)	(0.460)
$\text{Clicks}^{0.5}$	3.63***	3.83***	3.36***
	(0.231)	(0.247)	(0.327)
Clicks^{2}	-0.003***	-0.005***	0.003
	(0.001)	(0.001)	(0.002)
σ_ω	0.118***	0.233***	0.111
	(0.054)	(0.058)	(0.124)
σ_μ	1.256***	0.877***	1.533***
	(0.026)	(0.025)	(0.046)
Observations	1162	599	563

Notes: All period dummies are included and are insignificant except period 6 in (1); period 2 and 6 in (2). *** p<0.01, ** p<0.05, * p<0.1.

Table C.2: Comparisons between predictions and observed team averages

	Low Complexity		**High Complexity**	
	Catches	Clicks	Catches	Clicks
Prediction	25.97	23.12	19.39	6.34
Observed (individual)	23.23	14.39	19.06	6.96
Difference (t-test)	-2.73***	-8.72***	-0.32	0.62***

Notes: *** p<0.01, ** p<0.05, * p<0.1.

C.3 Figures

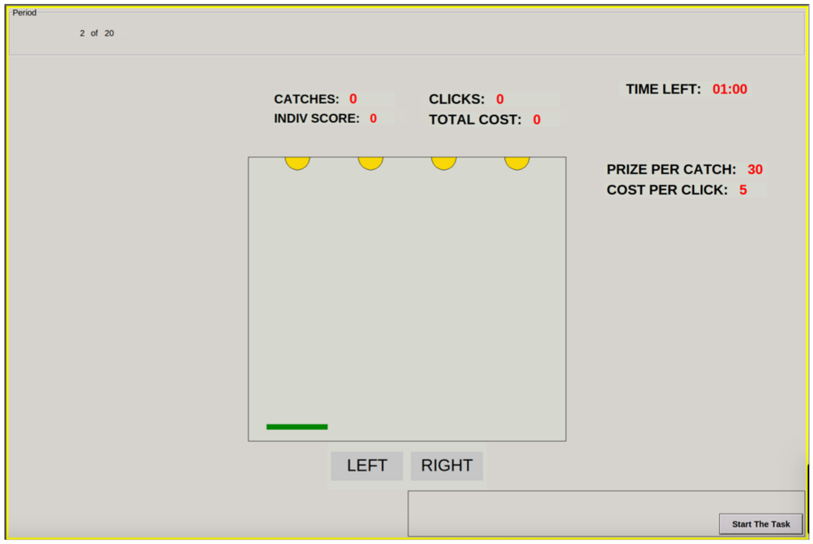

Figure 3.1: Decision Screen, No Goal Treatment

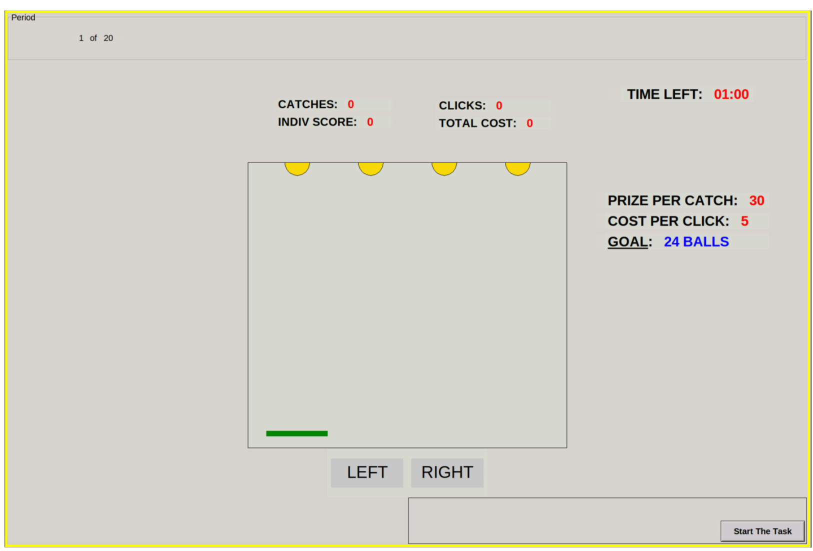

Figure 3.2: Decision Screen, Goal Treatment

C.4 Experiment Instructions

Thank you for participating in today's study.

You will make decisions using a computer, and your decisions will be associated with a randomly assigned ID number. You will never be asked to reveal your identity to anyone. Your name will never be associated with any of your decisions. In order to keep your decisions private, please do not reveal your choices or otherwise communicate with any other participant. Importantly, please refrain from verbally reacting to events that occur.

Today's session has four parts: Experiment 1, Experiment 2, Experiment 3, and a short questionnaire. You will have the opportunity to earn money in all experiments based on your decisions. In addition, you will receive a show-up fee of $7 for completing today's session. You will be paid your earnings privately, and via an amazon gift card, at the end of the experiment session. We will proceed through the written materials together. Please do not enter any decisions on the computer until instructed to do so.

Instructions for Experiment 1

In this experiment, all money amounts are denominated in US dollars. Please refer to your experiment screen while we read the instructions.

We would like you to make a decision for each of 10 scenarios. Each scenario involves a choice between playing a lottery that pays either $4 or $0 according to specified chances (Choice A) or receiving $2 for sure (Choice B).

You will notice that the only differences across scenarios are the chances of receiving the high or low prize for the lottery. At the end of the today's session, ONE of the 10 scenarios will be selected at random and you will be paid according to your decision for this selected scenario ONLY. Each scenario has an equal chance of being selected.

Please consider your choice for each scenario carefully. Since you do not know which scenario will be played out, it is in your best interest to treat each scenario as if it will be the one used to determine your earnings.

Before making decisions, are there any questions?

Please proceed to entering decisions on your computer. Once you are ready to submit your decisions, please click the "Submit" button.

Instructions for Experiment 2

In this experiment, all money amounts are denominated in US dollars. Please refer to your experiment screen while we read the instructions.

We would like you to make a decision for each of the 6 scenarios. Each scenario involves a choice between playing a lottery or not. In each scenario, if you choose to play the lottery (Choice A), there is a 50% chance you will you win $3 and a 50% chance you will lose a specified amount. If you do not play the lottery, (Choice B), you earn $0.

You will notice that the only difference across scenarios is the amount at stake to lose by playing the lottery. At the end of the today's session, ONE of the 6 scenarios will be selected at random, and you will be paid according to your decision for this selected scenario ONLY. Each scenario has an equal chance of being selected.

Please consider your choice for each scenario carefully. Since you do not know which scenario will be played out, it is in your best interest to treat each scenario as if it will be the one used to determine your earnings.

Note that, in contrast to the previous experiment, if you choose to play the lottery there is a 50% chance of losing money. If this happens, the amount of the loss will be subtracted from your overall earnings in the experiment (i.e., show-up fee, earnings from Experiments 1 and 3).

Before making decisions, are there any questions?

Please proceed to entering decisions on your computer. Once you are ready to submit your decisions, please click the "Submit" button.

Instructions for Experiment 3

In this experiment, all money amounts will be denominated in tokens. At the end of each experiment tokens will be converted to US dollars at a rate of 100 tokens = $1. This experiment has a total of 20 decision rounds. At the start of each round, you will be randomly placed into a group of three players. The members of your group will vary from one round to the next. In each round, you and the other members of your group will be asked to work on a computerized ball-catching task.

Ball-catching task

In each round, there will be a task box in the middle of the task screen like the one shown below:

<insert Figure 3.1 (Figure 3.2) here for no goal (goal) treatment>

Each round lasts one minute. Once you click on the "Start the Task" button, the timer will start, and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the "LEFT" or "RIGHT" buttons.

To catch a ball, your tray must be below the ball before it touches the tray. When the ball touches the tray, your CATCHES increase by one.

Your individual score is calculated as the number of balls you catch multiplied by 30 tokens.

For each mouse click YOU make, you will incur a cost. YOUR individual cost will be 5 tokens per click for the first 10 rounds. In the last 10 rounds, you will incur a cost of 20 tokens per click. Your total cost is equal to the total number of clicks multiplied by 5 or 20 tokens depending on the round.

In each round, the number of balls YOU have caught so far (displayed as CATCHES) and the number of clicks you have made so far (CLICKS) will be shown right above the task box. Also shown above the task box will be your individual score (displayed as INDIV SCORE), which is CATCHES multiplied by the prize per catch and TOTAL COST, which is CLICKS multiplied by the cost per click.

Your Earnings

When you and the other members of your group have finished the task, the computer will calculate your earnings, which will depend on the group score. The group score is equal to the lowest individual score among your group members (including yours).

Your earnings for a decision round will be calculated as follows:

Round Earnings = group score (in tokens) – YOUR total cost (in tokens)

Example. Suppose YOU catch 10 balls by making 5 clicks, then your individual score is 30 tokens $\times$ 10 balls = 300 tokens. Your total cost is 20 tokens $\times$ 5 clicks = 100 tokens.

Suppose further that your group members each catch 20 balls by making 10 clicks i.e., their individual score is 30 tokens $\times$ 20 balls = 600 tokens each. In this case, the group score is 300 tokens because YOU were the lowest scoring member in your group.

Your earnings in this example would then be 300 (group score) – 100 (your total cost) = 200 tokens.

Group performance goal ***(only included in the goal treatments)***

Your group will be assigned a performance goal – a recommended number of balls you and your group members should catch within a round. The goal will be displayed to the right on your decision screen.

Please know that whether you meet the goal will not impact your earnings. Your earnings will depend on the group score (lowest individual score in your group) as well as your total cost of clicking, as described in the instructions.

At the end of each decision round, the computer will display whether your group met the goal, your individual score, your total cost, the group score and your earnings. Any questions?

Proceeding through the experiment

You will now go through a total of 20 decision rounds. At the start of each round, you will be randomly placed into a group of three players. This means that the members of your group will vary from one round to the next.

The number of catches, total cost, your score, and the group performance goal will be displayed on your screen in every round. Your decision screen will look like the example provided in the instructions.

To determine the amount of money you earn from this experiment, the computer will randomly select two of the 20 rounds (one for the first 10 rounds and the other from the last 10 rounds). Your earnings from the two selected rounds will be converted into dollars and added to your earnings total for the experiment.

Each decision round is separate from the other rounds, in the sense that the decisions you make in one round will not affect the outcome or earnings of any other round.

Since you do not know which rounds will be selected, you should make choices in each round carefully.

Before we continue, do you have any questions?

Before continuing, we would like you to answer a few questions to make sure you understand the procedures. Here is the good news: for each question you answer correctly, you will earn 25 cents. You will have a total of 150 seconds (2.5 minutes) to answer all questions. You may use a calculator if you wish.

Vita

Vasudha Chopra was born and brought up in New Delhi, India. She completed a Bachelor of Arts degree in Economics from Hansraj College, University of Delhi in 2014 and obtained a first division. She later went on to complete her post-graduate studies at TERI University and graduated in 2016 with a Masters of Science degree in Economics. As a doctoral student at the University of Tennessee, Vasudha's research focused on using applied microeconomic theory and experimental methods to topics in environmental economics and organizational economics. She has also taught intermediate macroeconomics at the undergraduate level and received glowing reviews. Whether it is research or teaching, she particularly enjoys working with students. In the past, she has also worked as an assistant policy officer at Vasudha Foundation, an environmental think-tank based in New Delhi. Vasudha will join Thapar School of Liberal Arts & Sciences as an Assistant Professor starting August 2022.