Tiger Women: An All-Pay Auction Experiment on the Gender Heuristic of the Desire to Win*

DAVID ONG1 AND ZHUOQIONG (CHARLIE) CHEN2

Prior work has shown that lack of familiarity could prevent people from playing mixed strategy Nash equilibrium (MNE) as well as drive “overbidding” in first price and all-pay auctions, and Tullock contests. However, these prior works did not reveal anything about opponents and thus, could not thereby induce the systematic variations necessary to separate “mistakes” from rational reactions to beliefs about opponents’ risk attitudes and valuations. We hypothesized that gender was a familiar heuristic for risk attitude and valuation in real life competitions. We informed subjects of the opponent’s genders when we tested for gender differences in “desire to win” by measuring willingness to pay to win in female (F) vs. male (M), M vs. F, F vs. F, and M vs. M all-pay auctions. We found a very tight fit with MNEs. Though, like prior Tullock contest and first price auction experiments, women bid higher than men. We found that their payoffs were weakly higher than men’s, consistent with strategic high bidding rather than overbidding. We derived theory to separate the effects of gender differences in valuation and risk attitude on average bids across treatments. Our data indicate that women have a higher valuation for winning than men. We discuss risk attitude and other possible confounds in standard measures of gender differences in competitive attitude and our possible contributions to that literature.

Key Words: Nash equilibrium, experiments, overbidding, gender differences, all-pay auction, risk aversion, heuristic, stereotype

JEL Codes: C91, D44, J16, I20

* We thank Kai Konrad for his generous support of David Ong’s visit to the Max Plank Institute of Tax Law and Public Finance, where the initial idea for this experiment was developed. We thank participants at the Economics Department seminar at Florida State University, Rady School of Business Management seminar, University of California at San Diego; Economics Department seminar at the University of Western Ontario; Economic Science Association Meeting 2011 at Xiamen; Contests, Mechanisms and Experiments Conference - Exeter 2012; Oxford University Nuffield Seminar; and at the Game Theory Society World Congress at Bilgi University for their comments. We would also like to thank in particular Uri Gneezy, Vincent Crawford, Charles Zheng, Roman Sheremeta, Nathaniel Wilcox, Charles Naussaire, Shahar Kariv, Svetlana Pevniskaya, Lise Vesterlund, Linda Babcock, Jane Zhang, Daniela Rutzler, Gregory Pavlov, Hamid Sabourian, Aner Sela, Dan Kovenock, Dylan Minor, Ron Seigel, John Wooders, Donna Harris, Dirk Engelmann, Alexander Smith, George Georgiadis, and Zhong (Zachary) Zemin for their helpful comments, and Hong (Hannah) Lin who assisted in the experiment, and students in the experimental economics class of Peking University HSBC Business School.

1 Email: dvdong@gmail.com, Tel: +86-755-2603-2655, Fax: +86-755-2603-5344, Address: C402, Peking University, University Town, Shenzhen, P.R.C, 518055.

2 Email: z.chen16@lse.ac.uk, London School of Economics and Political Science

I. Introduction


Kovash and Levitt (2010) pointed out that the failure of minimax in the lab might be due to a lack of familiarity with lab games, low stakes, or the absence of real world competitive selection pressures. In contrast, professional sports players do seem to play MNE in the field (Walker and Wooders 2001, Hsu; Huang, and Tang 2007; Chiappori, Levitt, and Groseclose 2002; Palacios-Huerta 2003; Azar and Bar-Eli 2009; Hirschberg et al. 2009). However, even they appear not to follow MNE in the lab with the exception of Palacios-Huerta and Volij (2008), in which soccer players who were brought into the laboratory played minimax in both a 2x2 game and O’Neill’s (1987) 4x4 game. But, this last result was not replicated by Levitt, List, and Reiley (2010) using either professional soccer players or world class poker players.

Professionals could be using heuristics in the field which are not available in the abstract setting of lab experiments. Hsu, Huang and Tang (2003) showed that for a significant number of top tennis players, some simple low-information rules outperformed the minimax hypothesis in the field. Kovash and Levitt (2010) surveyed professional soccer players after a lab penalty kick game. None spontaneously recognized the analogy between the lab game and real penalty kicks. Even if they had, it is not clear that they should carry over the lesson. Loewenstein (1999) survey of psychology literature concluded that there was limited learning across tasks. Indeed, Wooders (2010) demonstrated that students in fact played closer to MNE than professionals in the lab.

However, the evidence that people are playing MNE anywhere could itself be problematic. Kovash and Levitt (2010) pointed out that field studies tended to have very
low power to reject (Levitt and List 2007). This is especially important because as Harsanyi (1973) pointed out, mixing is consistent with exogenous errors.

In other areas like 1st price auctions, though, the errors in playing even pure strategy equilibrium are so consistent that there is an ongoing debate as to whether robust high bidding is overbidding or reflective of higher valuation. The literature debating the possibility of an extra-value of winning in first price auctions dates back to at least Cox et al. (1988), including the comment by Harrison (1989), and Kagel and Roth (1992). More recently in contests: Price and Sheremeta (2012); all-pay auctions: Gneezy and Smorodinsky (2006); Rapoport and Amaldoss (2004), common value auctions: Charness and Levin, D. (2009); Chen, Y., Katuscak and Ozdenoren (2009); Casari, Ham, and Kagel (2007); Ham and Kagel (2006). Furthermore, women seem to overbid more than men. See the survey of the experimental contest and auction literature by Dechenaux, Kovenock and Sheremeta (2012). Such regularity in overbidding might have spawned the experimental literature on non-monetary rewards to status (Delfgaauw et al., 2009; Kosfeld and Neckermann, 2010; Charness et al., 2010; Tran and Zeckhauser, 2011; Levitt et al., 2012). Interestingly, gender differences were weak or mixed.

We note that in neither prior MNE experiments nor in first price auction experiments was any aspect of opponents revealed to players. This was to avoid the loss of control. As a consequence, they could not control beliefs about opponent characteristics and hence for the systematic variations in strategies to changes in opponent characteristics that would have been an indicator of rational adaptations to opponent characteristics rather than mere mistakes.

We hypothesized that gender was a familiar and important heuristic for risk attitude and valuations in real life competitions. If so, then by changing the gender of opponents, we could get the systematic variations we thought necessary to test whether people were making mistakes or reacting rationally to their beliefs. We tested for gender differences in “desire to win” (DTW, the non-monetary component to winning) by measuring willingness to pay to win in female (F) vs. male (M), M vs. F, F vs. F, and M vs. M common value all-pay auctions within and across universities. We follow Parreiras and Rubinchik’s (2009) model of DTW as an extra positive term in payoffs from winning. We found a very tight fit with MNE. Though, like in prior experiments in Tullock
contests and first price auctions, women bid higher than men, we found that their payoffs were weakly higher, suggesting strategic high bidding rather than overbidding. We derived theory to separate the effects of gender differences in valuations and risk attitude on average bids across treatments. Our data indicate that women have a higher valuation for winning than men. We discuss risk attitude and other possible confounds in standard measures of gender differences in competitive attitude and our possible contributions to that literature after our main results.

All-pay auctions have often been used to model real life contests. For complete information risk neutral cases, see Ellingsen (1991); Baye, Kovenock, and de Vries (1996); Hillman and Riley (1989); Rapoport and Amaldross (2004). For incomplete information examples, see Amann and Leininger (1996) for the risk neutral case, and Fibich, Gavious, and Sela (2006) for the risk averse case. Siegel (2009), (2010) derived general properties of all-pay contests, including the equilibrium for all-pay auctions with perfect information. Incentives of all-pay auctions (everyone pays their bid, win or lose) are prevalent, e.g., curved exams. Thus, we were confident of our subjects’ familiarity.

Gender has been used as a heuristic for risk attitude in experiments. Differences in risk attitude was suggested by visual contact (Cadsby and Maynes 1998; Eckel and Grossman, 2001, 2008; Gneezy, Niederle and Rustichini 2003; Niederle and Vesterlund 2007). Gender was explicitly mentioned in the instructions (Holm 2000; Slonim and Garbarino 2007; Sutter, Bosman, Kocher and van Winden 2009). It was implicitly and explicitly indicated through nicknames: Gupta, Poulsen and Villeval (2013). To our knowledge, gender stereotypes have not been used in an auction setting.

We indicated the gender of opponents explicitly in the instructions. However, we can rule out possible demand effects with across auction treatments. Each subject faced essentially a simple binary choice of bidding higher or lower depending on the gender of opponent. With two genders across 4 auctions \{FF, FM, MF, MM\}, there could be 701 orderings \{e.g., FF>MM, FF<MF….\}. But, gender differences in valuations or risk attitude would generate orderings that are consistent with only 6% of these 701 orderings. That is also the probability of getting orderings consistent with theory by chance. Thus,

---

3 However, since these papers used a single piece of notation for utility, valuation, and bids, they are not amenable to our goal of separating the effects of risk attitude and valuations on bids.
though we did not have a unique MNE prediction, our theory is still quite falsifiable. Demand effects also seem unlikely drivers of our results. The theory is not intuitive, e.g., women’s bids against men must be the same as men’s bids against men for women to have a higher valuation. Furthermore, even if subjects could somehow derive our theory and calculated the equilibria they thought we wanted, we cannot see how subjects of both genders could coordinate on all those equilibria across 6 auctions that we used to identify preferences except through mind reading. Our highly consistent data and significant differences across schools (see female bidder male opponent (FM) auctions below) alone rule out gender based demand effects. In any case, our results are in the opposite direction to our initial expectations.

Our design can also rule out gender differences in calculation ability as the driver of our results. It has been found in Tullock contests that women overbid more than men though they are more risk averse and less competitive (Price and Sheremeta (2012); Ham and Kagel (2006); Casari, Ham, and Kagel (2007); Charness and Levin (2009); Chen, Katsušcák, and Ozdenoren (2009)). Heterogeneity in numerical ability (i.e., errors) has also been suspected as the driver of within subject inconsistency in the most widely used measure of risk aversion. Charness and Viceisza (2012) and Dave, Eckel, Johnson, and Rojas (2010) compared complex (Holt and Laury 2002) and simple elicitation methods (Eckel and Grossman 2002, 2008, Gneezy and Potters 1997). See Dave, Eckel, Johnson and Rojas (2010) for a review of earlier work on problems of risk attitude elicitation. As observed in Charness and Viceisza (2012), “Most studies in the Western world have inconsistency rates between 10 and 15 percent 13 percent in Holt and Laury 2002, 11 percent in Stockman 2006, and 12 percent in Meier and Sprenger 2010). Lammers, Lau and Verbon (2006) find an inconsistency rate of 66.5% in South Africa; Galarza (2009) finds an inconsistency rate of 52% in Peru; and Doerr et al. (2011) find an inconsistency rate of 39 percent in Ethiopia. An exception is de Brauw and Eozonou (2011) who find an inconsistency rate of 14 percent in Mozambique. Jacobson and Petrie (2009) and Engle Warnick et al. (2011) use an instrument that is different from Holt and Laury (but can also be used to assess inconsistencies) and find a rate of 52 percent in Rwanda and Peru respectively.” In fact, in a survey paper Eckel and Grossman (2008) concluded that “the findings thus far shed serious doubt on the existence of risk attitude as a measurable,
stable personality trait, or as a domain-general property of a utility function in wealth or income.” However, unlike 1st and 2nd price auctions, gambles in risk attitude experiments, and Tullock contests, there is no right answer (per individual) to calculate for all-pay auctions due to their mixed strategy equilibrium.

A. Outline of experimental results

Our initial aim was to eliminate confounds possibly present in prior studies of gender differences in competitive attitude and to test the idea that gender could be used to induce changes in perceived and actual valuations and risk attitudes in experiments. Our results were very surprising, even to us. We came to them iteratively by first having our original hypothesis falsified, then endogenously generating a hypothesis from across school bids, which we tested and confirmed with new subjects in within school bids. In the interest of giving the reader a full picture of our path to our unexpected conclusions, we now outline the process by which we arrived at our results with the results.

We began with the hypothesis that men had higher desire to win (DTW) and that this was signaled to the opponents by maleness. We found what initially seemed marginally significantly positive results with 92 subjects (46 male vs. female pairs) that men bid higher when they knew they were bidding against women\(^4\). Subjects only bid once in any treatment in our across subject design. We recruited subjects with posters from 2 top tier graduate schools (Peking and Tsinghua universities) and a few students from Harbin University in Shenzhen University Town (UT)\(^5\). However, these initial results became insignificant when we added 64 more subjects on the suggestion of some conference participants. With a total pool for this first study of 156 subjects, we found no difference in bids between men and women in either control or treatment. This startling result, which was contrary to the whole gender difference literature as we knew it at that time, led us to consider the following possible causes: a) a selection effect from the admission process to top grad schools (our original fear), or b) self-selection bias from poster recruitment substantiating a concern of Croson and Gneezy (2009). See also for

\(^4\) This actually implied that men believed that women had higher DTW or were more risk averse. We are grateful to Charles Zheng for pointing this out.

\(^5\) These are the results we presented at the summer Canadian Economic Association in Ottawa, the Economic Science Association Meeting in Chicago, and the Game Theory Society Meeting at Stony Brook GTC 2011.
Krawczyk (2011); Falk and Heckman (2009); Eckel and Grossman (2000), and for other self-selection effects. We explored further the effects of selection in a follow-up paper which includes new data. We discuss how this initial null result helps rule out possible confounds in the discussion of errors after the main results.

To rule out the above selection confounds, we recruited entire classes (of economics students) at Shenzhen University (SZ), a mid-tier university. There we found, ignoring significance, that SZ subjects bid higher against their men than against their women. Using our theory of all-pay auctions with two risk averse bidders and complete information, our experimental results implied that women had higher desire to win, though they could be more risk averse. With another set of SZ students, we also tested whether SZ student’s stereotypes of UT students was consistent with our surprising initial finding at UT of no gender difference by having these new SZ students bid against a new set of UT students. Contrary to our expectations, SZ bid higher against UT women than against UT men. Using our theory, this meant that SZ students believed UT women were either less risk averse or had higher DTW. This was not a school effect because SZ subjects were not informed about how we recruited at the three graduate schools of UT. To test this new “SZ hypothesis” and to rule out possible selection effects in the initial study at UT, we retested at UT by recruiting entire classes of new students this time. Again, contrary to our expectations but consistent with SZ hypothesis, UT men and women bid higher against UT women than against UT men. Using theory, this meant that UT women were less risk averse and had higher DTW than UT men. These results became even more significant when we pooled the data. We completed the experiment by having a new set of UT students bid against SZ students. Combining these across school bids by UT students against SZ students with the across school bids of SZ students against UT students, we found that SZ women were more risk averse but had higher valuations than UT women. We found no gender differences in quitting either within or across schools with our bidding zero measure of quitting. This was consistent with Zhang (2011a) and (2011b), the only other study of gender differences in competitive attitudes.

with Han Chinese subjects. She found no gender differences. We now summarize our contributions.

**B. Summary of contributions**

Our results contribute to the experimental, auction, and gender differences literatures. To the experimental literature, the significant differences between our initial and final results for UT women suggest that prior conflicting results among studies of adults in measuring gender differences in risk attitude or competitive behavior, for example Price (2010)’s failed replication of NV (2007), and between adults and children (who are usually recruited by classrooms), could have been driven by selection effects from poster recruitment. This is addressed more fully in Ong and Chen (2012). We extend all-pay theory by separating the effects of risk aversion from valuation under complete information for 2 players. To the experimental all-pay auction literature, our contribution is to conduct the first all-pay auction experiment which used knowledge of the opponent’s gender as a treatment. We thought that gender could be taken as a ‘signal’ (in the sense of an exogenous indicator) of the desire to win. SZ students revealed through their bids in the across school treatments their belief that UT women were more competitive in either DTW or risk attitude than UT men. We found them to be both in the within UT treatments. Our data showed systematic differences in bidding for a common value auction, conditional on the gender of the bidder and the opponent, indicating a non-common value component to an apparently common value auction. To the non-monetary value of winning literature, we provide evidence that gender can be a heuristic for value of winning. See Gigerenzer (2007) for survey of literature on heuristics. We also provide evidence suggesting people might play MNE in all-pay auctions. Our results could help explain systematic “overbidding” found in auction experiments. See Heyman, Orhun, and Ariely (2004); Gneezy and Smorodinsky (2006) for examples. We discuss our possible contributions to gender difference in competitive literature after the main results.

**II. All-Pay Auction Theory**

**Theoretical Result 1:** The Equilibrium for an All-Pay Auction with 2 Risk Averse Bidders and Perfect Information.
We chose a perfect information setting in order to test whether we could naturally induce changes in risk attitude and valuations in a lab experiment using revealed gender. As mentioned, we expected that different gendered subjects should react differently (on average) to the same gendered opponent according to their own gender. Fixing the gender of the subject, they should react differently (on average) to different gendered opponents according to the opponent's gender. We hypothesized that this weak information requirement would be met in real life and in the lab. Furthermore, an imperfect information setup would add complexity and degrees of freedom which we did not need to explain the data. Having that extra complexity would also make our theory less falsifiable. We chose a one shot game because we were not sure whether repeated play, even with rematching, would meet the requirement of independence across rounds, necessary for a mixed strategy equilibrium. Rapoport and Amaldress (2004) had already shown that subjects were not mixing according to theory at the individual level with repeated play.

Baye, Kovenock, and de Vries (1996); Ellingsen (1991); and Hillman and Riley (1989) showed that the mixed strategy equilibrium for asymmetric all-pay auction with complete information and risk neutral players is:

\[(V_2 - b_2)G_1(b_2) + (-b_2)(1 - G_1(b_2)) = 0\]

\[(V_1 - b_1)G_2(b_1) + (-b_1)(1 - G_2(b_1)) = V_1 - V_2\]

where \(V_1 > V_2\)

where \(V_i\) is the value of prize, \(b_i\) is the bid, and \(G_j(b_i)\) is the probability that player \(j\) bids lower than \(b_i\). The basic idea of their proof is as follows. In any equilibrium, the higher valuation player, player 1 (he), can always bid the lower valuation \(V_2\) and win getting with \(V_1 - V_2\) certainty. However, player 1 cannot get more than this with certainty by bidding lower than \(V_2\), since then, the lower valuation player, player 2 (she), could always

---

7 Though a mixed strategy equilibrium is conceptually more difficult to game theoretically analyze, subjects need not find it more difficult to implement in experiments. A mixed strategy equilibrium can be interpreted as pure strategy equilibrium in which players are unsure of what strategy the opponent chooses (Aumann and Brandenburger 1995) or where bidders make exogenous errors (Harsanyi 1973). The subject's choice conditional on uncertainty in effect randomizes. Thus, the mutual indifference condition of a mixed strategy equilibrium can be met by common knowledge of uncertainty of opponent's bid. We discuss the possible effects of errors and the restrictiveness of the complete information assumption after the main results.
bid slightly higher and beat player 1. In a mixed strategy equilibrium, player 1 must be indifferent on the interval \([0, V_2]\), getting an expected prize of \(V_1 - V_2\). The lower valuation player must get at least 0 in equilibrium since she can always bid that for sure. However, she cannot get more than zero for sure because player 1 would then bid slightly higher and always win. Thus, player 2 must be indifferent on the interval \([0, V_2]\), getting an expected prize of 0. The same argument goes through in the risk averse case. But now, player 1 gets the utility of \(V_1 - V_2\) and player 2 gets the utility of 0 in equilibrium.

\[
U_2(V_2 - b_2)G_1(b_2) + U_2(-b_2)(1 - G_1(b_2)) = U_2(0)
\]

\[
U_1(V_1 - b_1)G_2(b_1) + U_1(-b_1)(1 - G_2(b_1)) = U_1(V_1 - V_2)
\]

Another way to think about this is first, we can easily rule out any pure strategy equilibrium. Nash’s theory for concave continuous utilities implies that there must be at least one mixed strategy equilibrium. In a mixed strategy equilibrium, players must make each other indifferent between what they can get for sure, their certainty equivalent (the right hand side of the equation), and what they can get in the gamble (the left hand side).

Therefore, the cumulative distribution functions (CDF) of equilibrium strategies or “bidding function” are:

\[
G_1(b) = \frac{U_2(0) - U_2(-b)}{U_2(V_2 - b) - U_2(-b)}
\]

\[
G_2(b) = \frac{U_1(V_1 - V_2) - U_1(-b)}{U_1(V_1 - b) - U_1(-b)}
\]

Note that because player 1 has the higher valuation, \(G_1(b)\), his equilibrium strategy, is a decreasing function of \(V_2\). Thus, the higher valuation player’s bid (from now on understood as expected or mean bid) is an increasing function of the lower valuation player’s valuation. Correspondingly, player 2’s bid decreases with \(V_1\), the higher valuation player’s valuation, but increases with \(V_2\). Furthermore, both players’ bids decrease with the increase in the opponents’ risk aversion only. We build up to this counterintuitive result first by noting that one’s own risk aversion only affects the opponent’s bid, not one’s own. To see why, note that when the bidder’s risk aversion increases, the mixed strategy to keep the opponent indifferent remains unchanged. However, the opponent’s strategy must change to compensate for the greater risk aversion of the bidder. Hence, if the opponent’s risk aversion increases, the bidder’s bid
must change. We now give an abbreviated proof of the effect of changing risk attitudes on bids.

**Theoretical Result 2:** Separating risk attitude and valuation.

We want to show that if the opponent’s risk aversion increases, the bidder must bid lower. The algebraic expression of the equilibrium bidding functions $G_i(b)$ for $i=1$ or $2$ above allows for an easy proof. Note that $G_i(b) = \text{ratio of the utility of the opponent’s certainty equivalent over his/her utility of the value of winning, both measured from the utility of losing.}$ If the opponent’s risk aversion increases, i.e., becomes more concave, then the utility of winning must increase even less proportionally than the utility of the certainty equivalent, an intermediate point. (That’s because concave functions increase at a decreasing rate. Greater concavity means that this increasing rate must decrease even faster.) Therefore, the ratio must go up and bids must decrease.

To make this claim more intuitive, we now give a graphical analysis in Figure I to show what happens when the opponent’s risk aversion increases. In a mixed strategy equilibrium for an all-pay auction with complete information, players are indifferent among all bids; all bids give them the same expected value. In particular, they are indifferent to the bid which gives them their maximum minimal value 0 for sure for the lower valuation player, (or the difference in valuations for the higher valuation player) which would be given by a bid of 0 (or $V_2$, respectively). This is their certainty equivalent (CE in Figure I) of the gamble.

Thus, for the lower valuation player (player 2 in our example), in a mixed strategy equilibrium, where she gets an expected value

---

8 For differentiable functions, $g(x)$ is more concave than $f(x)$ iff $f''(x) > g''(x)$
from all other bids, indifference to 0 requires that the expected value from all other bids should also be 0. See part A in Figure I. Graphically, an increase in player 2’s risk aversion increases the curvature of player 2’s utility from Figure IA to IB. For concave functions, which increase at a decreasing rate, an increase in the risk aversion must increase this rate of decrease in the slope. Thus, the gap between the utility of winning for player 2, $U_2(V - b)$, and the expected utility of winning, $G_1 * U_2(V - b)$, given the equilibrium strategy of player 1, must decrease. Therefore, the probability of winning $G_1$ must increase to $G'_1$ to keep player 2 indifferent to the CE. Thus, player 1 (player 2’s opponent) must bid lower when player 2’s risk aversion increases.

**Theoretical Result 3:** Inferring valuations and risk attitudes from different auctions with different gender pairs.

The ranking of mean bids of auctions will be the same as the ranking of the cumulative distribution functions if they do not cross. Since our predictions are only for cases where valuations and risk aversion would move bids in the same direction, i.e., cumulative distribution functions do not cross, or are identical, we can use the mean bids in auctions. Using the above theory and the average bids of different combination of genders {MM, FM, MF, FF} indifferent auctions, we can infer the relative risk attitudes and valuations of different genders. The ranking of average bids would be the same as the ranking of cdfs of actual bids if these cdfs did not cross. Our predictions are only for those cases where valuations and risk aversion moves bids in the same directions, i.e., cdfs do not cross, or are identical. Here are the basic rules of inference we follow.

For convenience, we will use DTW synonymously with “valuation”, interpreting valuation as the positive term in “total valuation” on top of the money value of winning. First, if all bids are the same, no difference in either valuations or risk aversion can be inferred. Now, for same gender pairs, e.g., if FF>MM, we can rule out M having both higher valuation and lower risk aversion. In other words, M must have either higher valuation or be less risk averse. For different bidders but fixed opponent, e.g., if MF<FF, M must have the lower valuation since the effect of risk aversion is fixed by the fixed opponent F. Less intuitively, but no less crucially MF=FF iff M has higher valuation. F is the highest valuation player in FM and M is highest valuation player in MM. According
to the bidding function derived above, the highest valuation players bid is not a function of own valuation. (See Proposition 6 in Appendix A for a proof.) This artifact arising from our use of across auction bids makes the theory as a whole falsifiable for particular data. For bidders and opponents with different genders, e.g., if MF>FM, either M has higher valuation or F has lower risk aversion. The reasoning can be seen in the following. We rewrite the equilibrium bidding function for F against M when F has a lower valuation than M. Here $MU_i$ stands for the marginal utility of player $i \in \{M,F\}$.

$$G_{FM}(b) = \frac{U_M(V_M - V_F) - U_M(-b)}{U_M(V_M - b) - U_M(-b)}$$

$$= \frac{U_M(0) + MU_M(V_M - V_F) - U_M(-b)}{U_M(V_F - b) + MU_M(V_M - V_F) - U_M(-b)}$$

$$> \frac{U_M(0) - U_M(-b)}{U_M(V_F - b) - U_M(-b)}$$

$$\frac{U_F(0) - U_F(-b)}{U_F(V_F - b) - U_F(-b)} = G_{MF}(b)$$

The complete set of implications is below in Tables 2 and 3. Proofs are in Appendix A.

One important advantage of only attempting to identify relative valuations and risk attitudes is that it does not require players to know their opponent’s absolute valuations and risk attitudes. Subjects need only have different reactions to the same opponent according to their own gender, and react differently to different opponents according to the opponent’s gender. Stochastic dominance of reactions is sufficient for differences in the averages bids. Differences in average bids across gender pairs of opponents merely need be in a consistent direction for us to identify risk attitudes and valuations.

Our methodology of using orderings across auction allows us a way to estimate the falsifiability of our theory. Recall that we rank the results of four auctions: MM, FM, MF and FF. These can be selected two at a time to yield six pairs. Each pair can have three orderings: $<, =, >$. The number of unique possible full orderings is: $3^6 = 729$. To calculate the number of relations which are implied by transitivity alone, we note that transitivity requires three elements. We can choose three things from four in four ways. The pairs of
relations consistent with transitivity are: > >, < <, < =, > =, = >, = <, and = =. Hence, the total number of transitive relations is 4*7 = 28. Thus, the total number of independent relations is 729-28=701. However, only 40 orderings are consistent changes in MNEs due to changes in valuations and risk attitudes. (See Tables 2 and 3.) Thus, the probability of our data being consistent with the theory by chance conditional on significant ordering = 40/701=0.06. This is the upper bond of significance since it is conditional on finding any significant full ordering. To get the unconditional probability of a particular significant ordering by chance, the 0.06 must be multiplied by the probability of getting any significant ordering at all by chance.

III. Experimental Design

We recruited a total of 582 subjects. The first 156 (78 of each gender) from UT were to test our initial hypothesis. 92 of these gave us what seemed to be preliminary confirmation that men had higher DTW. This was disconfirmed when we added 64 more subjects in UT. See Figures IV and V in Appendix E (available in supplementary material). All these subjects were recruited with posters around UT. For our 2nd set of studies, where we controlled for selection effects, one coauthor, (Charlie Chen) contacted the class monitors (student elected teaching assistants who don’t grade or lecture) at SZ (economics majors) and UT (Tsinghua Law School for our UT within school treatments and Peking University Law School for our UT across school treatments) to see if they would agree to announce the possibility of participating in the experiment at the end of class. These monitors sent all their students a message asking them to stay for 10 minutes longer after class to participate in our paid experiment on the following day. They also asked students not to leave unless they had an emergency. We recruited 416 subjects (201 male, 215 female), 213 of them (120 male, 93 female) from SZ and 203 of them (81 male, 122 female) from UT. We did not have perfect balance in genders because monitors were given envelopes to hand out based upon estimates of the gender mix in their area of the lecture hall. 19 (2 male, 17 female) in SZ out of 232 and 21 (18 male, 3 female) out of

---

9 We decided on these across-school treatments (UT against SZ) after seeing the within UT data which corroborated the “SZ hypothesis”. See Figure III.
224 in UT from each class were dropped from our pool because the class monitors ran out the envelopes we had given them. We thus had no endogenous self-selection. The gender compositions were 52.6% male and 47.4% female in SZ, 44.2% male and 55.8% female at UT.

Each envelope contained a bidding sheet with instructions in which subjects were informed that they had 10 CNY $^{10}$ and could bid for 10 CNY in the all-pay auction. The instructions are in Appendix C. In the cases in which we did not have gender balance, unmatched players faced a random draw from 0-10 CNY. Subjects were paired in randomly, secretly and anonymously in the following treatments.

The complete information treatments were:

1. With same gender opponent: (M,M) and (F,F), where the first coordinate is the bidder and the 2nd coordinate is the opponent. We will write these pairs without brackets for short.
2. With opposite gender opponent: MF and FM.
3. Subjects always knew the school of the opponent. However, to avoid clutter, we put SZ or UT before the gender only when bidding is with an opposing school’s student.

The incomplete information controls were MC and FC.

On the bidding sheet (in Appendix G available in supplementary material), they could mark a bid ranging from 0-10 CNY in ½ CNY increments. The winner got the prize of 10 CNY. Following Gneezy and Smorodinsky (2006), we split the prize if there was a tie. Bids were necessarily discrete since money was discrete. Our budget also necessitated a bidding cap of 10 CNY. Discrete bids with caps should decrease equilibrium bids slightly and should not have systematic effects on relative bids. See Appendix D (available in supplementary material) for the theory. As mentioned in the introduction, to allow for comparison with an NV (2007) type designs, we had “exit option”; bids of zero always gave subject endowment of 10 CNY. The only effect that this should have on equilibrium bids is to shift them away from zero to the next highest bid of ½ CNY since these zero bids will lose the value of ties. Our results are based upon inequalities of average bids generally above 4 CNY bids, so should not be affected.

--

$^{10}$ As a benchmark, student assistants make 10-15 CNY per hour.
We gave subjects 10 example bids and payoffs (see Appendix C), allowing 2 mins for questions; no one asked. There was a place on the bidding sheet for students to write down their name and bank account information. The instructions told students to put the bidding sheet back into the envelope. We transferred payment to their account after we finished all sessions of the experiment\textsuperscript{11}. See Table 1 for a detailed breakdown of subjects for our 2\textsuperscript{nd} set of experiments with whole class recruitment.

### Table 1: Number of subjects per treatment

<table>
<thead>
<tr>
<th>(SZ) Treatments</th>
<th>Number of Subjects</th>
<th>Date</th>
<th>(UT) Treatments</th>
<th>Number of Subjects</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZ-MM</td>
<td>26</td>
<td>Sep 26\textsuperscript{th} 2011</td>
<td>UT-MM</td>
<td>16</td>
<td>Oct 6\textsuperscript{th} 2011</td>
</tr>
<tr>
<td>SZ-FF</td>
<td>22</td>
<td></td>
<td>UT-FF</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>SZ-MF</td>
<td>25</td>
<td></td>
<td>UT-MF</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>SZ-FM</td>
<td>25</td>
<td></td>
<td>UT-FM</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>SZ-MC</td>
<td>23</td>
<td></td>
<td>UT-MC</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>SZ-FC</td>
<td>23</td>
<td></td>
<td>UT-FC</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>(SZ-M, UT-M)</td>
<td>23</td>
<td></td>
<td>(UT-M, SZ-M)</td>
<td>23</td>
<td>Oct 19\textsuperscript{th} 2011</td>
</tr>
<tr>
<td>(SZ-M, UT-F)</td>
<td>23</td>
<td></td>
<td>(UT-F, SZ-M)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>(SZ-F, UT-F)</td>
<td>23</td>
<td></td>
<td>(UT-F, SZ-F)</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

### IV. Main Results

Figure II shows the across gender bidding behavior within each school. Within SZ, women bid higher against men (FM=7.4 CNY) than men against women (MF=4.8 CNY). In contrast, within UT, women bid lower against men (FM=3.7 CNY) than men against women (MF=5.2 CNY). These across school differences are significant at the p<0.01 level of significance using the Hotelling joint test. See Figures VI and VII Appendix F (available in supplementary material) for the distribution of these bids. Figure II also shows the ‘same gender’ bidding behavior within each school. Within SZ, women and men bid about the same against the same gender opponent: FF=6.8 CNY and MM=6.2 CNY. In contrast, within UT, women bid higher against each other (FF=6.7 CNY) than

\textsuperscript{11} To avoid possible trust issues at SZ, with which neither of us were affiliated, we gave the money to class monitors to give to subjects.
men against men (4.8 CNY). From the drop in MM and FM from SZ to UT, it seems that the main change as the tier of the school increased was that males in UT were much less competitive than at SZ. See Figures VI and VII Appendix F (available in supplementary material) for the distribution of bids. However, the bidding behavior of SZ students against UT women in Figure III shows that women were at least perceived to be different in top graduate schools. Figure II includes the within school bidding behaviors when the bidder did not know the gender of the opponent. Within SZ, women bid between the perfect information cases at FC=7.3 CNY, as one would expect. In contrast, men bid lower than either perfect information cases at MC=4.5 CNY. Within UT, women and men bid about the same: MC=6.1 CNY and FC=6.0 CNY, but again, men bid higher than in either of the perfect information cases. We are not sure how to interpret male behavior in either SZ or UT for the imperfect information cases. In any case, none of these differences between male bidding with incomplete information and the closest point of male bidding with perfect information are significant.

Figure II: Incomplete information bidding behavior within SZ and UT.

Figure III shows how SZ women and men bid higher against UT women, (SZ-F, UT-F)=7.7 CNY, (SZ-M, UT-F)=6.5 CNY, than against UT men, (SZ-M, UT-M)=5.1 CNY. Due to our budget constraint, we did not have (SZ-F, UT-M) treatment. See Figure VIII Appendix F (available in supplementary material) for the distribution of bids.
Figure III: SZ bid higher against UT women than against UT men.

Figure IV shows how UT women bid lower against SZ men, (UT-F, SZ-M)=4.8 CNY, than against SZ women (UT-F, SZ-F)=4.9 CNY.

The rough pattern of across school bidding, ignoring statistical significance, is as follows. Everyone bid higher against males at SZ. Everyone bid higher against females at UT. See Figure IV for all the data.

The payoffs for all players are in Figure X in Appendix H (available in supplementary material). UT males (UT-M in the left most column) made the most money with an average payoff of 11.4 CNY when they bid against UT females. This payoff is
significantly higher (p=0.68%) than UT female bid against themselves (UT-FF earn 8.1 CNY). However, as can be seen from Figure IV, though UT males tend to bid low, UT females bid even lower against them at 3.7 CNY. Furthermore, women do not always earn less than men. In SZ, women earn significantly higher than men in the incomplete information treatment: SZ-MC=7.5 CNY < SZ-FC= 10.8 CNY (p=0.05%). In fact, for the subject pool of our main results (416 subjects) women earned 9.4 CNY while men earned 9.2 CNY on average (p=0.35%). If we include both our initial and main subject pool (582 subjects), men and women earned 9.4 CNY. We discuss the possibility of overbidding by women after these data results.

A. Identification of DTW and Risk attitude

Below in Table 2, we worked out all of the predictions of standard all-pay auction theory for each gender pair (left most column) given assumptions of their risk attitude (R) and valuations (V) (top of columns 2-5). V:M>F denotes that the valuation of males is greater than that of females. R:M>F denotes that the risk aversion of males is greater than females. Table 3 shows the theoretical predictions when valuations and risk aversion can be different across genders. The theoretical prediction for the relation between bids when women have higher DTW (V:M<F) is marked by red boxes. This uniquely predicts FM=MM and MF<FF. These predictions are marked in the smaller boxes within the larger boxes. The predictions marked by small boxes only are not consistent with any of the other columns. They are only consistent with these columns marked by larger red boxes which show all of the orderings where women having higher valuations then men.
Data Result 1: Women have higher desire to win than men.

Table 2: Theoretical predictions for the a) different valuations, risk neutral, b) same valuation different risk aversion cases.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Diff valuation (V), risk neutral</th>
<th>Same valuation, diff risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V:M&gt;F</td>
<td>V:M&lt;F</td>
</tr>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
<td>=</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
</tbody>
</table>

Table 3: Theoretical predictions for the different valuations and different risk aversion case.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Different valuation (V), different risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;=&gt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>&lt;=&gt;</td>
</tr>
</tbody>
</table>

The pooled data showing that women have higher DTW is shown in Table 4. FF-MF is significantly different from zero at less than the 1% level, while the difference FM-MM is not significantly different from zero. These are one tailed tests. P-values are always from the Mann-Whitney (MW) test.

Table 4: Pooled data showing that women have higher DTW.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>SZ</th>
<th>SZ-across</th>
<th>UT</th>
<th>UT-across</th>
<th>Pooled</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF-MF</td>
<td>&gt;0 (15 percent)</td>
<td>&gt;0 (25 percent)</td>
<td>&gt;0 (8.7 percent)</td>
<td>&gt;0 (p=0.03 percent)</td>
<td>&gt;0</td>
<td></td>
</tr>
<tr>
<td>FM-</td>
<td>&gt;0 (28 percent)</td>
<td>&lt;0 (42 percent)</td>
<td>&gt;0 (45 percent)</td>
<td>&gt;0 (p=32)</td>
<td>=0</td>
<td></td>
</tr>
</tbody>
</table>
Data Result 2: Everyone bid higher against UT women. Everyone bid lower against UT men.

In Table 5, we show all of the pairs relevant to the prediction that everyone bids higher against UT women than UT men. Here we merely combined all the data into pairs which should yield a positive value if the prediction were correct. We will write the bidder opponent pair as a function of the school. (UT-M,UT-F) – (UT-M,UT-M) should be greater than zero if UT men bid higher against UT women than against themselves. (SZ-M, SZ-M) – (SZ-M,UT-M) should also be greater than zero if SZ men bid higher against UT men than they bid against themselves. (UT-M,SZ-M) – (UT-M,UT-M) should be greater than zero if UT men bid higher against SZ men than against themselves. (UT-F,UT-F) – (UT-F,UT-M) should be greater than zero if UT women bid higher against themselves than against UT men. We exhibit the level of significance of the one sided MW in the right most column. The p-value is 0.8%. Similarly, (SZ-F,UT-F) – (SZ-F,SZ-F) should be greater than zero if SZ women bid higher against UT women than against themselves. (UT-F,UT-F) – (UT-F,SZ-F) should be greater than zero if UT women bid higher against themselves than against SZ women. (SZ-M,UT-F) – (SZ-M,SZ-F) should be greater than zero if SZ men bid higher against UT women than against SZ women. The p-value for these differences is 0.1%. This result is not consistent with UT women being less rational than UT men and bidding too much. Again, the lowest bid FM=3.7 CNY came from UT women against UT men.

Table 5: Everyone bids higher against UT women, lower against UT men.

<table>
<thead>
<tr>
<th>Relevant pairs</th>
<th>One sided MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>(UT-M,UT-F) – (UT-M,UT-M)</td>
<td>p=0.8 percent</td>
</tr>
<tr>
<td>(SZ-M, SZ-M) – (SZ-M,UT-M)</td>
<td>p=0.1 percent</td>
</tr>
<tr>
<td>(UT-M,SZ-M) – (UT-M,UT-M)</td>
<td></td>
</tr>
<tr>
<td>(UT-F,UT-F) – (UT-F,UT-M)</td>
<td></td>
</tr>
<tr>
<td>(UT-F,UT-F) – (UT-F,SZ-F)</td>
<td></td>
</tr>
<tr>
<td>(SZ-M,UT-F) – (SZ-M,SZ-F)</td>
<td></td>
</tr>
</tbody>
</table>
**Data Result 3:** SZ women have higher desire to win than SZ men, and might be more risk averse.

Tables 6 and 7 show how we inferred that SZ women had higher DTW than SZ men and might be more risk averse. First, note from Figure IV that FM=7.4>MF=4.8 (p=5%). This eliminates columns 2 and 5 in Table 6 and column 3 in Table 7, ruling out the case where men have higher DTW and are more risk averse. Furthermore, FF=6.8>MF=4.8 (p=15%) is not consistent with column 4 in Table 6, where women are more risk averse and have the same valuation as men. Another way to see this is: fixing the opponent fixes the effect of risk attitude. The fact that SZ women bid higher against SZ women than SZ men against SZ women implies that SZ women have higher DTW. Column 3 in Table 6 and column 4 and 5 in Table 7 are what remain. These are consistent with women having higher DTW. However, MM=6.2=FF=6.8 is not consistent with column 5 in Table 7. Column 4 in Table 7 is uniquely consistent with all of the data, though some of the data which identify it are not significant. Thus, the unique intersection of these sets suggests women have higher DTW and are more risk averse: V:M<F,R:M<F.

Table 6: Identification for SZ: a) different valuations and risk neutral, and b) same valuation, different risk aversions cases.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Diff valuation (V), risk neutral</th>
<th>Same valuation, diff risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V:M&gt;F</td>
<td>V:M&lt;F</td>
</tr>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>&lt;</td>
<td>MF=4.8&lt;MM=6.2</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
<td>FM=7.4&lt;MM=6.2</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
<td>MF=4.8&lt;FF=6.8</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>&gt;</td>
<td>MF=4.8&lt;FM=7.4</td>
</tr>
</tbody>
</table>
Table 7: Identification for SZ: different valuations and different risk aversions case.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Different valuation (V), different risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;=&gt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>MF=4.8&lt;MM=6.2</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>MF=4.8&lt;FM=7.4</td>
</tr>
</tbody>
</table>

**Data Result 4:** SZ believed that UT women have higher desire to win or lower risk aversion than UT men.

Assume SZ students believe they have higher DTW, then we can make the following inferences. Recall from the theory section that the stronger (higher valuation) player’s bid increases with the weaker player’s valuation. Furthermore, recall that a less risk averse opponent will increase the bidder’s bid. Thus, SZ men and women bidding higher against UT women than against UT men implies that they believed that UT-women were either less risk averse or had higher DTW than UT men (p=8%). See Figure IV. This implication makes a within school prediction at UT shown in the large red boxes in Tables 8 and 9. These boxes include the columns where either men have lower DTW (V:M<F) or men are more risk averse (R:M>F), and excludes the columns with the assumption V:M>F and R:M>F in Table 9. If within school results are consistent, then they should be within the included columns. This is indeed what we see. This suggest across gender signaling. Furthermore, consistent with the mentioned assumption that SZ students have higher valuations than UT students: UT women bid (4.9) lower against SZ women (7.7) (p=4%), and UT men bid (5.1) lower against SZ men (3.9) (p=40%).

Data Result 5: UT women have both higher desire to win and lower risk aversion than UT men.

MF=5.2<FF=6.7 (p=9%) is only consistent with the 3rd column of Table 8 and the 4th and 5th columns of Table 9. MF=5.2>FM=3.7 (p=12%) is only consistent with column 5
of Table 9, which is the column that is uniquely consistent with all the data. This is also supported by FF=6.7>MM=4.8 (p=4%) which ruled out column 4 in Table 8.

Table 8: Identification for UT: a) different valuations and risk neutral, and b) same valuation, different risk aversions cases.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Diff valuation (V), risk neutral</th>
<th>Same valuation, diff risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V:M&gt;F</td>
<td>V:M&lt;F</td>
</tr>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
<td>=</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
<td>MF=5.2&lt;FF=6.7</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
</tbody>
</table>

Table 9: Identification for UT: different valuations and different risk aversions case.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Different valuation (V), different risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;=&gt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>&lt;=&gt;</td>
</tr>
</tbody>
</table>

The across school bidding of UT women also corroborated the pattern of bids of SZ students against them and UT men. UT-women bidding lower against SZ women (4.9) than themselves (6.7) (p=6%) implies that they believed SZ women were either more risk averse or had lower DTW than UT women. This is consistent with the estimate of SZ women against UT women mentioned just above. Furthermore, UT women bid higher against SZ men (4.8) than UT men (3.7) (p=23%) weakly implies that they believed that SZ men were either less risk averse or had higher DTW than UT men. However, we did not match SZ men against UT men due to our budget constraint. We do not have data to confirm this estimation. In any case, these bids from UT women do not have within
school implications for SZ, because unlike SZ bids against UT men and women, these relate UT women to SZ women and UT men to SZ men.

**Data Result 6:** The simultaneous test of: a) the assumption that SZ women have higher desire to win and are more risk averse than SZ men, and b) the assumption that UT women have higher desire to win and are less risk averse than UT men shows high significance.

In the above, we used separate data from different combinations of genders to identify which of the 4 combinations of DTW and risk attitudes were most likely. However, we can also test all of the data within each school against the null simultaneously. We first test the hypothesis which we indentified above in Data Result 3, that SZ women have higher DTW and are more risk averse (V:M<F, R:M<F in column 4 of upper part of Table 10). Note that the first two rows of this column are not predictive. The 4\textsuperscript{th} row predicts equality, but that cannot be tested by a test of significance. However, the 3\textsuperscript{rd} predicts MF<MM, the 5\textsuperscript{th} predicts MF<FF, and the 6\textsuperscript{th} predicts MF<FM. The one tailed MW test at the bottom of Table 10 shows a level of significance that is less than 1% (0.36%). We next test the hypothesis that UT women have higher DTW and are less risk averse (V:M<F, R:M>F in column 5 of upper part of Table 10). Note that the 3\textsuperscript{rd} and 6\textsuperscript{th} rows of this column are not predictive. Again, the 4\textsuperscript{th} row predicts equality, but that cannot be tested by a test of significance. However, the 1\textsuperscript{st} predicts MM<FF, the 5\textsuperscript{th} predicts MF<FF, and the 5\textsuperscript{th} predicts MF<FM. The one tailed MW test shows a level of significance that is less than 1% (p-value=1.8e-05). When we pool all of the data, we get an even greater level of significance <1e-06. This indicates that our model fits the data very well.
Table 10: Pooled within school data.

<table>
<thead>
<tr>
<th>Gender Pairs</th>
<th>Different valuation (V), different risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM VS. FF</td>
<td>&gt;</td>
</tr>
<tr>
<td>FM VS. FF</td>
<td>&lt;=&gt;</td>
</tr>
<tr>
<td>MF VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>FM VS. MM</td>
<td>&lt;</td>
</tr>
<tr>
<td>MF VS. FF</td>
<td>=</td>
</tr>
<tr>
<td>MF VS. FM</td>
<td>&lt;=&gt;</td>
</tr>
</tbody>
</table>

Treatments | Pairs | MW (one tail) | Pooled |
-----------|-------|--------------|--------|
SZ within  | MF < MM | MF < FF      | MF < FM | 0.0036  |
UT within  | MM < FF | FM < FF      | MF < FF | 1.8e-05 | <1e-06  |

**Data Result 7:** SZ women have higher DTW than UT women but are more risk averse.

We only have sufficiently significant results for SZ females and UT females. See Figure V for the p-values.

![Figure V: Female within and across treatments, where S=ST and U=UT.](image)

We can now use a modified version of our original rules of inference for across school inferences of risk attitudes and valuations in Table 11. To be able to talk about UT women against UT women bidding, and SZ women against SZ women bidding, we now use U for UT and S for SZ. From the fact that SU=7.7>US=4.9 (p=0.6%),
US=4.9<SS=6.8 (p=6.4%), US=4.9<UU=6.7 (p=6%), we can exclude everything except two possibilities in Table 11: a) V:S>U, assuming different valuations and risk neutrality, and b) V:S>U, R:S<U, V:S>U, R:S>U, assuming different valuations and different risk aversions. However, SS=6.8 ~ UU=6.7 is inconsistent with the prediction that SS > UU for the risk neutral assumption. This fact is supported by SU=7.7 > SS=6.8 (p=38%) which is weakly inconsistent with the SU < SS prediction of risk neutral assumption. We therefore conclude: SZ women have higher DTW than UT women but are more risk averse.

Table 11: Female within and across school data

<table>
<thead>
<tr>
<th>Gender pairs</th>
<th>Diff valuation(V), risk neutral</th>
<th>Same valuation, diff risk aversion (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V:S&gt;U</td>
<td>V:S&lt;U</td>
</tr>
<tr>
<td>SS VS. UU</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>US VS. UU</td>
<td>US=4.9&lt;UU=6.7</td>
<td>&lt;</td>
</tr>
<tr>
<td>SU VS. SS</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>US VS. SS</td>
<td>US=4.9&lt;SS=6.8</td>
<td>=</td>
</tr>
<tr>
<td>SU VS. UU</td>
<td>SU=7.7~UU=6.7</td>
<td>&lt;</td>
</tr>
<tr>
<td>SU VS. US</td>
<td>SU=7.7&gt;US=4.9</td>
<td>&lt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender pairs</th>
<th>Different valuation, different risk aversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS VS. UU</td>
<td>&gt;</td>
</tr>
<tr>
<td>US VS. UU</td>
<td>US=4.9&lt;UU=6.7</td>
</tr>
<tr>
<td>SU VS. SS</td>
<td>&lt;</td>
</tr>
<tr>
<td>US VS. SS</td>
<td>US=4.9&lt;SS=6.8</td>
</tr>
<tr>
<td>SU VS. UU</td>
<td>SU=7.7~UU=6.7</td>
</tr>
<tr>
<td>SU VS. US</td>
<td>SU=7.7&gt;US=4.9</td>
</tr>
</tbody>
</table>

Data Result 8: There are no discernible gender differences in the exiting rate (rate of 0 bids) in either gender, in either SZ or UT, or in the across school bids. With the exception of FF at UT, there is no discernible gender differences in maximizal bid rates.
See Tables 12-14. The exiting rates for each gender were less than 10% of the average of 25 subjects per treatment. See Table 1 above for the exact number of subjects per treatment.

**In SZ**

Table 12: 0 CNY and 10 CNY bids at SZ per gender for within school treatments.

<table>
<thead>
<tr>
<th>Treatments</th>
<th>MC</th>
<th>FC</th>
<th>MM</th>
<th>FF</th>
<th>MF</th>
<th>FM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0s</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>10s</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>23</td>
<td>26</td>
<td>22</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

**In UT**

Table 13: 0 CNY and 10 CNY bids at UT per gender for within school treatments.

<table>
<thead>
<tr>
<th>Treatments</th>
<th>MC</th>
<th>FC</th>
<th>MM</th>
<th>FF</th>
<th>MF</th>
<th>FM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0s</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>10s</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>11</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>22</td>
<td>16</td>
<td>27</td>
<td>23</td>
<td>28</td>
</tr>
</tbody>
</table>

**Across schools**

Table 14: 0 CNY and 10 CNY bids at SZ per gender and UT for across school treatments.

<table>
<thead>
<tr>
<th>Treatments</th>
<th>SZ against UT</th>
<th>UT against SZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MM</td>
<td>FF</td>
</tr>
<tr>
<td>0s</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10s</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

V. Discussion

A. “Overbidding”

Many prior studies found “overbidding”, in the sense that average bids were higher than the monetary value of the prize, especially among women. (In contests: Price and
Sheremeta (2012), all pay auctions: Gneezy and Smorodinsky (2006), and common value auctions: Charness and Levin, D. (2009); Chen, Y., Katuscak and Ozdenoren (2009); Casari, Ham, and Kagel (2007); Ham and Kagel (2006). We also found that average female bids was 6.1 CNY, significantly higher than average male bids of 5.2 (p-value=1.4%) for the 416 subjects of the main results, or the average for Nash equilibrium of 5 CNY. This became 5.1 CNY and 5.9 CNY (p-value 1.7%) respectively for the total subject pools of 582 subjects. However, these averages hide conspicuous exceptions. Recall for example that the relation between across gender bids flipped highly significantly across schools: FM=7.4 CNY and MF= 4.8 in SZ but FM = 3.7 and MF = 5.2 in UT, where UT women bid lower (though not significantly) against men than men against women: FM=3.7<MF=5.2 (p=12%), or about the same MC=6.1-FC=6.0. Even at SZ, women do not bid significantly higher than men in the within gender treatments, e.g., FF=6.8-MM=6.2. In fact, though UT men made the most money, and UT women made the least money (Though not significantly. See Appendix H in supplementary material.), as Figure II showed, it was because UT women bid too low.

“Overbidding” presupposes common value which a number of our results ruled out. Higher average bids are also consistent with a non-monetary component to winning. As shown in the main results section, these variations in average bids between genders conditional on the gender of the bidder and the opponent identify differences in valuation and risk attitude across genders. In contrast to prior work, we systematically varied the combination of valuations and risks attitude by varying genders of paired bidders. We found systematic variation. To our knowledge, prior studies of contests and auctions did not exploit these possible variations because they in effect fixed the distribution of opponents subjects faced by not informing the subject of the gender of opponents.

Apart from identification of higher valuation in higher bids through systematic variation, the overbidding hypothesis would predict lower earnings for overbidders.

12 Linear regressions of bids on bidder’s gender, opponent gender’s, bidder’s school, opponent’s school and interactions can further corroborate the systematic pattern of variations. We found that women bid about 2 CNY higher than men (p=4%). However, our regression also showed that students from the higher ranked school (UT) bid lower by 2 CNY (p=6%), and that though opponent’s bid was insignificant, opponent’s gender interacted with school to lower bids by 4 CNY (p=5%). As Figure II showed, these regression findings are likely due to the fact that SZ women bid high against themselves and SZ men, but UT women bid high only against themselves and low against UT men, while UT men always bid low. As shown above in Table 5, everyone bid higher against UT women than against themselves or UT men. If overbidding was due to mistakes, this suggests that everyone believes women with higher educational attainment will make more mistakes, which seems paradoxical.
whereas higher valuations interacting with risk attitude would not necessarily predict lower earnings. Consistent with the latter, we found that women earned 9.4 CNY while men earned 9.2 CNY with our main subject pool. Both earned 9.5 CNY with our total subject pool. Women’s significantly higher bids combined with weakly higher payoffs suggest that women were no less rational in their respective auctions than men.

The overbidding hypothesis would also predict a very intuitive result which would falsify both the hypothesis that women had higher valuations, and the theoretical framework we use of perfect information all-pay auction theory. Starting with Gneezy, Niederle, and Rustichini (2003), others have found that women appeared to compete harder against each other than against men, i.e., they have higher relative DTW.

Our data seem to corroborate this finding. Female DTW against women, as reflected in FF-FM=2.76 (p=0.03%), was higher than against men: FM-MM=0.89 (p=32%). However, due to the fact that we compare bidders’ bids across two different auctions fixing the opponent, our theory does not allow us to identify this effect.

In fact, the counter intuitive implication of our theory is, if women have higher valuations than men, then FM should not be significantly greater than MM. Fixing the opponent fixes the effect of risk aversion. When women have the higher valuation, their bids are only a function of their opponent’s valuation, instead of their own. If FM was significantly greater than MM, the theory we use to show that women have higher DTW would be disconfirmed. See Proposition 6 in Appendix A for a proof.

B. Errors and possible omitted variables

There are several factors that in our view diminish the likelihood of random errors or confounds from omitted preferences like envy or spite driving our results. First, the all-pay auction is a relatively simple mechanism. To be sure of understanding, we also gave subjects 10 examples after instructions (see Appendix C), and 2 minutes to ask questions. As might be expected, no one asked questions. Prior studies showed that people could play mixed strategy equilibrium if they were familiar with the game (Hsu, Huang, and Tang (2007); Palacios and Huerta (2003); Chiappori, Levitt, and Groseclose (2002); Walker and Wooders (2001)). We expected familiarity with all-pay auctions from the prevalent use of academic exams and curved grades in China. As also mentioned in the
identification section, we only required that subjects have different reactions to the same opponent according to their own gender, and react differently to different opponents according to the opponent's gender. In particular, we required a simple binary choice -- the subject to bids higher or lower against the opposite gender than they would with their own.

Stochastic dominance of reactions is sufficient for differences in the averages bids. Differences merely need be in a consistent direction for us to get identification. Our finding of no gender difference with our bidding zero measure of desire to dropout was consistent with Zhang (2011b), which showed that Han Chinese female high school students were not significantly less competitive than Han Chinese male high school students, using NV (2007) methods. More tellingly perhaps, we found no gender differences with poster recruitment. The lack of difference indicates a lack of gender differences in systematic errors, should they have existed\(^\text{13}\). Our confidence is further increased by the high level of consistency in our data: everyone bid higher against UT women, and the nice fit between our data and theory, even for its counter intuitive implications, for example that FM=MM, discussed just above. The identification of preferences using aggregate data was highly significant, which was further confirmed by the above mentioned significant results with regressions on all bidder characteristics (gender of bidder, opponent’s gender, bidder’s school, opponent’s school and their possible interactions). Recall furthermore that the upper bound of probability of a match by chance was 0.06. Combined with probability of getting any significant ordering at all, the chance occurrence of our results is very low. As noted already for the pooled data in Table 10, the aggregate data was highly significant. In addition, if errors or other preferences were important factors, we would not expect consistency within and across schools.

\(^{13}\) We cannot rule out that spite and envy were present. Given that UT women bid highest against themselves, our results would suggest that UT women were most envious, spiteful and inequity averse to themselves.
C. Competitions, competitive attitude, and desire to win

We now discuss the possibility of using DTW in all-pay auctions as an alternative measure of competitive attitude. We begin by discussing the notion of competitive attitude and the experimental literature testing for it.

Competition is pervasive in most spheres of life. In anyone of these, we can observe among the winners a combination of ability, confidence in ability, risk attitude and possibly other factors. However, one might ask whether there is a some factor or attitude which is essentially competitive, which would generally lead to victory, should other factors be not too unequal. Could this factor or attitude explain gender disparities in labor market outcomes, e.g., that only 2.5% of top five executives for a large group of U.S. firms were women (Bertrand and Hallock, 2001), or that only 17% of partners at major law firms in US were women (O’Brien 2006)? There has been accumulating evidence that such disparities are not due only to factors outside of individual choice. Flory, Leibbrandt, and List (2010) found that women tended to select out of competitive jobs for which they were equally qualified, and therefore, labor market disparities might not be due only to discrimination. Their results confirmed a large body of laboratory experimental work initiated by Gneezy, Niederle, and Rustichini (2003). These authors found that only men’s performance increased in real tasks (solving mazes) laboratory experiments under a tournament (or “competitive”) payment scheme, where only the best received a large prize, over piece rate payment scheme, where payment was based upon output. Niederle and Vesterlund (2007), from now on NV (2007), further strengthened this ‘competitive attitude’ hypothesis for labor market disparities by changing the task to simple multiplication. They demonstrated that though there was no gender difference in performance in either piece rate or tournament payment schemes, women, to their payoff detriment, were less likely to choose the competitive over the piece-rate scheme controlling for ability, confidence in ability, and risk attitude. Numerous follow-up studies suggest the robustness of these results. See Niederle and Vesterlund (2011), from now on NV (2011); Bertrand (2011); Croson and Gneezy (2009) for surveys of this literature.

That men are more competitive than women might seem obvious. However, as mentioned, the outcomes of real life competitions are the consequence of a confluence of
factors. As will be shown by our discussion of a small part of the extensive and still growing literature, isolating competitive attitude, and therefore, proving what might seem obvious is both nontrivial and worth doing. One no less obvious possible cause for the labor market disparities is occupational preferences related to motherhood. We discuss some of the non-competitive attitude evidence for labor market disparities in Appendix I.

We begin by placing the laboratory experiments within our framework for most real life competitions. The latter generally involves three distinct stages: entry, preparation and participation. To our knowledge, Duckworth and Seligman (2006) made this distinction between preparation and participation first in their highly influential paper, which showed that girls performed better than boys in exams for which they could prepare, and that this was due to girls’ greater self-discipline\(^{14}\). Outside of their results, there are many reasons to suspect that competitive attitude could be expressed differently in these different stages. Competition is more salient and often more publically visible in the participation than in the preparation stage. Prior work would seem to have measured gender differences in either effort levels, given participation in competition (the literature initiated by Gneezy, Niederle, and Rustichini, 2003), or the desire to participate (the literature which stemmed from Niederle and Vesterlund 2007). An exception is Cotton et al. (2009), which found that males had higher performance only in the initial round of a 5-round math tournament, and no significant difference in any rounds when there was no time limit for completion, i.e., when it could be understood as an “exercise”. To our knowledge, no experimental studies have allowed subjects to either select into the form of competition or to prepare for the competition in the laboratory. The omission of preparation would be important for the external validity of prior experiments if there were gender differences in the motivation to prepare, the desire to participate, or the responsiveness to the saliency of competition, e.g., the “love of challenge”. “Sportsmanship”, less flatteringly term “bravado”, or the Bushido code of honor, which drove Japanese soldiers armed only with swords into American machine guns in World War II, suggest an ideal of male participation across many cultures. However, these ideals of participation seem specific to traditionally masculine domains, and thus, might

\(^{14}\) Gender differences in willingness to prepare and self-discipline could have an important effect upon which competitions boys and girls enter into. We are not aware of studies which allow subjects to select into competitions for which they might prepare.
not be highly predictive of higher levels of effort in preparation on the part of males for high stakes competitions like entrance exams to universities in technology driven societies, which are supposed to be predictive of the labor market outcomes in question.

Preparation or participation without intrinsic motivation in the stark language of economics are costs. There is evidence outside of the laboratory that women have a higher willingness to pay to win. Jurajda and Münich (2011), and Ors, Palomino and Peyrache (2011) found that women did not shy away from applying to more competitive schools, though their performance did suffer on higher pressure entrance exams. These authors and Niederle and Vesterlund (2010) suggest that competitive pressures might affect females more than males. Risk aversion could also be a factor. However, even here there is controversy. Paserman (2010) found that forced errors of women did not increase significantly more than men for the final round of Grand Slam tournaments. In any case, the lower rate of success for women given similar levels of applications is consistent with women have greater competitive attitude.

Our measure of competitive attitude, higher DTW, would predict higher levels of preparation and/or participation. (In principle, DTW could be conditional on entry; the people who like contests the least might also have the strongest DTW. However, separating conditional from unconditional DTW is beyond the goals of this paper.) To motivate our use of a non-real task contest to measure DTW, we now highlight some possible problems in the prior literature for getting unambiguous conclusions on competitive attitude from real task laboratory contests.

D. Nonlinear effect of risk attitude on performance

Recall that currently gender difference in competitiveness is inferred indirectly from the choice of competitive payment through differences in the size of the residuals for each gender after controlling for everything else including ability, confidence in ability, and risk attitude linearly. Reduced form controls of possible confounds are often

---

15 NV (2007) had a 4 stage design. First the subject faced piece rate incentives. Second, they were exogenously assigned to a competitive payment scheme where only the top performer in their group of 4 was paid 4 times the piece rate. Third, they could choose competitive payment over piece-rate where they would be paid 4 times the piece-rate in the competitive scheme if their current performance was higher than everyone else in their group in the 2nd stage. Fourth, they could choose competitive payment over piece rate but this time, their piece rate performance from stage 1 would have to better everyone else’s in stage 1 for them to win 4 times the piece rate. This fourth stage contained the same incentives as stage three except for the incentive effects on performance of
acceptable in experimental papers. However, without a model, it would be unclear to what extent the conclusions would be affected if the linearity assumption was not valid. In the case of gender difference experiments, Cotton, McIntyre, and Price (2011) were the first to our knowledge to have modeled these real task/effort contests as Tullock contests (where the probability of winning is the ratio of a player’s bid over all player’s bids) for players with CARA utility. They showed that ability and confidence in ability entered non-monotonically into performance. Furthermore, for CARA utility, they showed that risk aversion entered nonlinearly into performance. However, the nonlinear effect of risk attitude on performance is generalizable to other utilities in other contests because it arises from the fact that the probability of winning such contests is the ratio of one’s own efforts over everyone’s effort. The possible bias of ability and confidence in ability on measured competitive attitude when controlled for linearly is difficult to estimate due to their non-monotonicity. However, we can predict the bias of the nonlinear effect of risk aversion when it has been controlled for linearly\textsuperscript{16}. A linear regression will find a constant multiplier for risk aversion (the 4\textsuperscript{th} task in NV (2007)). A gender dummy will only introduce a new intercept for one of the genders, but not deal with the possibly different impact of risk aversion on the performance of each gender. Hence, a linear regression would average the impact of risk aversion on men’s performance (who are less risk averse, and therefore should have one coefficient) and women’s performance (who are more risk averse, and because of nonlinearity should have another coefficient)\textsuperscript{17}. Thus, the regression would bias the impact of own risk aversion upwardly for men and downwardly for women. This would increase the apparent competitiveness of men and decrease the apparent competitiveness of women. Nonlinearity and greater risk aversion also allow for the impact of opponent’s risk aversion on women’s bids to be much larger than on men’s bids. Thus, the difference in risk aversion alone could lead to the

tournament incentives. In particular, the third and fourth stage should be identically affected by confidence and risk aversion. They found a significant gender gap in stage four choice of competitive payment, which could be explained by general factors such as differences in confidence. However, they also found that the gap in the choice of competitive payment between stage three and four was much larger for women than for men, i.e., that stage four choices were less predictive, as regressors of women’s choices than of men’s choices. Since stage four contains general factors like risk attitude, what is not explained by stage four choices are from the purely tournament factors. The difference in these residuals, by this reasoning, demonstrate differences in competitive attitude.

\textsuperscript{16} Gneezy, Niederle, and Rustichini (2003) did not find a gender difference in risk aversion in their non-competitive exogenous risk treatment and hence did not need to control for its possibly nonlinear effects.

\textsuperscript{17} We think that interacting gender and risk attitude should solve the problem and have put in requests for the data of prior experiments.
observation of the pattern FM<FF<MM<MF, where the 1st coordinate is the bidder and the 2nd is the opponent, and F=female and M=male. Noise in observations would weaken the effect, perhaps resulting in observations which Cotton, McIntyre, and Price (2011) saw as evidence contrary to the hypothesis that risk aversion drove prior gender difference results: FM<FF and MM~MF in the empirical data.

The asymmetric effect of changing the gender of the opponent would explain the finding that women competed harder against other women than against men found in Sutter et al. (2009); Gneezy and Rustichini (2004); Gneezy, Niederle, and Rustichini (2003) and other papers. Consistent with the differential impact of risk aversion, Gupta, Poulsen, and Villeval (2011) found that risk attitude had a strong impact on women’s behavior but not on men’s. Dohmen and Falk (2011) even showed that gender was insignificant in the choice of tournament payment for an NV (2007) task when performance and risk attitude were controlled for.

The uncontrolled part of nonlinear risk aversion could be the reason why gender differences in altruism experiments have been found sensitive to context (Croson and Gneezy, 2009) while gender differences in competition experiments have been relatively robust. Risk enters more significantly in competition experiments and risk aversion is not very sensitive to context. However, though the effect of own risk aversion might be robust to context, the impact of opponent’s risk aversion could increase the variance of performance through the gender composition of the group of opponents which the subject faced in a particular session. This sensitivity could help explain the variance found across prior confirmatory studies summarized in Table 1 in Price (2010). It would also predict less gender differences in measured competitive attitude among children, if their across-gender risk aversions were more similar than that of adults. Indeed, the results on gender difference in competitive attitude among children are very mixed, with the majority finding none. See Appendix B.

Our discussion of the possibility that risk attitude has not been properly controlled for in prior real task experiments illustrates the hazardousness of using real task contests to measure gender differences in competitive attitude without a model. An obvious solution to this problem is to remove the real task in these Tullock contests and have effort be payments. Price and Sheremeta (released 2012) and Morgan, Orzen, and Sefton (2008)
tested for gender differences by using non-real task Tullock contest and found evidence that women had a higher valuation for winning. We also adopted the monetizeability of effort assumption necessary for the external validity of a non-real task contest. However, we simplified by using a one-shot all-pay auction (where everyone pays their bid regardless of whether they won or not) with a single opponent and observable gender instead of a Tullock contest with many opponents and unobservable gender. First, we wanted to let bidders condition bids on their own gender and that of their opponent. This would allow us to identify valuations and risk attitudes. Second, we were concerned that gender differences in calculation ability, as indicated by SAT scores (Coley 2001), and therefore confidence in ability, could drive results if we used Tullock contests. These, unlike all-pay auctions, have an optimal strategy that can, in principle at least, be explicitly calculated. Third, we were worried about possible differences in probability distortions discussed in Baharad and Nitzan (2008). Both errors and probability distortions could be exacerbated by having more than one opponent. The presence of these or possibly other problems is suggested by the fact that though women have generally been found to be more risk averse and less competitive, they are also more likely to bid higher (Price and Sheremeta (2012); Ham and Kagel (2006), Casari, Ham, and Kagel (2007), Charness and Levin (2009); Chen, Katušcák, and Ozdenoren (2009)), the opposite direction predicted by either factor. A further possible problem with the use of non-real task Tullock contests is that a global closed form solution, which separates the effect of risk attitude from valuation on bids, has not been found. Risk attitude has only been separated for local perturbations of symmetric equilibria, in particular, only when risk attitude and valuation are independent (Cotton, McIntyre, and Price, 2011). Again, without a closed form solution, it would be difficult to gauge how far the reality is from the approximation and therefore, the validity of the approximation. The problem of separating risk attitude from valuation globally for non-real task Tullock contests would apply a fortiori to real task Tullock contests, since these might have other confounds, and no less to non-Tullock contest type competitive situations for which we do not currently have models.
E. All-pay auctions to measure desire to win

As with Tullock contests, all-pay auctions have often been used to model real life contests. All-pay auction experiments have some important advantages over real and non-real task Tullock contests. Winning is only a function of relative willingness to pay (WTP), and WTP depends on DTW and risk attitude. Since there is a unique mixed strategy equilibrium, and we cannot see how subjects could calculate their optimal bids, we do not see how ability or confidence could be issues. Furthermore, we were able to develop the all-pay auction theory to completely separate the effect of risk attitude from valuation. This avoids the need to assume an approximately linear structure used in the real-task contest literature and the approximately symmetric and independent structure used in the non-real task Tullock contest literature. The second important advantage of the all-pay auctions for experiments on a limited budget is that unlike Tullock contests, where higher bids only increase the probability of winning (and quite marginally for some bids), all-pay auctions give zero or full probability of winning for any differences in bids. This always imposes the most “cut-throat” incentives. Most importantly, all-pay auctions economize on the information and rationality requirements for identification of risk attitude and valuations for each gender. Different gendered subjects merely have to react differently (on average) to different gendered opponents according to their own gender and the opponent’s gender.

A further important advantage of using all-pay auctions to measure gender differences in DTW is that it avoids having the subject make a self-conscious decision to compete. Mere bidding should reduce the saliency and therefore self/social image implications of the decision to compete. Women might feel constrained to a “ladylike” uncompetitiveness and men into a “manly” competitiveness if they deliberately compete in more traditional contests. Charness and Rustichini (2011) observed in their study of audience effects that, “Males wish to signal that they are formidable, while females wish to signal that they are cooperative.” The task sensitivity of gender differences in competitive attitude found in prior studies could be due to image concerns. For example, gender differences in competitive attitude decreased for verbal tasks (Grosse and Rieper, 2010), when stakes were high (Antonovics, Arcidiacono, and Walsh, 2009), when women were competing for other people (by negotiating harder for them) rather than themselves
(Bowles, Babcock, and McGinn 2005), in matrilineal societies (Gneezy, Leonard, and John, 2009), in Muslim countries (Fryer and Levitt, 2009), among Han Chinese women (Zhang, 2011b), and among children, where there is no consistent pattern (See Table A.1 in Appendix B for a comparison). However, as mentioned, real life competitive choices might not be saliently competitive: have no effect on social or self image. In fact, foregoing leisure to study for an exam may look modest.

In addition, though an all-pay auction may be an econometrically direct measure of competitive attitude in so far as it avoids many possible confounds in Tullock type contests, it is logically indirect. Croson and Gneezy (2009) observed that women were more sensitive to context in altruism experiments. This would also suggest a greater susceptibility to experimenter demand since experimenter expectation is a part of the context. However, in an all-pay auction, the decision to quit can be hidden in the decision to bid zero (see below in our experimental setup for details), while the choice of how hard to compete can be camouflaged as the willingness to pay for the prize, which only theory can reveal after separating out risk attitude.

Many prior experiments on gender differences in competitive attitude test for selection into competitions after subjects have already selected into the lab experiment on the assumption that the first form of self-selection would not drive the second. In follow-up work, Ong & Chen (2012) showed that while the aggregate distributions of bids for our classroom and poster recruited subjects were nearly identical for women, the cdfs of the two comparable treatments (FF and FM) revealed that classroom recruited women bid higher than poster recruited women in the FF and lower in the FM treatments. They furthermore report on the bidding behavior of subjects in a follow-up experiment in which many of subjects in this experiment self-selected out of their experiment. Their experiment was during the break after the last lecture before finals, when students might be fatigued. They found that the women who stayed in our experiment had bid higher than those who left in their experiment, who in turn bid higher (not significantly) than men who stayed. In contrast, the men who stayed had bid lower (not significantly) than those who left. Using theory developed here, they separated the effect of valuation from risk aversion to show that the women who stayed had a higher valuation for winning than the women who left. Furthermore, the women who left bid similarly to women recruited
through poster recruitment, while the men who stayed bid similarly to men recruited through posters. Their results suggest that leisure and poster recruitment selects for women who have lower valuations for winning, and therefore, could further help explain the robust finding that women are less competitive than men in laboratory experiments, but more competitive.

VI. Conclusions

We hypothesized that gender was a familiar heuristic for risk attitude and valuations in real life competitions. We informed subjects of their opponent’s genders when we tested for gender differences in “desire to win” by measuring willingness to pay to win in female (F) vs male (M), M vs F, F vs F, and M vs M all-pay auctions. We found a very tight fit with mixed strategy Nash equilibriums. Though, like prior Tullock contest and first price auction experiments, women bid higher than men, we found that their payoffs were weakly higher, consistent with strategic high bidding rather than overbidding. We derived theory to separate the effects of gender differences in valuations and risk attitude on average bids across treatments. Our data indicate that women have a higher valuation for winning than men.

The value of winning in an all-pay auction can serve as a direct measure of willingness to bear the total cost of winning, which in principle could be related to the ability to forgo leisure and prepare for real life competitions, e.g., academic exams. This measure does not have ability, confidence in ability, or risk aversion confounds, and is also less likely to have cultural context/demand confounds. The all-pay auction precludes heterogeneity in calculation ability as the driver of results. This measure of competitive attitude has the further advantage of being logically opaque, which should minimize possible demand effects based upon social or self-image. We introduce a bidding measure of quitting with these same features which should be more comparable to the standard measures of competitive attitude pioneered by Gneezy, Niederle, and Rustichini (2003) and NV (2007). Consistent with the literature on gender differences, we found that women were most likely more risk averse, but only at the mid tier undergraduate institution. In contrast to the literature, in a sample of nearly 600 subjects, we found no evidence that women were less competitive -- had less desire to win and were more risk averse. Furthermore,
when selection into the experiment was controlled for with 400 subjects, we found that women had greater desire to win, even at the mid tier school. To our knowledge, we provide the first evidence that women, in particular, aspiring professional women, are even more competitive than men. This is consistent with other empirical results (discussed in Appendix I), which showed that women were more competitive than men in the sense that they have higher levels of educational attainment despite lower ability as measured by standardized tests. The contrast between our initial where we used poster recruitment and our final study where we recruited whole classes suggests self-selection could have a significant impact on measurements of gender differences in competitive attitude. That could help explain past conflicting results in lab experiments and between lab experiments and field data. Our finding of systematically different bids in a common value all-pay auction could also help explain overbidding behavior found in prior studies.

VII. References


Khachatryan, K. "Gender Differences in Preferences at a Young Age? Experimental Evidence from Armenia." (2011).


