

Volume 40, Issue 2

Do women shy away from competition? Do men compete too much? : A (failed) replication

Curtis R. Price
University of Southern Indiana

Abstract

Niederle and Vesterlund (2007) document a large disparity between male and female choices in payment scheme in a simple addition task. This note describes a pure replication conducted at Purdue University in 2008 that fails to replicate these results. Analysis suggests that, although other studies have replicated the previous result, a difference in confidence which is correlated with student status (i.e. undergraduate, graduate or other) is the main contributor to the non-replication. These results are in line with recent research which suggests that much of the gender difference for competition is attributed to differences in confidence and risk preferences and proposes a valuable policy avenue which may help to attenuate the gender competition divide.

I would like to thank Muriel Niederle & Lise Vesterlund for supplying the raw data, instructions, and software along with helpful comments. I would also like to thank Jack Barron and Tim Cason for their support, comments, and discussions. This project was funded by the National Science Foundation grant #0752849. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. All errors are my own. Send correspondence to: crprice1@usi.edu

Citation: Curtis R. Price, (2020) "Do women shy away from competition? Do men compete too much? : A (failed) replication", *Economics Bulletin*, Volume 40, Issue 2, pages 1538-1547

Contact: Curtis R. Price - crprice1@usi.edu.

Submitted: September 18, 2019. **Published:** June 02, 2020.

1. Introduction

The purpose of this paper is to report the results of a pure replication of Niederle and Vesterlund (2007), hereafter NV. As a pure replication, the design and methodology are followed as closely as possible.¹ However, the results of the original study are not replicated even though the original results of NV have held up to considerable scrutiny. Analysis suggests that a difference in confidence which is correlated with student status (i.e. undergraduate, graduate or other) is the main contributor to the non-replication.

Given that the preponderance of data suggests that the result of NV is well established, this replication serves three purposes. First, replication itself is a useful endeavor and is an important facet of scientific research (Coffman and Niederle 2015; Brandt et al. 2015). Second, we analyze the data from the original study and compare to the replication data in an effort to categorize the reason for the null result. We find that there exists a correlation between the student status of the subjects (i.e. undergraduate, graduate or other) and confidence which contributes to the null result. Third, the analysis lend support to research which suggests that much, if not all, of the gender difference in competition can be attributed to differences in confidence and risk preferences (van Veldhuizen 2019; Gillen, Snowberg & Yariv 2019) and highlights a link between education and confidence (Koch, Nafziger & Nielsen 2015; Filippin & Paccagnella 2012; Benabou & Tirole 2002) that may attenuate the gender competition divide.

2. Related Studies

NV has inspired a vast amount of subsequent research most of which has replicated the original results in a variety of situations and tasks.² The interested reader is guided towards the surveys of Niederle (2014); Niederle and Vesterlund (2011); Corson & Gneezy (2009); and Dechenaux, Kovenock, & Sheremeta (2014) for a more comprehensive survey of experimental studies of contests.

The reader is particularly guided towards the excellent review by Niederle (2014) where the author surveys the many papers which replicate the original study of NV with only slight design differences. Additionally, Niederle (2014) point to several studies which have results that do not fall completely in line with NV. Among these are studies which investigate culture (e.g. Anderson et al 2013; Gneezy, Leonard & List 2009), performance under pressure (Shurchkov 2012), and the impact of age (Anderson et al 2013; Mayr et al 2012; Charness and Villeval 2009; Samak 2013).

Two recent studies have questioned the existence of the gender disparity in competitive choices. Gillen, Snowberg, and Yariv (2019) show that if a control variable is measured with error, there

¹ We make the distinction between a pure replication and that of a study which has a joint hypothesis that also considers replication of a previous result. Coffman and Niederle (2015) discuss the value of replications and term these as “exact replications” (p.94). Brandt et al. (2014) term these “close replications” (p. 218) and outline 5 facets of a convincing close replication. We have endeavored to satisfy all of these facets in this manuscript. Also see Duvendack, Palmer-Jones, & Reed (2015) where the authors consider the dearth of replications published in journals. Particularly, the authors show that less than 8% (12 out of 162) of the published replication studies in their sample were experimental studies and that most of those replications also include both replication and extension of previous work.

² Two other studies were performed at the same laboratory as this replication: Price (2012), and Cason, Masters, and Sheremeta (2010). Both replicate NV’s findings but with substantially more data and smaller gender effects.

will be a downward bias of the impact of the control variable. Since choosing a competitive compensation scheme involves both risk and (over)confidence, this result implies that the measured gender gap in studies such as NV are overstated.³ In a similar vein, van Veldhuizen (2018) cleverly develop a series of tasks allowing for direct investigation of the impact of confidence and risk preferences separately and, also, separately from the preference for competition. In this way, the experimental design controls for the facets of confidence, risk preferences and competitiveness. The author finds that preferences for competition explain negative thirteen percent of the observed gender gap in choice. When using standard econometric techniques to measure the gender gap in competitiveness the author finds that females choose the tournament less often than males, replicating NV, although, as the author points out, this measurement is subject to the criticism of Gillen, Snowberg, and Yariv (2019).

3. Experiment Design

Copies of NV’s instructions were utilized as well as their software.⁴

The experiment consists of four tasks. In tasks 1-3 subjects have five minutes to solve a series of simple addition problems.

Table I: The Four Tasks

Task 1 – Piece Rate	For each correct answer the subject is paid \$0.50
Task 2 – Tournament	Subjects are matched in groups of four. Groups consist of two men and two women. The subject with the largest number of correct problems earns \$2.00 per correct problem. Otherwise the subject earns zero for this task. In the case of ties, the winner is determined randomly.
Task 3 – Choice	Subjects have the option of choosing the tournament or the piece rate. Subjects choosing the tournament have their performance gauged against the performance in the task 2-tournament of the other three subjects in their group.
Task 4 – Submit Piece Rate	Subjects have the choice to submit their task 1 performance to be paid either by the piece rate or the tournament.

At the beginning of each session, the experimenter hands out instructions and reads them aloud. Before the first task, the experimenter informs the subjects that there are four tasks and how they are paid varies across the four tasks. No details of the payment schemes for each task are revealed until prior to the beginning of each task.

Subjects are informed that they are part of a four person group. At no point are the genders of the group members discussed. Groups consist of two males and two females seated directly next to each other so gender is observable. Subjects receive no feedback on the task outcomes (other than their own performance) until the end of the session. After finishing the four tasks, subjects are asked to guess their ranking (1st to 4th) in both task 1 and task 2. Finally, subjects are asked a series of demographic questions.

³ Gillen, Snowberg, and Yariv (2019) suggest a simple way to counter this problem is to design experiments to include more controls and to include these in the econometric investigation of the results. They go on to replicate the results of NV using their own data and show that by using an abundance of control variables the gender difference in competitive choices becomes insignificant.

⁴ The program was written in Z-tree. Fischbacher, U. (2007)

At the conclusion of the experiment, a number from 1 to 4 is drawn from a bingo cage to select the task for payment. The subject's compensation is the sum of a \$5 show-up fee, a \$7 participation fee, and the performance pay from the selected task. The average payoff was \$20.40 and the experiment lasted approximately one hour.⁵

4. Replication Results & Power Tests

In task 3, females and males choose the tournament 57% of the time. On the other hand, NV document a large disparity with male subjects entering the tournament 73% of the time and females choosing the tournament only 35% of the time. NV uses a sample of 80 observations (40 males and 40 females) and the replication includes 60 observations (30 males and 30 females) a power test yields a probability of Type II error less than 1%.⁶

In task 4, females and males choose the tournament 40% of the time. NV document that male subjects choose the tournament in task 4 55% of the time compared to only 25% of the time for females. Again, a simple power test yields a probability of Type II error of less than 1%.

5. Analysis

There are three specific areas that may cause this non-replication: Subject performance in the addition exercise, confidence, and other demographic effects.⁷

Table II: Average Performance by Gender

		Replication	NV	(p-value)
Task 1	Male	10.1	10.7	(0.522)
	Female	10.6	10.2	(0.526)
	(p-value)	(0.585)	(0.459)	
Task 2	Male	11.9	12.1	(0.738)
	Female	12.7	11.8	(0.300)
	(p-value)	(0.408)	(0.643)	
Task 2 - Task 1	Male	1.7	1.5	(0.645)
	Female	2.1	1.7	(0.527)
	(p-value)	(0.641)	(0.673)	

Table II compares performance in the first two tasks across the replication and NV and also across gender. The p-values under the averages represent a test within the replication and NV but across gender.⁸ The p-values in the rightmost column test the difference in means across the replication and NV data but within gender. We find no significant differences in performance

⁵ NV's average subject payment was \$19.80.

⁶ Although the initial intention in the replication was to have the exact number of observations as NV, it was deemed prudent to stop collecting data when it was clear that it was highly unlikely that the results of NV would replicate.

⁷ It should be noted that we cannot rule out potential bias in the replication that is (inadvertently) introduced by the experimenter himself (see Hoffman, McCabe, & Smith (1996)). In light of other studies that replicate NV's finding, the gender of the experimenter is an unlikely source of bias but it should be noted that studies such as Innocenti & Pazienza (2004) demonstrate that this is still potential source of bias. Any unconscious cue from the experimenter which may prime the subjects on gender is also unlikely to bias the results since Niederle, Segal, and Vesterlund (2009) replicate the results of NV and they explicitly inform the subjects of the gendered nature of their research hypothesis.

⁸ Comparisons of means are, unless otherwise stated, for two-sided t-tests which is done to facilitate comparison to NV.

which suggest that the disparate results between this replication and that of NV are not due to performance differences in the subject populations.⁹

Table III: Comparison of Task 1 & Task 2 Rankings

Task 1 Ranking	Male		Female		Within-Study <i>p-value</i>
	Replication	NV	Replication	NV	
1st (best)	10	20	10	8	Replication: <i>1.000</i> NV: <i>0.020</i>
2 nd	12	13	13	21	
3 rd	7	5	6	10	
4th (worst)	1	2	1	1	
Across-Study: <i>p-value</i>	<i>0.426</i>		<i>0.629</i>		
Task 2 Ranking	Male		Female		Within-Study <i>p-value</i>
	Replication	NV	Replication	NV	
1st (best)	17	30	15	17	Replication: <i>0.853</i> NV: <i>0.016</i>
2 nd	8	5	10	15	
3 rd	4	4	3	6	
4th (worst)	1	1	2	2	
Across-Study: <i>p-value</i>	<i>0.342</i>		<i>0.920</i>		

To investigate gender differences in confidence, we first consider the ranking questions which are administered after task 4. Particularly, subjects are asked to guess their within group ranking in task 1 and task 2 from 1st (best) to 4th (worst). At the point that subjects are asked to make this determination, they have no information on relative performance. Table III summarizes the guesses made by the subjects. The across-study *p*-values report the results of Fisher’s Exact test across experiments but within gender, the *p*-values in the right most column test within study but across gender differences.

The results show no discernable differences across studies but within gender. Notably, there is a difference in the within-study comparisons in both task 1 and task 2 measures. For example, in task 2, NV find evidence (*p*-value = 0.016) that males are more (over) confident than females whereas the replication fails to find a statistically significant difference (*p*-value = 0.853). If confidence is a driver of the choice to enter into a competitive compensation scheme, then we would anticipate that less males and more females in the replication would enter into the tournament relative to NV.

Task 4 choices can also be used as a metric for confidence since the decisions by subjects are similar in task 3 and task 4. To the extent that confidence affects both task 3 and task 4 decisions, we would expect that the differences in choice in task 3 across NV and this replication continue in task 4. Indeed, this is the case with NV documenting a large disparity in choices in task 4 and this replication documenting no choice disparity.

This analysis suggests that confidence is a driver of the disparate results between this replication and NV. To measure the impact of confidence to the null result, we will use traditional econometric techniques.¹⁰

⁹ Male subjects in the replication who choose the tournament in task 3 solve more problems than male subjects who choose the piece rate in task 3 in all measures of performance (task 1, task 2, and task 2 minus task 1). Female performance in these measures does not differ when we compare across task 3 choice. This implies that male subjects may be more accurate in their evaluation of their own ability. NV found no such difference for either male or female subjects. We will rely on econometric techniques to control for this possible difference.

6. Regression Results

In Table IV, we utilize an interaction term in the probit model to measure the differences across the two studies in task 3 choice.¹¹ The variable “NV” is one when a record is from NV’s data and 0 otherwise.

The top portion of the table reports estimated coefficients of the model and p-values. Column 1 controls for performance, column 2 adds controls for confidence by adding the task 2 ranking guess, and the last column further controls for confidence by adding the decision to submit the task 1 performance to a tournament.

The bottom portion of Table IV reports the predicted probabilities of choosing the tournament in task 3. Marginal effects can be calculated by taking differences of the predicted probabilities. For example, the marginal effect of being in the replication while holding gender constant and controlling for any performance differences can be calculated by comparing the predicted probability of a female in the replication in column 1 with the predicted probability of a female of being in NV. Thus, a female in the replication has a 22.4 percentage point higher probability (58.0% versus 35.6%) of entering into the tournament than a female in NV after controlling for task performance.¹² This difference is statistically significant at the 5% level which is evidenced by the p-value on the interaction term for the female and NV indicator in column 1.

The results in column 1 suggest that performance only has a small impact on the difference between the two studies.

Considering column 2, we see that the NV*female interaction is statistically significant, although now with a p-value of 0.088. Thus, the inclusion of confidence as measured by the task 2 ranking reduces the gap from 22.4 percentage points to 16.4 ($=0.717-0.553$) percentage points and also reduces the statistical significance of this difference.¹³ Thus differences in confidence, as measured by the guessed tournament ranking, do not seem to explain the entire difference across the two studies.

¹⁰ As noted earlier, Gillen, Snowberg, and Yariv (2019) show that NV may have overestimated the gender disparity in task 3. Since this replication uses the same design as NV we are unable to follow all of the recommendations of Gillen, Snowberg and Yariv(2019) for measuring the gender disparity more precisely but we did employ some of the techniques they suggest. Particularly, we also analyzed the data using a linear probability model instead of the probit model and investigated the results when we introduce controls for performance and confidence non-parametrically. The results are similar to those contained in this manuscript where we follow the analysis of NV to facilitate comparison and are available upon request.

¹¹ The introduction of interaction terms with indicator variables has the effect of parsing the data which correspond to the indicators. This could be a potential issue with a small dataset as the standard errors are normally larger since the sample sizes in each partition decrease. As a consistency check, we ran the same regressions with only one interaction term (NV * female). The results are similar with slightly lower standard errors and slightly stronger differences in the two studies. These results are available upon request.

¹² For details of using predictive margins see Williams (2012).

¹³ Following the advice of Gillen, Snowberg and Yariv(2019) we also investigated the impact of adding task 1 rankings to the analysis along with the other variables in column 2. The result increased the p-value on the NV*Female indicator to 0.096.

Table IV: Probit Models of Compensation Choice - Coefficients and Predictive Margins

	Coefficient		
	(p-value)		
	(1)	(2)	(3)
Female	-0.101	0.010	-0.004
(1 if female)	(0.767)	(0.977)	(0.990)
NV	1.227	1.634	1.804
(1 if NV's data)	(0.198)	(0.234)	(0.209)
NV*Female	-0.899**	-0.812*	-0.650
(interaction term)	(0.047)	(0.088)	(0.184)
Tournament	0.116*	0.066	0.060
(task 2 # correct)	(0.053)	(0.317)	(0.375)
NV*Tournament	-0.070	-0.077	-0.116
(interaction term)	(0.390)	(0.389)	(0.221)
Tournament - Piece Rate	0.035	-0.008	0.025
(task 2 minus task 1)	(0.645)	(0.914)	(0.778)
NV*(Tournament – Piece Rate)	0.011	0.003	0.042
(interaction term)	(0.915)	(0.978)	(0.731)
Tournament Rank		-0.483*	-0.425
(guessed rank in task 2)		(0.066)	(0.118)
NV*Tournament Rank		-0.259	-0.306
(interaction term)		(0.459)	(0.409)
Submit Piece Rate			0.358
(1 if submit task 1 to tournament)			(0.360)
NV*Submit Piece Rate			0.593
(interaction term)			(0.272)
Observations	140	140	140
Predicted Probabilities			
Male in Replication	0.619	0.713	0.761
Male in NV	0.735	0.825	0.909
Female in Replication	0.580	0.717	0.760
Female in NV	0.356	0.553	0.751

** - indicates significance at the 5% level; * indicates significance at the 10% level. The predicted probabilities are calculated for a subject who solves 12 problems in the piece rate and 13 problems in the tournament, who chooses to submit her task 1 performance to a tournament in task 4 and guesses a tournament rank of 1st

Lastly, we include the decision to submit task 1 performance to a tournament which reduces the p-value of the interaction term to 0.184, making the difference no longer significant at traditional levels. As noted before, a subject's decision to submit her piece rate performance to a tournament in task 4 can be thought of as another metric for confidence since both decisions are subject to the impact of confidence. The fact that there are no measurable differences between NV and this replication when we add this piece of information suggests that confidence is a main driver of the disparate results of the two studies. In the next section, we investigate demographic differences in the two studies to see how they contribute to the null result of this replication and also highlight how demographic effects are related to confidence.

7. Demographic effects

Table V reports a summary of the collected demographic categories.

Investigating the difference in the two studies along the lines of student status is promising because previous research has highlighted the role of confidence in education (Benabou & Tirole 2002; Koch, Nafziger & Nielsen 2015; Filippin & Paccagnella 2012).

Table V: Demographic Data by Gender & Subject Pool

Major	Replication Data			NV Data		
	Total	Male	Female	Total	Male	Female
Arts	3	1	2	4	1	3
Business	19	6	13	17	9	8
Humanities	0	0	0	11	5	6
Natural Sciences	3	1	2	2	0	2
Social Sciences	5	2	3	16	7	9
Physical Sciences	6	5	1	12	8	4
Other	24	15	9	18	10	8
Race						
Asian	11	8	3	13	6	7
Black	2	1	1	2	0	2
Caucasian	41	19	22	60	30	30
Hispanic	3	1	2	2	2	0
Other	3	1	2	3	2	1
Highest Degree Obtained						
Highschool/GED	54	25	29	55	26	29
Bachelors	4	4	0	15	7	8
Masters	1	0	1	8	6	2
Ph.D.	0	0	0	1	1	0
None	1	1	0	1	0	1
Student Status						
Undergrad	59	30	29	59	27	32
Grad	1	0	1	14	8	6
Neither	0	0	0	7	5	2
Age						
18-21	46	25	21	47	22	25
22-25	13	5	8	19	9	10
26-29	1	0	1	3	1	2
30-33	0	0	0	3	3	0
34-37	0	0	0	3	2	1
38+	0	0	0	5	3	2

In NV's study, 70% of the male undergraduate subjects choose the tournament and 41% of the female undergraduate students choose the tournament (p-value = 0.035). If we restrict the sample to only undergraduate students, we get no discernable differences in task 3 choice across NV and the replication (NV male = 70%; Replication male = 57%; NV female = 41%; Replication female = 55%). A Fisher's Exact Test across the two studies for males (females) yields a two-sided p-value of 0.410 (0.310).

Given that confidence and student status seem to both dissipate the disparate results, we will now investigate if confidence and student status are related. To investigate this possibility we will focus on the subject's task 2 guessed ranking in Table VI below.

Using NV's data, we include an interaction of female and an indicator variable for undergraduate subjects. The significant and negative coefficient on the interaction term implies that female undergraduate students are systematically more confident, as measured by the tournament ranking, than non-undergraduate females. This, along with the fact that males and females do not differ in the tournament rank in the replication, suggest the differences in the subject pools captures systematic differences in confidence across the two studies.

Table VI: Ordered Probit of Gussed Tournament Rank

Female	1.175***
(1 if female)	(0.003)
Undergrad	0.616
(1 if undergrad)	(0.254)
Female*Undergrad	-1.361**
(interaction term)	(0.047)
Tournament	-0.140**
(task 2 # correct)	(0.025)
Tournament - Piece Rate	-0.170**
(task 2 minus task 1)	(0.014)
Observations	80

*** indicates significance at the 1% level; ** indicates significance at the 5% level.

8. Conclusion

This note describes the results of a failed pure replication attempt of Niederle & Vesterlund (2007). An analysis of the data suggests that differences in confidence across the two subject pools which is related to reported student status is the main factor that contributes to this non-replication. This is in line with research which highlights the role of confidence in education and, although more research is definitely warranted, suggests a valuable policy avenue which may help to attenuate the gender competition divide.

References

- Andersen, Steffen, Seda Ertac, Uri Gneezy, John A. List and Sandra Maximiano (2013) “Gender, Competitiveness, and Socialization at a Young Age: Evidence From a Matrilineal and a Patriarchal Society” *Review of Economics and Statistics* **95** (4), 1438-1443
- Benabou, R., and J. Tirole (2002) “Self-confidence and personal motivation” *Quarterly Journal of Economics* **117** (3), 871-916
- Brandt, Mark J., Hans Ijzerman, Ap Dijksterhuis, Frank J. Farach, Jason Geller, Rodger Giner-Sorolla, James A. Grange, Marco Perugini, Jeffrey R. Spies, Anna van’t Veer (2014) “The Replication Recipe: What makes for a convincing replication?” *Journal of Experimental Social Psychology* **50**, 217-224
- Cason Timothy N., William A. Masters and Roman M. Sheremeta (2010) “Entry into Winner-Take-All and Proportional Prize Contests: An Experimental Study” *Journal of Public Economics*, **94**, 604-611
- Charness, Gary B. and Marie-Claire Villeval (2009) “Cooperation and Competition in Intergenerational Experiments in the Field and the Laboratory” *American Economic Review* **99**, 956-978
- Coffman, Lucas C. and Muriel Niederle (2015) “Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible” *Journal of Economic Perspectives* **29** (3), 81-98
- Croson, Rachel, and Uri Gneezy (2009) “Gender Differences in Preferences” *Journal of Economic Literature* **47** (2), 1-27
- Dechenaux, E., D. Kovenock and R.M. Sheremeta (2014) “A survey of experimental research on contests, all-pay auctions and tournaments” *Experimental Economics* **18** (4), 609-669
- Duvendack, M., R.W. Palmer-Jones and W.R. Reed (2015) “Replications in Economics: A Progress Report” *Econ Journal Watch* **12** (2), 164-191
- Filippin, A. and M. Paccagnella (2012) “Family background, self-confidence and economic outcomes” *Econ. Educ. Rev.* **31**, 824-834
- Fischbacher, U. (2007) “z-Tree: Zurich toolbox for ready-made economic experiments” *Experimental Economics* **10** (2), 171-178
- Gillen, Ben, Erik Snowberg and Leeat Yariv (2019) “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study” *Journal of Political Economy* **127** (4), 1826-1863

- Gneezy, Uri, K.L Leonard and J.A. List (2009), "Gender Differences in Competition: Evidence from a Matrilineal and Patriarchal Society" *Econometrica* **77** (5), 1637-1664
- Hoffman, Elizabeth, Kevin McCabe and Vernon Smith (1996) "Social Distance and Other Regarding Behavior in Dictator Games" *American Economic Review* **86**, 653-660
- Innocenti, Alessandro and Maria Grazia Paziienza (2004), "Experimenter bias across gender differences" University of Siena Department of Political Economy working paper #438
- Koch, Alexander, J. Nafziger and H.S. Nielsen (2015) "Behavioral economics of education" *Journal of Economic Behavior & Organization* **115**, 3-17
- Mayr, Ulrich, D. Wozniak, C. Davidson, D. Kuhns and W.T. Harbaugh (2012) "Competitiveness across the life span: The feisty fifties" *Psychology and Aging* **27** (2), 278-285
- Niederle, Muriel (2014) "Gender" in the *Handbook of Experimental Economics, Volume 2* by John Kagel and Alvin E. Roth, Eds., Princeton University Press, 481-553
- Niederle, Muriel, C. Segal and L. Vesterlund (2009) "How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness" *Management Science* **59** (1), 1-16
- Niederle, Muriel and Lise Vesterlund (2011) "Gender and Competition", *Annual Review of Economics* **3**, 601-630.
- Niederle, Muriel and Lise Vesterlund (2007) "Do Women Shy away from Competition? Do Men Compete too Much?" *Quarterly Journal of Economics* **122** (3), 1067-1101.
- Price, Curtis R. (2012) "Gender Competition and Managerial Decisions" *Management Science* **58** (1), 114-122
- Samak, Anya (2013) "Is there a Gender Gap in Preschoolers' Competitiveness? An Experiment in the U.S." *Journal of Economic Behavior and Organization* **92**, 22-31
- Shurchkov, Olga (2012) "Under pressure: gender differences in output quality and quantity under competition and time constraints" *Journal of the European Economic Association* **10** (5), 1189-1213
- van Veldhuize, Roel (2018) "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?" Rationality and Competition Discussion Paper Series 14, CRC TRR 190 Rationality and Competition.
- Williams, Richard (2012) "Using the margins command to estimate and interpret adjusted predictions and marginal effects" *The Stata Journal* **12** (2), 308-331