IS THE GENDER DIFFERENCE IN COMPETITIVE BEHAVIOR HISTORY DEPENDENT?

Elaine Rhee*†, Charles Noussair ‡

Abstract

This study tests whether men and women differ in their willingness to challenge a competitor in response to a prior transgression. A laboratory experiment is conducted, in which a player can choose to behave unfairly toward another. The other player may then challenge the first to a contest. We investigate the extent to which previous interactions can explain individual differences in tournament initiation decisions. The results show that men, but not women, tend to challenge a competitor more when the prior outcome is unfair and the unfairness occurred through the competitor's intentional choice. In contrast, unfair outcomes that occur by chance do not influence the decision to challenge others.

JEL Classification: C9. D9, J1

Keywords: Competition, gender gap, unfair, history, intentions, emotions

*University of Arizona, Department of Economics. Email: rhee@email.arizona.edu

†We would like to express our appreciation to Julian Romero, Subhasish Chowdhury, Jin Sohn, participants at the 2020 Reading Experimental and Behavioural Economics Workshop on Contests: Theory and Evidence, and at the 2020 Economic Science Association World Meeting, for their comments. We thank Daniel Gotsman, Pavas Gupta, and Kierstin Seaback for research assistance. The research reported in this paper was supported by the Economic Science Laboratory and the Graduate and Professional Student Council of The University of Arizona [R0419-17].

‡University of Arizona, Department of Economics. Email: cnoussair@arizona.edu

I INTRODUCTION

What makes an individual want to challenge another to a contest? Presumably, the perceived probability of winning, the value of winning the contest, and the effort required to compete are important determinants. However, it is also likely that the prior history of interaction with the potential competitor might also influence the decision to challenge. It is very common for employees of law firms, real estate firms, or medical groups to leave their organization, branch out on their own, and try to draw away the clients from their former firm, in effect initiating a challenge to the former firm. A firm might decide to launch a price war or a patent race with a competitor whom they have interacted with many times previously. Professional athletes in some sports, such as boxing or golf, challenge others to matches[1]. They often have a history of past matches, and may have off-the-field friendships or feuds. Candidates choosing to contend for elections typically have had some form of professional or personal contact with their opponents prior to the campaign.

There is an extensive literature in economics investigating the correlates of the decision to *enter* a tournament against other individuals, a different but closely related topic. A number of laboratory experiments have been conducted specifically to establish such correlates (Niederle and Vesterlund, 2007; Balafoutas et al., 2012; Buser et al., 2017; Gneezy and Rustichini, 2004; Buser, 2012; Cárdenas et al., 2012). This literature has focused on characteristics, traits or states of the individual competitors, such as gender, distributional preferences, hormone levels, and culture. A number of authors have observed that behavior in experiments in which individuals can select into competition correlates with eventual labor market outcomes (Buser et al., 2015; Reuben et al., 2015), and investment decisions (Berge et al., 2015). A principal finding reported in this literature is that men have a greater tendency to select into a tournament than do women who are equally able. This gender difference is sensitive to whether the rewards are monetary (Healy and Pate, 2011), whether they are shared with others (Cassar and Rigdon, 2021), and other factors. To our knowledge, however, the history of prior interaction has not been studied as a potential determinant of the decision to enter a tournament.

While the literature described above focuses on the voluntary decision of whether or not to enter the tournament, we consider here the decision to challenge a competitor to a contest that the other party is then compelled to enter. The determinants of this decision, to our knowledge, have not been investigated previously using experimental methods, other than in a recent study by Charness et al. (2018). In a paper focused on strategic announcements of confidence, they include two treatments, Deter and Lure, in which one player can compel another to enter a tournament. They find that individuals are more likely to initiate a tournament the more confident they are about winning the tournament and the less confident is the opponent.

[1]Well known examples include Oscar de La Hoya's challenging of Conor McGregor to a boxing match in 2020. Khabib Nurmagodenov similarly challenged Floyd Mayweather in 2018. The golfers Tiger Woods and Phil Mickelson challenged each other through multiple repetitions of The Match: Tiger vs. Phil.

Specifically, we study, using a laboratory experiment, how the history of interaction between individuals can affect their willingness to challenge each other to a competition. The first question considered is how a prior opportunistic action by one individual that hurts another affects the willingness of the aggrieved party to challenge the offending individual. A decision to punish an offender by forcing them into a contest can be rationalized by social preference models that incorporate intentions and reciprocity, such as Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006).

A number of studies are related to ours in that they have investigated how prior activity within a competitive interaction affects subsequent competitiveness. Buser and Yuan (2019) consider how experience affects entrants in the Dutch Math Olympiad. They observe that women are more likely than men to stop competing if they lose a round of the competition. They also find that positive feedback makes men more likely to enter and stay in a competition, while it does not have a similar effect on women. In other work along these lines, Wozniak et al. (2014) find that detailed relative performance feedback reduces the gender gap in the choice of payment scheme. De Paola and Scoppa (2017) find that female, but not male, players perform worse after losing the first set of a tennis match. However, Wozniak (2012), reports a similar effect of prior tournament outcomes on entry into future tournaments for male and female tennis players. Kessel et al. (2021) observe that merely making participants aware of a gender difference in the willingness to compete reduces the difference.

The second question that we investigate is whether there is a difference in the behavior of men and women. We consider whether there are gender differences in the propensity to challenge others, and whether a history of prior interaction affects women and men differently. As mentioned earlier, number of experimental studies, beginning with Niederle and Vesterlund (2007), have found that men select into competition at a higher rate than women. Bertrand and Hallock (2001) showed that women account for only 2.5 percent of the five highest paid executives within U.S. firms. It has been argued that gender differences in preferences for competition can explain, at least in part, this lopsided gender ratio. If there are gender differences in the willingness to compete, it may be that men are also more likely to try to initiate a competition with another party. We consider whether being treated unfairly widens or narrows any gender gap in the tendency to compete. For example, if mistreatment causes men to increase their propensity to compete, but does not have the same effect on women, mistreatment would increase the gender gap in competitive behavior. If history makes a difference in such a manner, it constitutes an additional potential determinant of the inequality in labor market outcomes.

The third research question is whether the emotion of anger affects the decision to initiate a tournament. Individuals who become angry from previous interaction with another party may become more likely to challenge them. By employing FaceReader^{TM}, a software package that identifies the presence of a number of emotional states – neutrality, sadness, happiness, surprise,

fear, disgust, anger, and arousal – we study the role of emotions as correlates of the effect of history on preferences for competition. Emotions have been linked to many types of economic behavior, such as risk aversion (Nguyen and Noussair, 2014), risk assessments (Lerner and Keltner, 2001), productivity (Oswald et al., 2015), asset pricing bubbles (Andrade et al., 2016), etc. However, to our knowledge, this is the first study to examine how emotions affect one's decision to compete.

The experimental design also allows us to measure performance in the competition, and thus to investigate how performance is affected by history and emotion, and how the pattern may differ by gender. We employ the paradigm used by Niederle and Vesterlund (2007), in which participants perform a simple arithmetic task for which they are compensated according to the payment scheme chosen. However, in our design, one individual can challenge the other to a tournament, and compel the other party to compete. Prior to this choice of payment structure, two randomly matched participants play a dictator mini-game. One member of the pair is a dictator who can take all of an endowment provided to be shared between the two players. Afterwards, the receiver can choose one of two payment schemes for the arithmetic task: (1) a piece rate or (2) a tournament where their score in the task is compared to the dictator they were matched with. Our experimental design also allows us to obtain additional information that might correlate with one's tournament initiation decision, including the subjects' confidence level, distributional preferences, and risk tolerance.

We obtain a number of findings. First, an unfair outcome resulting from the competitor's choice causes men, but not women, to be more likely to choose a contest against that competitor. Second, when an unfair outcome occurred by accident (a computer's random draw), unfairness has no effect on the likelihood of challenging the beneficiary of the random draw. Third, when a fair outcome is intentional, there is no gender gap in the propensity to challenge. Fourth, emotional state does not correlate with the likelihood of challenging others, after the effect of unfairness is taken into account.

The rest of the paper is structured as follows: In Section 2, we describe the experimental design and hypotheses. In Section 3, we report the results. There, we discuss how tournament initiation decisions and performance are affected by interaction history, how distributional preferences are associated with behavior, and how emotions correlate with decisions. In Section 4, the conclusions are stated and summarized.

II EXPERIMENTAL DESIGN AND HYPOTHESES

Each session of the experiment consisted of four parts: (1) Practice, (2) History and Tournament, (3) Distributional Preference, and (4) Risk Attitude. The Practice phase gave the subjects the opportunity to familiarize themselves with the task that they had to perform in the History and Tournament part. The History and Tournament part consisted of four periods, and each period in turn was composed of two stages, History and Tournament. This was the principal portion of the experiment, and was designed to address the main research questions. The Distributional

Preference and Risk Attitude phases obtained measures of additional possible correlates of the tournament initiation decision.[2] There were two treatments, the Dictator Game and the Random treatment.

Random draws determined which decisions in the experiment counted toward participant earnings. To determine which part/stage/period of the experiment counted, two volunteers among the subjects were chosen at the end of the session, one to roll a ten-sided die, and the other to witness the roll. The probability of each stage being chosen was not the same. The History and Tournament stages each had a 40% chance, and the Distributional Preference and Risk Attitude parts each had a 10% chance, of being paid out. If History or Tournament was chosen for payment, an additional random draw determined which of the four periods counted. In addition to the payment for the randomly selected stage, all subjects received an additional payment of \$1 per correct guess in a questionnaire administered in the Tournament stage. Thus, the total payment was \$5 (show-up fee) + (one randomly selected part/stage/period when applicable) + \$1×(number of correct guesses). By not allowing for both the History and Tournament stages to be paid out, we can eliminate inequality aversion as a motive for initiating a tournament.[3]

At the beginning of the session, subjects took an unpaid comprehension quiz consisting of six questions regarding the payment structure. No participant answered more than one question incorrectly. Each part of the session was introduced and explained after the previous one was completed.

A. Practice

In this part of the experiment, participants practiced the addition task. They had three minutes to solve addition problems, in which they were required to sum five randomly generated two-digit numbers. They were asked to solve as many problems as possible. Participants were not allowed to use calculators, but were given scratch paper. They were informed at the end of the task of how many questions they answered correctly. They did not receive any information about the other participants' performance. This practice part could not count toward the subjects' payment.

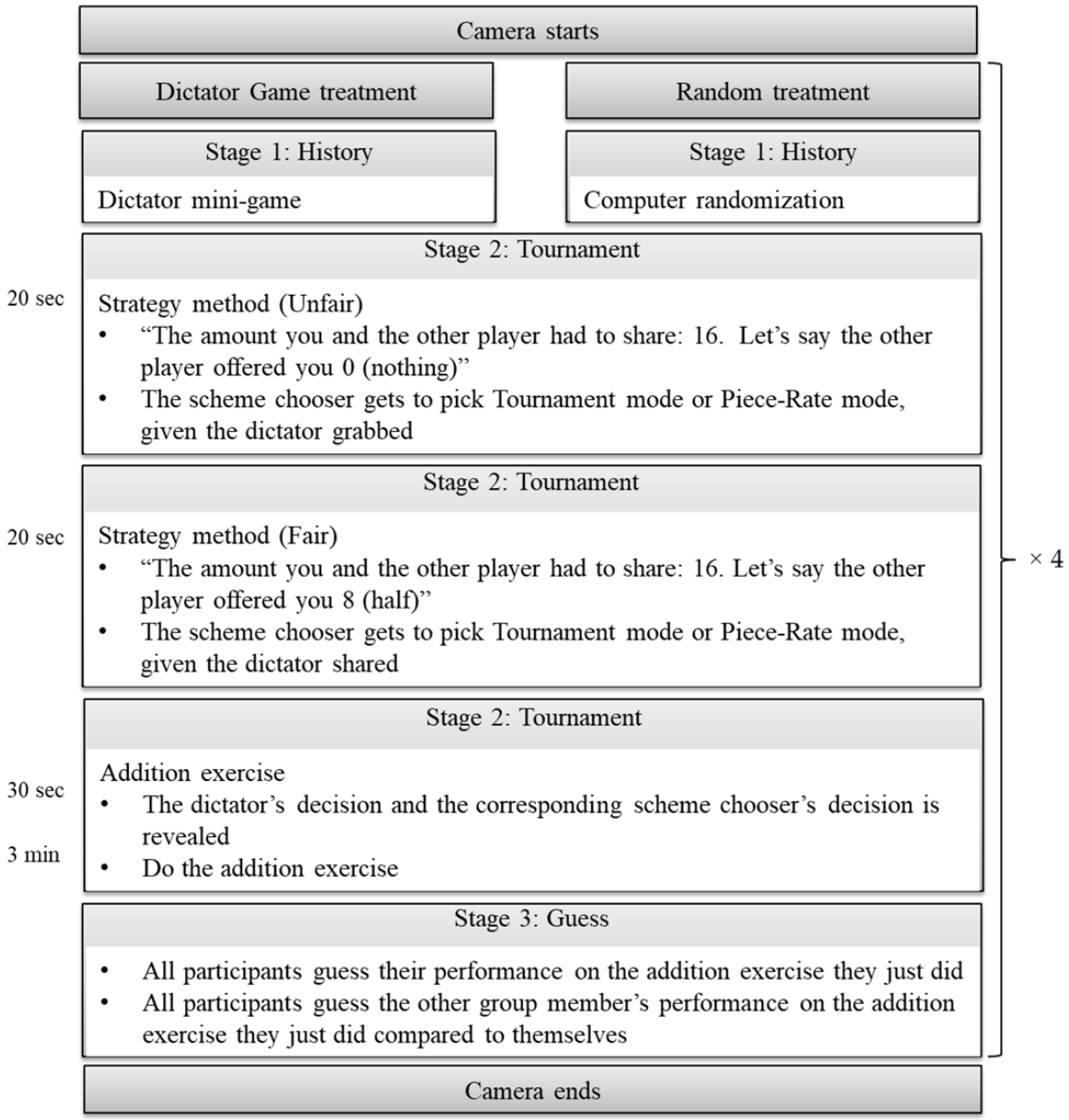
B. History and Tournament

The History and Tournament part consisted of four periods. Each period had three stages. In between the repetitions, there was a mandatory 30-second pause to allow participants to take a break. All participants were videotaped, with the event timing set with the Mucap software package (Doyle and Schindler, 2019). The videos were analyzed later with Noldus FaceReader^{TM}.

[2]One of the risk attitude measures was gathered at the beginning of the session and is thus unaffected by the subsequent activity in the session. We rely primarily on this measurement in our analysis.

[3]The tournament cannot be used to reduce ex-post inequality in payoffs. If the tournament is counted toward final earnings, it means that the History stage does not count, and current earnings are equal at the beginning of the tournament.

FIGURE 1
Timing within History and Tournament Part of a Session



Note: The table outlines the timing of events within the History and Tournament Part of the experiment for both treatments. Time flows from top to bottom in the figure. The only differences between the Dictator Game and Random treatment occur in the first (History) stage. In the Tournament stage, the scheme chooser first makes a strategy method choice between tournament or piecerate in response to an unfair and a fair allocation (the order of these two choices is counterbalanced). Then the tournament is conducted. To conclude the stage, participants guess their own and their matched player's performance. The timing of an entire experimental session is illustrated in Figure 2.

Each subject was randomly paired with another. Within each pair, one was assigned to the role of a dictator, and the other to the role of a scheme chooser. Throughout this part of the experiment an individual's role did not change, but subjects were randomly re-matched every period.

Stage 1: History. In the Dictator Game treatment, a simple dictator mini-game was conducted in the History stage. There were \$16 to divide between the two members of each pair. Each group's dictator had to choose between taking the whole \$16 (which we call here an *unfair* decision), or sharing it with the scheme chooser equally with each receiving \$8 (which we refer to as a *fair* choice). The dictator's decision was disclosed to the scheme chooser right before the Tournament stage. The stark setting, with only two possible divisions, equality and maximum inequality, makes it obvious which decision is considered "fair" and which is "unfair".

In the Random treatment, the dictator's decision was made by the computer randomly. The computer randomly decided whether the dictator received all \$16 or to shared it equally with the scheme chooser, with a 0.5 probability of each choice. The computer's random choice was revealed right before the addition exercise in the Tournament Entry stage. By comparing results from the Random treatment, where the dictator has no agency, to those of the Dictator Game treatment, the effect of dictator agency on the scheme chooser's response can be measured.

Stage 2: Tournament. Each scheme chooser was asked to choose between *Tournament mode* or *Piece-rate mode* with the matched dictator. The strategy method was employed. Before learning the outcome of the History stage, the scheme choosers had to decide which mode to implement with the matched dictator in response to the unfair dictator decision, and once again in response to the fair dictator decision. The sequencing of the two decisions was counterbalanced, with one-half of scheme choosers responding to a fair decision before an unfair one, and the other half doing so in the opposite sequence.

Under *Tournament mode*, the performance on the task was evaluated in comparison to that of the other group member. The group member who answered more questions correctly received \$2 per correct answer, and the other received nothing. If there was a tie, the winner was randomly chosen. Under *Piece-Rate mode*, the group members did not compete. Their payment was completely independent of the other group member's performance. Each received \$1 per correct answer.

To allow enough time for emotions to emerge in response to the dictator's decisions, the scheme choosers had to view a page that only showed "The amount you and the other player had to share: 16. Let's say the other player offered you 0 (nothing)", for 20 seconds. After that, they were able to choose which payment scheme to adopt with the matched dictator for that situation. Subsequently, the subjects viewed another page that only showed "The amount you and the other player had to share: 16. Let's say the other player offered you 8 (half)" for 20 seconds, and then they were able to pick the group's payment scheme for that situation. The emotional state data is retained from

the two 20-second segments when the scheme choosers were looking at each potential situation.[4]

After the scheme choosers made both choices, the dictator's decision and the corresponding scheme chooser's decision were revealed to both group members. At that point, the scheme choosers knew what they actually received from the dictator, and the dictators knew what mode was selected for the upcoming addition task. Then, both dictators and scheme choosers did the addition exercise for three minutes. They were not informed of their own or the competitor's performance at any time.[5]

Stage 3: Guess. The subjects were then asked how many questions they thought they answered correctly in the addition exercise they just performed. They were also asked to guess how many more/fewer questions the other group member answered correctly compared to their own score for the same exercise. This provides two measures of confidence. *GuessWin* is defined as a dummy variable that takes the value 1 when the subject guesses that the other group member got fewer questions correct than they did (Balafoutas et al., 2012). The *Overconfidence index* is defined as the guessed number of correct answers minus actual performance (Gillen et al., 2019). If the subject expected to get more (fewer) questions correct than they actually did, the overconfidence measure is positive (negative). We use the Guesswin measure to identify those individuals who believed that they would win the tournament.

C. Distributional Preference

The scheme choosers' distributional preference was measured in this part of the experiment. The elicitation procedure of the distributional preferences was based on Balafoutas et al. (2012), and Kerschbamer (2015).[6]

The scheme choosers were exposed to a series of binary choices between allocations that specified both an own payoff for herself and a payoff for another participant who had been a dictator in the History and Tournament part of the session. The scheme chooser then had to pick their preferred allocation. In each binary decision problem, one of the two allocations was symmetric – giving \$8 to each person – while the other one was asymmetric, involving unequal payoffs for the two subjects (Table 1). In the first five binary choices, the payoff of the matched dictator was higher than the scheme chooser. In the last five binary choices, the payoff of the scheme chooser was higher than the other party. If this part of the session was chosen to count toward participant earnings, only one randomly chosen decision (one line in Table 1) counted.

[4] We also consider the emotional state of scheme choosers at the time they found out the actual decision that the dictator made. The results are reported in Table A2 of Appendix A.

[5] Dictators knew that the tournament stage was coming up after the dictator game. It is possible that dictators might act strategically to try to avoid the tournament by choosing the fair allocation. While this is a possibility, it does not alter the main analysis of this paper, which concerns the behavior of scheme choosers rather than dictators.

[6] A number of other measures could have been used, such as the Social Value Orientation (SVO) protocol (Messink, 1968). However, the task we employed has been linked to competitive behavior in previous work (Balafoutas et al. (2012)) and is much simpler to administer than SVO.

TABLE 1
Choices in the Distributional Preference Elicitation Task

Task	LEFT	RIGHT
1	□ You: $5 Other: $10	□ You : $8 Other: $8
2	□ You: $6 Other: $10	□ You : $8 Other: $8
3	□ You: $7 Other: $10	□ You : $8 Other: $8
4	□ You: $8 Other: $10	□ You : $8 Other: $8
5	□ You: $9 Other: $10	□ You : $8 Other: $8
6	□ You: $5 Other: $6	□ You : $8 Other: $8
7	□ You: $6 Other: $6	□ You : $8 Other: $8
8	□ You: $7 Other: $6	□ You : $8 Other: $8
9	□ You: $8 Other: $6	□ You : $8 Other: $8
10	□ You: $9 Other: $6	□ You : $8 Other: $8

Note: This is the task that the scheme choosers complete in the Distributional Preference part of the session. This task is based on Balafoutas et al. (2012) and Kerschbamer (2015). Individuals make ten pairwise choices between allocations to themselves and another player. One of the decisions, chosen randomly at the end of the session, has the possibility of counting toward earnings. The choice pattern can identify a distributional preference type.

When faced with the first five tasks, a rational subject who is (at least weakly) benevolent decides for the asymmetric allocation for the first time in the third choice or earlier, while switching later is inconsistent with weak benevolence (and therefore counted as malevolence). Similarly, when faced with the last five tasks, a rational subject who is (at least weakly) benevolent decides for the asymmetric allocation for the first time in the fourth choice or later. Switching earlier is inconsistent with weak benevolence, and is therefore counted as malevolence (Balafoutas et al., 2012).

Following Balafoutas et al. (2012), we refer to a decision maker who is benevolent in both the first five and last five tasks as "Benevolent no matter what (BN)". A decision maker who is benevolent only in the first five tasks is "Benevolent only when behind (BB)", while one who is benevolent only in the last five tasks is "Benevolent only when ahead (BA)". A decision maker who is malevolent in both cases is "Malevolent (M)".

D. Risk Preference

To acquire the risk preferences, two Multiple Price Lists (MPL), similar to that used in Holt and Laury (2002) are used. The two MPLs were placed in different parts of the experiment to increase the isolation of the two tasks from each other. The first MPL, $Risk12$, was administered at the very beginning of the session. The second MPL, $Risk8$, was done at the very end of the session. In each MPL, the subjects were exposed to a series of binary choices between a gamble and a safe payoff (Tables 2 and 3). The gamble remained the same for the entire task: a 50% chance of receiving $12 ($8) and a 50% chance of $0 for $Risk12$ ($Risk8$). On the other hand, the

safe payment increased in increments of \$1 from \$0 to \$12 (\$8) for $Risk12$ ($Risk8$)[7].

TABLE 2
Choices in the Risk Preference Elicitation Task ($Risk12$)

Task	LEFT	RIGHT
1	□ \$0 for sure	□ 50% chance \$0 & 50% chance \$12
2	□ \$1 for sure	□ 50% chance \$0 & 50% chance \$12
3	□ \$2 for sure	□ 50% chance \$0 & 50% chance \$12
4	□ \$3 for sure	□ 50% chance \$0 & 50% chance \$12
5	□ \$4 for sure	□ 50% chance \$0 & 50% chance \$12
6	□ \$5 for sure	□ 50% chance \$0 & 50% chance \$12
7	□ \$6 for sure	□ 50% chance \$0 & 50% chance \$12
8	□ \$7 for sure	□ 50% chance \$0 & 50% chance \$12
9	□ \$8 for sure	□ 50% chance \$0 & 50% chance \$12
10	□ \$9 for sure	□ 50% chance \$0 & 50% chance \$12
11	□ \$10 for sure	□ 50% chance \$0 & 50% chance \$12
12	□ \$11 for sure	□ 50% chance \$0 & 50% chance \$12
13	□ \$12 for sure	□ 50% chance \$0 & 50% chance \$12

Note: This is the task that all of the participants had to complete in the very beginning of the session. This Multiple Price List was used to acquire the participants' risk preference data. Participants make 13 pairwise choices between a certain payment and a risky lottery. One of the decisions, chosen randomly at the end of the session, could count toward final earnings.

E. Emotions

The videos recorded in the History and Tournament part of the session were analyzed with the Noldus FaceReader^{TM} software. FaceReader is an automated system for the recognition of the six basic or universal emotions (Ekman and Keltner, 1997) – happiness, sadness, anger, surprise, fear, and disgust. It also provides measures of neutrality and arousal. It has been validated as effective in identifying the intended emotion of an individual (Den Uyl and Van Kuilenburg, 2005), and strongly correlates with observers' evaluations (Terzis et al., 2010). It reports the strength of each emotion on a scale from 0 to 1. In our analysis, we use the average emotional response during the 20 seconds when the scheme chooser viewed the page where it was assumed the dictator was fair or unfair and the participant chose the incentive scheme, as well as the 20 seconds subsequent to the scheme chooser learning the dictator's actual choice.

[7]The *risk tolerance measure* – $Risk12$ or $Risk8$ – was computed as the row number in which the participant first switches to the safe payoff, minus 0.5, then divided by 13 for $Risk12$, or 9 for $Risk8$. This index ranges from 0 to 1, with a higher value corresponding to a lower degree of risk aversion. For example, if a participant switches to the safe payoff at task 4 for the first time in $Risk12$, the value of the measure is (4-0.5)/13 = 0.269.

TABLE 3
Choices in the Risk Preference Elicitation Task ($Risk8$)

Task	LEFT	RIGHT
1	□ $0 for sure	□ 50% chance $0 & 50% chance $8
2	□ $1 for sure	□ 50% chance $0 & 50% chance $8
3	□ $2 for sure	□ 50% chance $0 & 50% chance $8
4	□ $3 for sure	□ 50% chance $0 & 50% chance $8
5	□ $4 for sure	□ 50% chance $0 & 50% chance $8
6	□ $5 for sure	□ 50% chance $0 & 50% chance $8
7	□ $6 for sure	□ 50% chance $0 & 50% chance $8
8	□ $7 for sure	□ 50% chance $0 & 50% chance $8
9	□ $8 for sure	□ 50% chance $0 & 50% chance $8

Note: This is the task that all of the participants had to complete in the very end of the session. This Multiple Price List was used to acquire risk preference data of participants. Participants make 9 pairwise choices between a certain payment and a risky lottery. One of the decisions, chosen randomly at the end of the session, could count toward final earnings.

F. Implementation

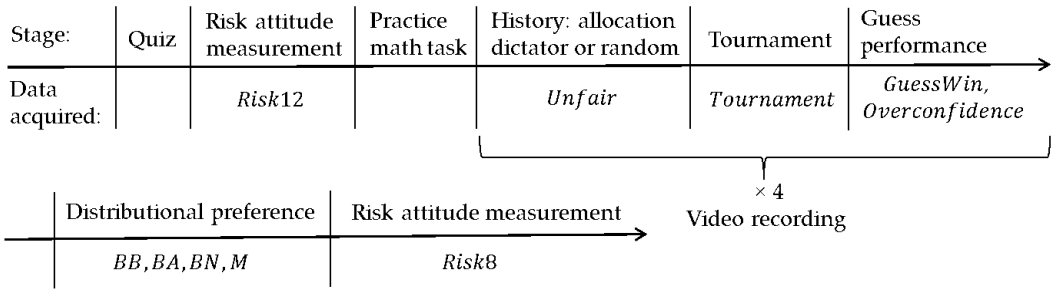
The experiment was conducted at the Economic Science Laboratory at the University of Arizona. All subjects were undergraduate students at the university. Each subject participated in only one treatment and received a $5 show-up fee. The experiment employed the z-tree platform (Fischbacher, 2007). On average, the sessions lasted 75 minutes. 124 people (65 males, 58 females, and 1 other) participated in the Dictator Game treatment, and 102 people (64 males, 37 females, and 1 other) participated in the Random treatment.

The experiment was implemented in the order shown in Figure 2. In Figure 2, each vertical line indicates when instructions were given[8]. Under the horizontal axis are the variables acquired from the data in the stage indicated above. The experiment started and ended with a Risk Attitude measurement, so that the two MPLs were separated in time. After the first Risk Attitude measurement, the Practice, History and Tournament, Distributional Preference, and second Risk Attitude measurement procedures were implemented in order.

Table 4 reports the average values of the risk aversion measures, and the confidence measures, by treatment. The table shows that there is no evidence that the mean of any variable differs between the two treatments (Wilcoxon rank sum tests and t tests were conducted). Therefore, the randomization into treatments was unbiased.

[8]The exception was that the instructions for both the History and Tournament stages were given before the History stage.

FIGURE 2
Timing of Events within a Session



Note: The timing of an entire session is described in the figure. The vertical lines indicate when new instructions were given. The session began with general instructions and a quiz. There followed the *Risk12* risk preference measurement task, and a practice period of the math task. The four periods of the History and Tournament stage came next. Instructions were given only before the first of the four periods. This was followed by the distributional preference measurement task and the second risk preference task. Video recordings were retained for the entire duration of the History and Tournament stage.

TABLE 4
Summary Statistics of Confident Scheme Choosers' Control Variables

Treatment	Risk12	Risk8	Overconfidence
Dictator Game	0.48	0.54	0.63
	(0.15)	(0.15)	(1.93)
Random	0.52	0.53	0.70
	(0.15)	(0.12)	(1.69)

Note: Averages given with standard deviations in parentheses. Risk12 and Risk8 range from 0 to 1, with higher values indicating more risk tolerance. $Overconfidence$ is equal to the guessed number of correct answers minus actual performance. The data only include scheme choosers who guessed that they got more questions correct than the matched dictator. t-tests of the differences in the variables between treatments yield no significant differences at $p < .05$.

G. Hypotheses

The first hypothesis considers how the decisions of scheme choosers depend on the prior history of interaction between the two parties. Reciprocal behavior is a well-established empirical phenomenon and constitutes the basis of Hypothesis 1. Unfair history would cause one who believes she is likely to win to prefer tournament to piece-rate mode, because the tournament can be used as a punishment against a dictator who was unkind. The punishment takes the form of lowering the dictator's payoffs to below what she would earn under a piece rate scheme, as well as potentially having to expend additional effort to try to outperform the scheme chooser. This behavior can only be considered as a punishment by people to believe that they can outperform the competitor in the tournament. Therefore, the hypothesis applies only to scheme choosers who are confident of winning their tournament.

Hypothesis 1: In the Dictator Game treatment, a scheme chooser who is confident of winning is more likely to choose the tournament after an unfair history than after a fair history.

The second hypothesis is that the effect of unfair history on the likelihood of choosing the tournament is specific to the Dictator Game treatment and does not exist under the Random treatment. Under reciprocity models (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk et al., 2008), it is the intention of one party that triggers reciprocation by another party. Since the unfair event happens through the competitor's choice in the Dictator Game treatment but not in the Random treatment, we hypothesize that the unfair history effect appears only in the Dictator Game treatment. In other words, people would want to challenge a competitor who caused an unfair event, but not one who benefited from, but did not cause, the event. Again, this hypothesis is applicable only to those who are confident of winning the tournament.

Hypothesis 2: In the Random treatment, there is no difference in the incidence of tournament choice of scheme choosers who are confident of winning between a fair history and an unfair history.

The third hypothesis is that there would be a gender difference in the effect of unfair history on tournament initiation, with men showing a greater effect of history than women. This hypothesis is justified by some preceding literature. Females have been found to be more forgiving than men (Root and Exline, 2011), as well as less vengeful (Miller et al., 2008). Since females are more forgiving, we hypothesize that women would use tournament mode as a punishment device less commonly than men would.

Hypothesis 3: The difference between the incidence of tournament initiation on the part of confident scheme choosers between unfair and fair histories in the Dictator Game is greater for men than for women.

The fourth hypothesis concerns the emotional underpinnings of the choice to challenge an individual to a tournament. We view the launching of the tournament as an act of negative reciprocity. Anger is often a correlate of negatively reciprocal behavior, such as rejecting an unfair offer in an ultimatum game (van Leeuwen et al., 2018) or applying punishment in a social dilemma (Hopfensitz and Reuben, 2009). The hypothesis is that anger is associated with the decision to launch a tournament in response to unfair dictator behavior. Those who get angrier in response to an unfair allocation would be more likely to initiate a tournament.

Hypothesis 4: Anger is a determinant of the decision to launch a tournament in response to

an unfair dictator decision.

III RESULTS

A. Summary of the Data

The percentage of scheme choosers opting for tournament or piece-rate mode after each history, by treatment, is displayed in Figure 3. The the figure, TT indicates a choice of tournament after both a fair and unfair allocation, and PP a choice of piece-rate in both instances. TP is a choice of tournament after an unfair, and piece-rate after a fair, history, while PT is the opposite pattern. The upper panels display the data for all scheme choosers, and the lower two panels the data for confident scheme choosers only. In the Dictator Game treatment, the most common strategy was to launch a tournament after an unfair allocation, and to adopt a piece-rate scheme after a fair division. In the Random treatment, the most common strategy was to choose piece-rate mode regardless of history. The strategy TT is more popular and PP less popular among confident scheme choosers than for the overall sample, presumably reflecting the fact that confident people feel that they are relatively likely to win a tournament.

Table 5 shows the overall percentages of scheme choosers initiating tournament by gender, treatment, and history. The percentages for confident scheme choosers only are given in parentheses. In all comparable cases for the full sample, men choose tournament at least as often as women. The same pattern holds for confident scheme choosers, with the exception of the reaction to a fair allocation in the Random treatment. In the Random treatment, pooling the data from fair and unfair trials, there is a gender gap. 45.1% of men initiate a tournament, while 30.0% of women do. The difference is significant according to a two-sample test of proportions ($z = 3.06$, $p = .002$). The gap is in the same direction but is not significant when only confident scheme choosers are considered. It must be emphasized, however, that we are measuring a type of competitiveness that is different from most of the literature in which the debate about whether there is a gender gap is taking place. This common paradigm, used for example by Niederle and Vesterlund (2007) studies the tendency to enter a tournament voluntarily against other parties. Here, we are studying individuals challenging others to a competition, which the others are then compelled to enter.

In all cases, an unfair outcome in the history stage leads to more tournament choice than a fair outcome. Unfair outcomes lead to more tournament choice when the outcome was intentional than when it occurred randomly. In response to a fair outcome, more male scheme choosers select a tournament under Random than in the Dictator Game treatment, and this may reflect more positive regard for those dictators who actively made the decision to be fair. In all cases, confident scheme choosers are at least as likely to initiate a tournament as the over overall sample. [9]

[9]The correlation between GuessWin, the dummy variable that equals 1 if a dictator believes that they would

Table A11 in the Appendix presents some of the common textual responses of scheme choosers when asked to explain their tournament initiation decisions.[10] The biggest difference between the two treatments is that those who chose tournament after an unfair history and piece-rate after a fair history used the launching of a tournament as a punishment in the Dictator Game treatment, whereas it was used as more of a chance to reduce expected payoff differences under Random. In other words, reciprocity is often mentioned in the Dictator Game treatment (40.39%), but rationales equalizing expected payoff are more common in the Random treatment (27.47%). Those who chose tournament mode for both histories commonly indicated that they were either confident of winning (29.31% in the Dictator Game and 44.07% in the Random treatment), or thought of it as an opportunity to earn more money (20.69% and 22.03%). Those who chose piece-rate for both histories most commonly said that they preferred a guaranteed payoff.

Table 6 shows the choices of dictators, by gender. Overall 38% of choices were unfair and 62% were fair. Men were more likely to choose a fair allocation than women, 72.2% to 47.1%. Table 7 presents individuals' distributional preferences, and breaks the data down by gender. The most common profile was benevolent only when ahead (BA), followed by BN. The least common distributional preference was benevolent only when behind. The patterns are consistent with Balafoutas et al. (2012) and received models of inequality aversion. Significant gender differences in distributional preferences did not exist.

Table 5
Percentage of Individuals Initiating Tournament Mode, by Treatment, History and Gender

Gender	Male		Female	
Treatment / History	Dictator Game	Random Treatment	Dictator Game	Random
Unfair	0.83 (0.83)	0.48 (0.57)	0.59 (0.62)	0.34 (0.46)
Fair	0.29 (0.34)	0.42 (0.48)	0.29 (0.43)	0.26 (0.44)

Note: The table shows the percentage of scheme choosers initiating tournament subdivided by treatment, history and gender. All of the scheme choosers' decisions are included. The numbers in the parentheses are only from decisions of scheme choosers who guessed that they got more questions correct than the matched dictator.

win a tournament, and choosing an Unfair allocation is 0.2292 and significant (p-value=0.003). This means that those who thought that they were good at the task were more likely to risk the increased likelihood that they would be forced to compete following an unfair decision. This correlation does not affect our analysis, since the behavior of scheme choosers rather than dictators is the primary focus of the study. If scheme choosers were aware of this correlation, it would narrow the gap between the response to unfair vs. to fair dictator behavior in the Dictator Game treatment. Thus, the possibility of this correlation biases us toward finding smaller effects and nevertheless, we find that the effect is significant.

[10]We had three participants from our subject pool, who were not familiar with the purpose of the study, code the statements as belonging or not to the categories listed in Table A11. If two of the three coders agreed that the statement fell into one of the categories, we classified it as belonging to that category.

TABLE 6
Number and Percentage of Fair and Unfair Dictator Allocation, by Gender

	Male	Female	Total
Unfair	40	55	95
	(0.16)	(0.22)	(0.38)
Fair	104	49	153
	(0.42)	(0.20)	(0.62)
Total	144	104	248
	(0.58)	(0.42)	(1.00)

Number of instances of unfair and fair decisions by dictators in the Dictator Game treatment. Males and females tabulated separately. The percentages of each choice are given in parentheses.

TABLE 7
Distributional Preferences of Scheme Choosers by Gender, Number of Observations of Each Type of Preference

	Male	Female	Total
Benevolent only when behind (BB)	3	3	6
Benevolent only when ahead (BA)	29	28	57
Benevolent no matter what (BN)	18	13	31
Malevolent (M)	11	8	19

Number of scheme choosers exhibiting each of four types of distributional preference. The data are tabulated separately by gender. Benevolent (malevolent) individuals prefer to increase (decrease) the other player's payoffs. Behind (ahead) refers to current earnings being lower (greater) than the other players'.

B. When an individual is unfair, are they challenged more?

To consider this question, we estimate Equation (1) using a Linear Probability (LPM) specification. Only the scheme choosers who were confident that they performed better than the matched dictator are included in the estimations. In the equations, i denotes scheme chooser, j the matched dictator, and t the period. $Tournament_{it}$ is a dummy variable that equals one when scheme chooser i picks tournament mode in period t. $Unfair_{jt}$ is a dummy variable that equals one when the response is for the case where the matched dictator j chose unfairly in the dictator mini-game of period t. $Female_i$ is a dummy variable that takes value one when the scheme chooser i is a female. μ_i is a subject-level random effect. $Risk12$ is the risk tolerance measure. The greater the measure is, the more risk seeking one is. ε_{it} is the error term. X_{it} is a vector of control variables: risk attitudes, emotions, distributional preferences, etc. As in Balafoutas et al. (2012), the baseline distributional preference in all specifications is benevolent no matter what (BN). The regressions were run with a panel data specification with subject-level random-effects, and we exclude observations of benevolent only when behind (BB), due to the paucity of data. The regressions were conducted separately for each treatment. Probit estimates are given in the Appendix and yield very similar

results.

$$Tournament_{ijt} = \alpha + \beta_1 Unfair_{jt} + \beta_2 Female_i + \beta_3 Unfair_{jt} Female_i + \theta Risk12_i + \gamma X_{it} + \mu_i + \varepsilon_{it} \quad (1)$$

Table 8 shows the estimation results for the Dictator Game treatment. In all specifications in Table 8, the coefficient of $Female$ is indistinguishable from zero in both specifications. This means that the behavior of women in response to a fair outcome is no different from that of men (the baseline condition in the regression is the behavior or males after a fair allocation). The significantly positive coefficients of $Unfair$, at $p < .01$ in both specifications, reveal that males chose tournament mode significantly more when their opponent was unfair than when (s)he was fair. The significance of the negative coefficient of $Unfair * Female$ at the $p < .01$ level reveals a gender gap in the increment of tournament launching after the unfair event. A history with an unkind rather than a fair competitor induces a significantly stronger competitive response from men than from women.

A test of the significance of $\beta_1 + \beta_3$, that is of the joint effects of the interaction term $Unfair * Female$ and $Unfair$ shows whether female participants did not choose tournament mode significantly more after an unfair than after a fair allocation. This test is insignificant with, $p > .1$, indicating that confident women are not more or less likely to launch a tournament after an unfair than after a fair dictator decision. While the coefficients involved are substantial, the standard errors are also sizable. It may be the case that in a higher-powered study, we would observe that female participants would respond to an unfair allocation with a higher likelihood of tournament initiation. Indeed, as shown in Table A1, the effect is significant when only the data from period 1 is used in the estimation. The effect is also significant if the emotions at the time the outcomes become known are included in the estimation rather than at the time the tournament initiation decision is made (see Table A2), or if all scheme choosers are included in the specification (see Tables A5 and A7).

Table 8 also presents the LPM results for the Random treatment. In contrast to the Dictator Game treatment, in the Random treatment, the effect of unfair history ($Unfair$) is indistinguishable from zero. Thus, even when distributional preferences are controlled for, fairness intention is a factor that affects one's competitiveness. Scheme choosers respond only when the unfair event occurred through the competitor's choice, and not when it occurred by accident. The gender gap in the effect of unfair history that exists in the Dictator Game treatment disappears in the Random treatment. Similar results are obtained in Probit regressions.

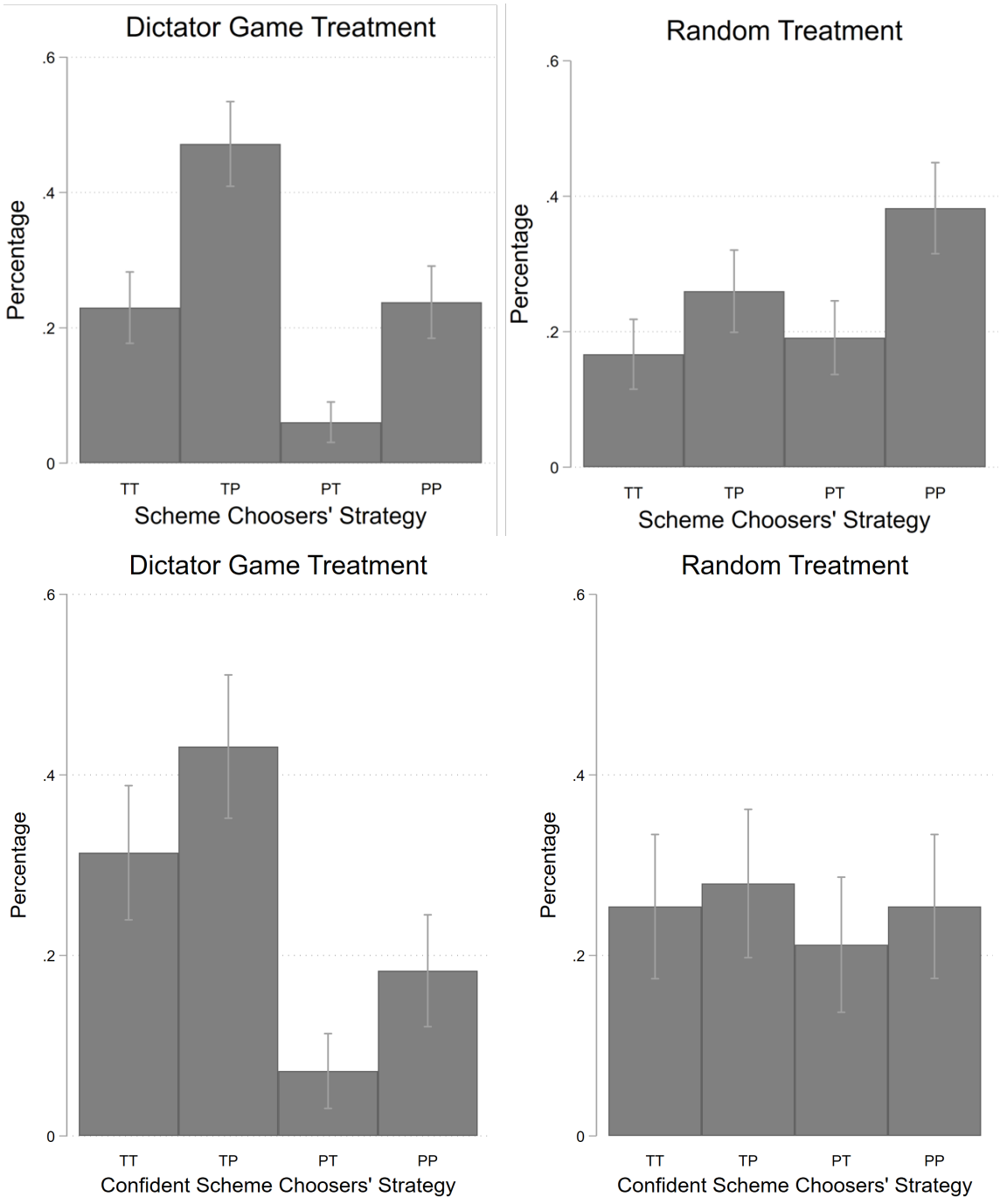
The estimates in Tables 8 provide the basis for the following three results:

Result 1: Hypothesis 1 is supported for males. A male individual who is confident of winning the contest is more likely to initiate a contest with an individual who previously behaved unfairly toward them than with one who has behaved fairly.

Result 2: Hypothesis 2 is supported. Unfair outcomes that were not intended by the competitor do not increase one's willingness to challenge compared to a fair outcome.

Result 3: Hypothesis 3 is supported. The willingness of men who are confident of winning a contest to compete is increased by a history of unfairness, while that of women is not.

FIGURE 3
Percentage of Individuals Initiating Tournaments, by Treatment and History



Note: The number of observations for the Dictator Game treatment and the Random treatment are 248 and 204, respectively. Data are pooled from all rounds. The top two graphs are from the full data set, and the bottom two graphs are from only the scheme choosers who guessed that they got more questions correct than the matched dictator. The error bars are 95% confidence intervals for the mean, assuming a binomial distribution on {0,1}. TT indicates a choice of tournament after both a fair and unfair allocation, and PP a choice of piece-rate in both instances. TP is a choice of tournament after unfair, and piece-rate after a fair, history, while PT is the opposite pattern.

TABLE 8
Determinants of Confident Scheme Choosers' Choice to Initiate Tournament

TREATMENTS	Dicator Game		Random	
	(1)	(2)	(3)	(4)
VARIABLES	Tournament	Tournament	Tournament	Tournament
Unfair	0.493***	0.426***	0.0836	0.116
	(0.0516)	(0.0751)	(0.119)	(0.0996)
Female	0.0890	0.0127	-0.0457	-0.0902
	(0.0742)	(0.118)	(0.153)	(0.133)
Unfair*Female	-0.309***	-0.252***	-0.0643	0.0422
	(0.114)	(0.0680)	(0.168)	(0.122)
Risk12	0.413	0.605**	0.203	0.134
	(0.263)	(0.275)	(0.227)	(0.219)
BA	-0.0207	-0.00658	-0.0987	-0.0652
	(0.107)	(0.104)	(0.134)	(0.150)
M	0.0410	-0.00410	-0.0516	-0.0108
	(0.176)	(0.176)	(0.108)	(0.0897)
Angry		-0.140		-0.0321
		(0.164)		(0.218)
Surprised		0.123		-0.271***
		(0.297)		(0.0839)
Scared		-1.251**		-0.0144
		(0.586)		(0.393)
Disgusted		0.176		
		(0.231)		
Happy				-0.118
				(0.292)
Arousal				0.212
				(0.300)
Period	0.0131	0.0193*	-0.00928	-0.00549
	(0.00876)	(0.0114)	(0.0121)	(0.0109)
Constant	0.0616	0.0195	0.431***	0.375**
	(0.181)	(0.221)	(0.105)	(0.148)
Observations	309	225	236	198

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

Note: The sample used for the regressions is the subset of scheme choosers who guessed that they answered more questions correctly than the dictator they were matched with. The dependent variable equals 1 if scheme chooser initiates tournament and 0 otherwise. Unfair and Female are dummies indicating unfair dictator choice and female gender of scheme chooser. Risk12 is a measure of risk tolerance ranging from 0 to 1. BA equals 1 if the scheme chooser was benevolent only when he was earning more than the dictator. M equals 1 if the scheme chooser was not benevolent either when earning less or more than the dictator. BN (benevolent when the scheme chooser was earning either less or more than the dictator) is the baseline. BB (benevolent only when the scheme chooser was earning less) is omitted, due to lack of data. The emotional state data are the average values of the emotions during the 20 seconds when the scheme choosers were observing a screen assuming that the dictator was either fair or unfair. Each of the seven emotions varies from [0, 1]. Some emotional state variables were eliminated from the regression, since they correlated significantly with other emotional states. Standard errors are clustered at session level. The p-values of F-tests of the restriction 'Unfair + Unfair*Female = 0' for specification 1, 2, 3, and 4 are 0.14, 0.16, 0.87, and 0.14 respectively.

C. Do emotions triggered by a certain history affect one's competitive behavior?

Table 8, in columns (2) and (4), shows the results of estimations of the determinants of tournament choice, for the both Dictator Game and the Random treatments, including the emotional states recorded by the FaceReader^{TM} program in the specification[11]. The emotional states are averaged over the 20 seconds that the scheme choosers were viewing one of the possible histories, right before they made their choice of incentive scheme. In the regressions, we exclude one of any two emotions that are significantly correlated[12].

Table 8 shows that in the Dictator Game treatment, an unfair history makes both men and women launch competition more than under a fair history. The gender gap in the effect of an unfair history identified earlier persists when we control for emotions, in that the impact of an unfair history remains significant only for men. Emotions are also insignificant in the Random treatment, and do not appreciably alter the lack of effect of history and gender. Similar results are found in Probit regressions (Table A4 in the Appendix) and if emotions at the time the scheme chooser learns if the actual outcome was fair or unfair is used (Table A2).

In Section 2, we hypothesized that anger would be associated with a greater likelihood of initiating a tournament. However, the effect of *Angry* is negligible and insignificant in all regressions. We therefore fail to support Hypothesis 4. This finding is stated as our Result 4.

Result 4: Hypothesis 4 is not supported. Anger does not explain tournament initiation after the effect of history is taken into account.

D. Does an unfair history or emotion affect one's performance?

We now consider how well individuals perform at the task. We analyze the effect of history, emotions, and distributional preferences on performance in the addition exercise. To this end, Equation (2) below was estimated. Since the number of correct answers is a count of the number

[11] In the Dictator Game treatment, the average level of *Neutral* was 0.46 after a fair history and 0.44 after an unfair history, average *Happy* was 0.05 after both fair and unfair histories, and average *Sad* was 0.11 after fair and 0.09 after unfair history. Average *Angry* was 0.15 after fair and 0.17 after unfair history, *Surprised* averaged 0.11 after fair and 0.13 after unfair history, and *Scared* averaged 0.02 after fair and 0.01 after unfair history. The average value of *Disgusted* was 0.07 after both fair and unfair histories, and average *Arousal* was 0.28 after fair and 0.29 after unfair history. Under the Random treatment, the average level of *Neutral* was 0.37 after fair and 0.035 after unfair history, average *Happy* was 0.06 after both fair and unfair histories, and average *Sad* was 0.07 after fair and 0.06 after unfair. Average *Angry* was 0.19 after fair and 0.20 after unfair history, average *Surprised* 0.14 after both fair and unfair histories, and average *Scared* 0.02 after both fair and unfair histories. *Disgusted* averaged 0.9 after fair and 0.11 after unfair history, and average *Arousal* was 0.33 after fair and unfair history. Wilcoxon signed-rank tests of the hypothesis that the strength of an emotion differs when a scheme chooser is deciding her action under the unfair vs. the fair scenario yield the following results. In the Dictator Game treatment we find that Neutrality is (borderline) greater under Fair than Unfair ($p = .055$), while Sadness ($p = .0852$), Surprise ($p = .0556$), and Arousal ($p = .0900$) are (borderline) greater under Unfair than under Fair. There are no significant differences in the Random treatment.

[12] If the Bonferroni-corrected p-value of the correlations was higher than 0.05, only one of the two emotions were included in the regression.

of problems solved correctly in each of several periods, a panel data Poisson count regression with random effects was specified.

$$\begin{aligned} Correct_{it} = \kappa + \theta_1 Unfair_{jt} + \theta_2 Female_i + \theta_3 Unfair_{jt} Female_i \\ + \theta_4 DidTour_{it} + \theta_5 DidTour_{it} Female_i + \sigma X_{it} + \mu_i + \varepsilon_{it} \end{aligned} \tag{2}$$

In Equation (2), i denotes scheme chooser, t is the period, and j indexes the corresponding dictator. $Correct_{it}$ indicates how many questions scheme chooser i got correct in period t. $DidTour_{it}$ is a dummy variable that equals one if scheme chooser i participated in a tournament in period t. X_{it} consists of control variables, such as emotions, distributional preferences, and time trend. μ_i is the subject random effect. ε_{it} is the error term.

Table 9 shows the estimates of Equation (2) for both the Dictator Game and the Random treatments. There are some relationships that are robust to specification. In both treatments performance improves over time, as the coefficient of $Period$ is consistently positive and significant. In the Random treatment, women are adversely affected by being in the tournament. The interaction term $Female * DidTour$ is significantly negative indicating that being in a tournament rather than under a piecerate benefits women less than men in terms of its effect on performance. Those who are benevolent when ahead perform better. Emotions do not affect performance and their inclusion in the specification does not affect the results appreciably.

TABLE 9
History, Gender, Emotions, Distributional Preferences and Confident Scheme Choosers' Performance

TREATMENTS	Dictator Game				Random			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
Unfair	-0.121	-0.154	-0.126	-0.158	-0.0461	-0.0399	-0.0433	-0.0335
	(0.258)	(0.218)	(0.239)	(0.224)	(0.0866)	(0.0555)	(0.0785)	(0.0572)
Female	-0.00680	0.144	0.00806	0.171	0.0627	0.0968	0.0659	0.107
	(0.396)	(0.341)	(0.395)	(0.323)	(0.133)	(0.136)	(0.131)	(0.113)
Unfair*Female	-0.00665	-0.00836	0.0120	0.0160	0.0631	0.111**	0.0661	0.0948*
	(0.246)	(0.236)	(0.218)	(0.230)	(0.0701)	(0.0551)	(0.0726)	(0.0554)
DidTour	0.0978	0.186	0.114	0.219	0.183	0.181**	0.208	0.198***
	(0.251)	(0.379)	(0.239)	(0.408)	(0.147)	(0.0775)	(0.127)	(0.0643)
Female*DidTour	-0.133	-0.254	-0.152	-0.297	-0.414*	-0.425***	-0.432**	-0.431***
	(0.279)	(0.344)	(0.293)	(0.413)	(0.213)	(0.124)	(0.182)	(0.148)
Angry		0.121		0.146		-0.00370		-0.0664
		(0.315)		(0.256)		(0.123)		(0.119)
Surprised		0.0733		0.0831		0.264		0.334
		(0.197)		(0.146)		(0.274)		(0.473)
Scared		3.051		3.025		-0.345		-0.131
		(2.777)		(2.975)		(0.216)		(0.228)
Disgusted		0.465		0.460				
		(0.447)		(0.362)				
Happy						-0.0401		0.0507
						(0.215)		(0.368)
Arousal						-0.00752		-0.215
						(0.263)		(0.384)
BA			-0.0329	-0.0670			0.188***	0.250**
			(0.228)	(0.176)			(0.0520)	(0.126)
M			-0.159	-0.192			0.00399	0.0495
			(0.436)	(0.298)			(0.121)	(0.175)
Period	0.0279***	0.0168	0.0277***	0.0162	0.0306***	0.0265***	0.0333***	0.0254***
	(0.00690)	(0.0144)	(0.00703)	(0.0147)	(0.0105)	(0.00836)	(0.00948)	(0.00798)
Constant	1.960***	1.814***	1.988***	1.852***	1.752***	1.729***	1.662***	1.678***
	(0.132)	(0.228)	(0.214)	(0.267)	(0.0798)	(0.0771)	(0.0723)	(0.0965)
Observations	153	113	153	113	118	101	118	101

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

Note: The sample used for the regressions is a subset of scheme choosers who guessed that they answered more questions correctly than the matched dictator. The dependent variable is the number of summation problems solved correctly. DidTour is a dummy variable that equals 1 if the data are from a tournament. Poisson count specification is employed. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, 4, 5, 6, 7, and 8 are 0.31, 0.17, 0.37, 0.22, 0.81, 0.28, 0.75, and 0.30 respectively.

IV Conclusion

We have considered how the prior history of interaction affects individual's willingness to compete and their performance in the competition. Specifically, we investigate how an earlier opportunistic action by one individual that hurts another affects the willingness of the aggrieved party to initiate a tournament against the offending individual.

In the Dictator Game treatment, both men and women challenge more when the competitor previously behaved unfairly, but the effect is only significant and robust to different specifications for men. An unfair outcome that results by accident does not trigger more challenges, on the part of either women or men, than would occur in response to a fair outcome.

In other words, when an unfavorable situation results from a rival's choice, the rival is challenged more often, but only by potential male challengers. If this pattern of gender difference is general, it would suggest that workplaces where there are opportunities to unfairly get ahead of others would trigger conflict among employees, and such conflicts would be particularly common among men. Of course, this is speculative since the gender of other parties is typically known in the workplace and not known in our experiment. The results also show that this competitive behavior does not improve performance and thus would not likely be amenable to being harnessed to achieve greater productivity.

We did not find a relationship between unfair outcomes and anger, or between anger and the subsequent likelihood of launching a tournament. There are a number of possibilities for the lack these relationships. The difference in anger levels was not significant between the two histories in the Dictator Game treatment, though it did go in the expected direction, with greater anger in response to an unfair history. It may be the case that Facereader was unable to detect the experienced anger with sufficient precision. Alternatively, unfair dictator behavior may not trigger anger. Moreover, it may be the case that anger is not a driver of tournament initiation and the decision may be primarily made for reasons that are not emotional in nature.

In our experiment, when women enter a competition after an unfair chance event, they perform worse than they would otherwise. Men are not affected in the same way by bad luck. These results suggest that unequal payment schemes, even when the inequality is unintentional, may adversely affect the performance of women. It would be interesting to test this conjecture in the field.

The gender difference in the response to unfair history operates in addition to the overall lower tendency of women to enter a competitive environment. Niederle and Vesterlund (2007) proposed that one reason why women are underrepresented in competitive jobs is because they shy away from competitive environments more than men. Our results suggest that an unfair history can aggravate this difference, as it appears to encourage men to become more competitive. If the possibilities for opportunistic behavior in the workplace could be reduced, it may reduce the gender disparity in the willingness to enter a competition.

V Appendix

Table A1 contains estimates of the determinants of scheme choosers' decisions, using the period 1 data only. In period 1, the data are unaffected by activity or outcomes in any previous periods. Compared to the estimates reported in the text, the result that unfair history in the Dictator Game treatment leads to more tournaments survives, as does the result that unfair history does not lead to more tournaments in the Random treatment. The result that females respond to unfairness in the Dictator Game treatment with greater choice of tournament becomes borderline significant, with both women and men more likely to launch tournaments in response to an unfair history.

Table A2 presents estimates of the determinants of scheme choosers' choice to launch a tournament, when the emotions in the 20 seconds after the outcome of the history stage is revealed are included in the specification, rather than the emotions at the time scheme choosers make their contingent tournament initiation decisions. The results are similar to those in Table 8. In the Dictator Game, Unfair outcomes trigger greater propensity to choose Tournament, and the strength of the effect is stronger for men than for women. These effects do not appear in the Random treatment.

Table A3 presents estimates of the of the determinants of scheme choosers' choice to launch a tournament, but using Probit estimation. The results are very similar to those in Table 8. In the Dictator Game, unfair outcomes trigger a greater propensity to choose tournament for men but not for women. These effects do not appear in the Random treatment. Table A4 contains estimates of equations that include emotional state variables, and obtain similar results. We do observe that more fearful individuals are less likely to launch a tournament in the Dictator Game treatment, and more surprised individuals are less likely to do so in the Random treatment.

Table A5 shows estimates of the determinants of tournament initiation when all scheme choosers are included in the estimation, rather than only those who believed that they had won the tournament. The results for the Dictator Game treatment in the pooled sample indicate that an unfair history increases tournament initiation for both women and men, with a greater effect for men. The results for the Random treatment are similar to those for the subsample of confident scheme choosers, with the additional result that those who guessed that they had won the tournament were more likely to have chosen to launch one.

Table A6 contains estimates of the determinants of task performance when all scheme choosers are included. It also yields results similar to those for the subsample of the scheme choosers who are confident. Under the Random treatment, unfair treatment results in lower performance by females, but not for males.

Table A7 and A8 contain Probit estimates of the similar specifications to those in Table A5 to estimate the probability of choosing the tournament. The results are very similar to those reported

in Table A5.

Table A9 shows estimates of factors that influence performers for all scheme choosers, rather than those for the confident scheme choosers reported in the text. The effects are similar to those for the subsample of confident individuals reported in Table 9. Under the Random treatment, women perform worse if the prior outcome has been unfair. Performance tends to improve over time as participants gain experience with the task. Table A10 reports the same analysis for dictators (and the players that the scheme choosers were paired with in the Random treatment), and shows that other than an improvement in performance over time, there are no consistently significant correlates of performance.

Finally Table A11 reports the text classification of the stated reasons that scheme choosers chose their actions. In both treatments, the modal reason for those who chose tournament in response to both fair and unfair history is that they were confident of winning. The most common responses for those who chose tournament when the history was unfair and piecerate when unfair were to reciprocate the dictator's action in the Dictator Game treatment and to equalize expected payoff in the Random treatment. In both treatments, those who chose piecerate for both histories most frequently cited a desire to equalize expected payoffs.

In Table A12, we conduct a regression analysis to identify some determinants of the use of the TP strategy. The data show that gender, risk attitude, and emotions are not correlates of the use of TP. On the other hand, there is a correlation between social preference and employing the strategy. In the Dictator Game treatment, those who are Benevolent when Ahead are less likely to play TP, while the they are more likely to do so under the Random treatment. This behavior in the Random treatment is consistent with Inequality Aversion. Those who are benevolent when ahead are averse to advantageous inequality and launching a tournament under the Random treatment is consistent with an aversion to disadvantageous inequality. The fact that the opposite pattern obtains in the Dictator Game treatment is consistent with the notion that those who are less Benevolent when Ahead (more competitive) are also more likely to respond with retaliation when subject to an intentional unfair division.

TABLE A1

Determinants of Confident Scheme Choosers' Decision to Initiate Tournament in Period 1

TREATMENTS	Dictator Game		Random	
	(1)	(2)	(3)	(4)
VARIABLES	Tournament	Tournament	Tournament	Tournament
Unfair	0.458***	0.413***	0.0435	0.104
	(0.0939)	(0.125)	(0.213)	(0.175)
Female	-0.106	-0.0906	0.00840	-0.0255
	(0.0893)	(0.125)	(0.207)	(0.246)
Unfair*Female	0.0123	-0.0321	0.0565	0.355
	(0.225)	(0.251)	(0.193)	(0.254)
Risk12	0.649*	0.608***	-0.0755	-0.408
	(0.372)	(0.214)	(0.301)	(0.371)
BA	0.0293	0.00958	-0.161	-0.0692
	(0.163)	(0.116)	(0.132)	(0.186)
M	0.200	0.219	-0.0539	0.0425
	(0.198)	(0.190)	(0.126)	(0.159)
Angry		-0.258		0.175
		(0.344)		(0.341)
Surprised		0.558**		-0.504
		(0.281)		(0.565)
Scared		-1.525		1.567***
		(1.768)		(0.495)
Disgusted		-0.258		
		(0.326)		
Happy				0.714**
				(0.294)
Arousal				0.0735
				(0.595)
Constant	-0.0287	0.0450	0.622***	0.592*
	(0.242)	(0.208)	(0.161)	(0.353)
Observations	82	64	66	56

Robust standard errors in parentheses

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

Note: The sample used for the regressions is the subset of scheme choosers who guessed that they answered more questions correctly than the matched dictator from Period 1. The dependent variable equals 1 if scheme chooser initiates tournament and 0 otherwise. Unfair and Female are dummies indicating unfair dictator choice and female gender of scheme chooser. Risk12 and Risk8 are measures of risk tolerance ranging from 0 to 1. BA equals 1 if the scheme chooser was benevolent only when he was earning more than the dictator. M equals 1 if the scheme chooser was not benevolent either when earning less or more than the dictator. BN (benevolent when the scheme chooser was earning either less or more than the dictator) is the baseline. BB (benevolent only when the scheme chooser was earning less) is omitted, due to lack of data. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specification 1, 2, 3, and 4 are 0.02, 0.08, 0.18, 0.02 respectively.

TABLE A2
Determinants of Confident Scheme Choosers' Choice, Emotional Response to Dictator's Action

TREATMENTS	Dictator Game		Random	
	(1)	(2)	(3)	(4)
VARIABLES	Tournament	Tournament	Tournament	Tournament
Unfair	0.494***	0.499***	0.0886	0.118
	(0.0510)	(0.0693)	(0.124)	(0.107)
Female	0.0876	0.0241	-0.0512	-0.105
	(0.0728)	(0.102)	(0.149)	(0.137)
Unfair*Female	-0.309***	-0.190**	-0.0630	0.0642
	(0.112)	(0.0959)	(0.167)	(0.120)
Risk12	0.423	0.624	0.178	0.119
	(0.270)	(0.482)	(0.236)	(0.180)
D_Angry		-0.0109		0.0909
		(0.132)		(0.115)
D_Surprised		0.0573		-0.266***
		(0.166)		(0.0761)
D_Scared		-0.0135		-0.0368
		(0.841)		(0.0945)
D_Disgusted		0.466		
		(0.357)		
D_Happy				0.0951
				(0.123)
D_Arousal				-0.234*
				(0.122)
BA	-0.0198	-0.0516	-0.0890	-0.0599
	(0.106)	(0.140)	(0.138)	(0.133)
M	0.0411	-0.105	-0.0456	-0.0580
	(0.177)	(0.223)	(0.111)	(0.110)
Constant	0.115	0.0297	0.397***	0.465***
	(0.169)	(0.273)	(0.126)	(0.155)
Observations	309	215	236	202

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

Note: The sample used for the regressions is the subset of scheme choosers who guessed that they answered more questions correctly than the matched dictator. The dependent variable equals 1 if scheme chooser initiates tournament and 0 otherwise. Unfair and Female are dummies indicating unfair dictator choice and female gender of scheme chooser. Risk12 and Risk8 are measures of risk tolerance ranging from 0 to 1. BA equals 1 if the scheme chooser was benevolent only when he was earning more than the dictator. M equals 1 if the scheme chooser was not benevolent either when earning less or more than the dictator. BN (benevolent when the scheme chooser was earning either less or more than the dictator) is the baseline. BB (benevolent only when the scheme chooser was earning less) is omitted, due to lack of data. The emotional state data are the average values of the emotions during the 20 seconds when the scheme choosers were observing a screen showing the dictator's action and their corresponding action. Each of the seven emotions varies from [0, 1]. Some emotional state variables were eliminated from the regression, since they correlated significantly with other emotional states. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, and 4 are 0.14, 0.01, 0.83, and 0.08 respectively.

TABLE A3
Probit Estimates of the Effect of Unfair History on Confident Scheme Choosers' Tournament Initiation

TREATMENTS	Dictator Game			Random		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament
Unfair	2.374***	2.377***	2.380***	0.317	0.317	0.318
	(0.363)	(0.361)	(0.367)	(0.426)	(0.426)	(0.426)
Female	0.402	0.469	0.475	-0.212	-0.214	-0.214
	(0.393)	(0.382)	(0.395)	(0.547)	(0.553)	(0.548)
Unfair*Female	-1.698***	-1.704***	-1.707***	-0.226	-0.226	-0.227
	(0.624)	(0.620)	(0.627)	(0.585)	(0.584)	(0.583)
BA	0.0428	0.0919	0.101	-0.337	-0.372	-0.371
	(0.452)	(0.504)	(0.516)	(0.554)	(0.525)	(0.518)
M	0.368	0.446	0.441	-0.205	-0.200	-0.187
	(0.913)	(0.911)	(0.916)	(0.407)	(0.404)	(0.376)
AverageRisk		2.046	2.037		0.903	0.913
		(1.277)	(1.278)		(0.889)	(0.890)
Overconfidence			0.0162			0.0240
			(0.0899)			(0.0542)
Constant	-0.872*	-1.996**	-2.011**	-0.00390	-0.471	-0.495
	(0.476)	(0.922)	(0.899)	(0.442)	(0.531)	(0.526)
Observations	309	309	309	236	236	236

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

Note: The sample used for the regressions is the subset of scheme choosers who guessed that they answered more questions correctly than the matched dictator. $AverageRisk$ is a risk tolerance measure that is an average of $Risk12$ and $Risk8$. $Overconfidence$ is own guessed performance minus own actual performance. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, 4, 5, and 6 are 0.14, 0.14, 0.14, 0.83, 0.83, and 0.83 respectively.

TABLE A4

Probit Estimates of the Effect of Emotions on Confident Scheme Choosers' Tournament Initiation

TREATMENTS	Dictator Game				Random			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament
Unfair	2.332***	2.341***	2.370***	2.369***	0.437	0.117	0.420	0.419
	(0.342)	(0.359)	(0.356)	(0.355)	(0.358)	(0.0995)	(0.361)	(0.358)
Female	-0.121	-0.0217	0.136	0.135	-0.342	-0.0926	-0.471	-0.479
	(0.636)	(0.608)	(0.595)	(0.610)	(0.518)	(0.130)	(0.507)	(0.490)
Unfair*Female	-1.643***	-1.640***	-1.670***	-1.670***	0.0814	0.0419	0.282	0.283
	(0.495)	(0.492)	(0.482)	(0.512)	(0.326)	(0.121)	(0.386)	(0.385)
Angry	-0.878	-0.734	-0.836	-0.836	-0.144	-0.0304	-0.0825	-0.0793
	(0.938)	(0.958)	(0.949)	(0.949)	(0.673)	(0.223)	(0.760)	(0.759)
Surprised		0.592	0.566	0.566		-0.261***	-1.183**	-1.167**
		(1.361)	(1.355)	(1.355)		(0.0828)	(0.460)	(0.466)
Scared		-6.517**	-5.824*	-5.830*		0.00512	0.0340	0.0175
		(2.944)	(3.220)	(3.205)		(0.374)	(1.436)	(1.450)
Disgusted		1.254	1.312	1.311				
		(0.987)	(0.987)	(1.065)				
Happy						-0.140	-0.781	-0.765
						(0.285)	(0.998)	(0.982)
Arousal						0.288	1.363	1.372
						(0.263)	(1.011)	(1.028)
BA	0.0294	-0.0591	0.145	0.145	-0.164	-0.0522	-0.287	-0.289
	(0.497)	(0.444)	(0.482)	(0.480)	(0.618)	(0.153)	(0.658)	(0.654)
M	0.209	0.108	0.270	0.270	-0.0885	-0.00494	-0.0367	-0.0193
	(1.009)	(0.996)	(0.942)	(0.958)	(0.398)	(0.0905)	(0.336)	(0.309)
AverageRisk			3.119**	3.119**			0.580	0.597
			(1.339)	(1.326)			(0.941)	(0.944)
Overconfidence				-0.000854				0.0355
				(0.167)				(0.0832)
Constant	-0.350	-0.415	-2.243**	-2.241**	-0.189	0.388**	-0.739	-0.778
	(0.612)	(0.518)	(0.977)	(0.881)	(0.528)	(0.162)	(0.790)	(0.786)
Observations	225	225	225	225	200	198	198	198

Robust standard errors in parentheses

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

Note: The sample used for the regressions is the subset of scheme choosers who guessed that they answered more questions correctly than the matched dictator. $AverageRisk$ is a risk tolerance measure that is an average of $Risk12$ and $Risk8$. $Overconfidence$ is own guessed performance minus own actual performance. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, 4, 5, 6, 7, and 8 are 0.20, 0.1, 0.18, 0.18, 0.02, 0.14, 0.05, and 0.05 respectively.

TABLE A5
Determinants of Tournament Initiation of All Scheme Choosers

TREATMENTS	Dictator Game				Random			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament
Unfair	0.533***	0.534***	0.503***	0.506***	0.0645	0.0645	0.0714	0.0677
	(0.0477)	(0.0477)	(0.0780)	(0.0771)	(0.107)	(0.107)	(0.102)	(0.103)
Female	0.00946	0.0193	0.00508	0.0199	-0.149	-0.132	-0.181	-0.153
	(0.0616)	(0.0619)	(0.0834)	(0.0874)	(0.139)	(0.130)	(0.116)	(0.108)
Unfair*Female	-0.236***	-0.236***	-0.212***	-0.214***	0.0105	0.0105	0.0971	0.102
	(0.0680)	(0.0680)	(0.0701)	(0.0701)	(0.127)	(0.128)	(0.111)	(0.111)
GuessWin		0.0413		0.0559		0.114***		0.147***
		(0.0379)		(0.0549)		(0.0312)		(0.0489)
Risk12	0.441	0.428	0.590	0.576	0.371**	0.374**	0.356**	0.317**
	(0.323)	(0.320)	(0.387)	(0.383)	(0.150)	(0.155)	(0.166)	(0.143)
BA	0.0162	0.0191	-0.0287	-0.0220	-0.103	-0.0936	-0.0529	-0.0440
	(0.0882)	(0.0870)	(0.105)	(0.103)	(0.130)	(0.124)	(0.136)	(0.131)
M	0.0785	0.0774	0.00357	0.00438	-0.00653	-0.00458	0.00741	0.0235
	(0.150)	(0.148)	(0.150)	(0.146)	(0.0611)	(0.0528)	(0.0658)	(0.0526)
Angry			0.0746	0.0702			-0.0699	-0.0721
			(0.121)	(0.126)			(0.164)	(0.168)
Surprised			0.280	0.292			-0.144	-0.128
			(0.173)	(0.178)			(0.104)	(0.0945)
Scared			-0.550	-0.604			0.390	0.343
			(0.497)	(0.543)			(0.358)	(0.337)
Disgusted			0.223	0.229				
			(0.208)	(0.203)				
Happy							-0.0739	-0.125
							(0.218)	(0.222)
Arousal							-0.00211	0.0754
							(0.260)	(0.272)
Constant	0.0548	0.0280	-0.0158	-0.0556	0.262***	0.184***	0.257**	0.149
	(0.172)	(0.178)	(0.208)	(0.210)	(0.0606)	(0.0618)	(0.122)	(0.114)
Observations	496	496	384	384	408	408	338	338

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

Note: All scheme choosers were included in these regressions. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, 4, 5, 6, 7, and 8 are 0.001, 0.001, 0.006, 0.005, 0.84, 0.84, 0.32, and 0.25 respectively.

TABLE A6
Determinants of Performance of All Scheme Choosers

TREATMENTS	Dictator Game				Random			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
Unfair	-0.0607	-0.108	-0.0627	-0.111	0.0108	0.0227	0.00784	0.0199
	(0.230)	(0.173)	(0.218)	(0.162)	(0.0604)	(0.0383)	(0.0602)	(0.0378)
Female	0.0326	0.181	0.0379	0.197	0.0240	0.0448	0.0206	0.0470
	(0.636)	(0.582)	(0.629)	(0.576)	(0.138)	(0.146)	(0.133)	(0.132)
Unfair*Female	-0.0608	-0.0645	-0.0554	-0.0529	-0.122***	-0.107**	-0.120***	-0.104*
	(0.226)	(0.122)	(0.219)	(0.135)	(0.0405)	(0.0539)	(0.0424)	(0.0541)
DidTour	0.0112	0.0741	0.0175	0.0893	0.118	0.0980	0.136	0.116
	(0.283)	(0.427)	(0.286)	(0.438)	(0.169)	(0.0927)	(0.174)	(0.0998)
Female*DidTour	-0.0956	-0.212	-0.102	-0.232	-0.236	-0.224*	-0.247	-0.245
	(0.423)	(0.508)	(0.423)	(0.525)	(0.196)	(0.127)	(0.202)	(0.156)
Angry		0.136		0.154		0.0868		0.0349
		(0.339)		(0.312)		(0.110)		(0.119)
Surprised		0.0925		0.1000		0.0770		0.0652
		(0.116)		(0.0798)		(0.121)		(0.132)
Scared		2.125		2.103		0.222		0.252
		(4.348)		(4.417)		(0.191)		(0.197)
Disgusted		0.580		0.588				
		(0.398)		(0.375)				
Happy						0.129		0.131
						(0.158)		(0.150)
Arousal						-0.119		-0.198
						(0.354)		(0.419)
BA			-0.00250	-0.0623			0.138	0.197**
			(0.252)	(0.289)			(0.0984)	(0.0915)
M			-0.0878	-0.151			-0.0579	-0.0211
			(0.517)	(0.462)			(0.158)	(0.168)
Period	0.0308***	0.0240**	0.0307***	0.0238**	0.0229**	0.0198**	0.0230**	0.0187**
	(0.00424)	(0.00998)	(0.00452)	(0.0105)	(0.0106)	(0.00961)	(0.0104)	(0.00937)
Constant	1.891***	1.743***	1.899***	1.784***	1.731***	1.732***	1.685***	1.692***
	(0.300)	(0.417)	(0.368)	(0.530)	(0.0982)	(0.106)	(0.122)	(0.127)
Observations	248	194	248	194	204	171	204	171

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

Note: All scheme choosers are included in these regressions. The dependent variable is the number of summation problems solved correctly. *DidTour* is a dummy variable that equals 1 if the data are from a tournament. Poisson count specification is employed. Standard errors are clustered at the session level.

TABLE A7

Probit Estimates of the Effect of Unfair History on Tournament Initiation of All Scheme Choosers

TREATMENTS	Dictator Game				Random			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament
Unfair	2.203***	2.204***	2.185***	2.208***	0.208	0.208	0.205	0.208
	(0.276)	(0.279)	(0.267)	(0.283)	(0.346)	(0.348)	(0.346)	(0.347)
Female	0.0175	0.103	0.163	0.115	-0.731	-0.698	-0.636	-0.685
	(0.284)	(0.260)	(0.260)	(0.277)	(0.575)	(0.573)	(0.539)	(0.571)
Unfair*Female	-1.154***	-1.157***	-1.134***	-1.161***	0.123	0.122	0.126	0.121
	(0.399)	(0.398)	(0.375)	(0.399)	(0.456)	(0.456)	(0.456)	(0.456)
BA	0.137	0.219	0.0567	0.233	-0.392	-0.454	-0.406	-0.455
	(0.362)	(0.388)	(0.257)	(0.400)	(0.592)	(0.568)	(0.529)	(0.568)
M	0.429	0.521		0.512	-0.0392	-0.0175	-0.0130	-0.0120
	(0.779)	(0.796)		(0.796)	(0.284)	(0.269)	(0.237)	(0.259)
AverageRisk		2.119	1.979	2.062		1.848**	1.796**	1.792**
		(1.512)	(1.376)	(1.464)		(0.841)	(0.851)	(0.734)
GuessWin			0.148				0.336***	
			(0.155)				(0.113)	
Overconfidence				0.0339				0.0304
				(0.0657)				(0.0328)
Constant	-0.911**	-2.102**	-1.988***	-2.110**	-0.0903	-1.058***	-1.260***	-1.056***
	(0.427)	(0.952)	(0.740)	(0.930)	(0.411)	(0.273)	(0.274)	(0.255)
Observations	496	496	496	496	408	408	408	408

Robust standard errors in parentheses

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

Note: All scheme choosers are included in these regressions. *AverageRisk* is a risk tolerance measure that is an average of $Risk12$ and $Risk8$. $Overconfidence$ is own guessed performance minus own actual performance. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, 4, 5, 6, 7, and 8 are 0.03, 0.03, 0.02, 0.03, 0.02, 0.03, 0.05, and 0.03 respectively.

TABLE A8

Probit Estimates of the Effect of Emotions on Tournament Initiation of All Scheme Choosers

TREATMENTS	Dictator Game					Random				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament	Tournament
Unfair	2.206***	2.224***	2.240***	2.235***	2.250***	0.243	0.0725	0.260	0.243	0.259
	(0.395)	(0.391)	(0.398)	(0.395)	(0.405)	(0.340)	(0.102)	(0.353)	(0.356)	(0.350)
Female	-0.152	-0.0300	0.125	0.191	0.156	-0.886	-0.185	-1.130*	-1.031*	-1.115**
	(0.436)	(0.438)	(0.392)	(0.412)	(0.430)	(0.595)	(0.116)	(0.585)	(0.546)	(0.567)
Unfair*Female	-1.075**	-1.092**	-1.106**	-1.102**	-1.110**	0.447	0.0957	0.691	0.723*	0.686
	(0.477)	(0.496)	(0.496)	(0.485)	(0.503)	(0.331)	(0.111)	(0.421)	(0.408)	(0.418)
Angry	0.346	0.546	0.444	0.461	0.475	-0.356	-0.0761	-0.336	-0.388	-0.312
	(0.558)	(0.599)	(0.589)	(0.597)	(0.611)	(0.557)	(0.158)	(0.614)	(0.627)	(0.596)
Surprised		1.197	1.188	1.251	1.258		-0.140	-0.325	-0.380	-0.303
		(0.817)	(0.827)	(0.866)	(0.794)		(0.0999)	(0.530)	(0.518)	(0.539)
Scared		-4.016	-3.373	-3.225	-3.041		0.422	2.382**	1.990**	2.440**
		(2.888)	(2.834)	(2.791)	(2.847)		(0.361)	(1.065)	(0.980)	(1.037)
Disgusted		1.103	1.198	1.246	1.284					
		(0.871)	(0.845)	(0.843)	(0.872)					
Happy							-0.109	-0.981	-1.083	-0.929
							(0.219)	(0.904)	(0.782)	(0.917)
Arousal							0.0382	1.016	1.028	0.998
							(0.250)	(0.820)	(0.843)	(0.849)
BA	-0.0769	-0.231	0.00374	0.0258	0.00655	-0.248	-0.0512	-0.325	-0.275	-0.305
	(0.452)	(0.420)	(0.458)	(0.443)	(0.462)	(0.701)	(0.146)	(0.765)	(0.707)	(0.731)
M	0.0916	-0.0482	0.153	0.153	0.117	0.00267	-0.00175	0.0767	0.114	0.0865
	(0.843)	(0.836)	(0.816)	(0.781)	(0.834)	(0.337)	(0.0752)	(0.335)	(0.280)	(0.318)
AverageRisk			3.275**	3.261**	3.142*			1.629*	1.474*	1.639*
			(1.620)	(1.582)	(1.607)			(0.865)	(0.785)	(0.847)
GuessWin				0.237					0.443**	
				(0.239)					(0.177)	
Overconfidence					0.0640					0.0664
					(0.0871)					(0.0747)
Constant	-0.669	-0.812*	-2.766***	-2.966***	-2.779***	-0.181	0.429**	-1.334**	-1.538***	-1.397**
	(0.544)	(0.469)	(0.985)	(1.033)	(0.947)	(0.512)	(0.172)	(0.532)	(0.493)	(0.561)
Observations	384	384	384	384	384	342	338	338	338	338

Robust standard errors in parentheses

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

Note: All scheme choosers are included in these regressions. *AverageRisk* is a risk tolerance measure that is an average of *Risk*12 and *Risk*8. *Overconfidence* is own guessed performance minus own actual performance. Standard errors are clustered at the session level. The p-values of the test 'Unfair + Unfair*Female = 0' for specifications 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 are 0.06, 0.09, 0.10, 0.11, 0.12, 0.11, 0.12, 0.19, 0.30, and 0.18 respectively.

TABLE A9

Determinants of the Performance of All Scheme Choosers, Poisson Count Model Estimates

TREATMENTS	Dictator Game				Random			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
Unfair	-0.0607	-0.00660	-0.0627	0.00452	0.0108	0.0203	0.00784	0.0186
	(0.230)	(0.134)	(0.218)	(0.121)	(0.0604)	(0.0550)	(0.0602)	(0.0581)
Female	0.0326	0.00762	0.0379	0.0326	0.0240	0.0313	0.0206	0.0411
	(0.636)	(0.709)	(0.629)	(0.690)	(0.138)	(0.168)	(0.133)	(0.162)
Unfair*Female	-0.0608	-0.0849	-0.0554	-0.0955	-0.122***	-0.128**	-0.120***	-0.125*
	(0.226)	(0.167)	(0.219)	(0.169)	(0.0405)	(0.0611)	(0.0424)	(0.0649)
DidTour	0.0112	-0.0409	0.0175	-0.0411	0.118	0.102	0.136	0.125
	(0.283)	(0.355)	(0.286)	(0.362)	(0.169)	(0.112)	(0.174)	(0.131)
Female*DidTour	-0.0956	-0.0597	-0.102	-0.0638	-0.236	-0.189	-0.247	-0.220
	(0.423)	(0.467)	(0.423)	(0.478)	(0.196)	(0.150)	(0.202)	(0.175)
Angry		-0.166		-0.136		-0.00569		-0.0782
		(0.227)		(0.224)		(0.155)		(0.211)
Surprised		0.224		0.236*		0.0609		-0.00353
		(0.149)		(0.140)		(0.0907)		(0.100)
Scared		1.172		1.109		0.206		0.191
		(3.802)		(3.946)		(0.545)		(0.571)
Disgusted		0.279		0.308				
		(0.531)		(0.481)				
Happy						-0.0769		-0.114
						(0.294)		(0.304)
Arousal						0.224		0.219
						(0.374)		(0.374)
BA			-0.00250	-0.112			0.138	0.185
			(0.252)	(0.365)			(0.0984)	(0.121)
M			-0.0878	-0.0791			-0.0579	-0.0197
			(0.517)	(0.353)			(0.158)	(0.209)
Period	0.0308***	0.0330***	0.0307***	0.0331***	0.0229**	0.0230**	0.0230**	0.0236**
	(0.00424)	(0.00842)	(0.00452)	(0.00868)	(0.0106)	(0.0108)	(0.0104)	(0.0116)
Constant Correct	1.891***	1.785***	1.899***	1.838***	1.731***	1.647***	1.685***	1.589***
	(0.300)	(0.482)	(0.368)	(0.664)	(0.0982)	(0.168)	(0.122)	(0.206)
Observations	248	176	248	176	204	175	204	175

Robust standard errors in parentheses

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

Note: All scheme choosers are included in these regressions. Poisson count model estimation is employed. The emotional state data is the average value of the emotion during the 20 seconds during which the scheme choosers were observing a page that showed the dictator's actual allocation decision. The emotions of Neutrality and Sadness were eliminated from the regression, since they correlated significantly with other emotional states. Standard errors are clustered at the session level.

TABLE A10
Determinants of the Performance of Dictators, Poisson Count Model Estimates

TREATMENTS	Dictator Game			Random		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Correct	Correct	Correct	Correct	Correct	Correct
Unfair	-0.0726	-0.0754	-0.0758	0.0489***	0.000806	-7.36e-05
	(0.237)	(0.282)	(0.280)	(0.0132)	(0.0302)	(0.0350)
Female	-0.285	-0.288	-0.288	0.0202	0.0144	0.0155
	(0.403)	(0.522)	(0.521)	(0.141)	(0.105)	(0.101)
Unfair*Female	0.0614	0.133	0.134	-0.113**	-0.0768	-0.0748
	(0.290)	(0.436)	(0.435)	(0.0545)	(0.0501)	(0.0596)
DidTour	-0.0214	-0.00133	0.000233	-0.0545	-0.0538	-0.0517
	(0.0738)	(0.0809)	(0.0795)	(0.0574)	(0.0709)	(0.0777)
Female*DidTour	0.105*	0.00841	0.00521	0.0118	-0.00938	-0.0108
	(0.0588)	(0.145)	(0.141)	(0.117)	(0.158)	(0.158)
Happy		0.0361			-0.00355	-0.00113
		(0.193)			(0.194)	(0.174)
Angry		-0.0776	-0.0824		-0.100	-0.0993
		(0.420)	(0.414)		(0.0919)	(0.0957)
Surprised		-0.0769	-0.0828		-0.171	-0.168
		(0.237)	(0.236)		(0.177)	(0.181)
Scared		1.801	1.792		-0.0393	-0.0541
		(2.307)	(2.312)		(0.507)	(0.440)
Disgusted		0.318	0.318		-0.0362	
		(0.240)	(0.240)		(0.355)	
Arousal					0.0378	0.0383
					(0.324)	(0.320)
Period	0.0209***	0.0229***	0.0228***	0.00691	0.0129**	0.0128**
	(0.00498)	(0.00588)	(0.00606)	(0.00590)	(0.00625)	(0.00607)
Constant Correct	2.058***	2.004***	2.008***	1.900***	1.963***	1.959***
	(0.247)	(0.331)	(0.328)	(0.0555)	(0.140)	(0.157)
Observations	248	193	193	204	159	159

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

Note: Poisson count model results are presented. The emotional state data is the average value of the emotion during the 20 seconds when the dictators observed what the scheme choosers' corresponding decision was. This time interval was chosen to achieve comparability with the models for scheme choosers presented in Table A9. Neutrality and Sadness were eliminated from the regression, since they correlated significantly with other emotional states. Standard errors are clustered at the session level.

TABLE A11
Commonly Stated Reasons for Different Scheme Chooser Strategies

	Dictator Game Treatment		Random Treatment	
Strategy	Reasons	% for each strategy	Reasons	% for each strategy
Tournament after fair & unfair history	Confident of winning	29.31%	Confident of winning	44.07%
	Opportunity to earn more	20.69%	Opportunity to earn more	22.03%
	Other	50.00%	Other	33.90%
Tournament after unfair history & piece-rate after fair history	Reciprocity	40.38%	Equalize expected payoff	27.47%
	Guarantied payment	21.51%	Guaranteed payment	18.68%
	Other	38.11%	Other	53.85%
Piece-rate after unfair history & tournament after fair history	Enough money to take risk / too little to take risk	38.46%	Guaranteed payment	38.98%
	Opportunity to earn more money	23.08%	Opportunity to earn more money	20.78%
		38.46%	Other	40.24%
Piece-rate after fair & unfair history	Guaranteed payoff	50.00%	Guaranteed payoff	38.18%
	Reciprocity	17.02%	Controllability	22.73%
	Other	32.98%	Other	39.09%

Note: The number of observations for the Dictator game treatment and the Random treatment are 62 and 51 respectively.

TABLE A12

TREATMENTS	Dictator Game		Random	
	(1)	(2)	(3)	(4)
VARIABLES	TP	TP	TP	TP
Female	-0.117	-0.0505	-0.0608	0.0169
	(0.120)	(0.0832)	(0.114)	(0.0976)
Risk12	-0.112	0.367	0.0568	0.228
	(0.147)	(0.260)	(0.342)	(0.298)
BA	-0.281**	-0.384***	0.248*	0.295**
	(0.122)	(0.130)	(0.149)	(0.148)
M	-0.188	-0.202	0.305	0.394*
	(0.250)	(0.257)	(0.191)	(0.206)
Angry		0.0707		0.0831
		(0.123)		(0.203)
Surprised		-0.327		-0.130
		(0.222)		(0.250)
Scared		2.215*		0.190
		(1.236)		(0.415)
Disgusted		0.0998		
		(0.106)		
Happy				0.0967
				(0.188)
Arousal				-0.367
				(0.364)
Period	-0.0126	-0.0138**	-0.00771	-0.0150*
	(0.0125)	(0.00675)	(0.00770)	(0.00814)
Constant	0.778***	0.531***	0.162	0.152
	(0.150)	(0.201)	(0.240)	(0.240)
Observations	153	113	118	101

Robust standard errors in parentheses

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

Note: The sample used for the regressions is the subset of scheme choosers who guessed that they answered more questions correctly than the matched dictator. The dependent variable equals 1 if the scheme chooser initiates tournament after an unfair history and chooses piece-rate after a fair history (TP). Risk12 is a measure of risk tolerance ranging from 0 to 1. BA equals 1 if the scheme chooser was benevolent only when he was earning more than the dictator. M equals 1 if the scheme chooser was not benevolent either when earning less or more than the dictator. BN (benevolent when the scheme chooser was earning either less or more than the dictator) is the baseline. BB (benevolent only when the scheme chooser was earning less) is omitted, due to lack of data. Standard errors are clustered at the session level.

References

Andrade, E. B., T. Odean, and S. Lin (2016). Bubbling with excitement: an experiment. *Review of Finance 20*(2), 447–466.

Balafoutas, L., R. Kerschbamer, and M. Sutter (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization 83*(1), 125–135.

Berge, L. I. O., K. Bjorvatn, A. J. G. Pires, and B. Tungodden (2015). Competitive in the lab, successful in the field? *Journal of Economic Behavior & Organization 118*, 303–317.

Bertrand, M. and K. F. Hallock (2001). The gender gap in top corporate jobs. *ILR Review 55*(1), 3–21.

Buser, T. (2012). The impact of the menstrual cycle and hormonal contraceptives on competitiveness. *Journal of Economic Behavior & Organization 83*(1), 1–10.

Buser, T., A. Dreber, and J. Mollerstrom (2017). The impact of stress on tournament entry. *Experimental Economics 20*(2), 506–530.

Buser, T., L. Geijtenbeek, and E. Plug (2015). Do gays shy away from competition? do lesbians compete too much? *IZA Discussion Paper*.

Buser, T. and H. Yuan (2019). Do women give up competing more easily? evidence from the lab and the dutch math olympiad. *American Economic Journal: Applied Economics 11*(3), 225–52.

Cárdenas, J.-C., A. Dreber, E. Von Essen, and E. Ranehill (2012). Gender differences in competitiveness and risk taking: Comparing children in colombia and sweden. *Journal of Economic Behavior & Organization 83*(1), 11–23.

Cassar, A. and M. L. Rigdon (2021). Option to cooperate increases women's competitiveness and closes the gender gap. *Evolution and Human Behavior*.

Charness, G., A. Rustichini, and J. Van de Ven (2018). Self-confidence and strategic behavior. *Experimental Economics 21*(1), 72–98.

De Paola, M. and V. Scoppa (2017). Gender differences in reaction to psychological pressure: Evidence from tennis players. *European Journal of Work and Organizational Psychology 26*(3), 444–456.

Den Uyl, M. and H. Van Kuilenburg (2005). The facereader: Online facial expression recognition. In *Proceedings of Measuring Behavior*, Volume 30, pp. 589–590. Citeseer.

Doyle, L. and D. Schindler (2019). Mucap: Connecting facereadertm to z-tree. *Journal of Economic Science Association 5*, 136 – 141.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*(2), 268–298.

Ekman, P. and D. Keltner (1997). Universal facial expressions of emotion. *in Nonverbal communication: Where nature meets culture, Segerstrale U, P. Molnar P, eds.*, 27–46.

Falk, A., E. Fehr, and U. Fischbacher (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior 62*(1), 287–303.

Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and economic behavior 54*(2), 293–315.

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

Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy 127*(4), 1826–1863.

Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. *American Economic Review 94*(2), 377–381.

Healy, A. and J. Pate (2011). Can teams help to close the gender competition gap? *The Economic Journal 121*(555), 1192–1204.

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

Hopfensitz, A. and E. Reuben (2009). The importance of emotions for the effectiveness of social punishment. *The Economic Journal 119*(540), 1534–1559.

Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: The equality equivalence test. *European Economic Review 76*, 85–103.

Kessel, D., J. Mollerstrom, and R. van Veldhuizen (2021). Can simple advice eliminate the gender gap in willingness to compete? *European Economic Review*, 103777.

Lerner, J. S. and D. Keltner (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology 81*(1), 146.

Miller, A. J., E. L. Worthington Jr, and M. A. McDaniel (2008). Gender and forgiveness: A meta–analytic review and research agenda. *Journal of Social and Clinical Psychology 27*(8), 843–876.

Nguyen, Y. and C. N. Noussair (2014). Risk aversion and emotions. *Pacific Economic Review 19*(3), 296–312.

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

Oswald, A. J., E. Proto, and D. Sgroi (2015). Happiness and productivity. *Journal of Labor Economics 33*(4), 789–822.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 1281–1302.

Reuben, E., P. Sapienza, and L. Zingales (2015). Taste for competition and the gender gap among young business professionals. Technical report, National Bureau of Economic Research.

Root, B. L. and J. J. Exline (2011). Gender differences in response to experimental forgiveness prompts: Do men show stronger responses than women? *Basic and Applied Social Psychology 33*(2), 182–193.

Terzis, V., C. N. Moridis, and A. A. Economides (2010). Measuring instant emotions during a self-assessment test: the use of facereader. In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*, pp. 18.

van Leeuwen, B., C. N. Noussair, T. Offerman, S. Suetens, M. Van Veelen, and J. Van De Ven (2018). Predictably angry—facial cues provide a credible signal of destructive behavior. *Management Science 64*(7), 3352–3364.

Wozniak, D. (2012). Gender differences in a market with relative performance feedback: Professional tennis players. *Journal of Economic Behavior & Organization 83*(1), 158–171.

Wozniak, D., W. T. Harbaugh, and U. Mayr (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics 32*(1), 161–198.