



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

How Stress Affects Performance and Competitiveness Across Gender

Jana Cahlíková, Lubomír Cingl, Ian Lively

To cite this article:

Jana Cahlíková, Lubomír Cingl, Ian Lively (2020) How Stress Affects Performance and Competitiveness Across Gender. *Management Science* 66(8):3295-3310. <https://doi.org/10.1287/mnsc.2019.3400>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Management Science*. Copyright © 2019 The Author(s). <https://doi.org/10.1287/mnsc.2019.3400>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2019, The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

How Stress Affects Performance and Competitiveness Across Gender

Jana Cahliková,^a Lubomír Cingl,^b Ian Levely^c

^a Department of Public Economics, Max Planck Institute for Tax Law and Public Finance, 80539 Munich, Germany; ^b Department of Economics, University of Economics in Prague, 130 67 Prague, Czech Republic; ^c Development Economics Group, Wageningen University, 6708 BP Wageningen, Netherlands

Contact: jana.cahlikova@tax.mpg.de,  <http://orcid.org/0000-0002-5757-6145> (JC); lubomir.cingl@vse.cz,  <http://orcid.org/0000-0001-8842-4876> (LC); Ian.Levely@wur.nl,  <http://orcid.org/0000-0003-4361-9930> (IL)

Received: September 12, 2017

Revised: May 22, 2018; August 29, 2018;
April 24, 2019

Accepted: April 26, 2019

Published Online in Articles in Advance:
July 16, 2019

<https://doi.org/10.1287/mnsc.2019.3400>

Copyright: © 2019 The Author(s)

Abstract. Because many key career events, such as examinations and interviews, involve competition and stress, gender differences in response to these factors could help to explain the labor market gender gap. In a laboratory experiment, we manipulate psychosocial stress using the Trier Social Stress Test and confirm that this is effective by measuring salivary cortisol level and heart rate. Subjects perform in a real-effort task under both tournament and piece-rate incentives, and we elicit willingness to compete. We find that women under heightened stress perform worse than women in the control group when compensated with tournament incentives, whereas there is no treatment difference under piece-rate incentives. For men, stress does not affect output under competition or under piece rate. The gender gap in willingness to compete is not affected by stress, but stress decreases competitiveness overall, which is related to performance for women. Our results could explain gender differences in performance under competition, with implications for hiring practices and incentive structures in firms.

History: Accepted by Axel Ockenfels, behavioral economics.



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Management Science. Copyright © 2019 The Author(s). <https://doi.org/10.1287/mnsc.2019.3400>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Funding: This work was supported by the Grant Agency of Charles University [Grant 1156213].

Supplemental Material: Data and online appendix are available at <https://doi.org/10.1287/mnsc.2019.3400>.

Keywords: competitiveness • performance in tournaments • psychosocial stress • gender gap

1. Introduction

Evaluating employees and job candidates often involves competition in stressful settings. Indeed, the most important events that determine career success, such as interviews, university entrance examinations, and asking for promotion, typically take place under heightened psychosocial stress. For example, a job candidate might be required to speak publicly before a committee. These situations also involve competition with peers. If men and women react differently to competition under stress, this could have implications for efficiently selecting the best candidates and picking the right incentives to motivate male and female employees. Gender-specific responses to stress and competition could also help to explain the persistent gender gap in the labor market—especially for top positions in business, government, and academia. We study how stress and competition affect men and women differently by experimentally manipulating exposure to psychosocial stress and subsequently measuring performance in a real-effort task under both

competitive and noncompetitive incentive schemes. We also test whether exposure to stress affects willingness to compete and whether this differs by gender.

Previous research shows that women are less likely to enter competitive situations than men (Niederle and Vesterlund 2007, 2011; Sutter and Glätzle-Rützler 2015; Almås et al. 2016). This has been highlighted as a potential explanation for the female wage gap because lower willingness to compete could make women less likely to enter competitive fields than men with similar ability. Experimental measures of willingness to compete partially explain female students' choices to enter less prestigious academic tracks (Buser et al. 2014). A substantial part of the gender wage gap is not because of choice of profession but because of quality of employers and advancement within fields (Card et al. 2016). This is plausibly related to lower willingness to compete, which could also make women less likely to ask for promotions or apply for jobs that have a competitive application process or competitive compensation scheme (Flory et al. 2015).

In addition to preferences for competing, gender differences in performance under competition may also explain labor market outcomes. So far the evidence is mixed. Using data from university examinations, Ors et al. (2013) find that female students perform comparatively worse when competing against peers, and similarly, Jurajda and Munich (2011) show that women do worse than their male counterparts on entrance examinations only when applying to more competitive programs. In laboratory experiments, Gneezy et al. (2003) find that competitive incentives only increase performance among men, whereas Niederle and Vesterlund (2007) show that tournament incentives lead to higher performance for both men and women. Several studies have examined performance and competition in children, with similarly contradictory results.¹ The type of task and cultural setting may also play an important role.²

In this study, we examine how performance under tournament incentives and willingness to compete are affected when individuals are exposed to a psychosocial stressor and whether the effects are gender specific.³ Stressors are environmental factors that produce a stress response (Goldstein and McEwen 2002). We experimentally manipulate stress levels by exposing subjects to a psychosocial stressor, which is a stressor that involves social interaction, evaluation, or a threat to social status. Specifically, we do so through a modified version of a standard psychological procedure, the Trier Social Stress Test for Groups (TSST-G) (von Dawans et al. 2011).⁴ We focus on the psychosocial stress because the hiring process in nearly all fields involves some degree of social evaluation (a type of psychosocial stressor).⁵ If men and women respond differently to competitive incentives under such conditions, this could have implications for the hiring process even for jobs that are not themselves particularly competitive or stressful.

Another reason for deliberately focusing on gender differences is indications from the literature that the behavioral stress response can differ by gender. Although “fight or flight” (Cannon 1932) is the dominant model for understanding how humans (and other animals) respond to an immediate perceived threat, a more recent theory proposes that females, who are less physically adapted to fight off foes and less mobile when caring for offspring, may have evolved a tendency to “tend-and-befriend” by leveraging affiliation with social groups to avoid danger in certain situations (Taylor et al. 2000, Taylor 2006). Based on this and previous literature showing that a gender gap exists in willingness to compete measured both in laboratory experiments and in the labor market (Niederle and Vesterlund 2011), we hypothesize that this difference in stress response between men and women could lead to gender differences in

how stress affects performance under competition and willingness to compete. If men respond to stress with a “fight” (rather than “flight”) reaction, they would become more competitive, and if women respond with a “tend-and-befriend” reaction, we predict that they would become less competitive. This difference in stress response between men and women therefore could lead to stronger gender differences in performance under competition and willingness to compete under stress; we designed the study to test this conjecture.

Our design consists of an economic experiment with 95 male and 95 female university students using the TSST-G to manipulate stress. Subjects were assigned to either the *stress* or *control* treatment for the duration of the experiment. We measure salivary cortisol level—a hormone related to stress⁶—and heart rate in order to confirm that the stress manipulation was successful for both genders.

Using a laboratory experiment solves two problems that make causal inference of the effect of stress on behavior in naturally occurring competitions difficult. First, it avoids problems of self-selection into competitive and stressful situations. Second, because competitive situations can cause stress (Fletcher et al. 2008, Buckert et al. 2017, Buser et al. 2017), this makes it difficult to isolate the effects of stress and competition from one another in observational data. Although these elements often occur together, understanding the distinct effects of each and how they interact could lead to more nuanced implications for designing incentive structures and evaluating job candidates.

The experiment measures the change in performance and willingness to compete under stress using a design based on Niederle and Vesterlund (2007) (the NV 2007 design). Subjects were compensated for adding up sets of four two-digit numbers within a time limit. The payment scheme varied by condition: in the baseline condition, each correctly solved problem was rewarded with a fixed piece-rate payment. This condition was repeated after the stress/control procedure to test the effect of stress on individual performance. Subjects then completed the task under a tournament incentive scheme in which payoff depends on performance relative to another randomly selected participant. After gaining experience with both types of compensation, subjects chose a linear combination of the piece-rate and tournament payment schemes following Gneezy et al. (2018), and this choice measures willingness to compete. Subjects then performed the calculation task a fourth time and were compensated accordingly. Additional tasks allow us to rule out several channels through which stress affects willingness to compete, including risk aversion and confidence.

We find that women in the stress treatment perform significantly worse in the tournament than women in the control group. Interestingly, stress alone does not affect performance because we find no treatment difference for women's performance in the piece-rate rounds. Our findings indicate that stress mutes the effect of competitive incentives on women's performance: although most women in the control group performed better under tournament incentives than when paid a piece rate, most women exposed to the stress treatment did not improve. This leads to a strong treatment difference in women's tournament performance. In contrast, we find no statistically significant treatment difference in performance for men in either the tournament or piece-rate rounds. We also find that willingness to compete is lower in the stress treatment overall, but contrary to our predictions, the exposure to the stressor does not widen the gender gap in willingness to compete. For women, the drop in competitiveness can be explained by worse tournament performance under stress. For men, we eliminate several alternative explanations and conclude that the treatment effect is preference based.

Recent work shows that stress affects decision making and preferences (Starcke and Brand 2012, von Dawans et al. 2012, Cahlíková and Cingl 2017). Although our paper is the first of which we are aware to study the causal effect of stress on performance under competition,⁷ it contributes to an emerging literature on the link between stress and willingness to compete.⁸

A few recent articles study the effect of competition on stress levels using competition tasks based on the NV 2007. Buser et al. (2017) find that engaging in competition mildly raises cortisol levels. For women only, a stronger cortisol reaction to (forced) competition is associated with higher willingness to compete. Zhong et al. (2018) confirm that competition increases cortisol levels. Buckert et al. (2017) find that competition increases subjective measures of stress and heart rate, but unlike Buser et al. (2017) and Zhong et al. (2018), they find no effect on salivary cortisol. As in Buser et al. (2017), they find that the stress reaction to (forced) competition is positively correlated with willingness to compete in women but find the opposite effect in men. Halko and Sääksvuori (2017) find that stronger heart rate variability response to tournaments (taken as a proxy for acute stress response) is positively correlated with willingness to compete in men but not in women.

Most relevant to our study, Buser et al. (2017) and Zhong et al. (2018) exogenously manipulate stress before eliciting willingness to compete. In contrast to our study, the treatment in Buser et al. (2017) consists of a physical stressor (putting a hand in ice-cold water), which can induce a different physiological and psychological reaction (Allen et al. 2014, Haushofer and

Jang 2015). This might explain why their findings differ from our findings: they observe that women (but not men) who were exposed to the stressor were comparatively *more* willing to compete. Zhong et al. (2018) experimentally induce psychosocial stress using the TSST-G and measure subsequent willingness to compete, but in contrast to this study, they find no significant effect of stress on competitiveness. They report no significant gender differences, though their study was not designed to test for this. Goette et al. (2015) also induce psychosocial stress using the TSST-G procedure and subsequently measure competitive confidence. They find no average effect on competitiveness. However, they measure only decisions to compete based on the results of past performance—in this study, we replicate this (non)result. We discuss possible explanations for differences between these studies and our findings in Online Appendix B.

A key difference in design between these studies and our study is that we measure the effect of stress on performance under forced competition, not only among those who self-select into the tournament. This allows us to make causal claims about the effect of stress on performance under competition independent of the effect of stress on willingness to compete. Although Buser et al. (2017) examine the correlation between cortisol reaction and tournament performance, they find no relationship for either men or women. They also study how randomized exposure to a physical stressor affects tournament performance but only after subjects self-select into competition, and likewise, they find no effect. Similarly, Zhong et al. (2018) observe only the performance of subjects who self-selected into (or out of) competition. We expand this literature by showing how women respond differently to competitive incentives when under stress, and we are able to do so causally by first exposing subjects to a psychosocial stressor in the stress treatment and subsequently having them perform under both piece rate and (forced) competition. This provides important insight for two reasons. First, a number of situations require individuals to compete without a meaningful choice to opt out, and understanding how stress and competitive incentives affect the performance of men and women differently is key to understanding gender differences in such settings. Second, we show that, for women, performance under stress and competition can explain the treatment effect of stress on willingness to compete that we find.

Our results imply that women perform worse when required to compete in stressful settings. Women are underrepresented in a variety of prestigious industries, high-paid jobs, and leadership positions in politics and business. These careers involve both intense competition and stress. If women know that they do not

perform well in these environments, they may decide to stay out—even if they are otherwise qualified for those positions. This is especially true when hiring involves more competition or stress than the position itself; our results suggest that this could prevent employers from selecting women who are best suited for the job. If managers introduce competitive incentives as a means of boosting employee productivity, this may have the opposite effect for women in the presence of heightened psychosocial stress.

2. Design

All subjects completed incentivized tasks that measure performance under piece-rate and tournament incentive schemes and willingness to compete. At the beginning of the experiment, subjects were informed that they would complete seven tasks in total and that two would be randomly chosen for payment. Our experimental manipulation consists of two treatment conditions, the stress treatment group and a control group, applied between subjects using the TSST-G. For a timeline of all tasks, the stress/control procedures, and cortisol measurements, see Figure A.1 in Online Appendix A.

2.1. Experimental Tasks

We measure competitiveness using a design based on the NV 2007 and Gneezy et al. (2018). Subjects completed a calculation activity twice under a non-competitive piece-rate scheme and then again under tournament incentives, after which they were asked which combination of these compensation schemes they preferred for the subsequent calculation round.

The calculation activity consisted of a series of addition problems requiring subjects to add up four two-digit numbers in each. They had two minutes per task to solve as many problems as they were able to. Subjects familiarized themselves with the calculation activity in an unpaid practice round. In subsequent rounds, correct results were incentivized according to two compensation schemes.

Under the *piece-rate* compensation scheme, participants earned 25 Czech crowns (CZK)—worth about 1 euro—per correct answer. Performance under the piece-rate scheme serves as a baseline measure of ability and effort. Subjects performed twice under the piece-rate compensation scheme: once before the stress treatment/control procedure (Task 1: piece rate before treatment) and once after (Task 2: piece rate under treatment). Comparing the within-subject differences in performance in Tasks 1 and 2 across treatments allows us to measure the effect of the stress treatment on performance, controlling for baseline differences in ability.

In Task 3, tournament under treatment, correct answers were rewarded according to the *tournament* compensation scheme: each participant was informed

that he or she would be randomly matched with another participant in the room (there were always four men and four women present) and that whoever had more correct answers would receive CZK 50 per correct answer, whereas the subject with fewer correct answers would receive nothing.⁹ In case of a tie, each participant received CZK 25 per correct answer, as in the piece-rate scheme. In all tasks, subjects were informed of the number of correct answers but were not given information on the performance of others—including whether they had won the competition. Comparing outcomes in Tasks 3 and 2 allows us to assess how competitive incentives affect performance and whether this changes by treatment and gender.

In Task 4, choice of compensation scheme for future performance, subjects chose how they would be compensated before completing the calculation portion of the task. They did so by splitting 100 points between the tournament and the piece-rate compensation schemes as in Gneezy et al. (2018). For each point invested in the piece-rate scheme, they earned CZK 0.25 per correct answer. For each point invested in the tournament compensation scheme, they earned CZK 0.5 per correct answer but only if they had more correct answers in Task 4 than another randomly selected participant and received nothing for each point invested in the tournament scheme if they answered fewer questions. In case of a tie, each point invested in the tournament account was rewarded according to the piece-rate scheme. In order to make the decision easily understandable, before making their final choice, subjects could experiment with different tournament investments, and the resulting payoffs per correct question if they won and lost were displayed.

It is important to note that the choice of compensation scheme in Task 4 cannot be driven by prosocial concerns or beliefs about who self-selects into competition because performance in Task 4 was compared with the Task 3 performance of a randomly selected subject. This information was highlighted in the instructions, and subjects knew that their decision to enter the tournament did not have payoff consequences for anyone else.

The choice of compensation scheme in Task 4 is our main measure of willingness to compete. To estimate the causal effect of stress, we compare the share of the 100 points invested in the tournament in Task 4 across the stress and control treatments. To determine the underlying mechanism, we implemented two additional tasks in which subjects competed on past performance. This isolates preferences and beliefs related to performing in a competitive environment (relevant in Task 4 but not in Tasks 5 and 6) from willingness to compete and other beliefs and preferences that are present in all three (the NV 2007).

In Task 5, choice of compensation scheme for past performance before treatment, subjects again split 100 points between the tournament and piece-rate schemes but were paid according to their performance in Task 1. Subjects were reminded that Task 1 was incentivized with the piece-rate scheme and that it took place in the first room—indicating that it was completed before the stress/control procedure. Additionally, they were reminded of how many problems they correctly solved in Task 1. The decision in Task 5 captures willingness to compete, but because the decision is made for past performance, which occurred outside the stress treatment, preferences for engaging in a competitive activity or (beliefs about) the potential negative effect of stress on performance should not be relevant.

In Task 6, choice of compensation scheme for past performance under treatment, subjects also split 100 points between the tournament and piece-rate schemes but were paid according to their performance in Task 2. Instructions for Task 6 reminded subjects of their performance in Task 2, that this task took place after the stress/control procedure, and that it was incentivized with the piece-rate scheme. Therefore, if stress negatively impacts performance and thus possibly changes subjective beliefs about relative performance, this should influence the subjects' decisions in both Tasks 4 and 6. However, preferences for engaging in a competitive activity and (beliefs about) performance in tournaments under stress are only relevant in Task 4.

In Task 7, we measure risk preferences using a design based on Holt and Laury (2002) and Dohmen et al. (2010). Subjects made a series of choices between a lottery, which paid CZK 240 or 0 with 50% probability each, and a safe payment. The safe payment varied across choices, gradually increasing from CZK 0 to CZK 240 in steps of CZK 20. Full experimental instructions for Tasks 1–7 are available in Online Appendix C.

To estimate the role of confidence in competitiveness decisions, we asked nonincentivized questions regarding subjects' perceived rank among all eight participants in the given session for Tasks 1–3. Furthermore, subjects filled out a questionnaire on demographics, personality traits, and stress measurement-related controls, including, for women, whether they took oral contraceptives and the current phase of their menstrual cycle. A complete overview of additional measures is available in Tables A.1 and A.2 in Online Appendix A.

2.2. Treatments

We experimentally induced stress in the laboratory using a modified version of the TSST-G (Kirschbaum et al. 1993, von Dawans et al. 2011). This procedure was intended to induce psychosocial stress in the stress

treatment, with a control procedure designed to similarly prime subjects but keep stress levels constant. The TSST-G has been shown to be the most efficient experimental method of inducing stress, as measured by cortisol response (Dickerson and Kemeny 2004). The TSST-G protocol consists of two parts: a public speaking task and a mental arithmetic task. In our experiment, the first part of TSST-G took place immediately before Task 2, and the second part took place immediately before Task 4 (see Figure A.1 in Online Appendix A).

In both parts of the TSST-G, subjects spoke one by one before a committee of two experimenters who sat at a table in front of the participants wearing white laboratory coats. In order to increase subjects' level of psychosocial stress, the committee did not give any feedback and maintained a neutral facial expression throughout the procedure. The setting of the room is depicted in Figure A.2 in Online Appendix A. Subjects were separated by dividers and wore headphones with ambient traffic noise during the entire TSST-G procedure, except when speaking to the committee, in order to prevent subjects from hearing others during the stress procedure and potentially developing subjective rankings in ability.

In the public speaking task, subjects were told to imagine a situation in which they had been caught cheating during an academic examination and that they should defend themselves in front of a disciplinary committee. This scenario required participants to talk extensively about their personal qualities, and they were instructed to do their best. They were interrupted and asked additional questions if they spoke too fluently for too long.

In the second portion of our modified TSST-G procedure, subjects in the stress treatment were again called individually and asked to recite the alphabet backward in steps of two, starting with a given letter. For example, if given Z, they were required to recite Z, X, V, and so on. Subjects engaged in this activity for a minute and were corrected if a mistake was made.

Our version of the TSST-G changes the standard protocol in several ways. We modified the speaking task to avoid possible priming effects: the original procedure is framed as a job interview, which could have influenced competitiveness and performance in the experiment independent of the stress reaction. In the second portion of the task, subjects were instructed to recite the alphabet rather than counting in intervals. Likewise, this was done to avoid contaminating performance in the experiment while still allowing us to use the addition of two-digit numbers as the real-effort task, consistent with previous work.¹⁰

The control procedure similarly primed subjects and involved a similar degree of physical activity but in a less stressful setting. Subjects were asked to read

an article about academic dishonesty silently for the first 14 minutes and then aloud for 2 minutes. In the second part of the procedure, they collectively recited the alphabet out loud for a minute. Two experimenters were again present in the room during the control procedure, but they wore casual clothes and behaved naturally. The subjects in the control group also wore headphones with ambient noise and were separated with dividers to mimic conditions in the stress treatment group.

The timing of the experiment tasks with respect to the stressors constitutes an important design choice. The physiological response to stressors is complex, and the autonomic nervous system responds within seconds when exposed to the stressor. Other hormones are produced before cortisol, resulting, among others, in an immediate increase of the heart rate.¹¹ Although heart rate returns to normal within a few minutes after the cessation of the stressor, the cortisol response is delayed by around 10 minutes, peaks only 20–40 minutes after the onset of the stressor, and typically lasts over 60 minutes after exposure ceases for psychosocial stressors, such as the TSST-G (Dickerson and Kemeny 2004). In principle, this entire interval can be perceived as an “acute stress response.” As shown in previous studies, the timing of stress exposure can affect behavior (Pabst et al. 2013, Vinkers et al. 2013, Margittai et al. 2015).

We argue that the immediate effect of being exposed to the stressor on competitive behavior is most relevant for our research question rather than the recovery response, which can last more than 60 minutes. This is because, in the labor market events that we aim to simulate, psychosocial stress and competition might be distinct, yet they often occur simultaneously. Therefore, our protocol purposefully minimizes the time between the stress intervention and decisions. To do so, each individual completed the tasks at computers in the same room immediately after they had finished the stress (or control) procedure. This means that the main experimental tasks, Tasks 2 and 3 performance, as well as the Task 4 decision, all took place within around five minutes after the subject completed part 1 or part 2 of the stressor. The full protocol for the TSST-G is included in Online Appendix C.

2.3. Sample and Procedures

The experiment was carried out in 2014–2015, with 24 sessions in total. Subjects were recruited using a standard recruitment database, ORSEE (Greiner 2004); no details about the nature of the experiment were mentioned in the invitation in order to avoid self-selection based on relevant personal characteristics, such as aversion to stressful or competitive situations. The stress and control treatments were randomized at the session level for logistical reasons. All sessions

took place after 3 p.m. to limit the impact of the circadian variability in cortisol levels. Each session consisted of eight subjects, four men and four women, and though the gender composition was not directly mentioned (following Niederle and Vesterlund 2007), it was easily observable—at the end of the experiment, 80% of subjects correctly reported the gender ratio.

The final sample is composed of 95 men and 95 women, primarily undergraduate students (82%) majoring mostly in economics, business, and related fields (61%).¹² The sample is drawn from several elite universities in Prague and thus consists of individuals who have already self-selected into a competitive environment. If individuals who are most strongly impacted by stress and competition never apply to competitive universities and are systematically excluded from our sample, then the treatment effects that we report are a lower bound. Tables A.1 and A.2 in Online Appendix A provide summary statistics. Of 85 treatment comparisons in Table A.1 in Online Appendix A, 2 (2.4%) are significant at the 5% level, which indicates that randomization was successful.

Decisions were made on computers using the program z-Tree (Fischbacher 2007). The experiment was conducted in the Czech language, and sessions were administered by one experimenter (male), one assistant (female), and two separate committee members for the TSST-G procedure (a male and a female). The average length of the experiment was around two hours, and the average payoff was CZK 516 (about 20 euro). Experimental procedures are described in detail in Online Appendix C.

The study was approved by the Internal Review Board of the Laboratory of Experimental Economics in Prague. We obtained informed consent from all participants, emphasizing that they were free to leave at any time. At the end of the session, subjects in the stress treatment were debriefed on the purpose of the stress procedure.

3. Results

3.1. Physiological and Psychological Stress Response

Though we are ultimately interested in the overall immediate effect of being exposed to a psychosocial stressor on behavior, we use salivary cortisol level and heart rate to confirm that the TSST-G manipulation was effective. Salivary cortisol level is a good proxy for the degree of the physiological stress response, and we analyze saliva samples taken throughout the experiment.¹³ Results are presented in Figure A.3 in Online Appendix A. Baseline cortisol was measured in sample 1 before the stress procedure, and samples 2 and 3 were taken afterward (see Figure A.1 in Online Appendix A).¹⁴ Because the cortisol response is typically delayed by 10–20 minutes after initial

exposure to the stressor and peaks after 21–40 minutes (Dickerson and Kemeny 2004), cortisol sample 2 was collected only after the second portion of the TSST-G procedure. Sample 2, collected around 18 minutes after the onset of TSST-G part 1, therefore primarily captures the physiological response to the first part of the TSST-G procedure. Sample 3, collected on average 31 and 16 minutes after the starts of TSST-G part 1 and part 2, respectively, reflects responses to both parts 1 and 2.

Although cortisol levels for subjects in the control group actually decreased over the course of the experiment, levels for those in the stress treatment group more than doubled. For men in the stress treatment, cortisol levels in samples 2 and 3 increased by 130% and 113% of baseline, respectively (signed rank test, $p = 0.000$ for both). For women, there was an increase of 109% on average, which remained constant in samples 2 and 3 (signed rank tests, $p = 0.000$ for both). We find no evidence that the TSST-G was relatively more successful in either men or women ($p = 0.200$ for the percentage increase between samples 1 and 2 and $p = 0.407$ for the percentage increase between samples 1 and 3).¹⁵ Although we do not measure the entire span of the cortisol response, the return to baseline cortisol levels typically takes over 60 minutes after the cessation of psychosocial stressors such as the TSST-G (Dickerson and Kemeny 2004). In our case, Task 7 ended on average 13 minutes after the end of TSST-G part 2 (i.e., all tasks took place shortly after direct exposure to the stressor).

The time delay of the cortisol response makes it difficult to precisely link to a specific stressor. For this reason, we also continuously measured heart rate, which increases with stress (Kirschbaum et al. 1993). Although heart rate can also increase because of other factors, such as arousal, physical activity, and strong emotions, a joint increase in cortisol level and heart rate is consistent with a stress reaction (Everly and Lating 2013, chapter 5). The heart rate data allow us to see an immediate reaction to the TSST-G procedure in contrast to the delayed cortisol response. During part 1 of the TSST-G protocol, the heart rate of the stress treatment group increases sharply and is significantly higher than in the control group ($p = 0.000$). It stays significantly higher during Task 2 (piece rate under treatment), Task 3 (tournament under treatment), and the willingness to compete decision in Task 4 (see Table A.2 in Online Appendix A). This is in line with the interpretation that subjects experienced the immediate reaction to the stressor when completing the main experimental tasks. Heart rate stabilizes only during the performance part of Task 4.¹⁶

We also measure psychological response to the stress treatment using a multidimensional mood questionnaire (Steyer et al. 1997) administered both

before the treatment procedure and again at the end of the experiment (after Task 7).¹⁷ The questionnaire measures mood across three dimensions. Both men and women in the stress treatment group have a stronger shift from “good” to “bad” mood than subjects in the control group. For men only, subjects in the stress treatment reported a larger change from “calm” to “nervous” (see Table A.2 in Online Appendix A). However, the effect on mood that we observe should be understood as the joint effect of being exposed to the stressor and completing the tasks under stress.

3.2. Performance and Competitive Incentives

We next analyze the effect of stress on performance under competition and whether this differs by gender. Recall that Tasks 1–4 included a calculation activity. Because Task 1 took place before the treatment and was incentivized using the piece-rate scheme, this serves as the baseline for ability and motivation. Performance in Task 2 (piece-rate payment after treatment) isolates the effect of stress on performance, and Task 3 measures how both stress and competition affect performance. If competitive incentives lead to increased performance—a common assumption—then subjects should be expected to complete more problems in Task 3 than in Tasks 1 and 2.

Results from performance in the calculation portions of Tasks 1–4 are presented in Table 1 and Figure A.4 in Online Appendix A. Under the piece-rate incentive scheme in Task 1, there is virtually no difference in the number of correctly solved problems between the stress treatment and control groups ($p = 0.931$). The same holds for both the male and female subsamples independently. This demonstrates that randomization was successful. At baseline, men outperform women by 1.38 correctly solved problems on average ($p = 0.002$).¹⁸ This holds independently in each treatment group, and men continue to outperform women in Tasks 2 and 3 in both treatments. In the following analysis, we show that all treatment effects hold when controlling for baseline performance in Task 1.

In Task 2, we do not find a statistically significant difference between treatments: subjects in the stress and control treatments correctly answered an average of 6.37 and 6.56 problems, respectively ($p = 0.560$). As before, this result holds for both the male and female subsamples. Regression analysis confirms that there is no gender-specific effect ($p = 0.652$; column (4) of Table A.3 in Online Appendix A). We reach the same conclusion when taking the difference between the number of problems that subjects solved correctly in Tasks 2 and 1 as the main outcome variable (columns (5)–(8) of Table A.3 in Online Appendix A).¹⁹ These results indicate that stress alone does not affect

Table 1. Summary Statistics: Performance in the Calculation Task

Sample	Number of problems solved correctly				Rank sum <i>p</i> -value (5)	<i>N</i>
	All (1)	Treatment group Stress (2)	Control (3)	Difference (4)		
Task 1: piece rate before treatment						
All	6.69	6.66	6.72	−0.06	0.931	190
Male	7.38	7.26	7.50	−0.24	0.943	95
Female	6.00	6.06	5.94	0.13	0.880	95
Task 2: piece rate under treatment						
All	6.47	6.37	6.56	−0.19	0.560	190
Male	7.14	7.02	7.25	−0.23	0.761	95
Female	5.80	5.72	5.88	−0.15	0.632	95
Task 3: tournament under treatment						
All	6.69	6.24	7.14	−0.89	0.018	190
Male	7.46	7.26	7.67	−0.41	0.562	95
Female	5.93	5.23	6.60	−1.37	0.003	95
Task 4: chosen scheme under treatment						
All	7.11	7.05	7.17	−0.11	0.650	190
Male	8.03	8.17	7.90	0.27	0.538	95
Female	6.19	5.94	6.44	−0.50	0.087	95

Notes. Mean performance in the tasks under different compensation schemes by treatment and gender. All differences are tested using a Wilcoxon rank sum test.

performance in the calculation task for either gender and therefore does not widen the gender gap in performance.

In contrast, for performance in the tournament in Task 3, we see a significant treatment effect, with lower performance among the stress group, who solved only 6.24 problems correctly, equivalent to 2.98 standard deviations (sd), compared with 7.14 (sd = 2.74) in the control group ($p = 0.018$). This difference is driven by female subjects: women in the stress treatment correctly solved 5.23 (sd = 2.43) problems on average compared with 6.60 (sd = 2.08) in the control group ($p = 0.003$). The corresponding treatment difference for men is less than one-third the size, 0.41, and it is not statistically significant ($p = 0.562$).

In Table 2, we confirm this pattern by regressing performance under tournament incentives in Task 3 on a dummy that equals 1 if the subject was assigned to the stress treatment, controlling for gender and baseline performance in Task 1, with standard errors clustered at the session level.²⁰ In column (2) of Table 2, we add an interaction term, *Stress treatment* × *Female*, and the results indicate that the effect of the stress treatment is indeed different for the female subsample ($p = 0.019$).

Next, we consider the average difference in the number of problems that each subject solved in Tasks 3 and 2. Figure 1 demonstrates that tournament incentives influence performance within individuals across treatment and gender. Overall, subjects in the control group solved 0.57 more problems under the

competitive compensation scheme in Task 3 than under the piece-rate scheme in Task 2 (signed rank test, $p = 0.001$). This holds independently for both men and women, who answered 0.42 (signed rank test, $p = 0.042$) and 0.73 (signed rank test, $p = 0.007$) more questions correctly in Task 3 than in Task 2, respectively.

However, in the stress treatment, only 25.5% of female subjects did better in Task 3 than in Task 2 compared with 56.3% in the control group, whereas 44.7% of women in the stress treatment group did worse in Task 3 than in Task 2 compared with only 20.8% of those in the control. These proportions differ significantly across treatments (chi-squared test, $p = 0.007$).²¹ On average, women in the stress treatment group solved 0.49 fewer problems in Task 3 than in Task 2, which is marginally insignificant (signed rank test, $p = 0.122$). For men, there is no treatment difference in the proportion of subjects who improved or did worse in Task 3 compared with Task 2 (chi-squared test, $p = 0.453$), and men in the treatment group solved 0.23 more problems on average in the tournament than they did under piece-rate incentives in Task 2, though this is not statistically different from zero (signed rank test, $p = 0.290$).

The regression results confirm this clear pattern (columns (5)–(8) of Table 2). The dependent variable is the difference between correctly answered problems in Tasks 3 and 2, which can be interpreted as the effect of the tournament incentive scheme on performance. The results confirm that the

Table 2. The Effect of Stress, Competition, and Gender on Performance

Dependent variable	Problems solved correctly							
	Task 3				Task 3–Task 2			
	Tournament incentives				Effect of tournament incentives			
	All	All	Men	Women	All	All	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Stress treatment</i>	−0.84*** (0.21)	−0.21 (0.29)	−0.19 (0.28)	−1.45*** (0.34)	−0.70*** (0.24)	−0.17 (0.30)	−0.15 (0.28)	−1.21*** (0.33)
<i>Female</i>	−0.41 (0.27)	0.22 (0.34)			−0.14 (0.28)	0.39 (0.36)		
<i>Stress treatment × Female</i>		−1.26** (0.50)				−1.05** (0.43)		
<i>Solved Task 1 (baseline)</i>	0.81*** (0.04)	0.82*** (0.04)	0.91*** (0.05)	0.63*** (0.12)	0.04 (0.06)	0.05 (0.06)	0.11** (0.05)	−0.09 (0.10)
<i>Constant</i>	1.88*** (0.36)	1.53*** (0.38)	0.85 (0.50)	2.85*** (0.75)	0.34 (0.48)	0.05 (0.51)	−0.44 (0.52)	1.24** (0.53)
Observations	190	190	95	95	190	190	95	95
R ²	0.61	0.62	0.72	0.39	0.04	0.06	0.04	0.10

Notes. Ordinary least squares; standard errors are clustered at the session level. The dependent variable is the number of addition problems that were correctly completed within the time limit in the specified task. Both Tasks 2 and 3 were completed under treatment and under piece rate and tournament incentives, respectively.

*** $p < 0.01$; ** $p < 0.05$.

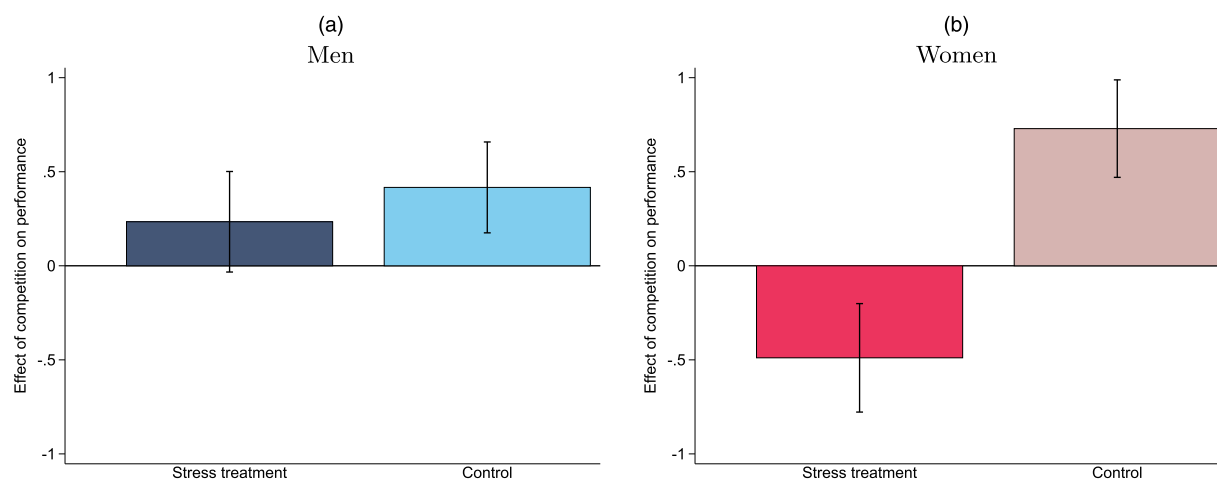
negative effect of the stress treatment is driven by women (column (6) of Table 2; $p = 0.021$). In Table A.4 in Online Appendix A, we confirm that these results are stable with respect to additional controls.²²

Because women demonstrate lower baseline ability than men, a concern is that female subjects respond differently to the stress treatment because of low ability rather than gender differences related directly to stress and competition.²³ In the main specification in Table 2, we control for baseline ability, which

mitigates this concern. In Table A.5 in Online Appendix A, results are the same if we allow for a gender-specific effect of baseline performance. Table A.6 in Online Appendix A further rules this out by showing that the interaction between gender and treatment is unaffected by controlling for a treatment effect specific to low-ability individuals.

Overall, these results show that exposure to psychosocial stress widens the gender gap in performance under competition. We find that both men and

Figure 1. (Color online) Effect of Tournament Incentives on Performance by Treatment and Gender



Notes. The graphs depict the change between Tasks 3 and 2 performance. Both tasks took place after the first part of the TSST-G under tournament and piece-rate incentives, respectively. Bars indicate mean $\pm$ standard error.

women in the control group respond positively to competitive incentives because they perform significantly better in Task 3 than in Task 2. However, the combination of stress and competition decreases performance for a large portion of female subjects. We do not find any such pattern for men, whose performance under tournaments is not significantly affected by the stress treatment.²⁴

3.3. Willingness to Compete

We now turn to investment in the tournament payment scheme in Task 4, which serves as our principal measure of willingness to compete. This decision captures both preferences for competitive outcomes and those for engaging in a competitive activity plus expectations of one's future performance under competition. Recall that in Task 4, subjects allocated 100 points between a tournament and a piece-rate incentive scheme before completing the calculation portion of the task. The results from Task 4 are presented in Table 3(a) and Figure A.6 in Online Appendix A. Overall, subjects allocated slightly less than one-half of the total amount, 46.68 points, in the tournament incentive scheme. We find that stress does indeed affect competitiveness: subjects in the stress treatment group invested 7.72 fewer points in the tournament scheme than those in the control group ($p = 0.046$).

We confirm this by regressing the points invested in the tournament in Task 4 on the stress treatment. We control for gender and baseline performance in Task 1 (i.e., before the treatment intervention) and cluster standard errors at the session level. As reported in column (1) of Table 4, we find that the stress treatment was associated with investing 7.59 fewer points in the tournament scheme ($p = 0.024$). Consistent with the literature, we also find that gender has a strong influence on

choices in Task 4, with women investing 22.06 fewer points in the tournament than men after controlling for treatment and baseline performance ($p = 0.000$).

In contrast to our *ex ante* prediction, the stress treatment does not widen the gender gap in willingness to compete.²⁵ The negative effect of stress on investments in the tournament that we find on average in Task 4 holds separately for both the male and female subsamples, though the treatment differences are not statistically significant owing to smaller sample sizes (Table 3(a)). In column (2) of Table 4, we add an interaction term between the female and stress treatment dummies and observe no statistically significant gender difference ($p = 0.926$).²⁶ In columns (3) and (4) of Table 4, we run regressions separately on the male and female subsamples and find that the coefficients for the stress treatment are virtually identical, though both coefficients are marginally insignificant: $p = 0.123$ and $p = 0.124$ for the male and female subsamples, respectively.

Because most studies use binary measures to measure willingness to compete, we perform a robustness test in which we classify subjects as competitive if they invest at least 50 of 100 points in the tournament incentive scheme in Task 4 and estimate the effects of the stress treatment and gender using a probit model. Results are similar to the linear measure (see Table A.8 in Online Appendix A). We also confirm that results are robust to allowing for a gender-specific effect of baseline performance and including additional controls (see Tables A.9 and A.10 in Online Appendix A).

Furthermore, willingness to compete in Task 4 is connected to tournament performance, subjective confidence in tournaments, and risk preferences in an intuitive way: subjects who perform better in tournaments, subjects who are more confident (which is highly

Table 3. Summary Statistics: Willingness to Compete

Sample	Number of points invested into competition (of 100)					
	Treatment group			Difference (4)	Rank sum p -value (5)	N (6)
	All (1)	Stress (2)	Control (3)			
(a) Task 4 choice: future performance under treatment						
All	46.68	42.78	50.50	−7.72	0.046	190
Men	59.32	55.02	63.52	−8.50	0.190	95
Women	34.04	30.53	37.48	−6.95	0.159	95
(b) Task 5 choice: past performance before treatment						
All	40.69	41.20	40.19	1.01	0.826	190
Men	47.85	48.21	47.50	0.71	0.988	95
Women	33.53	34.19	32.88	1.32	0.688	95
(c) Task 6 choice: past performance under treatment						
All	40.39	39.64	41.14	−1.50	0.702	190
Men	51.09	51.60	50.60	0.99	0.967	95
Women	29.69	27.68	31.67	−3.99	0.710	95

Notes. Mean willingness to compete across tasks, treatments, and gender. All differences are tested using a Wilcoxon rank sum test.

Table 4. The Effect of Stress and Gender on Willingness to Compete

Dependent variable	Willingness to compete			
	Points invested into tournament ex ante (Task 4)			
	All		Men	Women
	(1)	(2)	(3)	(4)
<i>Stress treatment</i>	−7.59** (3.14)	−7.93 (5.11)	−7.99 (5.00)	−7.31 (4.58)
<i>Female</i>	−22.06*** (4.14)	−22.41*** (6.23)		
<i>Stress treatment × Female</i>		0.69 (7.38)		
<i>Solved Task 1 (baseline)</i>	2.33** (0.95)	2.32** (0.95)	2.07* (1.15)	2.85** (1.34)
<i>Constant</i>	45.89*** (8.05)	46.08*** (8.33)	48.02*** (9.53)	20.55** (8.06)
Observations	190	190	95	95
<i>R</i> ²	0.21	0.21	0.05	0.09

Notes. Ordinary least squares. Standard errors are clustered at the session level. The dependent variable is investment in the tournament compensation scheme in Task 4, where the choice was made before the calculation portion of the task; 0 (100) indicates that all points were invested in the piece-rate scheme (tournament).

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

correlated with actual performance), and subjects who are less risk averse have a higher willingness to compete (Table A.11 in Online Appendix A). In Section 4.2, we discuss whether these factors mediate the treatment effect on willingness to compete.

The decisions in Tasks 5 and 6 provide additional insight into the mechanism behind the treatment effect in Task 4, which measures willingness to compete for future performance. Recall that in Tasks 5 and 6, subjects decided how much to invest in the tournament payment scheme based on the result of their past performance in Tasks 1 and 2, respectively. In contrast to the competition decision in Task 4, we do not find a significant difference between the treatment groups for investment in the tournament in Task 5 ($p = 0.826$) or Task 6 ($p = 0.702$). This holds for both men and women (Table 3, (b) and (c)).²⁷ If the decrease in competitiveness in the treatment group in Task 4 was driven by the effect of stress on performance (or confidence), independent of competition, we should see a similar result in Task 6. The fact that we do not suggests that the treatment effect in Task 4 is related to the combination of stress and competition rather than either element alone.

As in Niederle and Vesterlund (2007), the gender difference in competitiveness that we observe in both treatments in Task 4 holds in Tasks 5 and 6 as well.

4. Discussion and Additional Results

In this section, we explore possible mechanisms that drive our results.

4.1. Physiological Stress Response

To begin, we use the relative increase in cortisol between the first and second samples as a proxy for the physiological stress response. We examine the relation between cortisol response and the outcomes using the stress treatment as an instrument to estimate the average treatment effect on the treated. Results are robust: the stronger the physiological stress response, the lower is the willingness to compete and the worse is the tournament performance for women (see Tables A.13 and A.14 in Online Appendix A). This is evidence that the stress treatment indeed affects behavior through stress.

4.2. Stress and Willingness to Compete

Next, we consider potential channels through which stress might lower competitiveness. First, our results suggest that the lower willingness to compete that we observe among women in the stress treatment group is driven by weaker performance in the tournament. A sensitivity analysis reported in Table A.15 in Online Appendix A reveals that when controlling for Task 3 performance, the stress treatment does not additionally explain willingness to compete among women. Because tournament performance and confidence are endogenous and closely correlated (Table A.11 in Online Appendix A), it is impossible to identify the effect of each independently. Nevertheless, our results suggest that the stress treatment does not additionally explain confidence after controlling for actual performance (Table A.16 in Online Appendix A). In contrast, exposure to the stressor does not affect the performance or confidence of men, and therefore, these results cannot explain the drop in competitiveness.

Second, it is unlikely that stress influences competitiveness through risk preferences. A change in risk preferences would also affect willingness to compete for past performance, and we do not observe any effect in Tasks 5 and 6. Moreover, in our sample, we fail to find any significant relationship between the stress treatment and risk preferences elicited in Task 7 (see Table A.2 in Online Appendix A) in contrast to Cahlíková and Cingl (2017).²⁸

Third, another mechanism that we consider is a change in competitive preferences under heightened stress. We find a treatment difference only for the willingness to compete in Task 4 (future performance) but not for the willingness to compete in Tasks 5 or 6 (past performance). Therefore, our results are consistent

with changed preferences for engaging in competition under stress but rule out an effect of stress on preferences for competitive outcomes.

To summarize, for men, we do not find evidence that lower willingness to compete is related to a change in performance, confidence, or risk preferences, and by process of elimination, we conclude that lower willingness to compete among men is driven by preferences for engaging in competition under stress. For female subjects, we conclude that psychosocial stress lowers willingness to compete because women under stress perform worse under tournament incentives.

4.3. Stress and Tournament Performance

Why do tournament incentives lead to worse performance for women in the *stress treatment* group? Here we consider several mechanisms and discuss why these might be specific to women.

Potentially, the results are related to the phenomenon of “choking under pressure.” In the psychology literature, this is loosely defined as performing below one’s ability in high-pressure situations. Specific aspects of the situation, such as task complexity, practice, and social environment, can influence whether choking occurs or not (Beilock et al. 2004, Dohmen 2008). Distraction theories, which explain choking through attention, are especially relevant for math tasks. If attention is diverted by irrelevant thoughts and worries, it can become more difficult for an individual to focus on the task at hand (DeCaro et al. 2011).²⁹ “Pressure,” “stress,” and “high stakes” are often used interchangeably in this literature, and experiments typically combine several sources of pressure, such as monetary incentives, peer pressure, and social evaluation (Beilock 2008). Our experiment separates these sources of pressure; we show how psychosocial stress affects performance both with and without competition. Why does a combination of stress and tournament incentives, but neither element alone, cause women to choke under pressure?

First, tournament incentives might induce additional stress. There could be a threshold level beyond which performance suffers. We do not find support for this explanation. The stress response to tournament incentives is much weaker than the response to the psychosocial stressor, and exposure to the stressor alone does not reduce performance.³⁰ This would require an extremely sensitive threshold and one that is specific to women.

Second, the results might be explained by the higher stakes inherent to competition.³¹ Ariely et al. (2009) show that raising the stakes can lead to choking under pressure. Previous work indicates that women might be more sensitive to increased stakes than men (Azmat et al. 2016), which could explain the gender-specific effect that we find. The problem with this

explanation is that the stakes in our study are relatively low and that tournament incentives actually improve performance in the control group in our study. Nevertheless, Ariely et al. (2009) link their results to the law of Yerkes and Dodson (1908), according to which arousal increases executive function up to a point, but ability declines with increased stimulation after passing a threshold. If psychosocial stress also stimulates subjects, this may explain why performance decreases only with the combination of stress and competition.

Third, another possibility is that tournament incentives could accentuate social comparison, which has been shown to decrease women’s performance in other contexts (Ashraf et al. 2014, Schram et al. 2019). The calculation task plausibly produces a “stereotype threat” among female subjects owing to negative attitudes about women’s math ability (Spencer et al. 1999). As with increased stakes, it is possible that stress further lowers executive function already taxed by stereotype threat and competition. Schmader et al. (2008) provide a theoretical framework for understanding how stereotype threat (social identity threat) affects performance and argue that this could be exacerbated by increased cortisol and that the “physiological stress response could play a direct role in impairing task performance under stereotype threat” (Schmader et al. 2008, p. 343). Our results are in line with this hypothesis.

5. Conclusion

This article presents new evidence on the effects of stress on performance under competition and individual willingness to compete. Understanding the effects of stress on competitive behavior is vital for designing optimal hiring and management practices in firms, interpreting the gender gaps in examination performance, and analyzing possible sources of gender gaps on the labor market.

We experimentally induce psychosocial stress in the laboratory using a modified TSST-G protocol and find that subjects in the stress treatment are less competitive, investing less in the tournament compensation scheme than those in the control group. This result holds jointly for men and women, and exposure to the stressor does not widen the gender gap in willingness to compete. By examining salivary cortisol, which increases after exposure to the stress treatment, we confirm that the treatment difference in the willingness to compete is driven by stress; the cortisol response is negatively correlated with willingness to compete. Stress only lowers competitiveness if the decision is made before the competitive task; in the tasks for which subjects made willingness to compete decisions for past performance, we find no treatment effect.

For women, we find that performance under competition is worse in the stress treatment group than in the control group. Although most female subjects in the control group perform better under tournament incentives than under the piece-rate scheme, many women in the stress treatment group actually do worse. Interestingly, stress alone does not have a negative effect on performance in either women or men. It is the combination of stress and tournament incentives that is detrimental to women's performance, and this explains the lower willingness to compete among women in the stress treatment group. We do not find such a link among men, for whom there is no treatment difference in tournament performance. The lower willingness to compete among men in the stress treatment seems to be driven by a link between stress and preferences for engaging in competition.

We propose that the most plausible explanation for decreased performance under stress and competition among women is that the calculation task was perceived as a male-dominant activity, and women faced a stereotype threat. Stress, higher stakes in the tournament, and stereotype threat are all factors that can tax executive function, and the combination of all three could decrease certain abilities, including the impairment of working memory (Schmader et al. 2008).

Our findings help to explain past results regarding the effect of tournament incentives on performance. Although some studies have found a positive effect for both genders, others show a positive effect for men only (Gneezy et al. 2003, Gneezy and Rustichini 2004, Niederle and Vesterlund 2007). Potentially, the environments differed in the degree of stress involved. Moreover, our results also support the claim made in Niederle and Vesterlund (2010) that gender gaps in math test scores may not necessarily reflect differences in math ability. When test results come from highly competitive and stressful settings, such as university entrance examinations, women's performance may fall below their ability.

Although a gender-neutral or female-dominated task might produce different results, the labor market settings that motivate this research, such as business or academia, are similarly perceived as stereotypically male. Many competitive situations that affect one's career trajectory—examinations, job interviews, and asking for a promotion—are also stressful. If women perform worse under competition and stress, this will directly affect labor market outcomes and could dissuade women from entering competitive environments in the first place.

If employers make hiring decisions in stressful, competitive settings (e.g., assessment centers), our results suggest that this will lead to inefficient outcomes because women may underperform. This is

especially relevant if the position itself is not particularly stressful or competitive compared with the hiring process. Moreover, a combination of tournament incentives and increased pressure to perform is often used to boost output in firms, and although this incentive structure may be effective in motivating male workers, our results show that such policies could have unintended consequences when applied to female workers.

This phenomenon could produce path dependence across sectors: if the initial composition of a particular field is dominated by men or women, the optimal management and hiring strategies might differ with respect to stress and competition. Both hiring and self-selection of workers into the field would then exacerbate gender disparities in response to the dominant practices and norms.

Although in some settings both competition and psychosocial stress are inevitable, in others, active measures can be taken to reduce one element or the other. Our results suggest that, for example, in organizations where competition is inherent, such as in many firms or universities, reducing stress could be a viable policy for improving performance because workers—especially women—might be performing below their potential.

Acknowledgments

The authors thank Vojtěch Bartoš, Michal Bauer, Erwin Bulte, Alexander W. Cappelen, Belle Derks, Lorenze Goette, Michel Handgraaf, Johannes Haushofer, Štěpán Jurajda, Muriel Niederle, Haki Pamuk, Maroš Servátka, Simeon Schudy, Bertil Tungodden, and Bernadette von Dawans for helpful conversations at different stages of the project. All remaining errors are the author's own.

Endnotes

¹For example, Gneezy and Rustichini (2004) find that competition increases performance for boys but not for girls, whereas Sutter and Glätzle-Rützler (2015) find no such gender differences.

²In an experiment with Swedish children, Dreber et al. (2011) find that task matters: in two tasks (running and skipping rope), competition increased performance, but competition actually decreased performance in a third task (dancing). Gneezy et al. (2009) find that men are more competitive than women in a patriarchal society but not in a matriarchal society; Booth and Nolen (2012) show that the gender gap is much smaller for students from single-sex high schools; and Almás et al. (2016) find that the gender gap is only present in students from a higher socioeconomic background.

³In particular, we study acute (short-term) stress, which has been shown to have psychological, neurological, and behavioral effects distinct from those of long-term stress (McEwen 2012). We concentrate on the former because we consider it more relevant for the labor market events that motivate this study.

⁴A second category, physical (biogenic) stressors, includes noise, extreme temperature, or prolonged exercise, for example. Although both psychosocial and physical stressors produce a physiological stress response, including elevated heart rate and levels of the

hormone cortisol, each is associated with a distinct psychological and physiological response (Baum and Grunberg 1997, Haushofer and Jang 2015). For details on response to different stressors, see Kemeny (2003), Everly and Lating (2013), and Allen et al. (2014). The economically relevant question that we consider is the overall effect of the stressor on behavior rather than the particular physiological response. Understanding the effect of both psychosocial and physical stressors on behavior is scientifically important. One advantage to studying the latter is that physical stressors are comparatively straightforward to induce experimentally.

⁵ Similarly, key educational outcomes (e.g., entrance examinations) involve both social evaluation and competition.

⁶ Although the physiological effects of stress are complex, cortisol is released into the bloodstream at greater levels after exposure to psychosocial stressors as the final product of the hypothalamus–pituitary–adrenal axis. Because of the ease in measuring salivary cortisol levels, it is the most commonly used biomarker for measuring physiological stress response (Hellhammer et al. 2009, Everly and Lating 2013).

⁷ Angelucci and Córdova (2014) experimentally manipulate emotional load and uncertainty—which are potentially related to stress—by exposing subjects to short films and find that this lowered productivity as measured by correct responses to test questions for female subjects, especially those who had previously experienced stressful life events.

⁸ Several relevant studies examine the relationship between basal levels of stress markers and competitive behavior. Apicella et al. (2011) and Schipper (2015) both examine correlations between baseline cortisol levels and willingness to compete (after the NV 2007) and bidding behavior in an auction, respectively. Neither find a relationship. Halko and Sääksvuori (2017) use resting heart rate variability as a proxy for chronic stress and find that women under higher chronic stress are less willing to compete.

⁹ Note that we made several changes to the protocol used in the NV 2007 and subsequent papers [including Buser et al. (2017) and Zhong et al. (2018)]. First, we shortened the time available in the calculation task from five to two minutes so that Task 3 would take place soon after part 1 of the stress treatment procedure. Second, to compensate for this additional time pressure, we made the calculation task relatively easier (adding series of four numbers instead of five). Third, although subjects in the NV 2007 competed in groups of four, in our experiment we matched subjects in pairs (as in Gneezy et al. 2018). This was done to make winning the competition more likely and therefore, salient. Additionally, because subjects had more time to observe each other while moving from room to room, they may have formed beliefs about others or recognized other subjects, and with only two groups of four, this may have discouraged them from competing if they conjectured that a particular person would be of high ability. Fourth, in Task 4, instead of the binary measure used by the NV 2007 and others, we use the linear measure of competitiveness, again after Gneezy et al. (2018), in order to capture more variation in willingness to compete.

¹⁰ Additionally, to avoid deception, participants in the stress treatment were not told that the panel members were trained in behavioral analysis or that the video recordings would later be analyzed as in von Dawans et al. (2011).

¹¹ The physiological stress response proceeds as follows: (1) The autonomic nervous system is activated, (2) this causes the adrenal medulla to release epinephrin and norepinephrin into the bloodstream, (3) the hypothalamus–pituitary–adrenal axis is initiated by the release of corticotropin-releasing hormone in hypothalamus, which (4) stimulates the secretion of adrenocorticotropin hormone in the pituitary, and (5) eventually triggers the massive secretion of cortisol in the adrenal glands (Kemeny 2003).

¹² We dropped one female subject who left the experiment early and one male subject who did not meet the selection criteria.

¹³ Saliva samples were collected using Salivettes, a standard sampling device. The samples were stored at -20°C after each experimental session, and after each batch of the sessions (2014 and 2015), they were shipped for the analysis of salivary cortisol concentration to the laboratory of the Biopsychology Department at the Technical University in Dresden.

¹⁴ In contrast, a standard procedure in psychological studies aimed at cortisol reactivity is to collect six to eight cortisol samples at 10- to 15-minute intervals. The stress response is then calculated using repeated measures analysis of variance (von Dawans et al. 2011) or area under the curve measures (Pruessner et al. 2003) or with simple rule-of-thumb criteria (Miller et al. 2013). For a summary of the measurement of cortisol, see Nicolson (2008). In our experiment, cortisol primarily serves as a manipulation check, and we therefore only collected one sample before the TSST-G and two afterward. This is similar to Buser et al. (2017).

¹⁵ All comparisons are tested using Wilcoxon rank sum tests unless otherwise noted.

¹⁶ Although the TSST-G was designed specifically to induce stress, and although we took steps to reduce confounding factors (such as keeping physical activity similar between treatments), there are other possible explanations other than stress that could lead to an increase in cortisol and heart rate (e.g., differences in arousal). If this were the case, it could change the interpretation of our results. However, because the TSST-G involves a situation that is qualitatively similar to the labor market events that motivate this study, the results would nonetheless be informative about the relationship between gender and competition under social evaluation if not stress per se.

¹⁷ The questionnaire has two parts administered in a random order, each with similar questions.

¹⁸ This result is in contrast to the NV 2007, which finds no gender difference in baseline performance, although the direction is the same (men solved 10.68 problems, and women solved 10.15 problems; $n = 80$, $p = 0.459$). In other studies, men have also outperformed women, though the differences are not statistically significant: Halko and Sääksvuori (2017) find that men and women solve 9.10 and 7.93 problems, respectively ($n = 80$, $p = 0.127$). In Buser et al. (2017), in experiment 1, men outperform women 10.31 to 9.46 ($n = 101$, $p = 0.24$); in experiment 2, men and women answer 10.33 and 10.09 questions, respectively ($n = 103$, $p = 0.76$). Potentially, we find a larger baseline gender difference in performance owing to changes in the calculation task. Additionally, we have a sample size larger than any of these studies and thus more statistical power.

¹⁹ See Figure A.5(a) in Online Appendix A for cumulative distribution functions of Task 2–Task 1 performance.

²⁰ We cluster standard errors at the session level to account for shared experiences and because treatment was assigned at the session level. This results in 24 clusters. We confirm that the small number of clusters does not affect results by running a robustness check using the wild bootstrap method proposed by Cameron et al. (2008). Results are available on request.

²¹ See Figure A.5(b) in Online Appendix A for cumulative distribution functions of Task 3–Task 2 performance.

²² Using all controls presented in Table A.1 in Online Appendix A: descriptive characteristics, personality traits, baseline risk attitudes, controls related to stress measurement (including the intake of oral contraceptives and the phase of the menstrual cycle for women), baseline cortisol level, baseline heart rate, baseline mood, and potential problems with understanding.

²³ Booth and Nolen (2012) use a different task (solving mazes) and find that men significantly outperform women ($p < 0.01$). Despite this, they argue that treatment differences between men and women in

willingness to compete are valid after controlling for baseline performance. Similarly, we claim that baseline differences in performance do not explain gender differences in response to treatment.

²⁴ We also consider performance in Task 4, though interpretation is less clear, because the incentive scheme is endogenous. On average, the stress treatment does not have a statistically significant effect on the number of correctly answered questions in Task 4 ($p = 0.650$). However, women in the stress treatment group correctly completed 0.5 fewer correct problems ($p = 0.087$). Regression results confirm this; see Table A.7 in Online Appendix A.

²⁵ See Figure A.7 in Online Appendix A or cumulative distribution functions by treatment and gender.

²⁶ It is unlikely that the lack of statistically significant interaction effect between gender and treatment is the result of a small sample size. For a 2×2 design (two treatments $\times$ gender), 5% significance level, and 80% power, we are able to detect effect size of $f = 0.204$ [i.e., we can detect a medium-sized or larger interaction effect; Faul et al. (2009)].

²⁷ We present regression results for investment in the tournament incentive scheme in Tasks 5 and 6 in Table A.12 in Online Appendix A. We do not find any significant treatment effects for either men or women.

²⁸ Because we randomly chose two of seven tasks for payment, it is possible that decisions in the risk task were affected by decisions in previous rounds. Because stress lowered willingness to compete, leading to lower risk exposure in Task 3, this may have caused subjects in the stress treatment to make riskier decisions in Task 7 independent of risk preferences. Therefore, our measure of risk preferences should be interpreted with caution.

²⁹ In comparison, explicit monitoring theories suggest that pressure increases attention put on performing the task, which might be disruptive for otherwise highly automated processes, such as sport routines.

³⁰ In the control group, cortisol decreases throughout the experiment, including after the tournament. Thus, any increase in stress owing to competition is relatively small. Buser et al. (2017), using a similar design to our design, find that competition increases cortisol levels by 12%–15%. In comparison, we find that the TSST-G increases cortisol levels by 109% in women and 130% in men.

³¹ For example, if a subject was matched with a partner in Task 3 who had correctly solved the median number of six problems, the difference in payoffs between solving five and seven problems would be CZK 350. In the piece-rate tasks, the same two-problem difference in performance would only change payoff by CZK 50.

References

- Allen AP, Kennedy PJ, Cryan JF, Dinan TG, Clarke G (2014) Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. *Neuroscience Biobehavior Rev.* 38: 94–124.
- Almås I, Cappelen AW, Salvanes KG, Sørensen EO, Tungodden B (2016) Willingness to compete: Family matters. *Management Sci.* 62(8):2149–2162.
- Angelucci M, Córdova M (2014) Productivity and choice under stress: Are men and women different? Accessed June 24, 2019, <https://www.semanticscholar.org/paper/Productivity-and-Choice-Under-Stress-%3A-Are-Men-and-Angelucci-Cordova/3621386a1d6cc61c45d441fcb21aee194df065d8#citing-papers>.
- Apicella CL, Dreber A, Gray PB, Hoffman M, Little AC, Campbell BC (2011) Androgens and competitiveness in men. *J. Neuroscience Psych. Econom.* 4(1):54–62.
- Ariely D, Gneezy U, Loewenstein G, Mazar N (2009) Large stakes and big mistakes. *Rev. Econom. Stud.* 76(2):451–469.
- Ashraf N, Bandiera O, Lee S (2014) Awards unbundled: Evidence from a natural field experiment. *J. Econom. Behav. Organ.* 100(April):44–63.
- Azmat G, Calsamiglia C, Iriberrí N (2016) Gender differences in response to big stakes. *J. Eur. Econom. Assoc.* 14(6):1372–1400.
- Baum A, Grunberg NE (1997) Measuring stress hormones. Cohen S, Kessler RC, Gordon LU, eds. *Measuring Stress: A Guide for Health and Social Scientists* (Oxford University Press, Oxford, UK), 175–192.
- Beilock SL (2008) Math performance in stressful situations. *Current Directions Psych. Sci.* 17(5):339–343.
- Beilock SL, Holt LE, Kulp CA, Carr TH (2004) More on the fragility of performance: Choking under pressure in mathematical problem solving. *J. Experiment. Psych. General* 133(4):584–600.
- Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? *J. Econom. Behav. Organ.* 81(2):542–555.
- Buckert M, Schwioren C, Kudiella BM, Fiebach CJ (2017) How stressful are economic competitions in the laboratory? An investigation with physiological measures. *J. Econom. Psych.* 62: 231–245.
- Buser T, Dreber A, Mollerstrom J (2017) The impact of stress on tournament entry. *Experiment. Econom.* 20(2):506–530.
- Buser T, Niederle M, Oosterbeek H (2014) Gender, competitiveness and career choices. *Quart. J. Econom.* 129(3):1409–1447.
- Cahlíková J, Cingl L (2017) Risk preferences under acute stress. *Experiment. Econom.* 20(1):209–236.
- Cameron A, Gelbach J, Miller D (2008) Bootstrap-based improvements for inference with clustered errors. *Rev. Econom. Statist.* 90(3):414–427.
- Cannon WB (1932) The wisdom of the body. *Amer. J. Medicine Sci.* 184(6):864.
- Card D, Cardoso AR, Kline P (2016) Bargaining, sorting and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quart. J. Econom.* 131(2):633–686.
- DeCaro MS, Thomas RD, Albert NB, Beilock SL (2011) Choking under pressure: Multiple routes to skill failure. *J. Experiment. Psych. General* 140(3):390–406.
- Dickerson SS, Kemeny ME (2004) Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psych. Bull.* 130(3):355–391.
- Dohmen BT, Falk A, Huffman D, Sunde U (2010) Are risk aversion and impatience related to cognitive ability? *Amer. Econom. Rev.* 100(3):1238–1260.
- Dohmen TJ (2008) Do professionals choke under pressure? *J. Econom. Behav. Organ.* 65(3–4):636–653.
- Dreber A, von Essen E, Ranehill E (2011) Outrunning the gender gap — boys and girls compete equally. *Experiment. Econom.* 14(4): 567–582.
- Everly GS, Lating JM (2013) *The Anatomy and Physiology of the Human Stress Response: A Clinical Guide to the Treatment of the Human Stress Response* (Springer, New York), 17–51.
- Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* 41(4):1149–1160.
- Fischbacher U (2007) Z-TREE: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.
- Fletcher TD, Major D, Davis DD (2008) The interactive relationship of competitive climate and trait competitiveness with workplace attitudes, stress, and performance. *J. Organ. Behav.* 29(7):899–922.
- Flory JA, Leibbrandt A, List JA (2015) Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions. *Rev. Econom. Stud.* 82(1):122–155.
- Gneezy U, Rustichini A (2004) Gender and competition at a young age. *Amer. Econom. Rev.* 94(2):377–381.

- Gneezy U, Leonard KL, List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica* 77(5):1637–1664.
- Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quart. J. Econom.* 118(3):1049–1074.
- Gneezy U, Pietrasz A, Saccardo S (2018) On the size of the gender difference in competitiveness. *Management Sci.* 64(4):1541–1554.
- Goette L, Bendahan S, Thoresen J, Hollis F, Sandi C (2015) Stress pulls us apart: Trait anxiety modulates self-confidence under stress. *Psychoneuroendocrinology* 54:115–123.
- Goldstein DS, McEwen B (2002) Allostasis, homeostats, and the nature of stress. *Stress* 5(1):55–58.
- Greiner B (2004) An online recruitment system for economic experiments. Kremer K, Macho V, eds. *Forschung und wissenschaftliches Rechnen GWDG Bericht* 63 (Ges. für Wiss. Datenverarbeitung, Göttingen, Germany), 79–93.
- Halko ML, Sääksvuori L (2017) Competitive behavior, stress, and gender. *J. Econom. Behav. Organ.* 141:96–109.
- Haushofer J, Jang C (2015) Stress and temporal discounting: Do domains matter? Working paper, Princeton University, Princeton, NJ.
- Hellhammer DH, Wüst S, Kudielka BM (2009) Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology* 34:163–171.
- Holt CA, Laury SK (2002) Risk aversion and incentive effects. *Amer. Econom. Rev.* 92(5):1644–1655.
- Jurajda S, Munich D (2011) Gender gap in performance under competitive pressure: Admissions to Czech universities. *Amer. Econom. Rev.* 101(3):514–518.
- Kemeny ME (2003) The psychobiology of stress. *Current Directions Psych. Sci.* 12(4):124–129.
- Kirschbaum C, Pirke KM, Hellhammer DH (1993) The Trier Social Stress Test—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28(1–2):76–81.
- Margittai Z, Strombach T, van Wingerden M, Joëls M, Schwabe L, Kalenscher T (2015) A friend in need: Time-dependent effects of stress on social discounting in men. *Hormone Behav.* 73:75–82.
- McEwen BS (2012) Brain on stress: How the social environment gets under the skin. *Proc. Natl. Acad. Sci. USA* 109(Suppl 2):17180–17185.
- Miller R, Plessow F, Kirschbaum C, Stalder T (2013) Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: Evaluation of salivary cortisol pulse detection in panel designs. *Psychosomatic Medicine* 75(9):832–840.
- Nicolson NA (2008) Measurement of cortisol. Luecken LJ, Gallo LC, eds. *Handbook of Physiological Research Methods in Health Psychology* (SAGE Publications, Thousand Oaks, CA), 37–74.
- Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122(3):1067–1101.
- Niederle M, Vesterlund L (2010) Explaining the gender gap in math test scores: The role of competition. *J. Econom. Perspect.* 24(2):129–144.
- Niederle M, Vesterlund L (2011) Gender and competition. *Annual Rev. Econom.* 3(1):601–630.
- Ors E, Palomino F, Peyrache E (2013) Performance gender gap: Does competition matter? *J. Labor Econom.* 31(3):443–499.
- Pabst S, Brand M, Wolf OT (2013) Stress and decision making: A few minutes make all the difference. *Behav. Brain Res.* 250:39–45.
- Pruessner JC, Kirschbaum C, Meinlschmidt G, Hellhammer DH (2003) Two formulas for computation of the area under the curve represent measures of total hormone concentration vs. time-dependent change. *Psychoneuroendocrinology* 28(7):916–931.
- Schipper BC (2015) Sex hormones and competitive bidding. *Management Sci.* 61(2):249–266.
- Schmader T, Johns M, Forbes C (2008) An integrated process model of stereotype threat effects on performance. *Psych. Rev.* 115(2):336–356.
- Schram A, Brandts J, Gerxhani K (2019) Social-status ranking: A hidden channel to gender inequality under competition. *Experiment. Econom.* 22(2):396–418.
- Spencer SJ, Steele CM, Quinn DM (1999) Stereotype threat and women's math performance. *J. Experiment. Soc. Psych.* 35(1):4–28.
- Starcke K, Brand M (2012) Decision making under stress: A selective review. *Neuroscience Biobehavior Rev.* 36(4):1228–1248.
- Steyer R, Schwenkmezger P, Notz P, Eid M (1997) *MDBF-Mehrdimensionaler Befindlichkeitsfragebogen* (Hogrefe, Göttingen, Germany).
- Sutter M, Glätzle-Rützler D (2015) Gender differences in the willingness to compete emerge early in life and persist. *Management Sci.* 61(10):2339–2354.
- Taylor SE (2006) Tend and befriend: Biobehavioral bases of affiliation under stress. *Current Directions Psych. Sci.* 15(6):273–277.
- Taylor SE, Klein LC, Lewis BP, Gruenewald TL, Gurung RA, Updegraff JA (2000) Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psych. Rev.* 107(3):411.
- Vinkers CH, Zorn JV, Cornelisse S, Koot S, Houtepen LC, Olivier B, Verster JC, et al. (2013) Time-dependent changes in altruistic punishment following stress. *Psychoneuroendocrinology* 38:1467–1475.
- von Dawans B, Kirschbaum C, Heinrichs M (2011) The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology* 36(4):514–522.
- von Dawans B, Fischbacher U, Kirschbaum C, Fehr E, Heinrichs M (2012) The social dimension of stress reactivity: Acute stress increases prosocial behavior in humans. *Psych. Sci.* 23(6):651–660.
- Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. *J. Comparative Neurology Psych.* 18(5):459–482.
- Zhong S, Shalev I, Koh D, Ebstein RP, Chew SH (2018) Competitiveness and Stress. *Internat. Econom. Rev.* 59(3):1263–1281.