The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices

Wozniak, David and Harbaugh, William T. and Mayr, Ulrich

Eastern Michigan University, University of Oregon

August 2009

The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices[1]

David Wozniak[2]

William T. Harbaugh[3]

Ulrich Mayr[4]

March 25, 2011

[1]We are grateful for support and direction from Chris Minson and Paul Kaplan of the University of Oregon Human Physiology Department. Seminar participants at the (June) 2009 ESA and (July) 2009 IAREP/SABE meetings provided useful ideas and comments. Correspondence should be addressed to David Wozniak (Eastern Michigan University) at dwozniak@gmail.com. We would also like to thank Jack Wozniak for his programming assistance and Cathleen Leue, the director of the Social Science Instructional Labs and her staff at the University of Oregon. A final thanks to Sarah Wunderlich for all her help scheduling and running experiments.

[2]Eastern Michigan University

[3]University of Oregon

[4]University of Oregon

The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices.

March 25, 2011

Abstract

We use a within-subjects experiment with math and word tasks to show that feedback about relative performance moves high ability females towards more competitive forms of compensation such as tournaments, moves low ability men towards piece rate and group pay, and eliminates gender differences in choices. We also examine choices for females across the menstrual cycle, and find that women in the high-hormone phase are more willing to compete than women in the low phase, though somewhat less willing to compete than men. There are no significant differences between the choices of these groups after they receive relative performance feedback.

Introduction

Economic experiments have shown when given the choice between piece rate and winner-take-all tournament style compensation, women are more reluctant than men to choose tournaments (e.g. Niederle and Vesterlund 2007). These gender difference experiments have all relied on a framework where subjects were not informed of their abilities relative to potential competitors. We use a within-subjects design and replicate these previous findings for a math task, and show they also exist for a word task. We then show that feedback about relative performance moves high ability females towards more competitive compensation schemes, moves low ability men towards less competitive schemes such as piece rate and group pay, and removes the average gender difference in compensation choices. We also examine between and within-subjects differences in choices for females across the menstrual cycle. We find that the relative reluctance to choose tournaments on the part of women comes mostly from women in the low-hormone phase of their menstrual cycle. Women in the high-hormone phase are substantially more willing to compete than women in the low phase, though still somewhat less willing to compete than men. There are no significant differences between the choices of any of these groups after they receive relative performance feedback.

In low information settings the effects of gender and menstrual phase are large. A female has a 0.14 lower probability of choosing a tournament compared to a male, even when controlling for performance and confidence. For a female to be as likely to choose a tournament as an average male she must believe she is 40% better than average in performance. We find that the within-gender menstrual phase effect is larger than the across-gender effect. Females in the low-hormone phase of their cycle have a 0.16 lower probability of choosing a tournament than females in high-hormone phase. A low phase female must believe she has 50% better performance to be as likely to compete as a female in the high-hormone phase.

Without feedback, high ability females and males are both more reluctant to enter tournaments than expected value maximization would require. This effect is larger for high ability females. On the other hand, too many low ability types enter competitive environments, and this effect is larger for males. Relative performance feedback moves all these groups toward more optimal choices. This result suggests that the behavioral differences in the willingness to compete are driven less by stable preference differences than by differing reactions to the generally poor information concerning a person's relative rank.

One motivation for experiments on gender differences in competitive behavior comes from the job market where many careers involve a tournament aspect. An example of this is the corporate ladder, where females make up a small portion of top-level executive positions. Bertrand and Hallock (2001) found that in 1997 the fraction of females in top level management positions was 3% and only 15% of firms had at least one female in a top level executive position. There are many potential explanations for this underrepresentation. Discrimination is one explanation, the effect of traditional family roles and raising children on women's career choices and human capital investments is another (Polachek 1981). Part of this underrepresentation may be caused by a preference by females against competitive, tournament-like situations in favor of alternatives – or by a preference by men towards competition and tournaments.

Such a preference difference could have many causes. For example, Jirjahn and Stephan (2004) argue that the attractiveness of piece rate schemes for females is likely caused by reduced wage discrimination in such a setting, when performance can easily be measured. It could be for this reason that firms with a higher proportion of females are more likely to offer piece rate compensation (Brown 1990).

Another explanation could be a differential performance increase during competition. Gneezy, Niederle, and Rustichini (2003) find that females see lower performance gains from participating in competitive environments. Gneezy and Rustichini (2004) also find that this gender difference exists at a young age; by observing children's performance in running races, they find that competition increases the performance of boys, but not girls. These differences in performance seem tied to the gender of competitors. Gneezy, Niederle, and Rustichini (2003) also find that in mixed-gender competitive environments males have significant performance increases when an environment is made more competitive, while females do not. However, when females compete only against other females, their performance increases as the environment becomes more competitive. Gupta, Poulsen and Villeval (2005) find that females are more competitive when given the opportunity to choose the gender of a potential competitor. Specifically, females are more likely to choose to enter a tournament if they first choose to be paired against another female before making the tournament entry decision. These results suggest that the gender composition of groups may play a role in performance gains from competition, as well as in the selection into competitive environments. Overconfidence could be another explanation, but Niederle and Vesterlund (2007) find that the gender difference remains large even after controlling for the relatively larger overconfidence of men. Dohmen and Falk (2007) report similar results.

Different preferences for risk and ambiguity are another possible explanation. In most competition experiments the subjects have very little information about their relative ability when they make their competitive choices. They typically learn their own performance in trial runs, but not the performance of potential competitors. Grossman and Eckel (2003) provide a review of gender differences in risk preferences and find that results generally, but not always, find that females are more risk averse than men. However, most experimental studies examining gender differences for competition argue that gender differences in competitive choices remain after using various controls for risk preferences. Ambiguity aversion is also a possibility, but ambiguity aversion has not been found to vary systematically across gender. Moore and Eckel (2003) find that females are more ambiguity averse for specific contexts and domains, while Borghans et al. (2009) find that males are initially more ambiguity averse than females, but as ambiguity increases, males and females behave similarly.

None of the results cited so far show that these gender differences in competitive behavior are biologically determined. In fact, Gneezy et al. (2009) report results from experiments in a matrilineal society in India where women are more likely to compete than men in contrast participants in a patriarchical society performing the same type of tasks. Such a result suggests that socialization plays a large role in such gender differences. Cross cultural studies of this sort are one way to isolate the effects of biological factors. In this paper, we use systematic variations in the levels of hormones for females across the menstrual cycle to examine the same question. We find that women's competitive choices vary substantially across the cycle. Interestingly, the effect is such that during the low-hormone phase of the cycle the behavioral differences are quite large, while in the high-hormone phase females choices to compete are similar to those of males.

Hormones have been found to affect various economic behaviors in humans. The hormone oxytocin has been found to increase trusting behavior of individuals (Baumgartner et al. 2008). Fehr (2009) suggests that due to such results, preferences towards trust are affected by biological mechanisms. For males, testosterone levels of financial traders in the morning can predict their daily profits. Cortisol levels in these same traders were found to rise with increased volatility in their market returns (Coates and Herbert 2008). Testosterone levels are correlated with behaviors in economic experiments such as offers and acceptances in ultimatum games (Burnham 2007).

Financial risk taking has also been linked to circulating levels of testosterone in men (Apicella et al. 2008).

Males and females have very different levels of a number of hormones, including estrogen, luteinizing hormone (LH), follicle stimulating hormone (FSH), progesterone, and testosterone (Speroff and Fritz 2005); however, these hormones do not necessarily have the same effects across genders. Additionally, in a review article Shepard et al. (2009) conclude that the large literature on sex differences and brain organization indicates that the expression of hormonal effects within a gender may be very dependent on the particular environment under consideration. Women exhibit large and predictable hormonal variations across the menstrual cycle (Speroff and Fritz 2005). For females, estrogen and progesterone have received most of the attention in studies examining behavioral effects.[1]

The mechanisms by which hormones affect behaviors are explained by neuroendocrinological research on how hormones alter brain activity. Results show that major depression may be linked to reduced density of serotonin binding sites (Malison et al. 1998). By injecting estrogen in rats, Fink et al. (1996) find that estrogen stimulates an increase in the density of serotonin binding sites in certain areas of the brain, including the anterior cingulate cortex, anterior frontal cortex and the nucleus accumbens – areas that have been linked to the anticipation and receipt of monetary rewards (Fink et al. 1996, McEwen 2002, Bethea et al. 2002, Platt and Huettel 2008). Progesterone has been shown to inhibit neurotransmission, and as a result it may decrease anxiety and increase sedation (Vliet 2001). Other research suggests that progesterone may decrease the degradation rate of serotonin (Bethea et al. 2002).

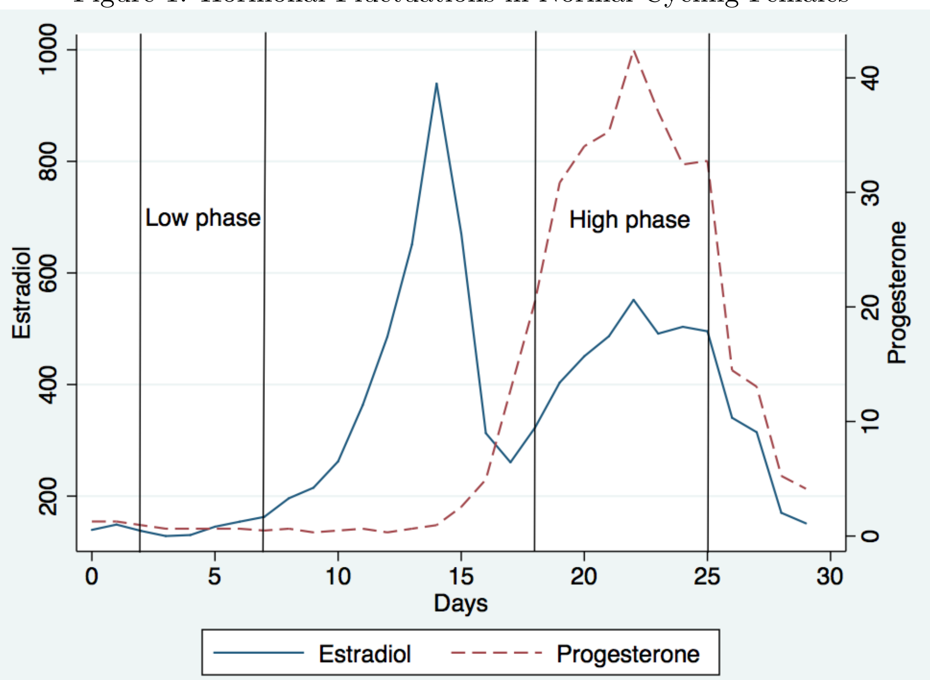
Sex differences in the brain develop during perinatal development where both females and male brains are organized differently from different exposure to steroid hormones (Gagnidze and Pfaff 2009). For female rodent brains, estrogen masculinizes aramatase-expressing neural pathways and also masculinizes territorial behavior (Wu et al. 2009). Aromatase is an enzyme that converts testosterone to estradiol. In the adult male brain testosterone acting through androgen receptors is necessary to complement male type behavior (Gagnidze and Pfaff 2009). For females, estrogen is required and received by estrogen receptors to express male-type aggressive and territorial behavior in mice (Gagnidze and Pfaff 2009). Thus, estrogen for females may lead to similar behavior for females as that induced by testosterone in males. Dreher et al. (2007) uses fMRI to assess brain

[1] Estrogen is the generic term for this hormone, estradiol is the form that is most often measured in humans.

activity during the anticipation of uncertain money payments, across different phases of the menstrual cycle. While the study does not include decisions, they find significant changes in activation in areas related to the processing of rewards (such as the striatum) and in the amygdala, an area that activates during fear and anxiety.

In this study, we exploit the large variations in estrogen and progesterone levels that occur in females over the menstrual cycle. As shown in Figure 1, both progesterone and estrogen remain low during the early part of the menstrual cycle. This first week of the cycle is when normal cycling females menstruate and can be considered a low-hormone phase. The later part of this is called the pre-follicular phase. Estrogen then rises quickly and spikes just prior to ovulation – this is referred to as the follicular spike. After ovulation (approximately day 14 in the graph), during what is called the luteal phase, females who ovulate experience heightened levels of both progesterone and estrogen. This second spike in both hormones may be referred to as the luteal spike or high-hormone phase (Speroff and Fritz 2005, Stricker et al. 2006). Testosterone levels also vary over the menstrual cycle, peaking just before the follicular estrogen spike (Sinha-Hikim et al. 1998). However the spike is much smaller than for estrogen, and testosterone levels are insignificantly different during menses and the luteal spike.

Figure 1: Hormonal Fluctuations in Normal Cycling Females



These are median values from Stricker et al. (2006)

Only a few studies have examined the economic effects of the menstrual cycle. Ichino and Moretti (2009) use detailed employee attendance data from a large Italian bank and find that absences for females below the age of 45 tend to occur according on a 28-day cycle. These 28-day cycle absences explain about one-third of the gender gap in employment absences at the firm. The female menstrual cycle is approximately 28 days and they focus on females below the age of 45, who are more likely to be pre-menopausal. In an experimental study, Chen et al. (2010) explore the possibility that menstrual cycle phase effects drive bidding differences between males and females in auctions. They find bidding differences in first-price auctions, with females in the low hormone follicular phase bidding more than females in the high hormone luteal phase, though most of this variation is found to be driven by contraceptive users. In direct contrast, Pearson and Schipper (2009) find that women bid more than men, and earn lower profits, only during the menstrual and premenstrual phases of the cycle when estrogen and progesterone levels are lower. There is one experimental study looking at competitive choices and the menstrual cycle, Buser (2010). This is a between-subjects study of choices of females to compete in all-female groups, and it finds that females participating during predicted high levels of progesterone tend to be less competitive. We compare these results with ours in the discussion section.

Not all economic studies have found support for hormonal effects on economic decision making. Zethraeus et al. (2009) examine 200 post-menopausal women in a double-blind study. Participants were given either estradiol (2 mg), testosterone (40 mg), or a placebo daily for a four week period. Then they participated in an economic experiment session that included a variety of different tasks looking at risk aversion, altruism, reciprocal fairness, trust and trustworthiness. No significant differences were found when comparing the behaviors between the three different treatment groups of females. Some research shows that neural receptors in post-menopausal women may have reduced sensitivity to hormonal changes, due to the effects of aging (Chakraborty et al. 2003). Such an aging effect could explain the lack of differences in such a study. Ideally, we would use a double-blind study using exogenous delivered hormones to examine the effects of hormonal differences, but such a study is not feasible with pre-menopausal women, since low hormone levels cause bleeding (Speroff and Fritz 2005).

1 Experimental Design

We use a within-subjects design with sessions about 2 weeks apart. An online pre-screening survey was used to recruit and schedule subjects for experiment sessions. We limited the sample of females to those using a monophasic hormonal contraceptive or not using hormonal contraceptives at all.[2] For normally cycling females, the sessions were scheduled during a low-hormone phase (days 2 to 7 in Figure1) and a high-hormone phase (days 18 to 25 in Figure1) of the menstrual cycle. These high and low phases are supported by research examining a drop in hormones during menses (Aden et al. 1998). We intentionally avoided the estradiol spike around day 14, because of its short duration and variability within and across females. Other phases were also avoided due to greater measurement error about the hormone changes that could be occurring during those times. Thus, we limit our study to examining the greatest contrast in hormones for females by using a scheduling design that has been successfully used in the field of neuroscience (Fernandez et al. 2003).

Females using a hormonal contraceptive experience suppression of endogenous hormone production when in the active phase of their contraceptive regimen (Speroff and Fritz 2005). Both progesterone and estrogen levels remain fairly constant as the body receives a daily dose of hormones exogenously (Aden et al. 1998). During the placebo or non-active phase of the contraceptive regimen, there are no exogenous hormones being provided to the body; this withdrawal leads to withdrawal bleeding (Speroff and Fritz 2005). Since menstrual bleeding is caused by low hormones, this allows for easy identification of the low-hormone phase. We scheduled contraceptive users and normal cycling females accordingly, so that both would be in the experiment during a low-hormone phase and during a high-hormone phase. The high-hormone phase coincides with the luteal spike for normal cycling females and a stable elevated hormone phase for contraceptive users. We avoided the follicular spike, because it is short and difficult to time and therefore difficult to correctly schedule subjects into sessions.[3]

Using the screening survey, females were first randomly scheduled during a predicted high or predicted low-hormone phase. Men were simply scheduled for two sessions about 2 weeks apart. To help minimize errors in classifying phases correctly, we also used an exit survey. The low-hormone

[2]Monophasic hormonal contraceptives, release the same level of exogenous hormones each day for the entire non-placebo phase of the hormonal contraceptive regimen. We excluded users of biphasic and triphasic pills, with varying daily hormone doses.

[3]Ovulation occurs twelve to forty-eight hours after the follicular peak is detected. Ovulation prediction kits capture this rise in hormones; thus, using such kits would not be helpful or cost effective.

phase is easily identified by the presence of withdrawal bleeding (in both normal cycling females and those in the placebo phase of contraceptives). The high-hormone phase is more difficult to pinpoint, particularly for subjects not on contraceptives, because of variability in the cycle. Rather than just asking the females that showed up for a session the date of their last or expected next menstruation, we focused on scheduling across two specific hormonal phases to minimize identification error, as the menstrual cycle has a large degree of variability, and females may have trouble accurately recalling and predicting menses (Crenin et al. 2004). The combination of pre-screening, scheduling, and the exit survey were designed to address this.

Previous studies on differences in competitive choices have used between-subjects designs (Niederle and Vesterlund 2007, Gneezy et al. 2009). In our within-subjects design each subject participated in one session of math tasks, and another of word tasks. We used two different tasks in part because we wanted to minimize behavioral spillovers from the first to the second session, and in part because it is generally believed that females may view themselves as having worse math skills than males (Niederle and Vesterlund 2010). For this reason females may be less likely to compete in math tasks than in word tasks. This design is the first to examine whether there are stable differences in competitive choices across genders between-subjects and within-subjects for math and word tasks.

Subjects were randomly assigned to start with a math or a word based session. In each session tasks were performed for five different treatments, one of which was randomly chosen for payment at the end of the experiment. Each treatment lasted 4 minutes. In the first treatment participants performed the task under a non-competitive piece rate compensation scheme, where pay was entirely dependent on the individual's own performance. In the second treatment, participants were randomly assigned to a winner-take-all tournament with a size of two, four, or six competitors. This second treatment provided participants with experience in a situation where their pay depended on their own performance as well as the performance of others. In the third treatment, participants performed the task with a group pay (revenue sharing) form of compensation. This treatment randomly paired participants and payment for the group's total production was split evenly. This third treatment can be considered the least competitive because of the possibility of freeriding. It can be shown that given some random assignment of competitors or group members, this design should lead low ability individuals to choose group pay and high ability individuals to choose a tournament.[4]

[4]A model justifying such predictions is provided in Appendix C.

In the final two treatments subjects were able to choose between piece rate, group pay, or a two, four, or six person tournament. Before the fourth treatment, subjects were told their own absolute performance from treatment 1, but were not told anything about the performance of others. Just before the fifth treatment, participants were shown how all individuals in the session had performed in the first treatment with their performance highlighted, and they then chose their compensation method and performed the task again.

The math task was similar to the one used in Niederle and Vesterlund (2007). Participants were asked to add four randomized two-digit numbers and complete as many of these summations as possible in 4 minutes. Equations were presented to participants on a computer screen and they typed in their answer and pressed the *Enter* key or clicked a *Submit* button on the screen. After each submission participants were promptly shown the next equation to solve, using scratch paper if they wanted. On the screen, the equations looked like the following:

$$12 + 57 + 48 + 52 =$$

The word task was similar to that used by Günther et al. (2008). In this task participants are shown a letter on a computer screen and have four minutes to form as many unique words as possible that begin with that specific letter. The letter remains on the screen for the entire four minutes and participants enter in their word submissions in a text box below the letter. The attempted word formations are then listed below the text box to help subjects minimize duplicate answers, since these are counted as incorrect. Common place names (cities, countries) are acceptable, but proper names are counted as incorrect. Plural and tense changes to root words are counted as separate and correct answers as long as these words still begin with the appropriate letter. In the experiment, participants were informed of the rules before beginning the task. All participants were informed that everybody in the same session and same treatment received the same letter, thus a task of equivalent difficulty for all participants in each treatment.

The word list used for grading words is a common English word list used by open source word processors.[5] We used a restricted group of letters for this study to limit the variation of difficulty between treatments and sessions (e,f,g,h,i,l,n,o). Between 2.7% to 3.8% of all words in the word list began with these letters.

[5] Spell Checking Oriented Word Lists (SCOWL), Revision 6, August 10, 2004 by Kevin Atkinson.

For the piece rate compensation, the payoff an individual receives is equal to the piece rate multiplied by the production of the individual for that particular treatment. Payoffs for both the math and verbal tasks were calculated in a similar manner though the base rate was different for word formation tasks ($0.25) and math addition tasks ($0.50), to adjust for generally higher performance in the word task. In a tournament, if an individual has the best performance in his group then he receives the piece rate multiplied by the size of the tournament, multiplied by his individual performance. If an individual does not have the best performance in his tournament then he receives nothing. In the event of a tie, the individual receives a fraction of the tournament winnings based on the number of individuals he tied with. Subjects were not informed about whether they won or lost a tournament until all five treatments were complete. After each treatment, and before seeing their score, subjects were asked how well they thought they did and how well they thought the average person in the session did, and they were paid for having accurate predictions.

Subjects were told that they could be randomly grouped with people that did not necessarily choose the same compensation option and that they therefore could be playing under different rules than their potential competitors or group members. This strengthens the incentive for high ability types to choose a more competitive tournament, since there is a positive probability that they may compete against lower ability individuals. This rule also creates an incentive for low ability individuals to choose group pay, as they may be matched with high ability individuals which would increase their expected payoffs.[6]

2 Results

Experiment sessions took place in a computer lab at a large public university, all IRB procedures were followed. The majority of the 219 participants were university students whose characteristics are in Table 1. The average size of the 26 sessions was 14.5 participants (with a standard deviation of 4.15). Sixty-two female and 64 male subjects participated in both sessions. Using the pre and post surveys we conservatively classified 45 females as participating in a session during a low-hormone phase of their menstrual cycle, and 34 during both a low and a high-hormone phase. The word task was used in 12 of the sessions and the math task was used in 14 sessions. Of the 345 individual subject sessions, 165 involved the use of the word task and 180 used the math task.[7]

[6] The text for experiment instructions is available in Appendix D.
[7] One female was removed from the data due to non-compliance with the task instructions.

Table 1: Summary statistics of session attendees

Variable	Mean	Std. Dev.	Min.	Max.	N
Participant Characteristics					
Age	20.52	2.81	18	33	218
Years PS	2.18	1.48	0	6	217
GPA	3.29	0.47	2	4.1	218
Live Independently	0.82	0.39	0	1	219
Female	0.5	0.5	0	1	219
Meds	0.09	0.28	0	1	219
Characteristics by Sessions*					
Low Phase	0.14	0.34	0	1	345
Word task	0.48	0.5	0	1	345
Session Size	14.54	4.15	7	21	345
Second session	0.37	0.48	0	1	345

*126 individuals attended a second session.

Table 2 shows that men and women were similar in terms of age, GPA, years of post secondary schooling (Years PS) and even have the same proportion taking psychological medication (Meds). Both genders were assigned to sessions with similar characteristics, except that on average females were in slightly larger sessions. The session female to male ratio ranged from 0.3 to 2.3 and averaged 1.01. Thus, all sessions had some degree of gender mix and on average this mix was about one-to-one.

Table 2: Mean values of individual and session characteristics by gender.

Sex	Age	Years PS	GPA	Indep.	Meds	Word	Size	Sess. 2
Male	20.70	2.14	3.25	0.83	0.09	0.47	13.79	0.37
Female	20.35	2.21	3.33	0.80	0.08	0.48	15.30	0.36
Total	20.52	2.18	3.29	0.82	0.09	0.48	14.54	0.37

Sessions took place three to four times a week and were held in the morning. Each session took slightly less than an hour, including approximately 10 minutes at the beginning of the session during which participants waited together in a foyer. This allowed participants to see that sessions included both males and females. Once participants entered the lab, partitions were used so that participants could not see each other's computer screens or facial responses from the feedback received. Competition and group memberships were also anonymous.

Payouts were based on one randomly chosen treatment, excluding the flat rate show-up payment, payouts averaged $7.38 for the math session and $15.01 for the word sessions. Participants who attended two sessions were later asked to perform a risk aversion task similar to that used in Holt and Laury (2002). The risk aversion tasks were performed a few days after the second session to avoid endogeneity with competition task earnings. A total of 112 participants (56 male and 56 females) participated in the risk aversion task. The average payout for the risk aversion task was $6.57.

2.1 Task Performance

Each individual participated in five different treatments in each session. For the first three treatments the compensation schemes were as follows:

Treatment 1: Piece rate ($0.50 per sum and $0.25 per word).

Treatment 2: Random sized tournament of 2, 4, or 6 individuals (the winner earned the piece rate multiplied by the size of tournament).

Treatment 3: Group pay: an individual was paired with a randomly chosen partner and the total production of the 2 individuals was multiplied by the piece rate and then split evenly.

Table 3: Performance Across Treatments and Gender

Math	T1	T2	T3	T4	T5	Word	T1	T2	T3	T4	T5
Male	10.9	12.1	12.3	12.7	12.8	Male	38.2	39.4	43.0	45.3	47.0
Female	9.9	11.4	11.8	12.3	12.1	Female	41.0	41.1	45.0	48.4	47.3
Both	10.4	11.8	12.0	12.5	12.5	Both	39.6	40.3	44.0	46.9	47.1

Table 3 shows mean performance by gender over treatments and tasks. The increasing mean values over the first three treatments in both the math and the word tasks suggest that subjects are learning to do the task better during the session. There are no statistically significant performance differences between males and females in either task.[8] This lack of a performance difference by gender, for either task, removes one obvious potential reason for gender differences in choices.

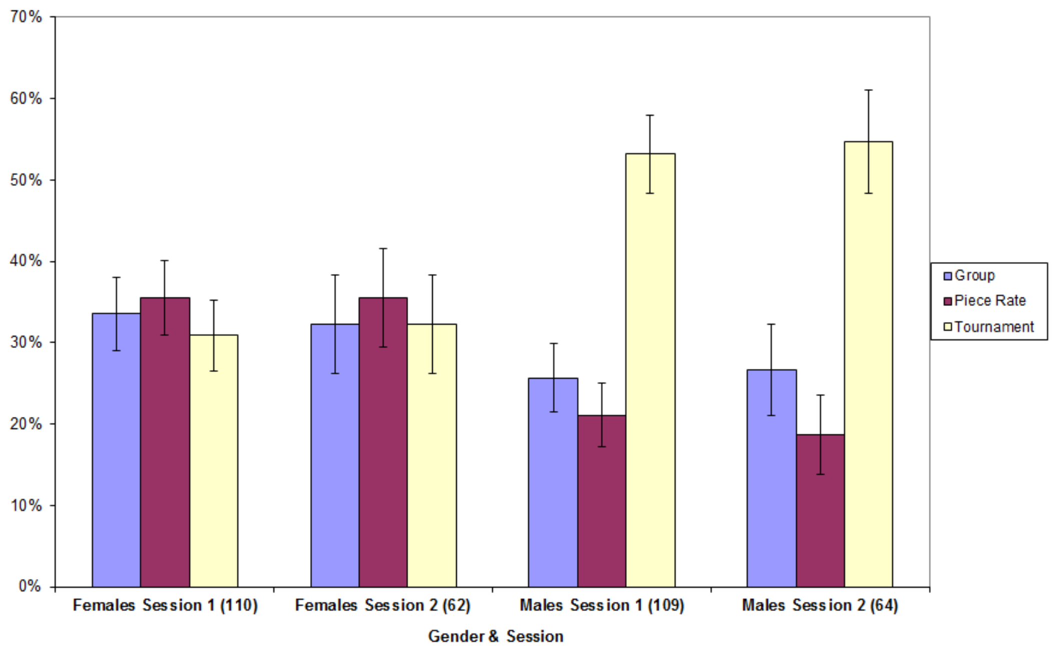
2.2 Gender Differences in Competitive Choices

Niederle and Vesterlund (2007) and Gupta et al. (2005) find that when given the choice between a tournament and piece rate females are less likely than males to enter tournaments. To test whether this basic result can be replicated with our protocol, we focus on choices made in Treatment 4. In those studies, individuals did not have information about their relative performance, and in

[8] An analysis of the performance effects that occur from learning and different competitive settings are available on request.

our study this feedback comes only after Treatment 4. The available choices of group pay, piece rate, and tournaments of increasing size can be ordered by increasing competitiveness, with sharing being less competitive than not sharing and larger tournaments being more competitive. In the figures and empirical analysis we lump the two, four and six person sized tournaments together though the results are robust when using an ordered scale for tournament size.

Figure 2: Choice Differences Before Feedback, by Gender and Session



Sample size in parentheses.

Figure 2 shows the distribution of choices made by males and females in the first and second sessions for Treatment 4. The gender differences are large: pooling over sessions we find that only 31% of females chose to compete in tournaments while 54% of males chose the tournaments. The difference persists for the piece rate: 36% of females chose the piece rate compared to only 20% of males. These differences are all significant at the 2% level or better with chi-square tests. This replicates earlier findings, and shows that gender differences for competitive choices are robust to the addition of a group pay option and different sized tournaments. We also find that, on average, males and females chose consistently across the two repeated sessions, despite the fact that these which are often separated by weeks.

While there are no significant differences in performance between males and females, other factors such as age and GPA might conceivably affect compensation choices. In this design, we predict that with full information about abilities individuals would sort according to ability, with the least able individuals choosing the least competitive environments and the higher ability individuals choosing more competitive tournaments. Given this we use an ordered probit to test whether the gender differences in the probability of selections remain after controlling for other potentially relevant factors.

Table 4 shows that the gender differences persist with these controls, along with the addition of control variables for confidence, performance, and improvement in the repetition of tasks in a tournament. Columns 1 to 3 use CompScale as the dependent ordinal variable, where group pay compensation is less competitive than piece rate which is less competitive than a tournament of any size.[9]. In the results, we include both pooled results and random effects estimations using an ordered probit model. For nonlinear estimations such as ordered probits, random effects models are often used to deal with the difficulties and bias involved with using fixed effects models (Arellano and Honoré 2001). Given that the experiment data is considered a short panel, any fixed effects estimation of a nonlinear model would also suffer from the well-known incidental parameter problem that may bias fixed effects results (Greene 2004). For these reasons we chose to use a random effects ordered probit for estimation purposes.

Table 4 replicates the results of Niederle and Vesterlund (2007) with Treatment 4, before relative performance feedback. Females are less likely than males to enter tournaments, even when controlling for individual confidence (Confidence (T1)) and relative rank of performance within the session (%-tile Rank (T1)) from the first treatment. The %-tile Rank (T1) variable gives the rank of an individual based on her or his performance in Treatment 1 in the session. Using rank allows us to have the same measure for both math and word tasks.[10] Confidence is measured by an individual's predicted performance at the end of Treatment 1 (prior to finding out their actual performance) divided by that individual's prediction of the average performance of all session participants.[11] To control for performance, we use the relative rank from Treatment 1, but the results are unchanged when using absolute performance along with an interaction term for word based tasks.

[9] Our results are consistent with a multinomial logit model and from using ordered probits with rankings that treat larger tournaments as more competitive.

[10] Using a variable that measures actual performance with an interaction term for the type of task, gives the same results as are presented here.

[11] Females tend to be less confident than males in the math task, but females and males have no significant differences in confidence in the word task and selection differences still remain.

Table 4: Ordered Probit Estimates: Choices for No Relative Information Treatment

VARIABLES	(1) Pooled CompScale	(2) RE CompScale	(3) RE Risk CompScale
Female	-0.36	-0.40	-0.49
	(0.13)	(0.15)	(0.19)
	(***)	(***)	(**)
Confidence (T1)	0.86	0.98	0.99
	(0.25)	(0.29)	(0.34)
	(***)	(***)	(***)
Improve (T2)	0.61	0.72	0.73
	(0.20)	(0.23)	(0.32)
	(***)	(***)	(**)
%-tile Rank (T1)	1.05	1.08	0.85
	(0.23)	(0.26)	(0.32)
	(***)	(***)	(***)
Characteristic Controls	Yes	Yes	Yes
Risk Controls	No	No	Yes
Observations	343	343	224
ll	-336.6	-335.6	-212.3
chi2	66.91	61.00	48.81

Pooled means pooled cross section. RE means that random effects were used.
Standard errors in parentheses. p<0.01, ** p<0.05, * p<0.10

As expected, both confidence and the actual percentile rank from the first treatment are positively correlated with the selection of more competitive environments. Improvement in performance between the first and second task (Improve (T2)) also has a significant positive effect. These regressions include controls for individual specific characteristics, including the number of years of college, psychoactive medication, GPA, and age.[12] The results are similar when using a random effects ordered probit, in column 2. Column 3 includes a measure of risk aversion for individuals that participated in a task similar to the one used by Holt and Laury (2003). We find that this measure of risk aversion is not significantly correlated with competitive choices in Treatment 4.

The marginal effects (calculated from column 1) show that a female has a 0.14 lower probability of choosing a tournament than a male, even when controlling for performance and confidence. For a female to be as likely to choose a tournament as an average male, we would have to increase her belief about her performance relative to the average by 40%, which is a significant increase in overconfidence. A ten-percentile improvement in actual relative performance would increase the

[12] Details on these controls are in the appendix.

probability of entering a tournament by 0.04. A female would have to improve her percentile rank by 34% to be as likely to enter a tournament as a male. Thus, these gender differences are not just significant, but they are also large.

After each treatment, before receiving any feedback, subjects were asked how many correct answers they believed they submitted. Subjects were paid ($0.25) for each correct answer to encourage accurate answers. We create a measure of confidence by dividing an individual's prediction of how well he or she did divided by his or her prediction of the session average for that treatment. Since the average individual should believe they did not perform any better than the session average, this confidence measure should have a mean of one – in the absence of overconfidence.[13] We could have asked for rank estimates instead of performance estimates, but rank is a poor measure of the degree of over or under confidence. Consider two individuals that think they are ranked first in their respective group. One may think that he is 10% better than the average while the other may think she is 50% better. Both these individuals would be treated as having the same level of confidence with the rank measure, but one individual is actually much more confident. We use the measure of confidence from the first treatment because every subject performed the task for this treatment under the same piece rate form of compensation. This confidence variable provides the earliest measure of overconfidence before experiencing any feedback or differing experimental manipulations.

Changes in performance as the experiment proceeds could also change confidence. The variable Improve (T2), measured as the ratio of the individual's performance from Treatment 2 divided by the performance in Treatment, captures the effect of individual improvement between Treatment 1 (piece rate) and Treatment 2 (tournament). There are two possible reasons that this variable should matter: First, individuals may feel that they improve more than the average individual or that they were unlucky in Treatment 1 compared to how others would have performed. Second, it may be the case that individuals become more motivated to put in greater effort because of the competitive nature of the tournament in Treatment 2. Individuals that improve a lot from competing in such settings would be more likely to choose to compete than individuals whose performances are not positively affected by competitive settings.

[13] We also asked how many correct answers they believed were submitted by the most productive person, the least productive person, as well as the average number of correct answers, for each session and treatment. We use the average instead of the prediction of the best or worst individual in the session because it provides a clean measure of overconfidence. In other estimations, not included here, these measures were separated and variations of using both the best performance and the worst performance as the denominator were used with little difference in our results.

Table 5: Gender Effects (Clustered Errors)

	(1)	(2)
VARIABLES	CompScale	CompScale
Female	-0.35	-0.35
	(0.135)	(0.114)
	***	***
Confidence (T1)	0.87	0.87
	(0.256)	(0.246)
	***	***
Improve (T2)	0.62	0.62
	(0.181)	(0.187)
	***	***
%-tile Rank (T1)	1.04	1.04
	(0.227)	(0.233)
	***	***
Clusters	Individual	Session
Risk Controls	No	No
Characteristic Controls	Yes	Yes
Observations	343	343
chi2	71.85	116.8
ll	-336.5	-336.5

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Niederle and Vesterlund (2007) found that part of the difference between male and female willingness to compete was driven by males being more overconfident than females. In their study, independent of confidence, females had a 0.16 lower probability of entering a tournament than males. Using our measure of confidence we find that the gender difference is nearly the same, 0.14.

Since we have multiple observations from the same individuals and individuals participate in the same sessions, we also run regressions where we cluster standard errors on experiment sessions and then also separately on individuals. Table 5 shows that the results concerning females being less likely to enter in tournaments without relative performance feedback remain consistent when using errors that are clustered on the specific experiment session or on the individual. In this table the dependent variable is the same ordered variable of competitiveness used previously where group pay is less competitive than piece rate and piece rate is less competitive than a tournament of any size.

Our within-subjects design includes one session of math treatments and one of word treatments. Günther et al. (2010) found that in a maze task, men increased performance in reaction to competitive pressures by more than women did. In a word task the improvements were the same. They attribute this to a "stereotype threat" arising from beliefs that women are not good at the maze task. This could logically lead to different choices by women to compete, with different tasks. Figure 3 shows that in our data there is little difference in the selection of competitive environments by females regardless of the type of task used. We also find little difference in choices by males as more than 50% of males chose to compete in tournaments in both math and word tasks.

Figure 3: Selection Differences for Females by Task Type for No Information Treatment

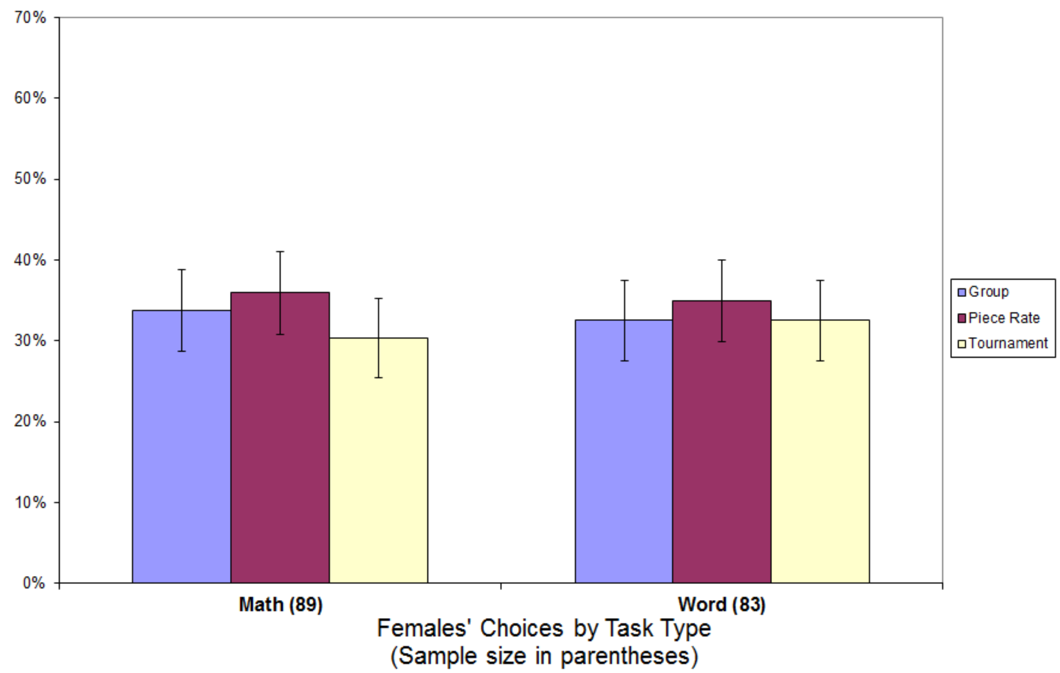


Table 6 looks at confidence differences by gender and task. Both genders are overconfident on average. Males are significantly more overconfident in their math abilities than females, and there is no significant difference in confidence between males and females in the word task. There is no significant difference among females between the math and word tasks, while males are significantly more confident in their math performance than in their word task performance. On average, males are slightly more confident in their abilities than females. This is partly driven by a few high ability males who are correct in believing they are better than the average, but overestimate the degree. For example, the highest level of confidence for a male is 3.38 times his prediction of the average. His actual performance is 2.29 times the actual session average. Overall, males and females are fairly consistent in their choices to compete in both types of task: males choose to compete more than females in both math and word tasks even though male overconfidence is higher in the math task. The type of task was not significant in regressions for choices, with or without confidence controls.

Table 6: Confidence Differences by Gender and Task Type

Task Type	Gender	Confidence (T1)	S.E.
Math	Female	1.081	0.030
Math	Male	1.209	0.037
Math	Both	1.145	0.024
Word	Female	1.046	0.029
Word	Male	1.039	0.030
Word	Both	1.043	0.021
Both	Female	1.064	0.021
Both	Male	1.128	0.025
Both	Both	1.096	0.016

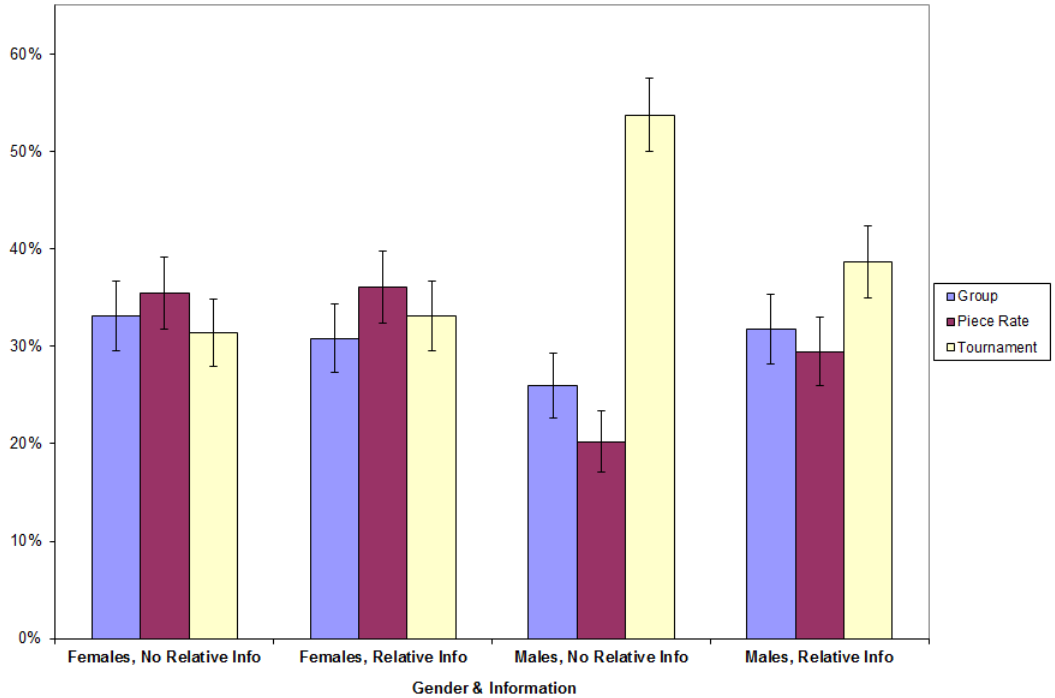
When comparing to Both genders, removing an outlier makes the gender difference insignificant at a 5% level.

2.3 Performance Feedback Eliminates Gender Differences to Compete

Providing information about the quality of possible competitors might reduce mistakes in competitive choices, but there is no obvious reason feedback should reduce the gender difference in choices, if that difference is primarily driven by preferences. We test the effect of performance feedback on choices by providing subjects with an ordered list of the performance of all the participants in their session from Treatment 1, with their own performance highlighted, before they choose their Treatment 5 compensation scheme. This provides information about the quality of their potential competitors, if they choose to enter a tournament.

The two groups of bars on the left side of Figure 4 suggest that females' choices are barely affected by information about the performance of potential competitors. The right side of the figure shows that males' choices change dramatically. There is a significant increase in the proportion of males choosing piece rate (5% significance level) and group pay (10% significance level), and a significant decrease in the proportion choosing tournaments (5% significance level). Comparing the distributions of men's and women's choices in Treatment 4 gives a Pearson chi-square statistic of 18.79 (p-value: 0.000). After relative performance feedback in Treatment 5 male and female's competitive choices are not significantly different (chi-square statistic is 1.91, p-value: 0.385).

Figure 4: Selection Differences by Gender and Information Treatment (Treatments 4 and 5)



Females (172), Males (173). Sample size in parentheses.

Table 7 shows the results from three different types of ordered probits for Treatment 5 choices, using the CompScale competitiveness definition from the Treatment 4 analysis. Columns 1 through 3 show, that once performance feedback is provided, there are no significant differences between male and female choices. Instead, we find that choices are very dependent on the relative performance information, and on the individual's improvement from Treatment 1 to Treatment 2. Risk aversion control variables are not significantly correlated with compensation choices on average; though risk aversion measures were significant when only examining high ability individuals' choices in Treatment 5. The one variable that consistently affects individual choices in Treatment 5 is an individual's percentile rank from Treatment 1, a summary statistic of the feedback information provided before the Treatment 5 choice.

The overall conclusion from Figure 4 and the probits in Table 7 is that there are no significant gender differences in competitive choices when subjects are fully informed of their relative performance compared to potential competitors. In the next section we consider the costs of the selection differences between men and women when they lack information about the quality of competitors and whether there are gender differences according to ability levels.

Table 7: Ordered Probit Estimates: Choices for Relative Information Treatment

VARIABLES	(1) Pooled CompScale	(2) RE CompScale	(3) RE Risk CompScale
Female	0.00	-0.02	0.13
	(0.13)	(0.18)	(0.21)
	()	()	()
Confidence (T1)	0.34	0.44	0.65
	(0.24)	(0.30)	(0.35)
	()	()	(*)
Improve (T2)	0.81	1.01	0.65
	(0.20)	(0.26)	(0.32)
	(***)	(***)	(**)
%-tile Rank (T1)	2.17	2.59	2.31
	(0.25)	(0.34)	(0.37)
	(***)	(***)	(***)
Risk Controls	No	No	Yes
Characteristic Controls	Yes	Yes	Yes
Observations	343	343	224
ll	-320.6	-316.7	-194.5
chi2	110.9	98.51	79.67

Pooled means pooled cross section. RE means that random effects were used
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.4 The Cost of Choices and the Effect of Feedback, by Gender and Ability.

To give some sense of the costs of gender differences in choices, we simplify and assume people maximize expected payoffs, keep effort constant across compensation choices, and take the choices and performance of others as given. Table 8 shows the average expected value losses for the suboptimal selections by males and females in Treatment 4 and Treatment 5.[14] Each column represents the optimal choice that should have been made. The numbers represent the average expected value cost for choosing something other than that optimal choice. For example, in the first row under column 6 (for the 6 person sized tournament), the 27.27 represents the average loss to females whose optimal choice was a tournament of six, but who instead chose a different form of compensation. The Avg Loss column provides the average loss by gender and treatment. The average loss of 6.78 in the first row means that females lost an average of $6.78 from their suboptimal choices in Treatment 4.

[14]The method used to calculate expected values is based on using the percentile rank as the probability of success within a tournament.

Table 8: Selection Losses

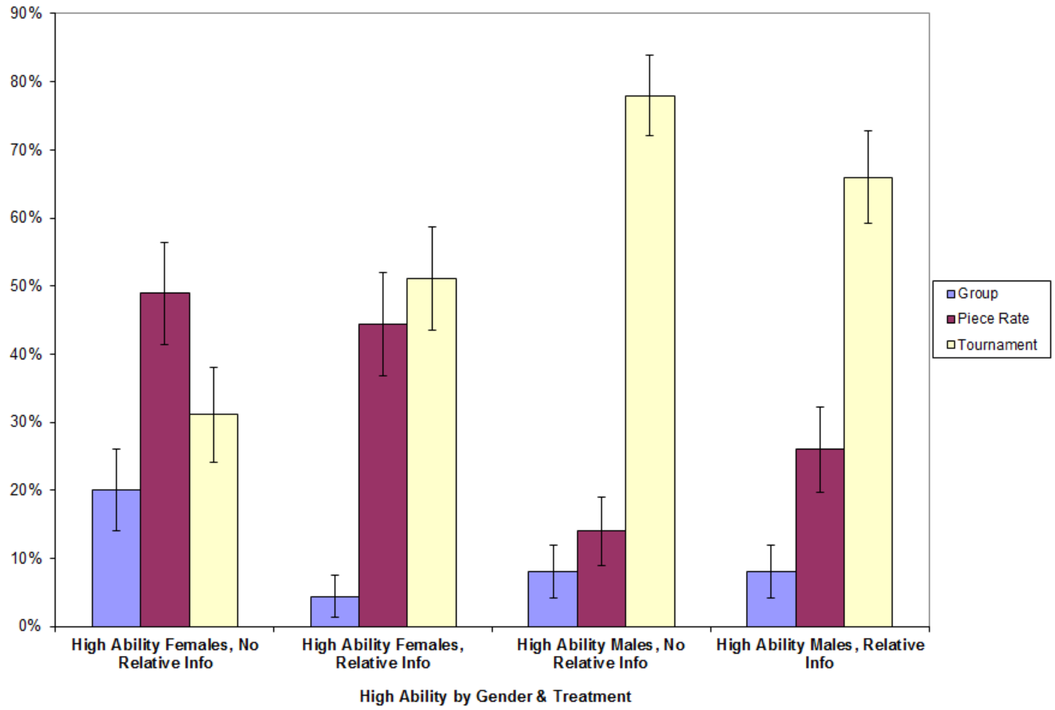
Average Loss from Suboptimal Decisions							
		Optimal Choice					
Treatment	**Gender**	Grp	PR	2	4	6	Avg Loss
4	Female Avg Loss	1.58	2.28	2.91	6.80	27.27	6.78
4	Male Avg Loss	2.42	2.97	2.31	3.29	12.60	4.91
5	Female Avg Loss	0.88	1.88	2.21	5.93	18.70	4.80
5	Male Avg Loss	1.39	1.49	2.02	4.79	10.98	3.95

In Treatment 4, the average expected value loss from selection mistakes was \$4.91 for males and \$6.78 for females, a statistically insignificant difference with a t-test. These loss differences are mostly driven by high ability females choosing not to compete, and to a lesser extent by low ability males choosing to compete. Column 6 shows that many high ability females (those who should select a tournament size of 6) are instead selecting smaller tournaments or group pay or piece rate, at a large cost. The top females lose \$27.27 in expected value compared to \$12.60 for the top males. In contrast, low ability males make only slightly more costly decisions than low ability females, averaging \$2.42 versus \$1.58 for the lowest types of each gender. We find that high ability females and high ability males are not entering competitive environments enough. But the high ability females overwhelming select the noncompetitive environments of piece rate and group pay, which are very costly decisions. In contrast, too many low ability males are entering competitive environments, but these mistakes are not particularly costly, on average, because low ability males would not perform well in the piece rate either.

Table 8 also shows that relative performance feedback decreases the average expected value losses for both males and females and shrinks the gender gap as well. The decreases in expected value losses are greatest for high ability females, whose average expected loss fell from \$27.27 in Treatment 4 to \$18.70 in Treatment 5, while losses for high ability males fell from \$12.60 to \$10.98. Low ability females and males tend to move towards group pay as they get performance feedback. While a gender difference remains, with low ability males making more expensive mistakes than women, the cost differences are small.

In Figure 5, we turn to the question of how relative feedback information affects the choices of high ability females and males. A high ability individual is defined as an individual who should enter a four person tournament or larger to maximize expected returns from competition. Figure 5 shows that the relative performance information leads to a large increase in the proportion of

Figure 5: Information Effects for Decisions by High Ability Types



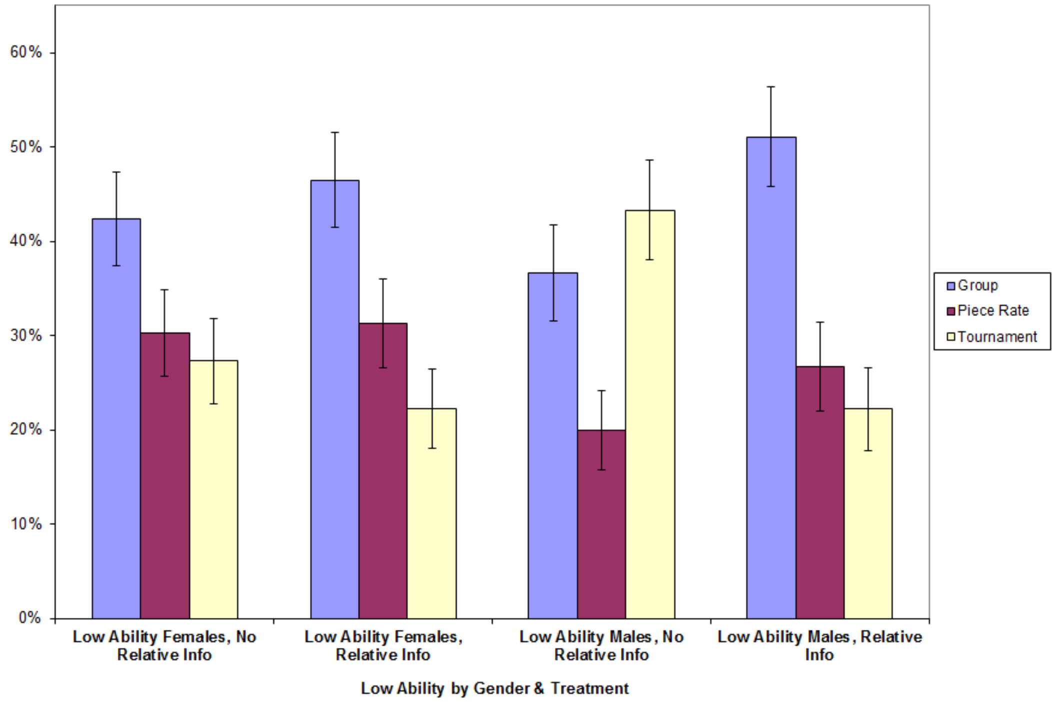
Females (45), Males (50). Sample size in parentheses.

high ability females entering tournaments. Over 50% of high ability females enter tournaments when given relative performance feedback, which is significantly more than the 31% that choose tournaments before receiving the performance feedback. In testing for distributional changes, we find that there is a significant difference in choices for females between Treatment 4 and Treatment 5; using a Pearson chi-square test the level of significance is $p = 0.034$.

With information, fewer high ability males enter tournaments (12% fewer), but this change in tournament selection is not statistically significant at the 5% level. The distributional difference of choices for high ability males coming from information feedback is not significant as a chi-square test comparing high ability males between treatments produces a level of significance of $p = 0.317$. Without feedback in Treatment 4, there is a significant difference in the distributions of competitive choices between males and females ($p = 0.000$). After receiving feedback as the level of significance using a χ^2 test is $p = 0.158$. Thus, relative performance feedback seems to eliminate most of the differences in choices between the high ability females and high ability males.

Figure 6 shows the effect of relative performance information on choices by low ability types, where low ability is defined as those individuals with performance below the median in their respec-

Figure 6: Information Effects for Decisions by Low Ability Types



Females (99), Males (90). Sample size in parentheses.

tive session from Treatment 1. The largest effects are for males. Information drops the percentage of low ability males choosing tournaments from 43% to 22%, and increases the percentage of low ability males choosing group pay from 37% to 51%. For low ability males, the difference in the distribution of competitive choices between Treatment 4 and Treatment 5 is significant at a $p = 0.010$ with a chi-square test. No such significant difference occurs for low ability females. The distributions of choices are significantly different for low ability females and males in Treatment 4 as chi-square test lead to a $p = 0.054$. But in Treatment 5 there are no significant differences between distributions for low ability females and males.

Information about relative performance moves high ability females towards more competitive choices and low ability males away from tournaments towards less competitive types of pay. Low ability females show only a small movement away from group pay towards piece rate. Overall, providing relative performance feedback information leads to more efficient sorting by both genders.

2.5 Competitiveness Differs Between High and Low Hormone Phases of Menstrual Cycle

Normal cycling women experience large changes in hormone levels across the menstrual cycle (Figure 1) and these variations are similar in women using hormonal contraceptives. As explained in the design section, we used a screening survey to schedule females for one session during a low-hormone phase and one during a high-hormone phase and exit surveys to confirm phases.

Table 9: Menstrual Cycle Regularity

Regularity of Period	Percent	Count
Identical	14.3%	55
Within 1-2 days	42.3%	163
Within 3-7 days	34.3%	132
Very Irregular (7+)	9.1%	35
Total		385
Missed Period in Last 3 Months	**Percent**	**Count**
Yes	14.7%	57
No	85.3%	330
Total		387

Numbers may not add up due to item non-response in screening survey.

Table 9 summarizes the screening survey responses of females. Of the females who completed the screening survey almost 15% missed a menstrual period during the previous 3 months. Over 43% of these females experienced menstrual cycle irregularity of 3 days or more, suggesting that predicted menstrual periods may have significant measurement error. Due to the potential inaccuracies introduced by this prospective survey, we also used an exit survey with both retrospective and prospective questions on menstruation to classify hormonal phases for our analysis.[15]

The screening survey also provided information on the proportion of females that use hormonal contraceptives. Over 54% of females in our sample used some form of hormonal contraceptive in the form of the pill or ring. This makes for easier predictability of low and high phases for these females, since hormonal fluctuations are exogenously determined by hormonal contraceptive use. To help identify hormonal phases for females using a hormonal contraceptive, we asked all female participants for the start day of their hormonal contraceptive regimen.

[15] Missed periods are a problem for identification purposes in normal cycling females as they imply that a female may not have ovulated during that month, and thus did not experience a mid-luteal peak in hormones. Furthermore, without a recent menstrual period it is difficult to determine the phase in the hormonal cycle.

Of the females that participated in experiment sessions, 62.7% of those attending a first session were following a hormonal contraceptive regimen, as were 62.9% of those at second sessions. The American College Health Association found that about 72% of sexually active females were using some form of hormonal contraceptive in 2008 . In examining contraceptive use by females in the United States, it was found that for women between the ages of 15 to 44, over 82% had at one time taken oral hormonal contraceptives (Mosher et al. 2004), suggesting that our sample is not unusual in terms of contraceptive use.

We hypothesize that the low-hormone phase, whether induced through endogenous or exogenous means, is associated with similar behavioral changes for both hormonal contraceptive users and normal cycling females. We tested this by controlling for hormonal contraceptive use and found no systematic significant difference in behavior between hormonal contraceptive users and normal cycling females. We therefore pool both groups of females and focus on similar differences across the two hormonal phases.

Figure 7: Competitive Choice by Gender and Hormonal Phase

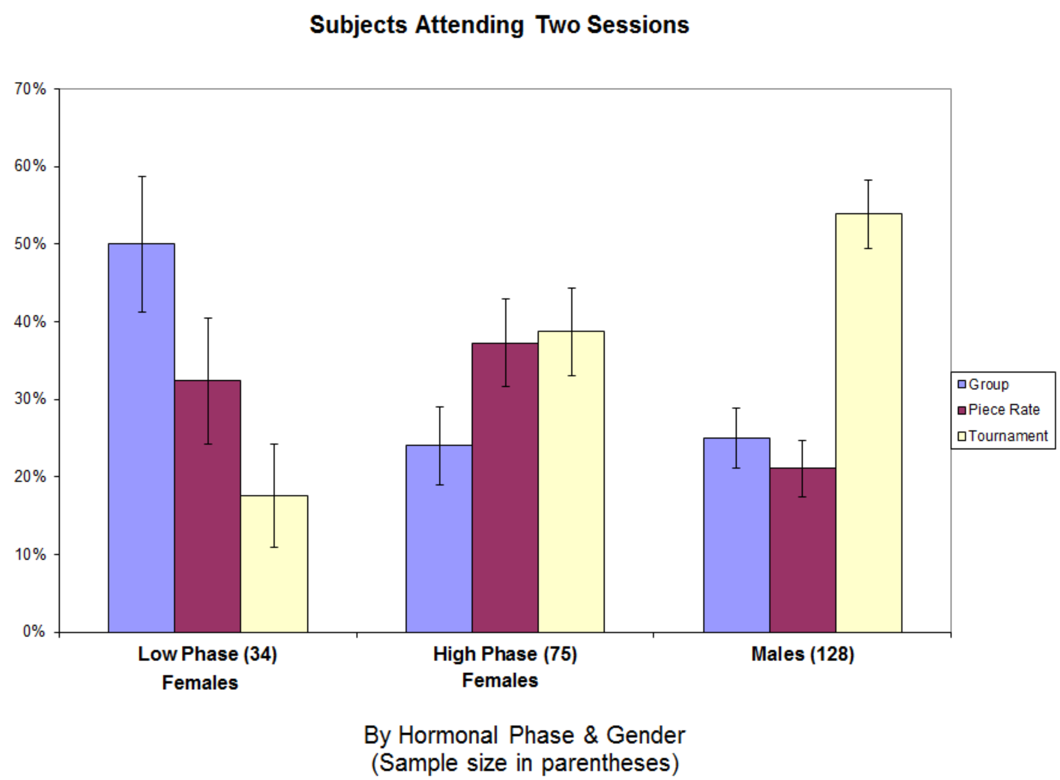


Figure 7 shows the distribution of competitive choices of females by phase, along with choices by males, before participants had relative performance feedback. Female behavior is very different in the two phases. They are more than twice as likely to choose group pay when they are in the low phase, and twice as likely to choose tournament when they are in the high phase, though still not as likely as men. When we include controls in regressions, this last difference will become insignificant. The data for the histogram includes all females and males that attended two sessions and all females who could be identified as being in the low or the high phase. Due to the difficulty of predicting the low phase, some females were identified by the exit survey as being in the same phase for both word and math tasks. As well, some phases could not be accurately identified and those subjects are not included in the analyses.

These differences in competitive environment choices across hormonal phases may result from differences in expected performance changes across the menstrual cycle, or from different preferences for competition. We find that for the most part, there are no significant performance differences between females in the low phase and those that are not in the low phase.[16] It is also possible that females in a specific hormonal phase might experience greater aversion to certain types of tasks; therefore, we separate out these results by math and word tasks. Figure 8 shows female compensation choices before feedback by hormonal phase and task type. Females that participated in a math or word task during the low phase were then scheduled for the other type of task when in a high phase, and vice-versa. The figure shows that the general correlation between competitive choice and menstrual phase holds across tasks: high phase females are less likely to choose group pay and more likely to choose tournaments in both word and math tasks.

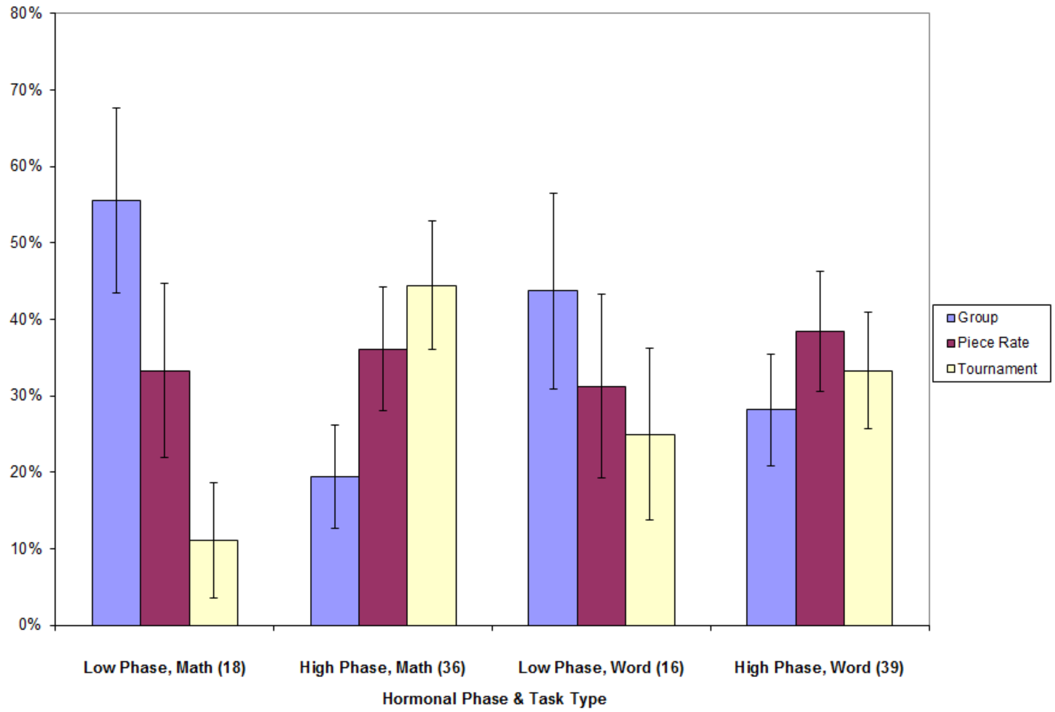
We use ordered probits to examine the statistical significance of gender and menstrual phase before feedback, while including control variables. Table 10 uses the CompScale variable, an ordered categorical variable with choices ranked from group pay, piece rate, to tournament. The first column provides pooled cross-sectional results including all subjects, the second to fourth columns provide estimates using random effects ordered probit. The second column includes all males and females, the third column consists of a female only sample and the fourth column takes into account only males and females for which risk aversion measures were available.[17]

We find that females in the low phase select noticeably less competitive compensation plans than females in the high-hormone phase. In fact much of the average difference in competitive choices between males and females is driven by the choices of the low phase females. This result holds even when controlling for confidence. It is worth noting that there are no significant differences in

[16] See Appendix B.1

[17] All regressions include controls for session ordering, GPA, age, education, and psychoactive medications.

Figure 8: Compensation Choice by Hormonal Phase and Task Type for Females.



We are comparing females that attended two sessions.
Sample size in parentheses.

confidence levels between low hormonal phase and high hormonal phase females, and yet females in the low phase avoid the competitive environments of tournaments and are more likely to choose the least competitive setting possible, group pay.

These differences could potentially result from discomfort during the low-hormone phase of menstruation. But females in the low-hormone phase do not behave differently from any other group once they receive relative performance feedback. Thus, physical discomfort is an unlikely explanation for these systematic differences in low information settings.

The magnitudes of the marginal effects (calculated using the pooled cross sectional estimates) of being in the low-hormone phase are substantial and are larger than the average gender effects. For group pay, females on average have a 0.08 higher probability of choosing group pay than males. Females in the low phase have an additional 0.16 higher probability of choosing group pay. For tournaments, females have a 0.10 lower probability of choosing a tournament when compared to males, and females in the low-hormone phase have an additional 0.16 decrease in the probability of choosing a tournament.

Table 10: Ordered Probit: Hormone Effects for No Relative Information (Treatment 4)

	(1)	(2)	(3)	(4)
Sample	All	All	Females Only	Risk
VARIABLES	Pooled	RE	RE	RE
Female	-0.26	-0.29		-0.26
	(0.14)	(0.16)		(0.21)
	(*)	(*)		()
Low Phase	-0.44	-0.46	-0.53	-0.76
	(0.21)	(0.22)	(0.26)	(0.27)
	(**)	(**)	(**)	(***)
Confidence (T1)	0.81	0.91	1.08	0.90
	(0.26)	(0.30)	(0.54)	(0.35)
	(***)	(***)	(**)	(**)
Improve (T2)	0.60	0.69	0.79	0.72
	(0.20)	(0.23)	(0.38)	(0.32)
	(***)	(***)	(**)	(**)
%-tile Rank (T1)	0.97	0.99	0.52	0.72
	(0.23)	(0.26)	(0.43)	(0.32)
	(***)	(***)	()	(**)
Risk Controls	No	No	No	Yes
Characteristic Controls	Yes	Yes	Yes	Yes
Observations	328	328	155	211
ll	-322.3	-321.7	-156.0	-197.4
chi2	64.32	58.60	19.76	51.31

Dependent variable is CompScale where -1 is group pay, 0 is piece rate, and 1 is a tournament.
The total low phase females that could be identified for data analysis is 45.
Pooled means pooled cross section. RE means that random effects were used.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The changes over the menstrual cycle are also large relative to the effects of confidence and performance. For a female in the low phase to have the same probability of entering a tournament as a female in the high phase we would have to increase her belief about her performance relative to the average by 50%. In terms of an equivalent performance effect, a female in the low-hormone phase would have to improve her percentile rank by 42% to be as likely to enter a tournament as a female in the high-hormone phase.

Table 11 shows that the results concerning females in the low hormonal phase persist when using standard errors that are clustered on the specific experiment session or on the individual for the no information treatment. In this table the dependent variable is the same ordered variable of competitiveness that was used previously. The results are entirely consistent with our previous findings. Although gender may have a prominent role in explaining females' reluctance to compete

Table 11: Hormone Effects (Clustered Errors)

	(1)	(2)
SE Clusters	Individual	Session
Female	-0.25	-0.25
	(0.149)	(0.137)
	*	*
Low Phase	-0.45	-0.45
	(0.200)	(0.156)
	**	***
Confidence (T1)	0.84	0.84
	(0.275)	(0.234)
	***	***
Improve (T2)	0.62	0.62
	(0.182)	(0.187)
	***	***
%-tile Rank (T1)	0.96	0.96
	(0.225)	(0.227)
	***	***
Clusters	Individual	Session
Risk Controls	No	No
Characteristic Controls	Yes	Yes
Observations	328	328
chi2	69.61	124.6
ll	-322.2	-322.2
N_clust	215	26

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

in mixed gender settings, it seems that hormonal phase may be a driving factor that needs to be considered in low information settings.

Interestingly, relative performance feedback makes these cycle specific effects disappear. Table 12 provides the results from ordered probit estimations for Treatment 5, where subjects were provided with relative performance information from Treatment 1 prior to making their competitive environment selections. Table 12 shows that when participants are informed of their relative performance compared to other potential competitors, then there is little difference in selection between genders or across the menstrual cycle.

As with the gender differences, we find that after participants are informed of the quality of potential competitors, choice differences across the menstrual cycle become insignificant. We find that choices after feedback mainly depend on the relative performance information provided prior to making the decision and, to a certain extent, on an individual's improvement from Treatment 1 to Treatment 2. Though females' choices to avoid competitive environments are most frequent in the low-hormone phase, these results suggest that this effect seems to be linked with the information available about the quality of potential competitors.

Table 12: Ordered Probit: Hormone Effects after Feedback (Treatment 5)

	(1)	(2)	(3)	(4)
Sample	All	All	Females Only	Risk
VARIABLES	Pooled	RE	RE	RE
Female	-0.07	-0.13		0.10
	(0.15)	(0.20)		(0.24)
	()	()		()
Low Phase	0.03	0.13	0.12	-0.12
	(0.21)	(0.25)	(0.27)	(0.29)
	()	()	()	()
Confidence (T1)	0.23	0.32	0.21	0.54
	(0.25)	(0.31)	(0.53)	(0.36)
	()	()	()	()
Improve (T2)	0.76	0.92	1.06	0.49
	(0.20)	(0.26)	(0.40)	(0.33)
	(***)	(***)	(***)	()
%-tile Rank (T1)	2.18	2.61	2.63	2.33
	(0.25)	(0.35)	(0.55)	(0.38)
	(***)	(***)	(***)	(***)
Risk Controls	No	No	No	Yes
Characteristic Controls	Yes	Yes	Yes	Yes
Observations	328	328	155	211
ll	-307.8	-303.9	-143.3	-183.4
chi2	104.7	93.82	45.89	75.79

We identified 45 females as low phase for this analysis.
Pooled is pooled cross section, RE is random effects.
Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

As discussed previously, there is a cost associated with high ability individuals avoiding competitive settings, and with low ability individuals choosing tournaments. We find that females in the low-hormone phase make more costly mistakes than high phase females, and males. The average expected value losses for males, low phase and high phase females in Treatment 4 are shown in Table 13. Low phase females sacrifice the greatest amount of expected value from making suboptimal choices, \$8.50. The expected value losses for high phase females and males are \$6.52 and \$4.91. The differences between low and high phase females and between males and high phase females are not statistically significant, but low phase females make more costly choices than males at the 5% significance level.

Table 13: Expected Value Loss in Treatment 4

	Mean	Std. Error
Male	4.91	0.72
Female Non-Low	6.52	1.30
Female Low	8.50	2.57

These results show that menstrual phase and the corresponding hormone levels are correlated with competitive choices, but only if the strength of the competition is not known. If there is little information about the relative abilities, then females in the low-hormone phase make more costly decisions than males and non-low phase females. But there are no significant differences in expected value losses between genders or between different hormonal phases for females if good relative performance information is available.

2.6 Systematic Absenteeism, Cancellations, and Tardiness

Absenteeism, cancellations, and tardiness are frequent in experiments, but their effect on sample composition and results are poorly understood. In this study, because of the screening and exit surveys, we know some of the characteristics of those who missed a scheduled session, canceled at the last moment, or showed up late. Our recruiting procedures were designed to ensure we would observe females during both the high phase and the shorter low-hormone phase. Due to the variation of the menstrual cycle, it is difficult to predict the low hormonal phase for females. Using the screening survey, female subjects were scheduled for their sessions according to their predicted cycle day, calculated using self reported data about the start of previous menstrual periods. Whenever possible these data were combined with self reported data concerning females' hormonal contraceptive regimens. Once the cycle day could be predicted, then a set of possible session days were provided to potential participants and they chose and confirmed a day with a research assistant. For individuals to be considered as scheduled, they had to confirm that they would attend a specific session.

Table 14 shows the proportion of participants that attended the experiment sessions as scheduled, meaning they were present and punctual. Females are noticeably less likely to show up than males: 79% of the males showed up as scheduled compared to 68% of females. Based on predicted phases, only 62% of low phase females attended as scheduled, while high phase females attended 72% of the time. These differences between attendance rates of low phase and non-low phase fe-

Table 14: Session Attendance After Confirmation

Gender		
	Proportion	N
Male	0.79	217
Female	0.68	243
Total	0.73	460

Females by Predicted Phase		
	Proportion	N
High	0.72	141
Low	0.62	102
Total	0.68	246

males were significant at the 5% level. High phase females were less likely to attend compared to males, but this was significant only at the 10% level. Since this study was partly focused on hormonal fluctuations we made attempts to incentivize more low phase females to attend. Part way through the study, the participation payment of $5 was raised to $10 for low phase females.[18] Even with this increase in participation payments a hormonal phase related difference in attendance remained. This result is consistent with Ichino and Moretti (2009), who found that female worker absenteeism was highly correlated with the female menstrual cycle.

These attendance results suggest there may be a systematic bias in the hormonal phase of females who show up to economic experiments in general. Coupled with our evidence of behavioral differences, this bias should be considered in interpreting the results of experiments that report gender differences. Additionally, if females who do not show up to sessions are the ones who have worse symptoms during the low phase, or are more likely to behave differently, then a selection bias may add a downward bias in the hormonal effects found in this study. Though one should still note that any such differences are removed with relative performance feedback.

2.7 Robustness Checks

It is possible that during the low-hormone phase females are not as confident in their abilities as during elevated hormonal phases. Here we show that confidence differences do not explain the choice differences across phases. Table 15 shows that there are no significant differences in confidence for females across phases. There are also no significant differences between females using or not using hormonal contraceptives. Thus, differences in confidence do not seem to explain why females in the

[18]These females were not told why they were receiving a higher participation payment.

low hormonal phase choose not to compete without relative performance feedback. We conducted

Table 15: Confidence Levels of Females by Hormonal Phase

		Confidence (T1)		
Phase	Task	Mean	Std. Error	N
High	Math	1.068	0.032	58
Low	Math	1.041	0.078	23
Both	Math	1.060	0.031	81
High	Word	1.021	0.031	54
Low	Word	1.042	0.057	22
Both	Word	1.027	0.028	76

The sample consists of all females whose phase could be accurately identified.

a variety of robustness checks. First, we checked to see whether the results for gender differences in competition and differences by hormonal phase remain consistent when clustering standard errors on experiment sessions and then also separately on individuals. We found that the results of both the gender and hormonal effects are robust to such error clustering.

Individuals in this study were asked to attend two sessions. Out of 219 total subjects, 126 attended two sessions and 93 attended only a single session. Session composition was random, sessions contained individuals that were attending sessions for either a first or a second time. The type of task individuals were using to compete was different for the first and second session, and the ordering was random. Neither order, the type of task, or group composition in terms of these measures had significant performance or behavioral effects.

The possibility exists that attendance in the second session was affected by how much subjects earned in the first session. Individuals that attended only a single session earned \$9.36 for their performance, while individuals that returned for a second session were paid \$11.73 for their first session performance. This difference in payouts between these groups of individuals is not statistically significant.

Another possibility is that the value of payment in the first session affects confidence in the second session. We exploit the random element in the payment scheme and estimate that, controlling for performance, a person who earns \$10 more than average in the first session increases their prediction of their performance relative to the session average by just 4%. Looking just at the second session females, we find that their confidence in the second session is more sensitive to how much they got paid in the first session than is that of males. But we find no significant difference between females' and males' confidence when attending a second session, or across menstrual cycle phase. Any such difference should in fact make females more competitive in second sessions, but we find no systematic behavior in this direction.

Table 16: Robustness Check for Gender: Controlling for First Session Payment

	(1)	(2)	(3)
	RE	RE & Risk	RE & Full Info (T5)
VARIABLES	CompScale	CompScale	CompScale
Female	-0.40	-0.50	-0.02
	(0.156)	(0.201)	(0.178)
	***	**	
Confidence (T1)	1.00	1.01	0.53
	(0.301)	(0.361)	(0.309)
	***	***	*
Improve (T2)	0.78	0.81	1.10
	(0.245)	(0.333)	(0.276)
	***	**	***
%-tile Rank (T1)	1.08	0.82	2.56
	(0.264)	(0.333)	(0.345)
	***	**	***
First Session Payout	0.02	0.02	0.00
	(0.009)	(0.009)	(0.008)
	*	*	
Risk Controls	No	Yes	No
Characteristic Controls	Yes	Yes	Yes
Observations	343	224	343
chi2	65.63	52.89	100.7
ll	-333.3	-210.2	-315.6

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

It is still possible that there is an interaction between payments and gender driving choice differences. Table 16 provides ordered probit estimations to examine whether females are still less likely to compete when controlling for payments in the first session. Though the first session payout variable is significant at the 10% level, its inclusion does not affect the significance of the gender difference of females to avoid competition.

Results seem consistent for both the sample where risk aversion measures were available and for the sample of participants that did not participate in the risk aversion task. The third column in Table 16 provides the coefficients from the ordered probit estimation for the full information treatment (Treatment 5). Again, with feedback we find that gender does not matter in this setting

Table 17: Robustness Check for Hormones: Controlling for First Session Payment

VARIABLES	(1) RE CompScale	(2) RE & Risk CompScale
Female	-0.29 (0.167) *	-0.27 (0.212)
Low Phase	-0.44 (0.224) **	-0.74 (0.275) ***
Confidence (T1)	0.95 (0.306) ***	0.94 (0.366) **
Improve (T2)	0.74 (0.245) ***	0.77 (0.328) **
%-tile Rank (T1)	0.98 (0.261) ***	0.68 (0.329) **
First Session Payout	0.02 (0.009) *	0.01 (0.009)
Risk Controls	No	Yes
Characteristic Controls	Yes	Yes
Observations	328	211
chi2	62.07	53.99
ll	-320.0	-196.0

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

and neither does the payout from the first session. Thus it seems that the gender difference for competition in low information settings is not affected by the payment received in a previous session.

We now use similar estimations to test whether the hormone specific results are sensitive to the inclusion of first session payments. We find no significant differences in confidence or payouts in the first session between females who could be identified as low phase and the females that were in a second session and were in the high-hormone phase. The first column of 17 provides ordered probit coefficients for the entire sample of individuals, including both those that attended a single session and those that attended two sessions. In this sample, we find that the first session payout variable is significant at the 10% level, but the effect is quite small. The effect of females being in the low hormonal phase remains large and significant at the 5% level.

In the second column in Table 17 we limit the sample to those individuals who participated in the risk aversion task and therefore attended two sessions. In this sample, while controlling for risk aversion, we find that the effect of the low hormonal phase is large and significant, but that the level of the payout from the first session is insignificant. As well, in this group of individuals who did the risk aversion task, we find that the gender effect is insignificant, but that the effect of the low phase for females remains significant.

These checks of the effects of first session payouts on second session attendance and behavior reinforce our conclusion that some of the gender difference in competitive choices is driven by the decisions of females in the low-hormone phase. The results also confirm that the gender and menstrual phase differences are only relevant in low information settings.

3 Discussion

We show that the gender differences in competitive choices in mixed gender groups that have been reported in other studies are robust to a variety of protocol changes, including different tasks and variations in the degree of competitiveness of the available choices. These gender differences are stable, persisting in a second session of the experiment performed days or weeks after the first. This persistence is perhaps more surprising, given that our experiment includes feedback which, in the last treatment of each session, makes men move away from competition and women move towards it. In the subsequent session with a different task (words or math), the gender gap for competition returns– and then goes away again, after the relative performance feedback for that task is provided.

We also find that female choices to compete vary across the menstrual cycle. In the low-hormone phase, females are less likely to enter tournaments than during a high-hormone phase. These differences in competitive choices also become noticeably smaller and statistically insignificant after subjects are provided with feedback on their relative performance. Thus, both the gender difference as well as differences to compete by females across different phases of the menstrual cycle are removed with relative performance feedback. We show the change comes from high ability females moving towards more competitive environments and low ability males moving towards less competitive choices.

The fact that information removes any significant differences in competitive choices has important implications. First, this result suggests that the decision not to compete by females may be more affected by the informational environment and the ambiguity of relative abilities than by any persistent preference for or against competition. Second, since females in the low-hormone phase are less willing to compete than females in the high-hormone phase, this suggests that the hormones linked with the phases of the menstrual cycle may be affecting areas of the brain that help evaluate the rewards and risks associated with these competitive tasks, as suggested by the findings of Dreher et al. (2007).

Progesterone has been found to decrease reported anxiety (Vliet 2001) and elevated progesterone could therefore reduce the disutility of worrying about the outcome of a competition– particularly before feedback information. Estradiol has been found to increase the density of serotonin binding sites in areas of the brain that are linked with evaluating rewards (Fink et al. 1996). Females may experience greater sensitivity to potential rewards during elevated levels of estrogen, which could lead them to be more competitive (Stanton and Schultheiss 2007). Thus, our results are consistent with previous endocrinological studies on hormones and behavior.

Our finding that women's competitive choices increase during a part of the menstrual phase when progesterone and estrogen are high is noticeably different than the result reported in the only other economics paper to look at this issue. Buser (2010) finds that when females compete against other females, females in menstrual phases with high levels of progesterone are less competitive. One major difference between our studies is that we use mixed-gender groups. Other studies have found that female competitive behavior in single gender groups is very different from that in mixed gender groups, with females being much more willing to compete when the group is all female (Gupta et al. 2005, Grosse and Reiner 2010). Being in a same gender group may promote the effect of progesterone on group affiliation motives reported in Schultheiss et al. (2004). As the gender composition experiments show, competitive decisions are context dependent, and the neuroendocrinological literature (Anders and Watson 2006) emphasizes that the behavioral effects of hormones are also context dependent. Thus, the gender composition of the particular labor market being examined may interact with the effects of hormones and relative performance feedback. Another difference is that we use a within-subjects design, which is particularly important in examining the effects of hormonal variations that are inherently within-subjects. A final difference between these studies is that our study includes a larger number of female subjects in both high and low-hormone phases. One potential disadvantage of the within-subjects design is that experiences in the first session might affect choices in the second. We use two different tasks, with the session order randomized to minimize this problem, and we show the results are robust to order and tasks.

The existence of regular and predictable within-subjects variation in behavior across the menstrual cycle suggests that there is a biological component to the gender difference in competitive choices – but not a simple one. We cannot conclude hormones explain the male female differences – there are too many other things that vary across genders. Instead, we show that biology drives differences in behavior for individual women, across time. This raises many interesting questions for optimal choice. Women who know that they are subject to this variation might well take steps to increase or decrease their competitive choices by changing when they make decisions involving competition, such as job choices and college applications.

Interestingly, hormonal contraceptives allow for control of the timing of the low-hormone phase, and in fact, there is no medical reason for the low hormone (placebo) phase of the hormonal contraceptive regimen (Anderson and Hait 2003). As a leading textbook on clinical gynecological endocrinology states: *Monthly bleeding, periodic bleeding, or no bleeding– this is an individual woman's choice (Speroff and Fritz 2005, pg 908).* It has been suggested that the placebo and withdrawal bleeding was a marketing effort to make the birth control pill seem less novel and more acceptable(Coutinho and Segal 1999). Females can entirely avoid hormonal fluctuations and the low-hormone phase by sustained contraceptive use, and some forms of contraception such as the oral contraceptive *Seasonale*, now eliminate the placebo phase and ensure that the low-hormone phase does not occur (Anderson and Hait 2003). If the decrease in hormones affects females decisions in a costly manner then there is an incentive for avoiding this phase; hence a further reason the existence of a market for sustained contraceptives. Of course, our results also show that such manipulation would only matter when there is poor information about ability relative to potential competitors.

Since information about relative ability removes all behavioral differences across the menstrual cycle and between genders, it would be natural to suggest that the differences in competitive choices might be the result of differences in overconfidence. But we find gender differences in competitive choices in both math and word tasks. In the word tasks, confidence levels of females and males are the same and a gender difference in choices persists. In the math task the gender difference persists even after controlling for confidence. The information about relative abilities may help subjects form more accurate beliefs about the possibilities of succeeding in competitive situations, but since females hold similar overconfident beliefs as males in word tasks, we would expect females to choose in a similar fashion as males. Instead, females choose not to compete in low information settings, irrespective of confidence or the type of task.

Another possible explanation for the differences in choices across gender with and without feedback would be a gender difference in risk aversion. However, when we control for risk aversion the gender difference in choices still remains in the uninformed treatment. Furthermore we find risk aversion is not significant in explaining competitive choices in the full information treatment. Another explanation that would be consistent with our results would be a gender difference in ambiguity aversion. Little research has been done on gender differences in ambiguity aversion and the results are mixed.

Firms, governments, and schools sometimes implement affirmative action policies to encourage females to apply for competitive jobs and scholarships. Affirmative action policies typically focus efforts on recruiting females or changing the acceptance or promotion process to favor females. Niederle et al. (2009) report on experiments showing that affirmative action can encourage competition by females with low efficiency costs. The performance feedback condition result from our study suggests an alternative to affirmative action – providing better information about relative abilities.

We find that information on relative ability plays a large role in choices to compete. Information reduces gender differences, reduces differences across the menstrual cycle, and improves sorting by ability. Schools and many large firms already collect relative performance rankings about students or employees (Pfeffer and Sutton 2006), and our work suggests that making these rankings more available to these agents and emphasizing relative comparisons will promote an environment where the best workers, both females and males, will seek out more competitive positions.

References

Aden, U., Jung-Hoffmann, C. and Kuhl, H.: 1998, A randomized cross-over study on various hormonal parameters of two triphasic oral contraceptives, *Contraception* **58**(2), 75–81.

Anders, S. and Watson, N.: 2006, Social Neuroendocrinology: Effects of Social Contexts and Behaviors on Sex Steroids in Humans, *Human Nature* **17**(2), 212–237.

Anderson, F. and Hait, H.: 2003, A multicenter, randomized study of an extended cycle oral contraceptive, *Contraception* **68**(2), 89–96.

Apicella, C., Dreber, A., Campbell, B., Gray, P., Hoffman, M. and Little, A.: 2008, Testosterone and financial risk preferences, *Evolution and Human Behavior* **29**(6), 384–390.

Arellano, M. and Honoré, B.: 2001, Panel data models: some recent developments, *Handbook of Econometrics* **5**, 3229–3296.

Association, A. C. H. A. C. H.: 2009, Acha national college health assessment ii: Reference group data report, *Technical report*, ACHA.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U. and Fehr, E.: 2008, Oxytocin shapes the neural circuitry of trust and trust adaptation in humans, *Neuron* **58**(4), 639–650.

Bertrand, M. and Hallock, K.: 2001, The gender gap in top corporate jobs, *Industrial and Labor Relations Review* pp. 3–21.

Bethea, C., Lu, N., Gundlah, C. and Streicher, J.: 2002, Diverse Actions of Ovarian Steroids in the Serotonin Neural System, *Frontiers in Neuroendocrinology* **23**(1), 41–100.

Borghans, L., Golsteyn, B., Heckman, J. and Meijers, H.: 2009, Gender Differences in Risk Aversion and Ambiguity Aversion, *Journal of the European Economic Association* **7**(2-3), 649–658.

Brown, C.: 1990, Firms' choice of method of pay, *Industrial and Labor Relations Review* pp. 165–182.

Burnham, T.: 2007, High-testosterone men reject low ultimatum game offers, *Proceedings of the Royal Society B: Biological Sciences* **274**(1623), 2327.

Buser, T.: 2010, The impact of female sex hormones on competitiveness, *Unpublished manuscript* .

Chakraborty, T., Ng, L. and Gore, A.: 2003, Age-Related Changes in Estrogen Receptor {beta} in Rat Hypothalamus: A Quantitative Analysis, *Endocrinology* **144**(9), 4164.

Chen, Y., Katuscak, P. and Ozdenoren, E.: 2009, Why Can't a Woman Bid More Like a Man?, *Working Paper, Ann Arbor* **1001**, 48109–1220.

Coates, J. and Herbert, J.: 2008, Endogenous steroids and financial risk taking on a London trading floor, *Proceedings of the National Academy of Sciences* **105**(16), 6167.

Coutinho, E. and Segal, S.: 1999, *Is menstruation obsolete?*, Oxford University Press, USA.

Creinin, M., Keverline, S. and Meyn, L.: 2004, How regular is regular? An analysis of menstrual cycle regularity, *Contraception* **70**(4), 289–292.

Dohmen, T. and Falk, A.: 2007, Performance pay and multi-dimensional sorting: productivity, preferences and gender, *Institute for the Study of Labor (IZA); CESifo (Center for Economic Studies and Ifo Institute for Economic Research)* .

Dreher, J., Schmidt, P., Kohn, P., Furman, D., Rubinow, D. and Berman, K.: 2007, Menstrual cycle phase modulates reward-related neural function in women, *Proceedings of the National Academy of Sciences* **104**(7), 2465.

Eriksson, T., Teyssier, S. and Villeval, M.-C.: 2009, Self-selection and the Efficiency of Tournaments., *Economic Inquiry* **47**(3), 530 – 548.

Fehr, E.: 2009, On the economics and biology of trust, *Journal of the European Economic Association* **7**(2-3), 235–266.

Fernandez, G., Weis, S., Stoffel-Wagner, B., Tendolkar, I., Reuber, M., Beyenburg, S., Klaver, P., Fell, J., de Greiff, A., Ruhlmann, J. et al.: 2003, Menstrual cycle-dependent neural plasticity in the adult human brain is hormone, task, and region specific, *Journal of Neuroscience* **23**(9), 3790.

Fink, G., Sumner, B., Rosie, R., Grace, O. and Quinn, J.: 1996, Estrogen control of central neurotransmission: effect on mood, mental state, and memory., *Cell Mol Neurobiol* **16**(3), 325–44.

Gagnidze, K. and Pfaff, D.: 2009, Sex on the Brain, *Cell* **139**(1), 19–21.

Gneezy, U., Leonard, K. and List, J.: 2009, Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society, *Econometrica* **77**(5), 1637–1664.

Gneezy, U., Niederle, M. and Rustichini, A.: 2003, Performance in Competitive Environments: Gender Differences, *Quarterly Journal of Economics* **118**(3), 1049–1074.

Gneezy, U. and Rustichini, A.: 2004, Gender and Competition at a Young Age, *American Economic Review* **94**(2), 377–381.

Greene, W.: 2004, The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects, *Econometrics Journal* **7**(1), 98–119.

Grosse, N. and Riener, G.: 2010, Explaining Gender Differences in Competitiveness: Gender-Task Stereotypes, *Jena Economic Research Papers* **2010**, 017.

Grossman, P. and Eckel, C.: 2003, Men, women and risk aversion: Experimental evidence, *Unpublished manuscript* .

Gunther, C., Ekinci, N., Schwieren, C. and Strobel, M.: 2008, Women can't jump?–An experiment on competitive attitudes and stereotype threat, *Technical report*, Mimeo.

Gupta, N., Poulsen, A. and Villeval, M.: 2005, Do (wo) men prefer (non-) competitive jobs, *Technical report*, GATE Working Paper.

Gupta, N., Poulsen, A., Villeval, M. and Gade, H.: 2005, Male and Female Competitive Behavior-Experimental Evidence, *IZA Working Paper* .

Holt, C. and Laury, S.: 2002, Risk Aversion and Incentive Effects, *American Economic Review* **92**(5), 1644–1655.

Ichino, A. and Moretti, E.: 2009, Biological Gender Differences, Absenteeism and the Earnings Gap, *American Economic Journal: Applied Economics* **1**(1), 183–218.

Jirjahn, U. and Stephan, G.: 2004, Gender, piece rates and wages: evidence from matched employer-employee data, *Cambridge Journal of Economics* **28**(5), 683–704.

Lazear, E. and Rosen, S.: 1981, Rank-order tournaments as optimum labor contracts, *The Journal* *of Political Economy* **89**(5), 841.

Malison, R., Price, L., Berman, R., van Dyck, C., Pelton, G., Carpenter, L., Sanacora, G., Owens, M., Nemeroff, C., Rajeevan, N. et al.: 1998, Reduced brain serotonin transporter availability in major depression as measured by [123I]-2β-carbomethoxy-3β-(4-iodophenyl) tropane and single photon emission computed tomography, *Biological Psychiatry* **44**(11), 1090–1098.

McEwen, B.: 2002, Estrogen Actions Throughout the Brain, *Recent Progress in Hormone Research* **57**(1), 357–384.

Moore, E. and Eckel, C.: 2003, Measuring Ambiguity Aversion. Virginia Tech.

Mosher, W., Martinez, G., Chandra, A., Abma, J. and Willson, S.: 2004, Use of contraception and use of family planning services in the United States: 1982–2002, *Advance data from vital and health statistics* **350**, 1.

Niederle, M., Segal, C. and Vesterlund, L.: 2009, How costly is diversity? Affirmative action in light of gender differences in competitiveness, *NBER Working Paper* .

Niederle, M. and Vesterlund, L.: 2007, Do women shy away from competition? do men compete too much?, *Quarterly Journal of Economics* **122**(3), 1067–1101.

Niederle, M. and Vesterlund, L.: 2010, Explaining the Gender Gap in Math Test Scores: The Role of Competition, *The Journal of Economic Perspectives* **24**(2), 129–144.

Pearson, M. and Schipper, B.: 2009, Menstrual cycle and competitive bidding, *MPRA Paper* .

Pfeffer, J. and Sutton, R.: 2006, Evidence-based management, *Harvard Business Review* **84**(1), 62.

Platt, M. and Huettel, S.: 2008, Risky business: the neuroeconomics of decision making under uncertainty., *Nat Neurosci* **11**(4), 398–403.

Polachek, S.: 1981, Occupational self-selection: A human capital approach to sex differences in occupational structure, *The Review of Economics and Statistics* pp. 60–69.

Schultheiss, O., Wirth, M. and Stanton, S.: 2004, Effects of affiliation and power motivation arousal on salivary progesterone and testosterone, *Hormones and Behavior* **46**(5), 592–599.

Shepard, K., Michopoulos, V., Toufexis, D. and Wilson, M.: 2009, Genetic, epigenetic and environmental impact on sex differences in social behavior, *Physiology & behavior* **97**(2), 157–170.

Sinha-Hikim, I., Arver, S., Beall, G., Shen, R., Guerrero, M., Sattler, F., Shikuma, C., Nelson, J., Landgren, B., Mazer, N. et al.: 1998, The use of a sensitive equilibrium dialysis method for the measurement of free testosterone levels in healthy, cycling women and in human immunodeficiency virus-infected women, *Journal of Clinical Endocrinology & Metabolism* **83**(4), 1312.

Speroff, L. and Fritz, M.: 2005, *Clinical Gynecologic Endocrinology and Infertility*, Lippincott Williams & Wilkins.

Stanton, S. and Schultheiss, O.: 2007, Basal and dynamic relationships between implicit power motivation and estradiol in women, *Hormones and Behavior* **52**(5), 571–580.

Stricker, R., Eberhart, R., Chevailler, M., Quinn, F., Bischof, P. and Stricker, R.: 2006, Establishment of detailed reference values for luteinizing hormone, follicle stimulating hormone, estradiol, and progesterone during different phases of the menstrual cycle on the Abbott ARCHITECT analyzer., *Clinical Chemistry and Laboratory Medicine* **44**(7), 883.

Vliet, E.: 2001, *Screaming to be heard: hormone connections women suspect and doctors still ignore*,

M. Evans and Co., New York, NY.

Wu, M., Manoli, D., Fraser, E., Coats, J., Tollkuhn, J., Honda, S., Harada, N. and Shah, N.: 2009, Estrogen masculinizes neural pathways and sex-specific behaviors, *Cell* **139**(1), 61–72.

Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., von Schoultz, B., Hirschberg, A. and Johannesson, M.: 2009, A randomized trial of the effect of estrogen and testosterone on economic behavior, *Proceedings of the National Academy of Sciences* **106**(16), 6535.

Appendices

Appendix A (Not For Publication): About controls

A number of control variables were used in this analysis including whether a subject was attending a first or second session, or a word based session, neither were ever significant. Also, other individual characteristic control variables such as the number of years of post secondary schooling, age and GPA were included, but were never significant with the exception of number of years of post secondary schooling, which was negative and significant. In estimations not included here categorical dummy variables were included for the majors of students, but none were significant and were removed due to their irrelevancy.

The other set of controls that are not listed are measures controlling for risk aversion. After participating in both a math and word session, subjects were invited to participate in a Holt and Laury (2002) task a few days after the competition experiment sessions. These risk aversion sessions were done separately so as not to affect behavior in the competition task. Consequently, the sample of these individuals is slightly different from the other specifications, but the results remain fairly consistent. The risk aversion coefficients were not found to be significant in the treatment with no relative information feedback.[19] The only specification where these measures are significant is in the full information treatment for a limited sample of high ability individuals. The fact this measure does not significantly predict behavior in the no information treatment is intuitive as individuals have little information regarding their probabilities of success; thus, beliefs may matter more in such situations than risk preferences. Since risk aversion is significant for high ability individuals in the full information treatment; this result suggests that risk appetites matter only to individuals that have higher returns from making an optimal choice. For those individuals of low ability, their expected value is higher by avoiding the tournament regardless of risk preferences.[20]

Three variables that are used and presented in the tables include, Confidence (T1), Improve (T2), %-tile Rank (T1). These variables deserve further explanation as they are measuring some very relevant factors in the individual decision making process. In other estimations, variations of these variables were used, but these variables provide the cleanest and most intuitive interpretation possible and results remain consistent with other possible measures of the same factors.

[19]For brevity these results are not included in this discussion, but are available upon request.

[20]There were no significant gender differences for risk aversion in the sample.

Appendix B (Not For Publication): Performance

The main focus of this paper is on the compensation choices that individuals made in the experiment; however, these choices may have been affected by performance differences. This section focuses on task performance for the different treatments. To consider how individuals are affected by the different incentives of each type of compensation, we focus on the performance of individuals in the first three treatments. In these first three treatments, individuals had no choice over the type of compensation they received for their efforts; thus, the performance effects from different compensation environments are exogenously determined.

According to the theory of piece rates and tournaments, one would expect greater effort for a higher piece rate. Similarly, an individual of higher ability and higher probability of winning should increase effort in a tournament. As the tournament gets larger and more competitive, one would expect that individuals would increase effort or set their effort levels to zero. Before considering the effects of tournament size on effort, we first focus on possible order effects and gender differences between treatments.

The regressions in Table 18 are used to consider gender differences in performance, learning effects and the incentive effects of increasing tournament size. The performance in the word task, but not the math task, is highly correlated with the GPA of participants. Regression estimates for both word and math show an order effect suggesting that subjects are learning in the first three treatments. Regression 1 in Table 18 shows that the tournament size has a statistically significant positive effect on individual performance in the math task. Increasing the competitiveness of a compensation environment from the piece rate to a tournament size of 6 should increase performance of an individual by 0.65 problems in a four-minute task. This is an increase of 5.7% for the average individual. In columns 2, 4, and 5, categorical variables are used to investigate whether tournament size is actually leading to the increase or whether just competing against someone in a tournament of any size leads to performance increases.

In the second column in Table 18, categorical variables were used for each of the possible competitive environments, for group pay and for tournaments. To test whether the tournament size matters a separate dummy variable, Tourney (ts>1) was used to identify if an individual had to compete against someone else. This categorization created a separate baseline for tournaments consisting of six individuals. Once controlling for tournament competition it was found that the size of the tournament does not matter and that group pay performance is not significantly different from the piece rate environment in the math task. There is also a positive effect for age as the average effect of a year of life leads to an increase in performance in the math task of 0.5 problems, though this may be offset by further post-secondary schooling. It is worth noting that overall the competitive environment and individual characteristics explain very little of the variation in performance for the math task in terms of goodness of fit measures such as R^2.

The results from the word task (Regressions 3 and 4) in Table 18 suggest that neither the tournament nor tournament size increase performance. There is a significant amount of learning that occurs with each treatment. GPA has a significant positive effect in terms of performance. This likely occurs because an individual's vocabulary expands with age and individuals with a higher GPA probably have richer vocabularies than individuals with lower GPAs. More of the variation in performance can be explained in regressions using the word task than the math task; this mainly stems from the inclusion of control variables for the random letters used for each task.

The math task results suggest that being in a tournament does increase performance, but the size of the tournament is irrelevant. One might expect that only high ability individuals would increase performance from the incentive effects from being in a tournament, but we find the opposite. In splitting the sample for high and low ability individuals, according to their performance in the first task and whether they are above or below the median, we find that the low ability individuals increase performance in response to being in a tournament (significant at 1%) in the math task. We find no significant effects from tournament size for high ability individuals in the math task.[21]

The competitiveness of the environment has a significant impact on performance only in math tasks and once an individual is participating in a tournament, then the number of competitors does not lead to further performance benefits. Competitive environments (tournament size) had no influence on the performance of individuals in the word task. Thus, depending on the type of task, competition between individuals may increase performance. Therefore, due to mixed results, one cannot conclude that tournament size increases performance or effort of agents.

Overall, we find that a more competitive work environment may not lead to performance increases as the incentive effects of competitiveness are not robust across different types of tasks. Another important result shown by these regressions is that there are no significant performance

[21]The estimation results for high ability individuals are not shown here, but none of the competitive environment variables were significant in these estimations.

differences between males and females. In terms of performance effects, some learning occurs across the different treatments and only low ability individuals tend to increase performance when they are put in a tournament of any size – it is enough to be competing against someone.

Table 18: RE Performance Regression of No Choice Treatments

VARIABLES	(1) Math	(2) Math	(3) Word	(4) Word	(5) Low Math
Task Order	0.88	0.82	3.49	3.50	1.10
	(0.09)	(0.09)	(0.39)	(0.38)	(0.11)
	(***)	(***)	(***)	(***)	(***)
Tourney Size	0.13		0.04		
	(0.04)		(0.15)		
	(***)		()		
Tourney(ts>1)		0.63		-0.16	1.18
		(0.27)		(1.02)	(0.31)
		(**)		()	(***)
Tourney Size=2		-0.07		2.08	-0.24
		(0.38)		(1.49)	(0.42)
		()		()	()
Tourney Size=4		-0.21		0.22	-0.73
		(0.37)		(1.40)	(0.45)
		()		()	()
Female	-0.45	-0.46	2.31	2.29	0.44
	(0.59)	(0.59)	(1.87)	(1.88)	(0.39)
	()	()	()	()	()
Years PS	-0.50	-0.50	0.16	0.15	0.17
	(0.29)	(0.30)	(0.87)	(0.87)	(0.23)
	(*)	(*)	()	()	()
Age	0.40	0.40	0.33	0.34	-0.07
	(0.15)	(0.15)	(0.49)	(0.49)	(0.14)
	(***)	(***)	()	()	()
GPA	-0.26	-0.26	6.60	6.57	0.32
	(0.63)	(0.64)	(1.94)	(1.95)	(0.41)
	()	()	(***)	(***)	()
Letter Controls	No	No	Yes	Yes	No
Observations	534	534	492	492	303
Number of id	178	178	164	164	101
R-sq	0.0745	0.0721	0.367	0.368	0.196
chi2	102.5	104.5	565.4	570.9	136.0

Standard errors in parentheses

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

Appendix B.1 Performance differences according to hormonal phase.

To examine performance differences across the menstrual cycle, we consider the word and math tasks separately and estimate effects using linear specifications similar to the ones used to examine exogenous performance effects of tournaments (Appendix B). Table 19 provides the random effects OLS estimates for a number of factors that may explain performance differences in both word and math tasks for all treatments where the participants could not choose their competitive environments. The estimations for math tasks are in columns 1 to 3 and the estimations for word tasks are in columns 4 to 6.

There seems to be no correlation with the low-hormone phase and performance in the word task (columns 4 to 6) for females. The Low Phase coefficient is insignificant for all the different samples in the word task. Focusing on performance in the math task (columns 1 to 3), there seems to be no effect from the low phase in the sample of both females and males (column 1), and only females (column 2). There seems to be some effect for low phase females when including controls for risk aversion for the portion of the particpants for which such measures were available. On average, performance in the math task decreases by about 2.2 correct answers for low phase females when controlling for risk aversion, which is a 20% decrease for the average female. Though the low phase effect is only significant at the 10% level when taking into account individuals for which measures of risk aversion can be used as controls. Therefore, if performance or confidence differences are driving selection differences, then controlling for performance and confidence in a discrete choice model should help isolate the effect of the low phase on selection choices.

Table 19: Hormonal Effects for Performance (t<4)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All	Female	Female	All	Female	Female
Task	Math	Math	Math	Word	Word	Word
VARIABLES	RE	RE	RE Risk	RE	RE	RE Risk
Task Order	0.87	0.94	1.01	3.46	3.92	3.09
	(0.09)	(0.13)	(0.18)	(0.39)	(0.57)	(0.70)
	(***)	(***)	(***)	(***)	(***)	(***)
Tourney Size	0.14	0.12	0.09	0.05	-0.03	-0.20
	(0.04)	(0.06)	(0.08)	(0.16)	(0.23)	(0.29)
	(***)	(**)	()	()	()	()
Low Phase	-1.07	-1.09	-2.21	2.22	2.09	4.57
	(0.98)	(0.71)	(0.89)	(3.00)	(2.98)	(3.58)
	()	()	(**)	()	()	()
Female	-0.32			1.72		
	(0.67)			(2.11)		
	()			()		
Years PS	-0.41	-0.70	-1.19	0.31	-0.87	-3.41
	(0.30)	(0.40)	(0.52)	(0.87)	(1.39)	(2.10)
	()	(*)	(**)	()	()	()
Age	0.38	0.55	0.89	0.31	1.48	3.52
	(0.15)	(0.23)	(0.27)	(0.49)	(0.90)	(1.22)
	(**)	(**)	(***)	()	()	(***)
GPA	-0.12	-0.43	-0.46	5.73	7.56	6.22
	(0.65)	(0.78)	(1.03)	(2.00)	(3.19)	(4.13)
	()	()	()	(***)	(**)	()
Letter Controls	No	No	No	Yes	Yes	Yes
Risk Controls	No	No	Yes	No	No	Yes
Observations	510	237	147	471	225	147
Number of id	170	79	49	157	75	49
R-sq w	0.206	0.242	0.248	0.624	0.647	0.665
R-sq b	0.0740	0.147	0.365	0.291	0.340	0.430
R-sq o	0.0909	0.168	0.339	0.369	0.410	0.496
chi2	100.5	62.28	54.45	563.4	289.1	209.3
df_m	9	8	10	16	15	17

Standard errors in parentheses

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

Appendix C (Not For Publication): Soting Theory

The goal of the experiment design was to mimic sorting by ability that a firm would want in terms of having individuals apply for competitive promotions. To do this we use a menu of competitive choices as a sorting mechanism where high ability individuals choose more competitive compensation schemes while low ability individuals choose less competitive compensation schemes. First let us consider a simple piece rate framework. Suppose an individual is paid a piece rate, w, for each unit of effort, e. Assume the individual has a linear utility function in terms of payoffs, but payoffs come at the cost of a convex function for effort, where the disutility of effort is determined by θ. Then an individual's utility function for working under a piece rate scheme is as follows:

$$\begin{aligned} U_{pr}(e) &= we - e^{\theta} \\ &\text{where } \theta > 1 \end{aligned} \qquad \text{(Appendix C.1)}$$

Optimal effort depends on the piece rate, w, and the disutility from effort, θ. This disutility of effort can be thought of as an ability parameter where an individual with higher ability will have a smaller θ. The optimal choice of effort using the piece rate is then:

$$e^{*}_{pr} = \left(\frac{w}{\theta}\right)^{\frac{1}{\theta-1}} \qquad \text{(Appendix C.2)}$$

An individual that chooses group pay receives half of the output that he provides from his own effort along with half the output that someone else provides. The amount an individual expects to get from his group member is the average of all the possible individuals that he may be grouped with. Thus, in making the decision to choose group pay the individual must consider the expected utility from such a selection. We will denote the expectation by individual i of others' $(-i)$ efforts as $E[e_{-i}]$. The expected utility from choosing group pay is as follows:

$$\begin{aligned} U_G(e_i) &= 0.5wE[e_{-i}] + 0.5we_i - e_i^{\theta_i} \\ &\text{where } \theta_i > 1 \end{aligned} \qquad \text{(Appendix C.3)}$$

In this model the optimal choice of effort under group pay is not dependent on the efforts of others provide. In essence, if an individual is performing under a group pay compensation scheme, then the element of potential free riding causes the individual to lower his efforts, regardless of the efforts of the other group member. Therefore, we can write the optimal choice of effort under the group pay compensation scheme in a general form, such as:

$$e_G^* = (\frac{0.5\, w}{\theta})^{\frac{1}{\theta - 1}} \qquad \text{(Appendix C.4)}$$

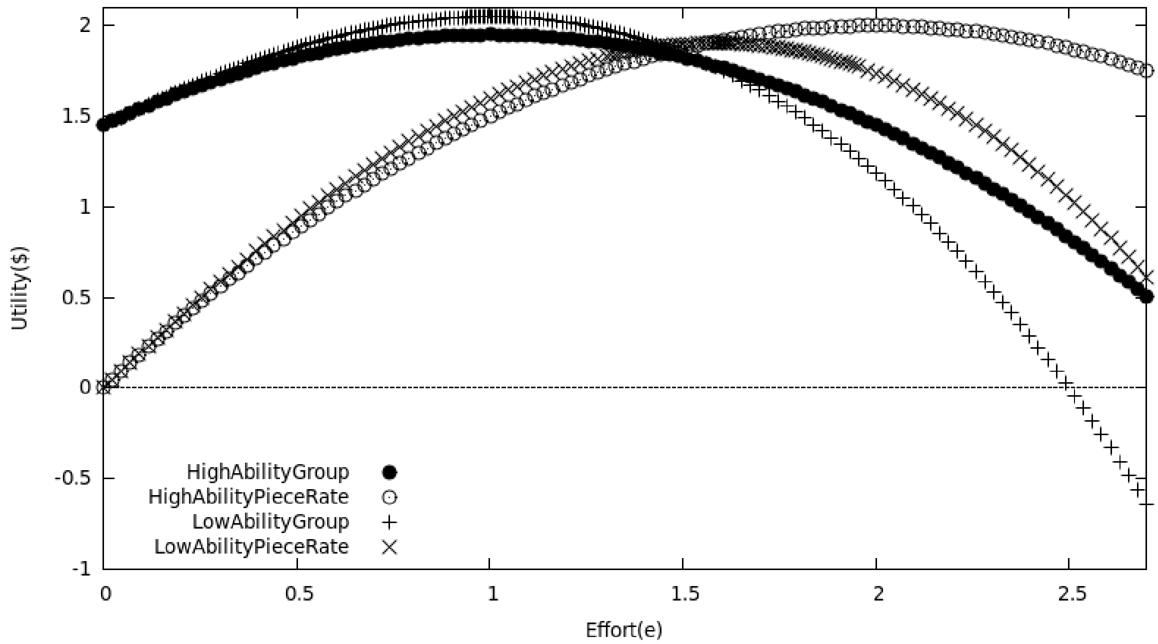
This optimal choice of effort under group pay is less than the level of effort from piece rate for individuals of all abilities. Thus, for an individual to choose group pay over the piece rate compensation it must be that the utility from the optimal choice of effort under group pay is greater than the utility found from the optimal choice of effort under the piece rate compensation. This decision will depend on whether the expected effort from others, $E[e_{-i}]$ is large enough to dissuade one from choosing piece rate. This tipping point for $E[e_{-i}]$ falls between an individual's optimal choice of effort under group pay and the optimal choice of effort under the piece rate compensation and is largely dependent on an individual's ability in the form of his disutility from effort θ. Normalizing the piece rate to one ($w = 1$), the value for $E[e_{-i}]$ such that the individual is indifferent between group pay and piece rate is:

$$E[e_{-i}] = 2(\theta^{\frac{1}{1-\theta}} - \theta^{\frac{\theta}{1-\theta}} - 0.5(2\theta)^{\frac{1}{1-\theta}} + (2\theta)^{\frac{\theta}{1-\theta}}) \qquad \text{(Appendix C.5)}$$

For an individual with $\theta = 2$, this indifference point is the mean value between e_{pr}^* and e_G^*. In terms of our experiment design, this type of choice menu, including piece rate and group pay, leads higher ability individuals to choose the piece rate and exert more effort, rather than choose group pay. Low ability individuals will gain greater utility from choosing group pay and free riding on someone else's effort. For these reasons of reduced effort and free riding, along with group pay being preferred by lower ability individuals, we consider the group pay compensation scheme to be less competitive than the piece rate. In terms of sorting by abilities between group pay and piece rate, we can see that in Figure 9 the high ability individual receives greater utility from choosing the piece rate while the low ability individual receives greater utility from working under the group pay compensation scheme. In this figure, we assume that the piece rate is equal to two, $w = 2$. The non-zero intercept reflects the expectation of the other group member's efforts, $E[e_{-i}] = 1.45$.

The other competitive choices in the experiment consist of entering a tournament of different possible sizes: with a total of two, four or six competitors (including oneself). In this experiment, along with others like it (see Niederle and Vesterlund, 2007), the tournament payoffs differ from

Figure 9: Utility of Group Pay vs. Piece Rate by Ability



what is found in the standard tournament literature. The standard tournament literature considers only fixed prizes for tournaments, which by design is meant to deal with difficult to measure output of workers (Lazear and Rosen, 1981). Our payoffs differ and are analogous to the flexible winner payoffs found in Niederle and Vesterlund (2007).

Furthermore, in the experiment output is measurable, but some of the output may be attributable to luck and not effort, due to the precision involved in the task. Output for individual i can be written as $y_i = e_i + \epsilon_i$. One can therefore think of the probability of winning the tournament for player i, when playing against player j, as $pr(e_i, e_j) = pr(\epsilon_i - \epsilon_j > e_j - e_i)$. But if one considers that the probability of winning a tournament leads to a Nash equilibrium solution with both individuals providing effort then they both must be of comparable abilities, θ, with comparable variance in terms of luck, ϵ. If the competitors were not of similar ability then the individual with greater ability, lower θ, could increase his effort and therefore increase his probability of winning while it becomes more costly for his competitor to do so. The best response of this increased effort by the lower ability individual would eventually lead to an effort of zero. Such a result would lead to only the highest ability individual entering the tournament. To ensure that individuals who choose to compete in a tournament always have competitors, we added an extra element in the experiment design for individuals that choose to compete in a tournament: *if there are not* *enough people for your chosen format, then individuals who have selected another pay format will be randomly selected to meet the participation requirements for your chosen format.*

In terms of expected utility from choosing a tournament, we follow a similar framework as Eriksson et al. (2009) though we use the flexible payoffs and simply an exogenous probability of winning. The exogenous probability of winning can be thought of as a function of ability and the best response towards the expected effort of other competitors. For simplicity, we focus on what the probability of winning a tournament of a certain size needs to be for a person to see entering a tournament as being superior to the other options of piece rate and group pay.[22]

Let n be the size of the tournament, where $n-1$ is the number of individuals someone competes against within a tournament. Suppose the exogenous probability of beating an average individual in a tournament for individual i is equal to π_i. The expected utility of competing in a tournament of any size n such that $n \in 1, 2, 3, 4, ..., N$ and an exogenous probability of winning is:

$$U_T(e) = nw\pi_i^{n-1}e - e^{\theta_i} \quad \text{where } \theta_i \geq 2 \tag{Appendix C.6}$$

This expected utility for tournaments of size n simplifies to the piece rate when $n = 1$, a tournament without competitors. The optimal choice of effort for a tournament of size n is as follows:

$$e_T^* = (\frac{nw\pi_i^{n-1}}{\theta_i})^{\frac{1}{\theta_i - 1}} \tag{Appendix C.7}$$

Consequently, an individual's choice of piece rate or tournament size in this experiment largely depends on her probability of success. In Table 20, we show the minimum probability of winning against an average individual that is needed to choose a tournament of size n over the piece rate form of compensation. As a tournament of size n is preferred to the piece rate as long as $\frac{1}{n} < \pi_i^{n-1}$.

Table 20: Minimum win probabilities for tournament entry

Tourney Size	**Min. Probability**
2	$\frac{1}{2} < \pi_i$
4	$\frac{1}{\sqrt[3]{4}} \approx 0.63 < \pi_i$
6	$\frac{1}{\sqrt[5]{6}} \approx 0.70 < \pi_i$

[22] Since we have already examined group pay versus piece rate, we focus our attention on tournament versus piece rate.

If we suppose that the win probabilities are correlated with abilities such that higher ability types have greater probabilities of winning then we can expect that higher ability individuals choose more competitive tournaments. Therefore, in combining the choices of group pay, piece rate, tournaments, we have choices that can be considered as ordinal across a spectrum of competitiveness. If one does not, believe that these categories can be considered competitiveness then one can consider them as being a type of menu pricing such that subjects should sort themselves by their abilities, with low ability individuals choosing group pay, followed by the next highest ability individuals choosing piece rate, followed by the highest ability individuals choosing tournaments.

Appendix D (Not For Publication) Experiment Instructions

Appendix D.1 Math Task Instructions

General Instructions

In this experiment you will be performing a task five different times. The task will consist of having you solve 2-digit 4-number addition problems in a 4 minute period.

The addition problems will look similar to the following equation:

$12 + 57 + 48 + 52 =$

In some cases, you will be asked to make decisions about how you will potentially be paid for your performance.

Only one of the five tasks will determine your payout for the experiment and it will be randomly chosen at the end.

To answer a problem, you will simply type the numbers on the keyboard, then press enter and another problem will appear. You can choose not to answer a question by pressing the Enter button or clicking on submit. The answer will then be recorded as being incorrect and you will be moved to the next problem.

To help with time management, there will be a clock counting down the seconds for the 4 minute duration.

Task 1

For Task 1, you will be paid $0.50 for each correctly answered addition problem during the 4 minute time limit.

You will not know how many problems you answered correctly until the end of the 4 minutes.

At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen showing your payout method and confirming that you understand.

The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the ”I understand” button below.

Task 2

For this task, you will be randomly placed in a tournament.

The tournament will have a size of *2 or 4 or 6* people, including yourself.

If you win the tournament you will be paid $0.50 multiplied by the number of people in the tournament for each correctly answered problem. For example, if you are in a 4 person tournament then you will be paid $2.00 for each correct answer so long as you win the tournament.

If you do not win the tournament then you will receive nothing.

In the event of a tie for first place, you will split your tournament winnings evenly with the number of people that tied for first place in the tournament.

You will not know who you are competing against.

You will not know how many problems you answered correctly until the end of the 4 minutes.

To help with time management, there will be a clock counting down the seconds for the 4 minute duration. At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen showing your tournament size and confirming that you understand.

The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the ”I understand” button below.

Task 3

For this task, you will be randomly put in a group with one other person.

The group has a total size of 2 people (including yourself) and group members will be paid an equal amount.

You will be paid according to the productivity of your group. The number of correct answers by you and your partner will be added together and as a group you will be paid $0.50 per correct answer. Your group earnings will be divided evenly between each partner ensuring you each get equal pay.

You will not know who you are a partner with.

At the end of the 4 minutes you will see a screen asking about your performance.

You will not know how many problems you answered correctly until the end of the 4 minutes.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen showing your payout method and confirming that you understand.

The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Task 4

You will now have the opportunity to choose the individual pay rate of $0.50 per correct answer, or to enter in a tournament, or to join a partner and receive the group pay rate.

Your possible choices are: *Group Pay, Individual Pay, 2 Person Tournament, 4 Person Tournament, 6 Person Tournament*

Whenever possible, you will be paired with other people that chose the same pay format as yourself. If there are not enough people for your chosen format, then individuals who have selected another pay format will be randomly selected to meet the participation requirements for your chosen format.

Tournament:

If you win the tournament, then you will be paid $0.50 multiplied by the number of people in the tournament for each correctly answered problem. In the event of a tie for first place, you will split your tournament winnings evenly with the number of people that tied for first in the tournament. If you do not win the tournament then you will receive nothing.

You will not know who you are competing against.

Individual Pay:

You will be paid $0.50 for each correctly answered problem.

Group Pay:

For the Group Pay format, you will be joining one other person in a group.

You will be paid according to the productivity of your group. The number of correct answers by you and your partner will be added together and as a group you will be paid $0.50 per correct answer. Your group earnings will be divided evenly between each partner ensuring you each get equal pay.

You will not know who you are a partner with.

You will not know how many problems you answered correctly until the end of the 4 minutes.

At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below, you will be taken to a screen to make your payout method choices and confirm that you understand.

The experiment will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Task 5

Every task that you have done so far has involved no feedback about other's performance. Now we will provide you with feedback regarding all session participants' performance from Task 1.

Your performance: *Individual performance.*

Everyone's performance: *List of everyone's performance (high to low).*

You will now have the opportunity to choose the individual pay rate of $0.50 per correct answer, or to enter in a tournament, or to join a partner and receive the group pay rate.

Your possible choices are: *Group Pay, Individual Pay, 2 Person Tournament, 4 Person Tournament, 6 Person Tournament*

Whenever possible, you will be paired with other people that chose the same pay format as yourself. If there are not enough people for your chosen format, then individuals who have selected another pay format will be randomly selected to meet the participation requirements for your chosen format.

Tournament:

If you win the tournament, you will be paid the rate of $0.50 times the number of people in the tournament for each correctly answered problem.

In the event of a tie for first place, you will split your tournament winnings evenly with the number of people that tied for first in the tournament.

If you do not win the tournament then you will receive nothing.

You will not know who you are competing against.

Individual Pay:

You will be paid $0.50 for each correctly answered problem.

Group Pay:

For the Group Pay format, you will be joining one other person in a group.

You will be paid according to the productivity of your group. The number of correct answers by you and your partner will be added together and as a group you will be paid $0.50 per correct answer. Your group earnings will be divided evenly between each partner ensuring you each get equal pay.

You will not know who you are a partner with.

You will not know how many problems you answered correctly until the end of the 4 minutes.

At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below, you will be taken to a screen to make your payout method choices and confirm that you understand.

The experiment will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Appendix D.2 Word Task Instructions

General Instructions

In this experiment you will be performing a task five different times. The task will consist of giving you an alphabetical letter and having you type as many words as possible that begin with that letter in a 4 minute period.

To enter a word, you will simply type the word using the keyboard, then press enter or submit. You will then be able to enter another word. Using an English dictionary, the computer program will verify that the words you have spelled are correct. Misspelled words will be counted as incorrect and will not be included in your task total. During the task, you will see the words you have typed, but you will not know if you have spelled them correctly.

- Capitalization will not affect spelling.
- Duplicate words will be counted as incorrect.
- Proper names will be counted as incorrect.
- Common place names (cities, countries) will be counted as correct.
- Plurals and tense changes to root words will count as separate correct answers.

Everybody in the session will be given the same letter.

In some cases you will be asked to make decisions about how you will potentially be paid for your performance.

Only one of the five tasks will determine your payout for the experiment and it will be randomly chosen at the end.

To help with time management, there will be a clock counting down the seconds for the 4 minute duration.

Task 1

For Task 1, you will be paid $0.25 for each correctly spelled word during the 4 minute time limit.

You will be informed of how many words you spelled correctly at the end of the task.

At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen showing your payout method and confirming that you understand.

The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Task 2

For this task, you will be randomly placed in a tournament.

The tournament will have a size of *2 or 4 or 6* people, including yourself.

If you win the tournament, you will be paid $0.25 multiplied by the number of people in the tournament for each correctly spelled word you enter beginning with the designated letter.

For example, if you are in a 4 person tournament, then you will be paid $1.00 for each correctly spelled word so long as you win the tournament.

If you do not win the tournament, then you will receive nothing.

In the event of a tie for first place, you will split your tournament winnings evenly with the number of people that tied for first place in the tournament.

You will not know who you are competing against. Everybody in the session will be given the same letter.

You will be informed of how many words you spelled correctly at the end of the task.

At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen showing your payout method and confirming that you understand.

The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Task 3

For this task, you will be randomly put in a group with one other person.

The group has a total size of 2 people (including yourself) and group members will be paid an equal amount.

You will be paid according to the productivity of your group. The number of correctly spelled words by you and your partner will be added together and as a group you will be paid $0.25 per word. Your group earnings will be divided evenly between each partner ensuring you each get equal pay.

You will not know who you are a partner with. Everybody in the session will be given the same letter.

You will be informed of how many words you spelled correctly at the end of the task.

At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen showing your payout method and confirming that you understand.

The task will begin once everyone in this session is ready to begin.
If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Task 4

You will now have the opportunity to choose the individual pay rate of $0.25 per correctly spelled word, or to enter in a tournament, or to join a partner and receive the group pay rate.
Your possible choices are: *Group Pay, Individual Pay, 2 Person Tournament, 4 Person Tournament, 6 Person Tournament*
Whenever possible, you will be paired with other people that chose the same pay format as yourself. If there are not enough people for your chosen format, then individuals who have selected another pay format will be randomly selected to meet the participation requirements for your chosen format.
Tournament:
If you win the tournament, then you will be paid $0.25 multiplied by the number of people in the tournament for each correctly spelled word. In the event of a tie for first place, you will split your tournament winnings evenly with the number of people that tied for first in the tournament.
If you do not win the tournament then you will receive nothing.
You will not know who you are competing against.
Individual Pay:
You will be paid $0.25 for each correctly spelled word.
Group Pay:
For the Group Pay format, you will be joining one other person in a group.
You will be paid according to the productivity of your group. The number of correctly spelled words by you and your partner will be added together and as a group you will be paid $0.25 per correctly spelled word. Your group earnings will be divided evenly between each partner ensuring you each get equal pay.
You will not know who you are a partner with. Everybody in the session will be given the same letter. You will be informed of how many words you spelled correctly at the end of the task. At the end of the 4 minutes you will see a screen asking about your performance.
Do not discuss your performance with anyone else at anytime.
Once you press the button below you will be taken to a screen to make your payout method choices

and confirming that you understand.

The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.

Task 5

Every task that you have done so far has involved no feedback about other's performance. Now we will provide you with feedback regarding all session participants' performance from Task 1.

Your performance: *Individual performance.*

Everyone's performance: *List of everyone's performance (high to low).*

You will have the opportunity to choose the individual pay rate of $0.25 per correctly spelled word, or to enter in a tournament, or to join a partner and receive the group pay rate.

Your possible choices are: *Group Pay, Individual Pay, 2 Person Tournament, 4 Person Tournament, 6 Person Tournament*

Whenever possible, you will be paired with other people that chose the same pay format as yourself. If there are not enough people for your chosen format, then individuals who have selected another pay format will be randomly selected to meet the participation requirements for your chosen format.

Tournament:

If you win the tournament, you will be paid the rate of $0.25 multiplied by the number of people in the tournament for each correctly spelled word.

In the event of a tie for first place, you will split your tournament winnings evenly with the number of people that tied for first in the tournament.

If you do not win the tournament then you will receive nothing.

You will not know who you are competing against.

Individual Pay:

You will be paid $0.25 for each correctly spelled word you submit.

Group Pay:

For the Group Pay format, you will be joining one other person in a group.

You will be paid according to the productivity of your group. The number of correctly spelled words by you and your partner will be added together and as a group you will be paid $0.25 per word. Your group earnings will be divided evenly between each partner ensuring you each get equal pay.

You will not know who you are a partner with.

Everybody in the session will be given the same letter. You will be informed of how many words you spelled correctly at the end of the task. At the end of the 4 minutes you will see a screen asking about your performance.

Do not discuss your performance with anyone else at anytime.

Once you press the button below you will be taken to a screen to make your payout method choices and confirming that you understand. The task will begin once everyone in this session is ready to begin.

If you feel like you understand the instructions and are ready to begin then press the "I understand" button below.