Tournament Structure and Effort in an Experimental Setting

Yan Lau*
Reed College

October 2013

Abstract

Tournament reward mechanisms, in which prizes are awarded to participants based on relative position, are a common feature in many parts of labor markets, as well as in other contexts. Tournament structure—in terms of the prize amount and the location of the relative winning cutoff—may affect participant effort decisions. This paper presents a tournament model which predicts that receiving larger prizes and being nearer to the relative cutoff incentivize tournament participants to exert more effort. I test these predictions experimentally by recruiting participants to perform real effort memorization tasks over multiple rounds. Between rounds, I manipulate the tournament structure in order to induce effort changes. I find that increasing the prize amount is effective in incentivizing additional effort. I also find that low-performing participants increase (decrease) effort when the relative winning cutoff is shifted nearer to (further from) them, but high-performing participants do not respond in this manner when the cutoff is shifted nearer (further). These experimental results provide suggestions on how to structure tournaments in labor markets to increase effort/productivity.

*[yanlau@reed.edu] I would like to thank Alex Mas for his helpful guidance and feedback. The Princeton Laboratory for Experimental Social Science (PLESS) played a crucial part in this project, providing the space and resources necessary to conduct the experiments. I am especially grateful to J. Baxter Oliphant, the lab manager at PLESS. I would also like to thank the attendees of the Public Finance Working Group and the Summer Labor Lunch Series for their comments and suggestions. Funding for this project was generously provided by the Industrial Relations Section at Princeton University. Lastly, I would like to thank all the participants of the experiment. All experimental procedures were approved by Princeton's Institutional Review Board for Human Subjects (#6230).

1 Introduction

Tournaments are reward mechanisms in which relative—rather than absolute—position determines the winner. Tournaments can be found in many everyday settings; they are a common feature in labor markets (e.g. personnel promotion within firms; compensation schemes based on on-the-job performance such as retail sales or productivity; employee of the month contests), but appear in other contexts as well, such as sports and education. In all of these settings, tournaments can be defined by three basic structural parameters: the prize amount, the relative cutoff past which winners receive the prize, and the absolute number of players competing in the tournament.

The structure of a tournament can affect effort decisions and outcomes. Considerations as to how to parameterize the structure of a tournament are especially relevant in settings in which tournament organizers want to induce effort exertion or specific outcome goals. For example, educators may structure a tournament in order to incentivize (certain) students to study harder or obtain higher test scores. Knowing how tournament structure affects these effort decisions and outcomes will be useful in formulating tournaments.

This paper uses experimental tournaments in a controlled lab setting to investigate how adjusting tournament structure affects the effort decisions of participants. In particular, I focus on manipulating the prize amount and the relative winning cutoff. In each round of the experiment, groups of participants were asked to memorize and recall lists of word-number pairs on a computer screen. There are a total of five rounds in each experimental session, and a different list of word-number pairs is presented in each round. At each participant's own discretion, an amount of time is spent memorizing this list. When satisfied, the participant clicks to the next screen, where the same list of words (but reordered randomly) is presented with blank text boxes next to each word. The goal of the task is to recall as many numbers as possible (each corresponding to separate words). A subset of participants with the most number of correct answers (relative to all participants in the session) wins a monetary prize. The prize dollar amount and proportion of participants receiving the prize is changed from round to round, thus varying the tournament structure. Effort is measured as the amount of time spent memorizing the list chosen by each participant.

In the economics literature, tournament theory was brought to prominence by the seminal work of Lazear and Rosen (1981), and later expanded upon by Green and Stokey (1983), among others. Since then, many empirical papers have analyzed tournaments in various natural experiment settings, including software programming contests[1], retail sales[2], and personnel economics within the firm[3].

A separate strand of literature has focused on using experimental methods to analyze tournaments in controlled settings. Most of this research has focused on how using a tournament reward mechanism compares to using a piecerate one, in which prizes are awarded proportional to outcomes. Bull et al. (1987) randomize participants into two such reward mechanisms to study how effort choices differ under the two schemes. They find that mean effort is similar in both cases, but variance in effort is higher in tournaments. Their study elicits effort by having pairs of participants choose "decision numbers" from cost and payoff tables; the "effort decisions" of competing pairs (plus a random component) are then compared against one another with the winner receiving a prize, less an effort cost. Higher decision numbers are associated with increasing costs, but offer higher expected payoffs. Use of such "cost and payoff tables" are common in the experimental tournament literature. In a subsequent paper, Schotter and Weigelt (1992) add unfair aspects to similar experimental tournaments and analyze the impact of fairness-restoring policies.

Research by Orrison et al. (1997) and Harbring and Irlenbusch (2003) focus on tournament (as opposed to piecerate) reward structures and investigate how varying the number of competitors and the prize structure affect effort decisions. Both these papers use cost and payoff tables in their experimental setup. Orrison et al. (1997) find that altering the number of competitors from 2 to 4 to 6 does not change mean effort. They also obtain results for changing the proportion of relatively larger prizes, and using discriminatory cost and payoff tables. Harbring and Irlenbusch (2003) find that mean effort increases as the proportion of prize winners increases, but that variance in effort decreases.

Using cost and payoff tables, however, has the disadvantage of external validity. In many real-life tournaments, the decision to exert effort often incurs more tangible costs of the physically and cognitively taxing sort. More recent experimental research involving human subjects devise what experimentalists call "real effort tasks" in order to elicit effort. Dijk et al. (2001) use a grid computer game task in which participants search for prizes within a grid of cells as a real effort task to compare how different compensation schemes—including

[1] Boudreau et al. (2012)
[2] Casas-Arce and Martinez-Jerez (2009)
[3] Main et al. (1993); Eriksson (1999)

individual (i..e. piecerate), team, and tournaments—incentivize effort exertion. They find similar effort levels in individual and team compensation schemes, but higher and more variable effort levels in tournaments.

The contribution of this study is to use a real effort task to investigate the effects of changing tournament structure parameters—namely the prize amount and the relative winning cutoff—on effort decisions. To the best of my knowledge, previous papers looking at tournament structure have only used cost and payoff tables or similar non-real effort measures to elicit effort. In the experiment, I use time spent memorizing the list of word-number pairs as my measure of effort. This innovation allows me to measure effort precisely without the risk of significant Hawthorne effects, as it is not immediately apparent during the experiment that this is the outcome of interest.

The remainder of the paper is presented as follows. In Section 2, I present a tournament model and discuss the comparative statics of changing tournament structure. Next, I describe the experimental procedure in Section 3. I then present the results in Section 4 and conclude in Section 5.

2 Model

In this section, I explore how the structure of a tournament—in particular the (relative) location of the relative winning cutoff and the prize amount—influences effort decisions. I will set up a tournament model to describe the effort decisions of experiment participants. In this framework, the score a participant obtains will be a function of ability and effort. Participants choose effort to maximize their expected prize less an effort cost. Furthermore, a constraint will be added to account for the fact that there is a maximum obtainable score which cannot be surpassed—corresponding to the experiment conducted, in which the maximum score attainable was 18.

Let θ_i denote the ability of participant i, drawn from the ability distribution $F(\theta_i)$. Let e_i denote the effort choice of participant i. Let$(1-p)$ be the proportion of participant who win the prize α. A participant wins only if his or her score is at least as high as a certain absolute score cutoff $\overline{S}_p$.[4] This $\overline{S}_p$ depends on the p which is set in each round of the experiment. The participant's realized

[4]Note that this absolute score cutoff, which is the numerical score a participant needs to surpass in order to win the prize, while related, is not the same as the relative winning cutoff, which is the rank a participant needs to surpass in order to win the prize.

score is described by the function

$$S(\theta_i, e_i) - \eta_i$$

which contains the production function component $S(\theta_i, e_i)$, and a random error component η_i which is i.i.d. and orthogonal to effort and ability. The score production function depends on the participant's ability and effort choice as inputs.

A participant's utility depends on whether the prize is won, less an effort cost which must be paid prior to observing the realized score. This is given by

$$u(e_i) = \begin{cases} \alpha - c(e_i) & \text{if the participant wins} \\ -c(e_i) & \text{otherwise} \end{cases}$$

I make the following assumptions.

Assumption A. Score production is increasing in effort and non-decreasing in ability. That is,

$$\frac{\partial S(\theta_i, e_i)}{\partial e_i} > 0 \ \text{ and } \ \frac{\partial S(\theta_i, e_i)}{\partial \theta_i} \geq 0$$

However, score returns on effort are non-increasing in effort and increasing in ability[5]. That is,

$$\frac{\partial^2 S(\theta_i, e_i)}{\partial e_i^2} \leq 0 \ \text{ and } \ \frac{\partial^2 S(\theta_i, e_i)}{\partial \theta_i \partial e_i} > 0$$

Assumption B. Effort costs are increasing and strictly convex. That is,

$$c'(e_i) > 0 \ \text{ and } \ c''(e_i) > 0$$

These are reasonable assumptions, considering that common functional forms of $S(\theta_i, e_i)$ (e.g., Cobb-Douglas) and $c(e_i)$ (e.g., quadratic costs) satisfy them.

The participant's expected utility conditional on ability and effort choice is

$$\begin{aligned} E[u(e_i) \mid e_i, \theta_i] &= \alpha \Pr\left(S(\theta_i, e_i) - \eta_i \geq \overline{S}_p\right) - c(e_i) \\ &= \alpha H\left(S(\theta_i, e_i) - \overline{S}_p\right) - c(e_i) \end{aligned}$$

[5]This would be so if higher ability participants make better use of their effort exertion.

where $H(\eta_i)$ is the c.d.f. of η_i (and $h(.)$ is the corresponding p.d.f.).

Assumption η. Let the distribution of η be single peaked. Let this peak be (without loss of generality) at zero.

Furthermore, because of the existence of maximum obtainable score S_{max}, any optimization problem will be subject to the constraint

$$S(\theta_i, e_i) - \eta_i \leq S_{max} \tag{1}$$

In choosing effort to maximize expected utility subject to the above constraint, the participant's first order condition (FOC) is

$$\alpha \left[h\left(S(\theta_i, e_i) - \overline{S}_p\right) - \lambda\right] \frac{\partial S(\theta_i, e_i)}{\partial e_i} - c'(e_i) = 0 \tag{2}$$

where λ is the Lagrange multiplier for the constraint. This equation defines the participant's optimal effort decision e_i^* as a function of ability θ_i. In particular, the closer a participant is to the absolute score cutoff (i.e., the higher $h\left(S(\theta_i, e_i) - \overline{S}_p\right)$ is), the more effort he or she will exert in order to overcome the potential of an adverse error, get above the score cutoff, and receive the prize α. This generates a "bump" in effort near the absolute score cutoff.

Moreover, as ability increases and the realized score approaches the maximum score S_{max}, these increasingly higher performing participants exert less and less effort. This is because the λ term increases until $\left[h\left(S(\theta_i, e_i) - \overline{S}_p\right) - \lambda\right]$ becomes lower (possibly negative), and optimal effort decreases due to the convexity of $c(e_i)$ supposed in Assumption B.[6]

The second order condition (SOC) is given by

$$\alpha h'p\left(S(\theta_i, e_i) - \overline{S}_p\right)\left(\frac{\partial S(\theta_i, e_i)}{\partial e_i}\right)^2 + \alpha\left[h\left(S(\theta_i, e_i) - \overline{S}_p\right) - \lambda\right]\frac{\partial^2 S(\theta_i, e_i)}{\partial e_i^2} - c''(e_i) < 0 \tag{3}$$

This is negative as I will only consider interior solutions with $e_i^* > 0$.[7]

[6] I abstract away from the fact that the distribution of the error term $h(.)$ will become positively skewed for participants with high score production $S(\theta_i, e_i)$ near the maximum S_{max}. Only the magnitude, but not the sign, of the comparative statics changes in this case.

[7] In the case where the only solution to the FOC has a positive SOC (i.e., the FOC solution is utility-minimizing), then the corner solution $e_i^* = 0$ is utility-maximizing.

Using the implicit function theorem,

$$\frac{de_i}{dp} = -\frac{-\alpha h'\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right) \frac{\partial S(\theta_i, e_i)}{\partial e_i} \frac{d\overline{S}_p}{dp}}{\alpha h'\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right) \left(\frac{\partial S(\theta_i, e_i)}{\partial e_i}\right)^2 + \alpha \left[h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right) - \lambda\right] \frac{\partial^2 S(\theta_i, e_i)}{\partial e_i^2} - c''\left(e_i\right)} \tag{4}$$

This equation describes the comparative statics of an increase in the proportion of non-winners p. For illustration purposes, I will assume for now that $\frac{d\overline{S}_p}{dp}$ is positive; that is, that the absolute score cutoff increases with an increase in the proportion of non-winners. (Later, I will discuss the possibility that $\frac{d\overline{S}_p}{dp}$ is negative.) Note also that the denominator is the SOC, which is negative.

Since $\frac{\partial S(\theta_i, e_i)}{\partial e_i} > 0$ under Assumption A, $\frac{de_i}{dp}$ is positive if and only if $h'\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right)$ is negative. Recall that under Assumption η, $h(.)$ is single-peaked at zero. This implies that initially (i.e. prior to the change in p), high-performing participants (with values of $h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right)$ to the right of the peak) will increase their effort in response to an increase in the proportion of non-winners; correspondingly, initially low-performing participants (with values of $h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right)$ to the left of the peak) will decrease their effort in response to an increase in the proportion of non-winners.

The intuition behind this result is that when the proportion of non-winners increases, initially high-performing participants see the absolute score cutoff move closer to them, thus increasing their risk of falling below the score cutoff due to the error term; in response, they "step up" effort to compete for the relatively fewer number of prizes. On the other hand, initially low-performing participants see the absolute cutoff score move further away from them, decreasing the probability of them ever surpassing it; in response, they "give up" competing for the relatively fewer number of prizes as effort exertion is costly.

Similarly,

$$\frac{de_i}{d\alpha} = -\frac{\left[h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right) - \lambda\right] \frac{\partial S(\theta_i, e_i)}{\partial e_i}}{\alpha h'\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right) \left(\frac{\partial S(\theta_i, e_i)}{\partial e_i}\right)^2 + \alpha \left[h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right) - \lambda\right] \frac{\partial^2 S(\theta_i, e_i)}{\partial e_i^2} - c''\left(e_i\right)} \tag{5}$$

This equation describes the comparative statics of an increase in the prize amount α. For participants not experiencing censoring of scores (i.e. λ is zero), the derivative is positive.[8] This implies that all participants increase effort in response to an increase in the prize amount. The intuition is that the marginal

[8]The denominator is the SOC, which is negative. Furthermore, $\frac{\partial S(\theta_i, e_i)}{\partial e_i} > 0$ is assumed, and $h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right)$ is a p.d.f., so it is positive. With the negative sign in front of the fraction, this implies the derivative is positive.

benefit of winning (for every unit increase in the probability of winning) is now higher, so participants are more willing to exert additional effort and incur a higher marginal cost of effort. In particular, participants around the absolute score cutoff $\overline{S}_p$ will have more to gain (or lose) if they just surpass (or fall below) the cutoff. Hence, their effort increase will be highest (as seen by higher values of $h\left(S\left(\theta_i, e_i\right) - \overline{S}_p\right)$) compared to participants further away from the absolute score cutoff.

Should top performing participants be capped by S_{max}, then λ becomes positive and $\frac{de_i}{d\alpha}$ decreases. The effort response to the increase in prize amount is smaller because there is no benefit to the capped participant's score in exerting additional effort.

These two sets of predictions from the separate changes in tournament structure relating to $\frac{de_i}{dp}$ and $\frac{de_i}{d\alpha}$ are tested in the experiment to follow.

3 Experimental Procedure

Experiment sessions were conducted at the Princeton Laboratory for Experimental Social Sciences (PLESS) over the course of two weeks in May 2013. A total of 173 participants were recruited through the PLESS mailing list[9] to participate in 21 sessions.[10] Participants included undergraduates, graduate students, and staff members at Princeton University, as well as local community members not necessarily affiliated with the institution. Each session consisted of multiple participants competing against one another for prizes. Additionally, participants were paid a base amount of $15 for showing up, regardless of performance. The number of participants in each session varied between five and twelve.[11] Figure 1 is a histogram of the size of the 21 sessions.

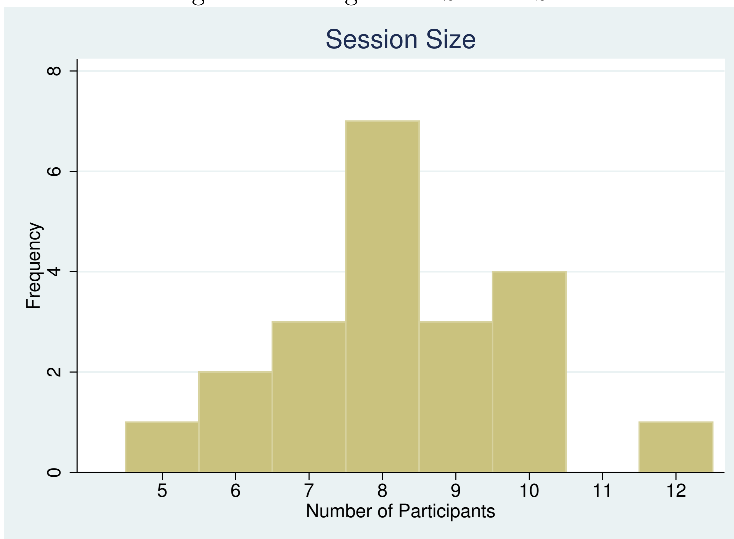
As they arrived, participants were seated at computer terminals with a screen, keyboard, and mouse. Cubicle-like blinders in between each terminal ensured privacy. Once most participants who had signed up for that timeslot had arrived, the session began. Those who had signed up but were not there

[9]The exact text of the recruitment email can be found in Appendix A.

[10]These numbers exclude two sessions which were dropped from the data. In chronological order, the very first session was dropped because of a change in experimental procedure implemented in all subsequent sessions. (The prompts after 7 minutes were added.) The third session in chronological order were also dropped because two of the participants cheated, and ejecting them in the middle of the experiment would have meant only 4 participants would remain in that session.

[11]This variation in session size arose because for each timeslot, ten or twelve seats were made available for sign up, but either not all were filled, or participants who had signed up did not show up. For all intents and purposes, the session size can be considered random.

Figure 1: Histogram of Session Size



were assumed to be no-shows; latecomers were informed that the experiment had already started and encouraged to sign up for another timeslot. Participants who arrived on time and completed any remaining part of the session were disqualified from signing up for another subsequent timeslot by the computer system.

A session consisted of five rounds of memorization tasks. Prior to these rounds, instructions were given to participants regarding their task, which is as follows.[12] In each round, participants are asked to memorize a list of 18 word-number pairs on the first screen. (These numbers are three digits long.) Participants are allowed to choose the amount of time spent on memorizing this list, after which they would click to the next screen. In this second screen, the list of 18 words (reordered randomly) are shown next to text boxes. Participants are asked to fill in as many of the corresponding 3-digit numbers into the text boxes. After submitting their answers, participants are then taken to a third "waiting" screen where they wait for the results of that round to be graded.[13]

While participants were free to choose the amount of time spent memorizing and responding to the questions, prompts were given to all participants at the 7, 9, 10, 11, 12, 13 and 14 minute marks, or up until all participants had submitted

[12]The exact wording of the instructions given to participants at the beginning of each session can be found in Appendix B.

[13]Screen-shots of the computer interface can be found in Appendix C.

their responses, whichever being sooner. (Less than 1% of observed time spent were more than 12 minutes.) These prompts were necessary to ensure that the sessions ran on time, because grading of responses could not proceed until answers from all participants had been submitted.

The participants were ranked by the number of correctly recalled word-number pairs; ties were broken based on time spent on the entire task, in favor of faster participants. There was no penalty for guessing. Each participant was then informed privately of their rank (but not the absolute number of correct responses) using small slips of papers. After passing out these "result slips" the participants were then asked to click on a link on the waiting screen to proceed to the next round.

Out of the total of five rounds in each session, only in the last four were participants given the chance to win monetary prizes. The first round (Round 0) was an example round in which no prize was awarded; this round was used to allow participants to familiarize themselves with the interface. In the later four actual rounds (Rounds 1 through 4), prizes of $X were given to the top Y participants, where X and Y varied with each round. These two parameters were also projected onto a screen at the front of the room during the rounds so participants could remind themselves of the current tournament structure in the middle of any round. The result slips also contained this information for the round just completed. Prior to the beginning of the next round, new values of X and Y were announced verbally and then projected on the screen. Changing X and Y from round to round constitutes changing the parameters of the tournament structure. Data from example Round 0 will not be used in the analysis except in some cases as lags.

The integer Y is the relative winning cutoff, and varies both between sessions and within sessions between rounds. Table 1 lists the relative winning cutoffs used in each session depending on its size.[14] For each session, there were two cutoffs: the low cutoff and the high cutoff. Within a session, two out of four rounds were assigned the low cutoff, while the other two rounds were assigned the high cutoff. For example, suppose at the beginning of the session, 9 participants show up. Then during the rounds assigned the low cutoff, the top 6 out of 9 participants receive the prize; and during the rounds assigned the high cutoff, the top 3 out of 9 participants receive the prize. The row for session size

[14] These are different from the absolute score cutoffs, which will vary from session to session and round to round depending on the absolute performance of participants.

of 11 is struck through because none of the sessions had 11 participants.[15]

Table 1: Relative Winning Cutoffs by Session Size

Session Size	Low Cutoffs	High Cutoffs
5	4 / 5	2 / 5
6	4 / 6	2 / 6
7	5 / 7	2 / 7
8	5 / 8	3 / 8
9	6 / 9	3 / 9
10	6 / 10	3 / 10
~~11~~	~~7 / 11~~	~~3 / 11~~
12	8 / 12	3 / 12

Note: No sessions had 11 participants. "3/10" means that the top 3 out of 10 participants received prizes.

As for the size of the prize $X, two amounts were used: $5 and $10. Again, two out of four rounds were assigned the $5 prize, while the other two rounds were assigned the $10 prize. As such, each participant (in addition to the $15 "show-up" payment) had the chance to win an additional $30, making the maximum possible total payment $45. Table 2 summarizes the four possible combinations of prize amounts and cutoff levels. Since each session comprised four rounds (excluding the example round), there was one round for each combination, as represented by one cell in the table.

Table 2: Possible Prize-Cutoff Combinations

	Low Cutoff	High Cutoff
$5 Prize	Low $5	High $5
$10 Prize	Low $10	High $10

As the experiment transitioned from one round to the next, I manipulated (only) one of the two tournament structure parameters—either $X or Y—holding the other parameter constant. A "treatment" is the change in the tournament structure experienced by participants from one round to the next. There are a total of eight possible treatments. Table 3 lists these eight possi-

[15] There is a "break" in the assignment of the cutoffs as session size increases from 9 to 10. This is done because of funding limitations.

bilities, which I label as +1, -1, +2, -2, etc. I refer to these as treatment codes.

Table 3: Treatments

Treatment Code	Description
+1	Cutoff from low to high @ $5 prize
-1	Cutoff from high to low @ $5 prize
+2	Prize from $5 to $10 @ low cutoff
-2	Prize from $10 to $5 @ low cutoff
+3	Cutoff from low to high @ $10 prize
-3	Cutoff from high to low @ $10 prize
+4	Prize from $5 to $10 @ high cutoff
-4	Prize from $10 to $5 @ high cutoff

Within a session, the experiment was structured such that participants experienced exactly three treatments as follows:

- From Round 1 to Round 2, I moved the cutoff Y (either up or down)
- From Round 2 to Round 3, I moved the prize $X (either up or down)
- From Round 3 to Round 4, I moved the cutoff Y again (in the opposite direction compared to the first treatment)

The decision to move the cutoff twice per session and the prize amount only once is intended to give such cutoff-movement treatments a larger sample size. This is done because the theoretical model predicts that high-performing participants will respond in the opposite direction compared to low-performing participants. This pattern will require additional precision to identify, compared to the comparative statics of an increase in the prize amount, which is in the same direction for both low- and high-performing participants.

In this manner, there are only four possible sequences of treatments. Table 4 lists these four permutations of treatment sequences. A session was assigned to one of four "treatment groups"—A through D—and each group experienced a corresponding treatment sequence, as defined in the table.[16]

Table 5 shows summary statistics (means and standard deviations) of participants by treatment group. While sample sizes for treatment groups are small,

[16] For example, after the example round, participants in group A first experienced a low cutoff with a $5 prize in Round 1, then a high cutoff with a $5 prize in Round 2, next a high cutoff with a $10 prize in Round 3, and finally a low cutoff with a $5 prize in Round 4.

Table 4: Treatment Sequences and Groups

Group	A	B	C	D
Round 0		Example round, no prize		
Round 1	Low $5	Low $10	High $5	High $10
Round 2	High $5	High $10	Low $5	Low $10
Round 3	High $10	High $5	Low $10	Low $5
Round 4	Low $10	Low $5	High $10	High $5
Treatment Sequence	+1, +4, -3	+3, -4, -1	-1, +2, +3	-3, -2, +1

other than an unusually low number of whites for group A, the statistics are very similar across the groups, suggesting that randomization was successfully carried out.

Table 5: Summary Statistics by Treatment Group

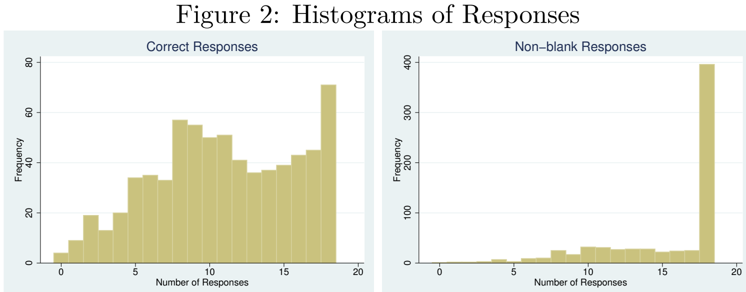
Group:	Overall	A	B	C	D
Male	0.384	0.422	0.35	0.333	0.422
	(0.488)	(0.499)	(0.483)	(0.477)	(0.499)
White	0.6	0.395	0.757	0.703	0.579
	(0.491)	(0.495)	(0.435)	(0.463)	(0.5)
Black	0.11	0.14	0.054	0.135	0.105
	(0.314)	(0.351)	(0.229)	(0.347)	(0.311)
Age	25.06	23.04	27.59	25.36	24.57
	(10.5)	(6.54)	(13.63)	(10.98)	(9.97)
% College	0.831	0.864	0.886	0.805	0.775
Father	(0.376)	(0.347)	(0.323)	(0.401)	(0.423)
% College	0.747	0.778	0.743	0.756	0.707
Mother	(0.436)	(0.42)	(0.443)	(0.435)	(0.461)
N	173	45	40	42	46

Note: Means and proportions within groups shown. Standard deviations in parentheses.

4 Results

The ability of participants to successfully complete the memorization and recall task varied greatly. The first panel in Figure 2 shows the distribution of correct responses in all rounds. There does seem to be top-coding towards the top of the distribution, which suggests that it may be important to consider a constraint such as the inequality in (1) in any theoretical framework. However, a majority of participants obtained results well below this, and adding further word-number pairs beyond the 18 would have increased the time necessary to conduct each experimental session, possibly to an amount that would have been infeasible. The second panel in Figure 2, a histogram of non-blank responses, supports this; because there was no penalty for guessing, almost all participants filled in all 18 text boxes in the second screen.[17]

Figure 2: Histograms of Responses



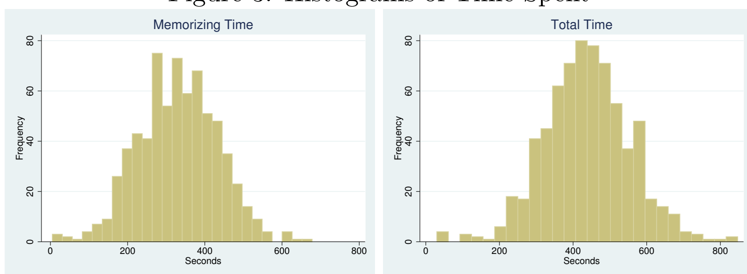
The main outcome of interest is effort as measured by the time taken by participants to memorize the word-number list. Timestamps of exactly when participants clicked from screen to screen are used to obtain the time spent memorizing the list, as well as the time spent filling in the text boxes in the next screen. I call the sum of these two times the total time taken by the participant to complete a round. The two panels in Figure 3 show histograms of memorizing time and total time spent in seconds by all participants. Despite the first prompt given at 7 minutes (420 seconds), and subsequent prompts every minute starting at 9 minutes (540 seconds), there does not seem to be distinct mass points in either distribution at these two particular points in time.[18]

From round to round, as participants are treated with changes in tournament structure, there is considerably variation in the change in time spent. The two panels in Figure 4 show histograms of these changes in both memorizing time and total time spent. This variation, which allows for the identification of

[17]This lack of variation in non-blank responses also invalidates the variable as a measure of effort.

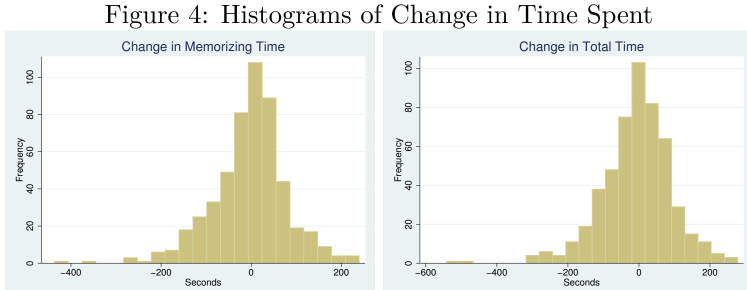
[18]Since prompts were consistently applied to all groups, it can merely be thought of as an additional cost of effort for high levels of effort (longer time spent), insofar as the prompts create discomfort for those still working on their responses.

Figure 3: Histograms of Time Spent



the estimates to follow, can arise from two sources. Firstly, the variation can come from the treatments (changes in tournament structure) induced in the experimental procedure, which is what we are trying to identify. Secondly, as participants proceed from round to round, there may be learning in the sense that they get better at recalling the word-number pairs or more comfortable with the interface and environment. While the former source of variation will cause both positive and negative changes in time spent, the latter source causes only negative changes as participants require less time to complete the same task. Thus, in all the regressions to follow, I include round fixed effects to control for such learning.

Figure 4: Histograms of Change in Time Spent



4.1 Does effort affect outcomes?

Additional effort exertion can affect outcomes in two ways: spending more time on the task increases the number of correct responses; it can also potentially improve a participant's relative ranking. In order to investigate whether these effects exist, and whether memorizing time or total time spent (in seconds) is a better measure of effort, I regress both score and percentile rank on these two time measures, using specifications of the form

$$y_{it} = \alpha + \beta Time_{it} + \mu_t + \mu_i + \varepsilon_{it}$$

where

- y_{it} is the score (correct responses out of 18) or percentile rank (out of 100) of participant i in round t
- $Time_{it}$ is one or more measures of time
- μ_t are round fixed effects
- μ_i are participant fixed effects
- ε_{it} is an error term

The inclusion of round fixed effects controls for learning effects in the coefficient estimates of the time measures[19], and the relative size of the fixed effects can also shed light on how learning evolves as the rounds progress. Furthermore, including participant fixed effects accounts for individual ability as well as any effects which are fixed within session.

Table 6 shows the results for regressions of score (i.e. number of responses correct) on the time measures. The specification of column (1) regresses score on memorizing time. As expected, increasing memorizing time by 1 second increases a participant's score by 0.012 on average (significant at 1% level). The specification of column (2) regresses score on total time. Again, increasing total time spent by 1 second increases a participant's score by 0.009 on average (significant at 1% level).

In order to determine which time measure is a better measure of effort, the specification in column (3) regresses score on both memorizing and total time. Here, we see that only the coefficient of 0.016 on memorizing time is statistically significant (at a 1% level), implying that it is really time spent memorizing—and not the time spent entering responses into the text boxes—that has an effect on score.[20] As such, using memorizing time as the measure of effort seems most appropriate.

[19] This assumes that learning improves outcomes in level terms (e.g. each additional round increases the number of correct responses by Z responses), rather than improving outcomes through increases in the marginal benefit of each additional second of time spent on the task.

[20] This also suggests that the significant coefficient in column (2) arises from the correlation between memorizing and total time.

Table 6: Score Regressions

Dep. Var.: Score	(1) Mem. Time	(2) Total Time	(3) Both	(4) Top / Bottom
Mem. Time	0.012***		0.016***	0.012***
	(0.001)		(0.004)	(0.001)
Total Time		0.009***	-0.003	
		(0.001)	(0.004)	
Mem. Time ×Top				-0.001
				(0.001)
Round				
2	0.758***	0.834***	0.732***	0.737**
	(0.204)	(0.210)	(0.202)	(0.334)
3	0.778***	0.892***	0.740***	0.505
	(0.202)	(0.205)	(0.199)	(0.313)
4	1.167***	1.213***	1.144***	0.955***
	(0.201)	(0.206)	(0.200)	(0.309)
Round×Top				
2×Top				0.038
				(0.447)
3×Top				0.531
				(0.431)
4×Top				0.435
				(0.400)
Constant	6.217***	6.247***	6.465***	6.399***
	(0.475)	(0.541)	(0.536)	(0.474)
N	692	692	692	692
R-square	0.877	0.871	0.877	0.878

Legend: Significance level: *** = 1% ; ** = 5%; * = 10%.

Note: Time variables are in seconds. All specifications include participant fixed effects. Robust standard errors are in parentheses.

To understand learning effects from round to round, we look at the coefficient estimates on the round fixed effects. Note that Round 1 is the omitted category.[21] Estimates in columns (1) through (3) are qualitatively similar, all indicating that there is an initially steep learning curve which tapers off in later rounds. Moving from Round 1 to 2, participants' scores improved by about 0.7 to 0.8 responses on average, ceteris paribus. However, moving from Round 2 to 3 only improves scores by about 0.01 to 0.06 responses on average. There seems

[21] Also recall that there is an example round, so Round 2 is actually the third round encountered by the participant.

to be a further improvement from Round 3 to 4 of about 0.3 to 0.4 responses. All these significantly positive coefficients indicate that there is a good degree of learning as the rounds progress.[22]

We can also check whether there is heterogeneity in learning or returns to effort. To do so, I create a "top" indicator which takes on the value of 1 when a participant's percentile rank from the previous round is above 50%. Henceforth, I will refer to such "top" participants as high-performing participants (as opposed to low-performing participants). I then interact this top indicator with all the right hand side variables of interest. Column (4) in Table 6 shows the coefficient estimates for the specification where score is regressed on memorizing time, its interaction with the top indicator, round fixed effects, and their interactions with the top indicator.

The baseline coefficient on memorizing time is identical to the one in column (1), the specification without the added interaction terms. The coefficient on the interaction of memorizing time and the top indicator is small and statistically insignificant. This implies that the returns to effort of low-performing participants (as reflected in results from the previous round) is no worse than that of high-performing participants.

Furthermore, while the coefficients on the baseline round fixed effects are quantitatively similar to those in column (1) (except for the one for Round 3, which becomes statistically insignificant), none of the coefficients on the interacted fixed effects are statistically different from zero. This implies that both high- and low-performing participants learned at similar rates as the rounds progressed.

Instead of focusing on absolute score, we can also examine relative percentile rank as an outcome of effort exertion and learning. Table 7 shows the results for regressions of percentile rank (out of 100) on the time measures. The specification of column (1) regresses percentile rank on memorizing time. The coefficient estimate suggests that spending 1 more second on memorizing the word-number list increases a participant's percentile rank by 0.03% on average.

However, regressing percentile rank on total time instead in column (2) yields a statistically insignificant coefficient estimate. The reason for this is explained by the results in column (3), for the specification in which percentile rank is

[22]Notice that the estimates are average learning effects averaged over all treatments and treatment groups. Also note that these estimates are conditional on memorizing time, which means that if memorizing time is affected by the various treatments (as will be shown later), then these coefficients are measuring learning independent of any round to round treatment effects.

Table 7: Percentile Rank Regressions

Dep. Var.: Percentile Rank	(1) Mem. Time	(2) Total Time	(3) Both	(4) Top / Bottom
Mem. Time	0.030***		0.118***	0.081**
	(0.010)		(0.030)	(0.034)
Total Time		0.012	-0.080***	-0.048*
		(0.008)	(0.025)	(0.027)
Mem. Time				0.101**
×Top				(0.040)
Total Time				-0.085***
×Top				(0.030)
Round				
2	-0.013	0.094	-0.678	0.436
	(1.589)	(1.599)	(1.569)	(2.347)
3	-0.020	0.139	-1.005	0.353
	(1.460)	(1.478)	(1.472)	(2.161)
4	0.330	0.236	-0.290	0.560
	(1.628)	(1.657)	(1.617)	(2.281)
Round×Top				
2×Top				-2.441
				(3.184)
3×Top				-2.585
				(2.897)
4×Top				-1.899
				(3.091)
Constant	45.901***	50.705***	52.352***	52.499***
	(3.464)	(3.824)	(3.748)	(3.943)
N	692	692	692	692
R-square	0.811	0.808	0.817	0.823

Legend: Significance level: *** = 1% ; ** = 5%; * = 10%.

Note: Time variables are in seconds. Percentile rank is out of 100. All specifications include participant fixed effects. Robust standard errors are in parentheses.

regressed on both memorizing time and total time. The coefficient on memorizing time is positive and statistically significant at 0.12. Combining this with the coefficient on total time, this implies that a 1 second increase in memorizing time increases percentile rank by 0.04%, similar to the estimate in column (1). While spending more time memorizing does seem to move participants up the rankings, spending more time in total (memorizing and responding) has an averse effect on percentile rank. The negative and statistically significant coefficient suggests that 1 more second of total time (or, equivalently, time spent filling in text boxes) reduces percentile rank by 0.08%. This negative effect is most likely the result of ties being broken in favor of participants who complete the entire task (memorizing and responding) faster.

Column (4) in Table 7 shows results for the specification where all the right hand side variables in column (3) are interacted with the top indicator. The coefficient on the interaction term between memorizing time and the top indicator is positive and statistically significant (at a 5% level). Moreover, the interaction term between total time and the top indicator is negative and statistically significant (at a 1% level). These results suggest two things.

First, high-performing participants (as reflected in results from the previous round) have better relative ranking returns to effort compared to low-performing ones—by about 0.015% per second of memorizing time in percentile rank terms. The fact that this difference is not observed in returns to effort in score may imply that high-performing participants (especially the top-performing ones) are being censored at the top by the maximum score of 18. Thus, while they are not able to increase their score anymore through additional effort exertion, they are still able to improve their relative ranking through being faster and hence breaking ties in their favor.

This latter point is the second point suggested by the results. The -0.085 statistically significant coefficient on total time interacted with the top indicator implies that each additional second of total time (or, equivalently, response time) hurts high-performing participants much more in relative ranking terms than low-performing counterparts. This means that high-performing participants are more likely to encounter ties with other participants (in part due to top censoring), so those additional seconds in total time spent matter more for their relative ranking.

Throughout all the specifications, coefficients on round fixed effects, as well as their interaction with the top indicator, are statistically insignificant. This implies that learning does not affect relative ranking. This would be expected since we concluded previously that there seems to be no heterogeneity in learning. If every participant learns at about the same rate, regardless of relative position, then even as participants improve as the rounds progress, there is no relative gain for any one participant compared to other participants who improve at the same rate.

4.2 Effort Response to Changes in Tournament Structure

Having settled on using time spent memorizing the word-number list as the measure of effort, this section now examines the effect of tournament structure on effort. To recap, the tournament model in Section 2 predicts the following:

1. Having the relative winning cutoff shift towards (away from) a participant's own high or low location in the score distribution increases (decreases) effort.

2. Increasing (decreasing) the prize amount increases (decreases) effort.

Equivalently,

1. Participants near (far from) the relative winning cutoff (i.e. high-performing when high cutoff, low-performing when low cutoff) exert more (less) effort

2. Participants competing for higher (lower) prizes exert more (less) effort

A simplified way to test these predictions is to assume symmetry in "opposite" treatments. By opposite treatments, I am referring to treatment +1 being the opposite of treatment -1, in that the former treatment is a movement of the cutoff from low to high at a $5 prize, while the latter treatment is the movement of the cutoff in the opposite direction from high to low at a $5 prize, and similarly for other treatment pairs. To assume symmetry in opposite treatments means that the treatment effect of a movement in a tournament structure parameter is exactly the negative of the treatment effect of a movement in the opposite direction of the same tournament structure parameter.[23] This will be true if, for instance, there are no behavioral patterns which predict dissimilar treatment effects based on directionality (e.g. loss aversion, endowment effects, reference points, etc.).

To test the two predictions above under symmetry in opposite treatments, I run pooled regressions of the form

$$\begin{aligned} Effort_{it} = \beta_0 + \beta_1 LowCutoff_{it} \times Bottom_{i(t-1)} + \beta_2 HighCutoff_{it} \times Top_{i(t-1)} \\ + \beta_3 HighPrize_{it} + X_{it}\gamma + \mu_t + \mu_i + \varepsilon_{it} \quad (6) \end{aligned}$$

where

[23] For instance, if increasing the prize from $5 to $10 when the cutoff is low induces effort to increase by 10 seconds, then decreasing the prize from $10 to $5 (at the same low cutoff) should make effort decrease by 10 seconds.

- $Effort_{it}$ is the memorizing time in seconds of participant i in round t
- $LowCutoff_{it}$ and $HighCutoff_{it} \equiv 1 - LowCutoff_{it}$ are indicators for whether the relative winning cutoff is low or high respectively for participant i in round t
- $Top_{i(t-1)}$ and $Bottom_{i(t-1)} \equiv 1 - Top_{i(t-1)}$ are indicators for whether participant i was ranked in the top or bottom half (respectively) among all participants in the previous round $(t-1)$
- $HighPrize_{it}$ is an indicator for the prize being \$10 (as opposed to \$5) for participant i in round t
- X_{it} is a set of other possible controls which may be included in some specifications

All regressions also contain round and participant fixed effects.

Column (1) of Table 8 shows coefficient estimates of the regression using specification (6) where no additional controls (X_{it}) are included. Low-performing (i.e. bottom) participants spend on average of 24.4 seconds more memorizing the list during rounds where the cutoff is low. On the other hand, high-performing (i.e. top) participants do not seem to be induced by the nearness of the cutoff to exert more effort. Furthermore, increasing the prize amount from \$5 to \$10 induces all participants to exert on average 33.0 more seconds of effort.

Two possible concerns arise from this first specification. First, high-performing participants (as determined by the previous round) may be encouraged by their performance independent of any change in the tournament structure. Second, high-performing participants may respond differently to a higher (\$10) prize compared to low-performing participants. In order to address these two concerns, I augment the specification in column (1) with two controls: the top indicator, and its interaction with $HighPrize_{it}$.

Column (2) shows the results of this regression. The coefficient estimates of interest (β_1, β_2, and β_3) are qualitatively similar to those of column (1). Being low-performing near a low cutoff increases one's effort by 29.8 seconds. Being high-performing near a high cutoff has no significant effect on effort. Lastly, the larger \$10 prize induces 40.6 seconds more of effort. Interestingly, high-scorers spend an average of 27.4 seconds more memorizing. (This estimate is significant at a 5% level.) Since participant fixed effects are included, this

Table 8: Pooled Regressions (Assumes Symmetry)

Dep. Var.: Mem. Time	(1) Pooled	(2) Pooled Top	(3) Log	(4) No Round 1	(5) No Middle
Low Cutoff	24.448***	29.759***	0.127***	25.706**	30.381**
×Bottom	(8.15)	(8.953)	(0.047)	(11.603)	(12.878)
High Cutoff	0.983	-3.307	0.003	-0.271	-1.397
×Top	(5.567)	(5.379)	(0.018)	(6.182)	(6.848)
$10 Prize	32.97***	40.621***	0.101**	32.349***	33.555**
	(4.996)	(8.888)	(0.048)	(10.694)	(13.782)
$10 Prize		-14.705	-0.030	-2.373	-7.139
×Top		(10.589)	(0.053)	(12.672)	(15.517)
Top		27.356**	0.095**	14.713	19.023
		(11.641)	(0.047)	(14.035)	(22.121)
Round					
2	0.296	0.216	-0.003		-2.435
	(6.955)	(6.919)	(0.029)		(10.388)
3	0.324	0.029	0.008	0.004	-2.096
	(6.514)	(6.428)	(0.026)	(6.231)	(9.523)
4	-11.005	-11.219	-0.090**	-11.303	-21.996*
	(8.17)	(8.128)	(0.041)	(7.456)	(12.236)
Constant	311.109***	296.866***	5.637***	304.704***	304.551***
	(6.322)	(9.476)	(0.043)	(10.085)	15.838
N	692	692	692	519	472
R-Square	0.697	0.700	0.577	0.756	0.719

Legend: Significance level: *** = 1% ; ** = 5%; * = 10%.

Note: Memorizing time in seconds. All specifications include participant fixed effects. Robust standard errors are in parentheses.

estimate represents the "encouragement" effect of being a high-performer from the last round, independent of any time-invariant "ability" factors.

Column (3) shows the same regression except now, the dependent variable is the logarithm of memorizing time. The pattern of coefficient estimates, now interpretable as percentage changes, remains similar. In particular, being low-performing near a low cutoff increases one's effort by 13%, whereas being high-performing near a high cutoff has no significant effect. A larger prize of $10 increases effort by 10%.

Specifications in columns (1) through (3) contain top and bottom indicator variables coming from Round 0. However, since this was an example round without monetary prizes, participants may not have taken this round seriously[24], and the percentile rank outcomes reported to them may not reflect their true belief about their relative position. To address this concern, column (4) uses the same specification as column (2), except observations from Round 1 (which use lagged dependent variables from Round 0) have been excluded. The coefficient estimates are qualitatively similar to those in column (2), suggesting that this should not be too big of an issue.

One additional concern is that the group in between the low and high cutoffs (before and after treatment) are affected by two opposing forces as the cutoff moves "over" them. On the one hand, the cutoff is moving towards them, inducing them to increase effort. On the other hand, the cutoff then passes them and moves away from them, inducing them to decrease effort. Depending on the error structure (the shape of $h(.)$), their effort response may be positive or negative. Column (5) uses the same specification as column (2), except the middle third of participants (based on the percentile rank from the previous round) has been excluded. Again, the coefficient estimates are qualitatively similar, suggesting that any irregularity in effort response by this middle group is not a significant concern.

Learning through the rounds in terms of reductions in effort exerted does not seem to be taking place. This can be seen from the insignificant coefficient estimates on the round fixed effects in all three columns.[25] This is nonetheless consistent with the findings from the previous section. On average, from the results in this subsection, participants are not learning how to memorize the lists faster as they progress through the rounds; however, from the results previously, they do seem to be learning how to memorize the lists with better accuracy as they progress through the rounds, thereby increasing their scores—though again, in relative ranking terms, there is no improvement. And since only relative rank was reported to the participants in the result slips, they would believe they were not improving through learning, thus maintaining similar effort levels, ceteris paribus.

To summarize, these results suggest that all participants (regardless of performance) are induced by larger prizes to increase effort. However, only low-performing participants are incentivized by being near a low cutoff to work

[24] An attempt was made to encourage participants through the instructions to take this round seriously as practice for the real thing. However, some participants immediately clicked through the example round screens without completing a single response.

[25] These specifications treat the rate of learning as the same for all participants.

harder to surpass the nearby cutoff and win a prize. High-performing participants, it seems, are not threatened by the risk of falling below a nearby high cutoff, and do not alter their effort choice. I will discuss possible explanations for these findings in the discussion subsection.

But before that, it would be interesting to know whether these findings remain as they are without making the symmetry in opposite treatments assumption. To do this, I estimate the treatment effects for each of the 8 possible treatments (+1, -1,...,-4), and I do so separately for high-performing (i.e. top) and low-performing (i.e. bottom) participants. The regression specification I use is of the form

$$\begin{aligned}\triangle Effort_{it} = & \sum_{\tau} \beta_{Top,\tau} Top_{i(t-1)} \times Treatment\left(\tau\right)_{it} \\ & + \sum_{\tau} \beta_{Bottom,\tau} Bottom_{i(t-1)} \times Treatment\left(\tau\right)_{it} + \mu_t + \varepsilon_{it} \quad (7)\end{aligned}$$

where

- $\triangle Effort_{it} = Effort_{it} - Effort_{i(t-1)}$ is the change in effort from the previous round
- $Treatment\left(\tau\right)_{it}$ is a set of dummies indicating that treatment τ was administered to participant i in period t

and other notation are similarly defined as before. Note that because the perfectly colinear top and bottom indicators are both included with a full set of treatment dummies, no constant is necessary. For this specification, because neither lagged effort nor the top and bottom indicators in Round 0 can be taken seriously (for the same reasons as stated before for the top indicator), only observations from Round 2 onwards (with lag dependent variables from Round 1) are used. The inclusion of round fixed effects again controls for learning through the rounds.[26,27]

The two panels in Table 9 present the coefficient estimates from the regression using the above specification (7). Columns are labeled with treatment

[26] However, because Round 2 observations are used as the omitted category, and Round 3 treatments are always movements in the prize amount, and therefore omitted when including a full set of treatment dummies, only a coefficient estimate for Round 4 can be identified. This estimate is -11.170 with a standard error of 9.056, which is not statistically significant.

[27] I do not include participant fixed effects because the dependent variable is the first difference of effort, so any participant specific component of effort is differenced out already.

codes. The first panel shows treatments in which the cutoff is shifted; the second panel shows treatments in which the prize amount is changed. The sample size (N) for this regression is 519, and the R-square is 0.096.

Table 9: Regressions with Separate Treatment Effects

Dep. Var.: △Effort	(+1) Treatment	(-1) Treatment	(+3) Treatment	(-3) Treatment
Top	-4.106	5.400	6.110	12.773
	(11.201)	(8.809)	(9.035)	(11.812)
Bottom	-23.690*	23.468*	-44.928**	23.288
	(13.64)	(13.706)	(21.472)	(15.895)

Dep. Var.: △Effort	(+2) Treatment	(-2) Treatment	(+4) Treatment	(-4) Treatment
Top	22.405	-12.312	35.871***	-35.281***
	(14.165)	(10.732)	(10.871)	(13.12)
Bottom	24.881	-28.174	33.633*	-46.305**
	(19.81)	(20.93)	(17.757)	(18.383)

Legend: Significance level: *** = 1% ; ** = 5%; * = 10%.

Note: Change in effort measured in seconds. All specifications include round fixed effects (see footnote 26). Robust standard errors are in parentheses. N=519; R-square is 0.096.

The coefficient estimates for each separate treatment effect, while not as statistically significant as the corresponding estimates in the pooled regression[28], reveal similar findings. For treatments in which the cutoff shifts from low to high (+1, +3), effort for low-performers decreases. Correspondingly, for treatments in which the cutoff shifts from high to low (-1, -3), effort for low-performers increases (though only treatment -1 has a statistically significant effect at the 10% level). And similar to earlier findings, high-performers do not respond significantly to any shift in the cutoff, be it towards or way from their high relative position (see the top row of the first panel).

For treatments in which the prize amount increases (+2, +4), effort for all participants increases. (However, only for treatment +4 are the estimates statistically significant.) Correspondingly, for treatments in which the prize amount decreases (-2, -4), effort for all participants decreases. (Again, however,

[28]This is because by not assuming symmetry, each “cell” within each treatment by top/bottom combinations is identified over a smaller number of observations.

only for treatment -4 are the estimates statistically significant.) In general, the coefficient estimates for high-performers are more or less similar in magnitude to those of low-performers for each column of prize-amount moving treatments in the second panel.

From this set of estimates, it is hart to tell if the symmetry assumption holds. The joint significance test of the hypothesis that every pair of opposite treatment effects is the negative of one another cannot be rejected. However, given the small sample size of each cell, this is not all that surprising. In Appendix D, I present non-parametric local polynomial regressions which show the treatment effect across the entire relative percentile rank distribution.

4.3 Discussion

The results in the previous section suggest that all participants (regardless of performance) are induced by larger prizes to increase effort. However, only low-performing participants are induced by being near a low cutoff to work harder to surpass this cutoff and win a prize. High-performing participants do not seem to be affected by being near a high cutoff, and are not changing their effort choice to avoid falling below this cutoff. Relating these results back to the model, there are several possible explanations as to why this is so.

First, are the high-performers not responding because they believe they are being capped? While it is clear from the distribution of scores in Figure 2 that there are top performers who hit the maximum score of 18, since the absolute scores are never reported to them, their effort decisions may not take this into account. Equation (5) of the model suggests that if top performers believed they were capped, the constraint on score would bind, λ would be greater than zero, and the positive $\frac{de_i}{d\alpha}$ would be lower (and potentially be negative even). However, high performers and low performers responded similarly to an increase in the prize amount; the coefficient estimate on the interaction term between the high prize and top indicators, while negative, is not statistically different from zero. This suggests that at least a large fraction of top performers are not aware of hitting the cap (and thus consider $\lambda = 0$).

Second, should we be concerned about how the tie-breaking rule may affect behavior, in particular those high performers who are non-responsive to cutoff shifts? Ties were pretty common: of the 692 observations from Rounds 1 to 4 (173 participants multiplied by 4 rounds per participant), a total of 204 of these constituted two-way ties (102 pairs), and a total of 63 constituted three-way ties (21 triples). Recall that from the discussions regarding Table (7), the application of the tie-breaking rule did seem to affect the relative ranking of participants.

The tie-breaking rule basically generates an incentive for participants to finish faster (i.e. exert less effort), ceteris paribus (conditional on score), because doing so increases the probability of winning any tie. While it is possible to add in an additional term into the utility function to account for this, it can equivalently be thought of as an additional cost of exerting too much effort. From equations (4) and (5), such a cost would not affect $\frac{de_i}{dp}$ and $\frac{de_i}{d\alpha}$ differentially, since the cost function appears identically in the denominator. Thus, even if the tie-breaking rule were to affect high-performing participants differentially somehow, this would still not explain their significant effort response to a change in α, versus no response to a change in p.

Looking more closely at equations (4) and (5), we note that one difference between their numerators is the additional $\frac{d\overline{S}_p}{dp}$ term in $\frac{de_i}{dp}$. Hence, one possible explanation for the high-performing participants' non-response is heterogeneous expectations of the sign of $\frac{d\overline{S}_p}{dp}$. Since the exact location of the absolute score cutoff (or even one's own score) is not revealed, each participant will have individual beliefs as to what the expected value of $\frac{d\overline{S}_p}{dp}$ is. That it does not appear in the equation (5) explains why there is no difference between high- and low-performers when the prize amount is changed.

Suppose low performers always expect $\frac{d\overline{S}_p}{dp}$ to be positive, whereas some high performers, being possibly more sophisticated, expect $\frac{d\overline{S}_p}{dp}$ to be negative as well with some non-zero probability. Such negative expectations may arise because high-performers (correctly) believe that when p increases, some low performers give up and reduce their effort, thus exerting downward pressure on the cutoff $\overline{S}_p$. As such, $\frac{de_i}{dp}$ will be negative for these high performers, and any estimate of the average effort response will be less positive, and less statistically significant due to the greater variation in effort responses. Low performers, on the other hand, will always reduce effort when p increases, if they believe $\frac{d\overline{S}_p}{dp} > 0$ with probability 1. And because $\frac{d\overline{S}_p}{dp}$ does not appear in $\frac{de_i}{d\alpha}$, there would not be any heterogeneity in effort response to an increase in the prize amount.

5 Conclusion

The policy implication of these findings is that if one wants to incentivize participants to work harder, the best way to do this would be to increase prize amounts (i.e. dangle a bigger carrot). On the other hand, if one's goal is to encourage only low-performers to work harder, then the results suggest that the best way to go about doing this is to shift the cutoff closer to the lower end of the distribution (i.e. dangle the carrot further down). The experimental results suggest that this latter method would not discourage high-performers from slacking off, though by lowering the relative wining cutoff, more prizes have to be offered, the number of which may be constrained by some fixed budget. The results in this study are similar to those of Harbring and Irlenbusch (2003), who find that as the relative winning cutoff is lowered, mean effort increases.

One solution to overcome such a budgetary constraint is to create an eligibility criteria which excludes high-performers from winning prizes. However, this criteria cannot be a function of the score (or high-performers will be encouraged to "fake" a low score), so it must be based on some other factor correlated with performance. In this way, low-performing students are still encouraged by the low relative cutoff near them to step up effort, but there is no need to allocate scare resources to prizes for high-performing students. Examples of this include merit scholarships given only to low-income students, or ones given only to students in below-average schools (with a higher incidence of low-performing students).

One concern regarding identification in this experiment is that there is no control group of participants who remain untreated during any of the round transitions. Moving from one round to the next, one of the tournament structure parameters is always changed. (That is, there is no round transition in which $X and Y number of winners is held constant.) This was done to ensure fairness across sessions, so that in every session, the maximum amount of prize money (on top of the $15 base amount) any participant could earn stays the same at $30.[29] Thus, an underlying assumption which is being made in identification is that conditional on round fixed effects (possibly from learning), effort (or other outcomes) remain constant from round to round when the tournament structure remains unchanged.

[29]This feature of not having a control group actually occurs often in the experimental literature. For example, the experiment involving bicycle messengers in Fehr and Goette (2007) also has this feature.

While the sample size of this experiment is relatively small and unrepresentative, the findings are partially supported by tournament theory. As with any experiment, there is the question of external validity. But using a real effort task and a novel yet unobtrusive measure of effort hopefully addresses at least some of these concerns, especially those relating to the ability of this experiment to reflect the real world, or the possibility of Hawthorn effects.

Appendix

A Recruitment Email

Subject: Sign up for our Economics Experiment!

Hi!

We are recruiting participants for an economics experiment to be carried out at the Princeton Laboratory for Experimental Social Science (PLESS) located in Green Hall. The experiment will involve several recall tasks. You will also be asked to complete a short survey. You will receive a base amount $15 for your participation, as well as the potential to win up to an additional $30 depending on your performance on the tasks. The entire experiment should take approximately 60 minutes.

If you are interested in helping to further the frontiers of economics research, please visit [web sign up form] to sign up.

Thanks,

Yan Lau

Project PI

B Instructions Provided to Participants

The following instructions (in point form) were projected on a screen in front of the room at the beginning of each session. They were also read out to participants in full, who were given the chance to interrupt and ask clarifying questions.

- In this study, you will be asked to complete memorization tests involving words and numbers.
- There will be 5 rounds of these tests. In each round, the following will occur:
    1. You will be asked to memorize a list of 18 pairs of words and 3-digit numbers.

2. You may choose the amount of time you spend memorizing this list. (You will be prompted at 7 min.) After you are satisfied, you will then click to the next screen.
3. In the next screen, you will be presented with a list of the 18 words (reordered randomly) with text boxes next to them. Your task is to fill in the 3-digit numbers associated with each word.
4. Your score (out of 18) will be the number of 3-digit numbers you are able to recall correctly.
5. You don't have to answer all of them. (Most participants don't.) There is no penalty for guessing.
6. After all participants have finished, everyone will be ranked based on their score. Ties (which are common) will be broken based on who took less time to complete the entire task.
7. Based on this ranking, the top X participants will be awarded a monetary prize. (PLESS is non-deception) The number X and the size of the monetary prize will be announced at the beginning of each round.

- The initial round (round 0) is an example round to introduce the interface. As such, there will be no prize.
- However, we highly suggest you complete this round as if it were the real thing, as many have found it helpful for practicing their memory techniques.
- The remaining four rounds (rounds 1 to 4) will have prize rules announced at the beginning of the round.
- During each round when you are interacting with the computer, please
  - refrain from using electronic devices (including cell phones), and any writing or reading materials
  - do not communicate with other participants
  - do not use the "back" feature on the web browser (an error page will appear)
  - excludes "rest time" in between rounds

- Cheating, or any attempt to gain an unfair advantage, is grounds for disqualification.
- Please raise your hand at any point if you encounter issues or if you have any questions.
- You will now receive Participant ID numbers on sticky-notes. Please be sure to enter these correctly.

C Screen-shots of Computer Interface

In each round, the first screen encountered by participants was the list of word-number pairs shown in figure 5. After memorizing the list, participants click the "next" button to proceed to the second screen shown in figure 6, which contains text boxes along with the reordered list of words. Having filled in their responses, they click "submit" to proceed to the third screen shown in figure 7. This last screen of the round informs participants that they should wait for their results to be graded and relax until instructed to proceed to the next round.

Figure 5: First Screen: Word-Number Pairs

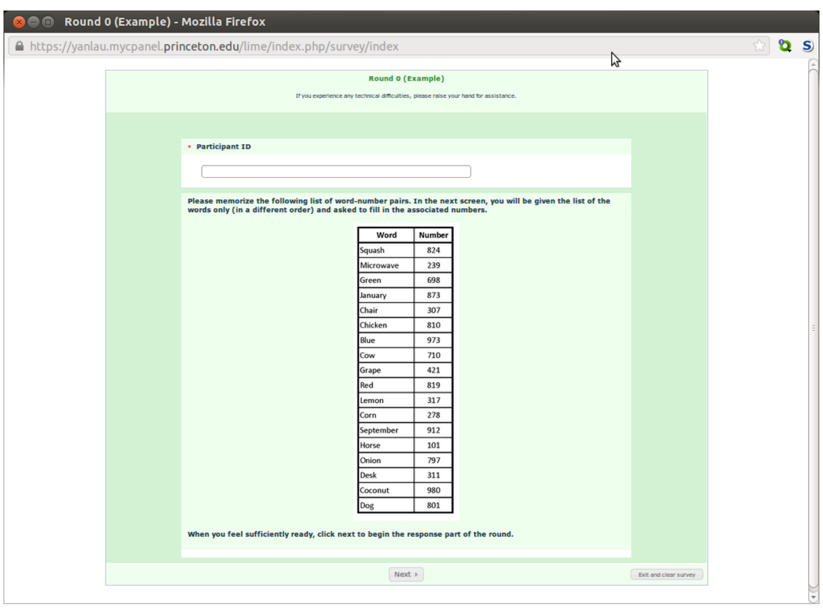


Figure 6: Second Screen: Words with Text Boxes

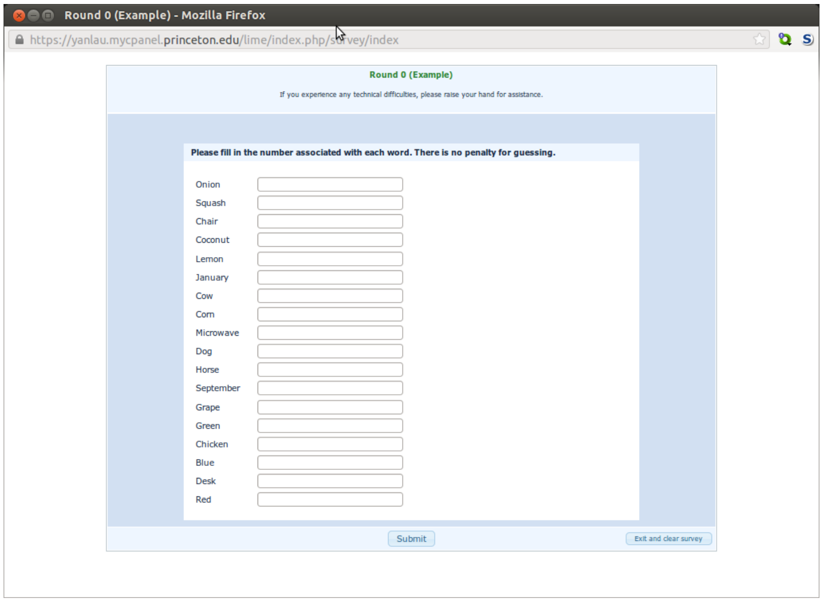
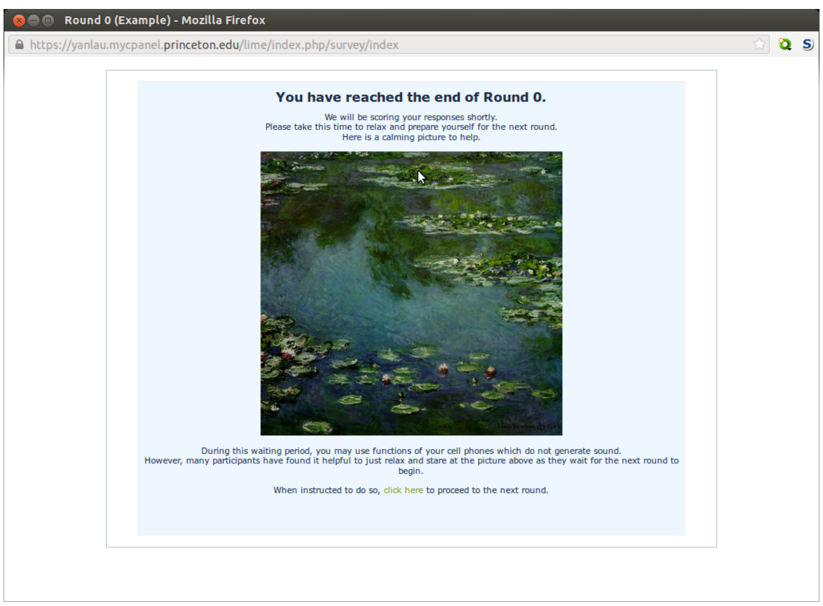


Figure 7: Third Screen: End-of-Round Waiting

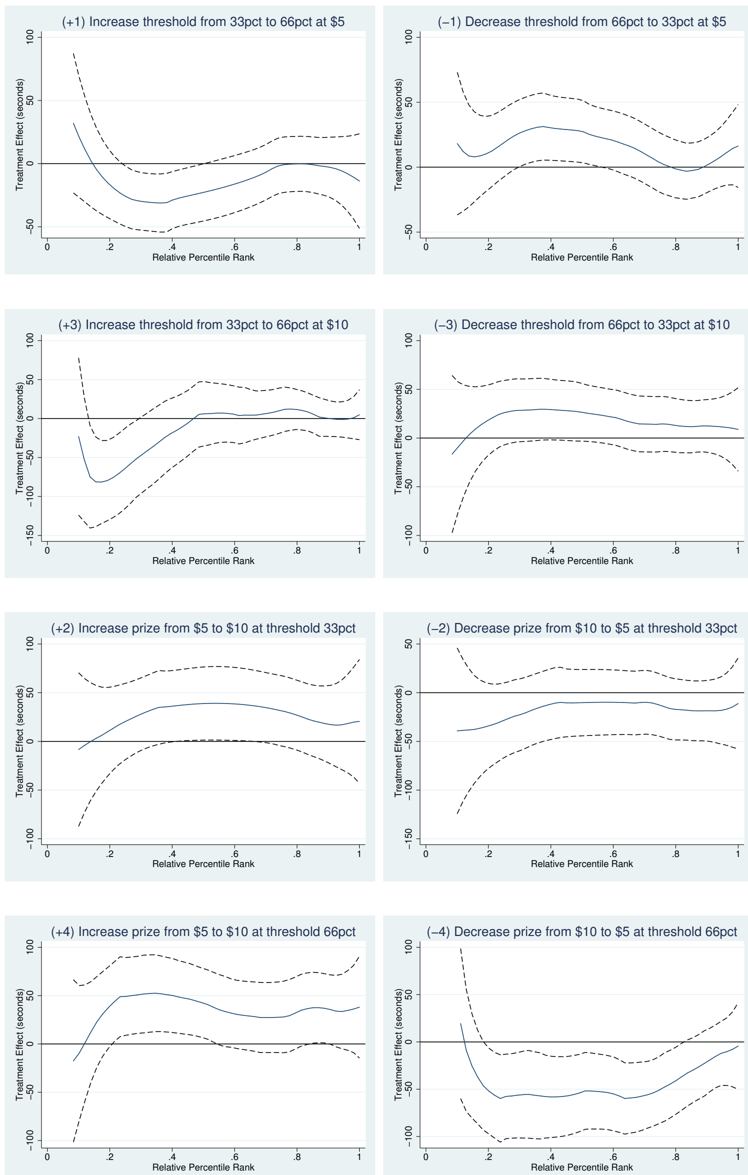


D Local Polynomial Regressions

While the regressions with separate non-symmetric treatment effects in Section 4 consider differences between the top- and bottom-performing participants using "top" and "bottom" indicator variables, it would be interesting to see how the treatment effect varied across the entire performance distribution. To graph this out, I run a non-parametric local polynomial regression with residual change in effort on the y-axis and relative percentile rank on the x-axis. (Residual change in effort is obtained by regressing change in effort on round fixed effects and a constant and then backing out the residuals using the coefficient estimates.)

I run these non-parametric regressions separately for each treatment. Figure 8 shows 8 panels corresponding to each of these regressions. The dashed lines are 95% confidence intervals. The graphs reinforce the results discussed in Section 4, and are qualitatively similar to those described in the previous regressions.

Figure 8: Non-parametric regressions



References

Boudreau, Kevin J., Constance E. Helfat, Karim R. Lakhani, and Michael Menietti (2012) "Field Evidence on Individual Behavior & Performance in Rank-Order Tournaments," *Harvard Business School Working Paper*, No. 13-016, August.

Bull, Clive, Andrew Schotter, and Keith Weigelt (1987) "Tournaments and Piece Rates: An Experimental Study," *Journal of Political Economy*, Vol. 95, No. 1, pp. pp. 1–33.

Casas-Arce, Pablo and F. Asis Martinez-Jerez (2009) "Relative Performance Compensation, Contests, and Dynamic Incentives," *Management Science*, Vol. 55, No. 8, pp. 1306–1320.

Eriksson, Tor (1999) "Executive Compensation and Tournament Theory: Empirical Tests on Danish Data," *Journal of Labor Economics*, Vol. 17, No. 2, pp. 262–280.

Fehr, Ernst and Lorenz Goette (2007) "Do Workers Work More if Wages Are High?: Evidence from a Randomized Field Experiment," *The American Economic Review*, Vol. 97, No. 1, pp. 298–317.

Green, Jerry R. and Nancy L. Stokey (1983) "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, Vol. 91, No. 3, pp. 349–364.

Harbring, Christine and Bernd Irlenbusch (2003) "An experimental study on tournament design," *Labour Economics*, Vol. 10, No. 4, pp. 443 – 464, August. Special Issue on Labour Market Experiments.

Lazear, Edward P. and Sherwin Rosen (1981) "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, Vol. 89, No. 5, pp. 841–864.

Main, Brian G. M., Charles A. O'Reilly III, and James Wade (1993) "Top Executive Pay: Tournament or Teamwork?" *Journal of Labor Economics*, Vol. 11, No. 4, pp. 606–628.

Orrison, Alannah, Andrew Schotter, and Keith Weigelt (1997) "On the Design of Optimal Organizations Using Tournaments: An Experimental Examination," Working Papers 97-26, C.V. Starr Center for Applied Economics, New York University.

Schotter, Andrew and Keith Weigelt (1992) "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results," *The Quarterly Journal of Economics*, Vol. 107, No. 2, pp. 511–539.

van Dijk, Frans, Joep Sonnemans, and Frans van Winden (2001) "Incentive systems in a real effort experiment," *European Economic Review*, Vol. 45, No. 2, pp. 187 – 214.