CESifo WORKING PAPERS

8147
2020

March 2020

Tournaments with Safeguards: A Blessing or a Curse for Women

Zhengyang Bao, Andreas Leibbrandt

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8147*

Tournaments with Safeguards: A Blessing or a Curse for Women

Abstract

Workplace tournaments are one likely contributor to gender differences in labor market outcomes. Relative to men, women are often less eager to compete and thrive less under competitive pressure. We investigate a competitive workplace environment that may produce more gender-neutral outcomes: tournaments with safeguards. In our experiments, participants take part in a tournament with a real effort task and choose whether they want to have a complimentary safeguard that guarantees higher wages for the low-ranked. As expected, we find that women are more likely than men to obtain such a safeguard. However, obtaining a safeguard comes at a cost. On average, the safeguard causes lower performance, creates a gender wage gap, and over-proportionally disadvantages women. Thus, we provide novel evidence that easing women into tournaments can backfire.

JEL-Codes: C920, J160, M520.

Keywords: workplace tournaments, gender differences, safeguard, experiment.

*Zhengyang Bao**
Department of Economics
Monash University
Australia – Clayton, VIC 3800
zhengyang.bao@monash.edu

Andreas Leibbrandt
Department of Economics
Monash University
Australia – Clayton, VIC 3800
Andreas.Leibbrandt@monash.edu

*corresponding author

This version: 17 Feb. 2020
Zhengyang is grateful to the faculty and staff at the University of Chicago for their hospitality during his visit, when this paper was written.

Introduction

There is little disagreement that gender equality is desirable in labor markets, yet there are still significant gender differences in key labor outcomes. For example, of the CEOs who lead the companies that make up the 2018 Fortune 500 list, less than 5% are women. Such differences may be partly attributed to tournaments because females tend to shy away from competitions (Niederle & Vesterlund 2007) and thrive less when they compete (Gneezy, Niederle & Rustichini 2003; Gneezy & Rustichini 2004). There are a few mechanisms studied in the literature to encourage women to join tournaments ranging from team instead of individual competitions (Healy & Pate 2011; Dargnies 2012; Flory, Leibbrandt & List, 2014), gender quotas (Niederle, Segal & Vesterlund 2012; Leibbrandt, Wang & Foo 2017) to making tournaments the default choice (Erkal, Gangadharan & Xiao 2019). However, luring women into tournaments may not be sufficient to narrow gender gaps. Once taking part in a tournament, women also need to find themselves in a competitive environment where they can thrive.

Many organizations employ tournaments, but they vary the risk exposure for low-ranked employees. On the one extreme, some use in-or-out partner schemes, while others use different types of safeguards to protect low-ranked employees. For example, organizations can provide tenure and significant base salaries so that the consequences for even the lowest-ranked workers are moderate.[1] However, it is very difficult to identify whether such safeguards are useful to tackle gender gaps in labor outcomes as their implementation is typically highly correlated with organizational and industry characteristics.

In this experimental study, we explore a tournament environment that may help women without hurting men. We design a real-effort rank-order tournament with an elective 'safeguard', a device which softens the consequences of being low-ranked. More precisely, workers are informed about the availability of this safeguard and decide whether to obtain it before the start of the tournament. The safeguard is complementary and increases the user's minimum wage if her relative performance falls into the lowest ranking category. Our conjecture is that the safeguard is particularly popular amongst women, improves women's relative to men's outcomes and alleviates psychological competitive pressure to perform.

[1] For example, in many countries, government jobs are considered to be more secured than jobs which require similar skills in private sectors because the government sector usually has a lower dismiss rate. In academia, there are also large differences in job security. In the US, for example, tenure is usually only awarded with promotion to the associate or full professor level, while tenured assistant professors are common in the UK and Australia. In China, some universities offer permanent positions already to fresh PhD graduates.

Our findings show that women are indeed more likely than men to select a complementary safeguard. However, we also find that this safeguard increases the gender wage gap as compared to more standard tournaments. Further, we observe that the safeguard reduces performance of both women and men per se, regardless whether voluntarily selected or automatically implemented. Our survey findings suggest that the safeguard does not to alleviate pressure more from women than from men and that the safeguard tempts both genders to slack off.

Our study is closely related to the literature on gender differences in the selection of incentive schemes (Eckel & Gorssman 2008; Gneezy, Leonard & List 2009; Dohmen & Falk 2011; Flory, Leibbrandt & List 2014). This literature provides evidence that women are more likely than men to sort out of competitive environments such as workplace tournaments (Niederle & Vesterlund 2007; Gneezy, Niederle & Rustichini 2003; Gneezy & Rustichini 2004). We depart from this literature by studying an environment where competition cannot be avoided altogether. Such environments are common in hierarchical organizations where promotion to higher levels is often based on competition.

We also contribute to the literature on gender quotas and other policies aimed at improving female labor outcomes (Healy & Pate 2011; Dargnies 2012; Niederle, Segal & Vesterlund 2012). While there is some evidence that they can address gender gaps (Erkal, Gangadharan & Xiao 2019), there is also evidence that they might backfire (Leibbrandt, Wang & Foo 2017; Leibbrandt & List 2018). Our intervention deviates from most affirmative action policies because it does not treat women differently than men but only provides an additional choice that is accessible to either gender. Nevertheless, we find that our instrument can also backfire because the difference in the selection process distorts the incentives disproportionately across genders. The results provide evidence that supports the need for caution when designing seemingly harmless policies that may attract one gender more than the other.

In addition, we contribute to the tournament literature.[2] A key topic in this literature is how individuals react to exogenously imposed variations in the prize structure (see e.g., Harbring & Irlenbusch 2003; Sheremeta 2011; Moldovanu & Sela 2001; Orrison, Schotter & Weigelt 2004) and how this relates to behavioural aspects (e.g., Sheremeta 2015; Hossain, Hong & List 2014; Delfgaauw et al. 2013). We complement this literature by investigating

[2] For a survey on the literature studying tournament, the reader is referred to Dechenaux, Kovenock & Sheremeta (2015).

reactions to safeguards that allow for endogenous selection of the lowest prize and how this relates to gender differences in tournament outcomes.

2. Experimental Design

2.1 The real-effort task

In this experiment, we use a real-effort task, which mimics tedious work assignments that require focus as well as mathematical and verbal skills (see e.g., Carpenter, Matthews & Schirm 2010; Gill & Prowse 2012; Leibbrandt, Wang & Foo 2018; Erkal, Gangadharan & Nikiforakis 2011). In this task, participants are asked to solve as many 'puzzles' as possible in a timespan of 40 minutes. Figure 1 provides one example of such a puzzle. Participants need to first decipher the value of five letters and then sum up the corresponding values. In this example, letter 1 through 5 have values 61 (g), 30 (c), 52 (h), 30 (c) and 50 (b), respectively and the correct answer is 223 (61+30+52+30+50). After typing their answers in the answer box and press the *Next* button, the computer displays a different set of numbers until the participants run out of time.[3] Participants were not allowed to use calculators but could write on scrap paper.

INSERT FIGURE 1 ABOUT HERE

2.2 The rank-order tournament

All participants take part in a rank-order tournament and their performance (number of correctly solved puzzles) relative to other participants determines their monetary compensation. There are three compensation levels: (*i*) the top 10% participants in a given session receive \$60, (*ii*) the top 10-50% participants receive \$30, and (*iii*) the remaining participants (bottom 50%) receive \$15 or \$20, depending on treatment and choice. Thus, the tournaments use a multiple-prize payment rule common to many workplaces (Kalra & Shi 2001; Vandegrift, Yavas & Brown 2007; Cason, Masters & Sheremeta 2010).[4]

2.3 The treatments

This study has three main treatments. The *safeguard choice* treatment allows participants to select their own tournament incentives. To properly identify the impact of this

[3] To make sure each quiz has similar difficulty across workers and treatments, all values are two-digit integers and all letters in the encode game are randomly drawn between a to j.

[4] When the cut off between prize levels is not an integer, we round up the number of workers who get the better prize in favor of workers. For example, if there are 28 workers in a group, then the top three performers receive \$60.

choice, there are two other treatments: *compulsory safeguard* and *no safeguard*. All participants took part in one treatment only. The experiment instructions are in the Appendix.

2.3.1 Safeguard choice treatment

In this treatment, workers choose whether to have a complimentary safeguard before the task starts. The safeguard guarantees a higher minimum payment if performance is in the lower half but does not affect payment if performance is in the upper half. More precisely, as can be seen in Table 1 if a participant chooses the safeguard, she will get $20 if her performance is in the bottom 50% instead of only $15 if she does not choose the safeguard. If her performance is anywhere in the top 50%, her payment is not affected.

INSERT TABLE 1 ABOUT HERE

2.3.2 Compulsory safeguard treatment

In this treatment, the safeguard is already embedded in the incentive structure for all participants. That is, the minimum payment is $20, just like in the safeguard choice treatment if participants choose a safeguard. The safeguard guarantees a minimum payment of $20 for all. We tested two versions of this treatment. In one frame, we make the safeguard explicit (i.e., we tell participants that the minimum compensation is only $15 in some of the other sessions). In the other frame, the safeguard is implicit and unknown to the participants (i.e., we give workers no reference about the other group's prize structures).

2.3.3 No safeguard treatment

In this treatment, there is no safeguard available and the bottom 50% receive $15. Participants are unaware that participants in other sessions had access to a safeguard.

2.4 Conjectures

The standard prediction for all three treatments is that all workers are incentivized to provide effort and thus increase their likelihood to increase their compensation. Thus, the first conjecture is that there are no treatment differences in effort across treatments. The second conjecture applies to the *safeguard choice* treatment, in which we conjecture that all participants choose the safeguard as it weakly dominates not choosing the safeguard.

These standard predictions do not take into account mental effort costs associated with fatigue and temptations to rest. While the tournaments provide significant incentives to work hard (possible tripling to quadrupling of compensation), it is possible that some participants' mental effort costs are sufficiently high to prevent them from providing

(maximal) effort. Accordingly, the alternative prediction is that effort is lower in the compulsory safeguard treatment than in the no safeguard treatment because the incentives are less pronounced to leave the bottom 50% in the former. Further, it is possible that some participants do not choose the safeguard if they believe that it may undermine their effort provision and thus chances to increases their payments. On the other hand, it is also possible that the safeguard reduces pressure and stress and thus affects performance. Whether less pressure and less stress increase or decrease performance is still an open question, although some progress has been made (e.g., Harbring & Irlenbusch 2003; van Dijk, Sonnemans & van Winden 2001; Hall & Lawler 1971; Allen, Hitt & Greer 1982; Compte & Postlewaite 2004).

So far, these predictions do not take into account gender. However, there is evidence that gender plays a crucial role in tournaments. In particular, there is evidence that women as compared to men are more risk-averse (Croson & Gneezy 2009; Charness & Gneezy 2012), less likely to enter tournaments (Niederle & Vesterlund 2007; Gneezy, Leonard & List 2009; Flory, Leibbrandt & List 2015), and underperform in competitions (Gneezy Niederle & Rustichini 2003, Gneezy & Rustichini 2004). In the safeguard choice treatment, we allow for endogenous selection into two different tournaments, which allows workers to limit their risk exposure to low compensation. Thus, our conjecture is that women are more likely than men to choose the safeguard. In turn, this may improve or harm their compensation relative to men, depending on whether it increases temptations to rest and has a gender dependent impact on pressure and stress.

2.5 Experimental procedures

The experiment is programmed in z-Tree (Fischbacher 2007) and workers are recruited with the software SONA. In total, 431 workers took part in the three treatments and they earned on average $32 for a roughly 70 minutes experimental session. Before the start of the experiment, workers read an information sheet and signed a consent form. Thereafter, we read the instructions aloud. After reading the instructions, workers had time to read the instructions on their own and ask questions. We then implemented a practice round, identical to the questions in the real effort task and quiz questions to make sure that workers understood the instructions and payment mechanism. After they answered these questions correctly, they started with the real-effort task. When all workers completed the task, we administered a short post-experiment questionnaire to conclude. In the questionnaire, we collected the workers demographics, self-evaluation of psychological wellbeing during the experiment and incentivised them to reveal their beliefs about their own performance and

group average performance (Hossain & Okui 2013) as well as their risk preferences (Eckel & Grossman 2002). There was no mentioning of gender throughout the entire experiment to prevent potential experimenter demand effects.

3. Findings

3.1 Experimental findings

We observe high effort levels in our experiment. Figure 2 illustrates the average number of attempted questions (effort) and correct questions (performance) across treatments. On average, 72.6 quizzes were attempted and 65.4 solved correctly during the course of 40 minutes and we observe few cases (7.9%) where participants attempted less than one quiz per minute. This suggests that the large majority of the participants was motivated to provide substantial effort.

INSERT FIGURE 2 ABOUT HERE

We find significant treatment differences in effort levels. In particular, we observe that workers attempt less quizzes in the compulsory safeguard treatment than in the no safeguard treatment (78.1 vs 66.7, $p<0.001$) and that the performance is also significantly lower (70.1 vs 60.7, $p<0.001$).[5] This finding squares well with the alternative prediction based on fatigue and temptation but is inconsistent with the standard prediction of no treatment differences in effort levels across treatments. In addition, we observe that participants also attempt less quizzes in the safeguard choice treatment than in the no safeguard treatment (72.5 vs 78.1, $p=0.037$) and that the performance is also lower (65 vs. 70.1, $p=0.062$). There are also significant differences between the compulsory and choice safeguard treatment in both effort (66.7 vs 72.5, $p=0.03$) and performance (60.7 vs 65.0, $p=0.095$).

Table 2 reports the estimation results of an OLS regression with effort and performance as dependent variables. We observe the following. The compulsory safeguard significantly reduces the effort and performance by around 14.5% ($p<0.01$) and 13.2% ($p<0.01$), respectively, compared with the baseline. For workers in the choice treatment, those who choose to opt against the safeguard perform similarly to the baseline ($p=0.61$ for effort and $p=0.37$ for performance); those who choose to use the safeguard perform like those in the compulsory treatment and their performance is 8.1-8.3% worse than the baseline in

[5] For simplicity, we pool in the main analysis the two versions in the compulsory safeguard treatment. The findings in the two versions are qualitatively similar and discussed in Section 3.3.3.

terms of effort and performance ($p<0.05$ for both cases). Finally, the male dummy is significant at the 5%-level and shows that men provide more effort and have a higher performance ($p<0.05$ for both effort and performance).

INSERT TABLE 2 ABOUT HERE

Result 1: *The safeguard reduces effort, regardless whether implemented as a choice or as default. Individuals who decide against using a safeguard provide more effort than those who decide using a safeguard. Only workers that decide against using a safeguard perform similarly well as workers who do not have a safeguard.*

The safeguard is popular despite the negative impact on effort and performance (87% choose the safeguard). Importantly, and as conjectured, we find that women are more likely to choose the safeguard as compared to men (93.1% vs. 81.8%, $p=0.0368$).

INSERT TABLE 3 ABOUT HERE

Table 3 reports the characteristics of the participants who chose the safeguard. We can see that the fraction of men who chose the safeguard is much higher in the sample of participants who do not choose the safeguard (76.2%) than in the sample of participants who choose the safeguard (51.8%). Characteristics other than gender play less important roles for the choice of the safeguard. In fact, only risk preferences are a marginal significant predictor of safeguard choice. Participants whose risk-taking behavior in a risk task is below the medium, are more likely to encounter in the sample of those who choose the safeguard (81% vs. 62.6%, $p=0.1$). Task ability (math skills) and confidence (the ratio of guessed own performance and guessed group performance) are not significant drivers for the choice of the safeguard. If we regress the choice of safeguard on these characteristics (risk preferences, math skills, confidence), only gender remains a significant predictor (Table 4, $p<0.05$).

INSERT TABLE 4 ABOUT HERE

Result 2: *Gender is the most important predictor for the choice of the safeguard. Women are more likely than men to choose a complimentary safeguard in tournaments.*

The safeguard benefits women only at first sight. Figure 3 illustrates for each gender how likely the safeguard is to materialize and increase the compensation in the safeguard choice treatment. The two bars on the left side show that in this treatment, 56.9% of the female workers and only 33% of the male workers receive the additional $5 from the materialization of the safeguard ($p<0.01$). In contrast, in the two bars on the right side for the

compulsory safeguard treatment, we can see that gender plays no role as an almost identical percentage of either gender receive the safeguard payment of $20.

INSERT FIGURE 3 ABOUT HERE

A closer look at the data reveals that women do worse than men when there is a voluntary safeguard. Figure 4 compares the compensation for male and female workers across treatments. Compared to the baseline, we observe that compensation is higher for both men ($27.0 vs $30.3, p=0.013) and women ($24.9 vs $26.0, p=0.024) in the choice treatment, however, the treatment impact is gender specific. Men's average wage increase by $3.3 (12.2%), while women's wage only increase by $1.1 (4.4%). In addition, compared with the compulsory treatment, giving workers the choice of safeguard slightly increases men's payoff ($29.6 vs $30.3, p=0.60) but somewhat decreases women's payoff ($27.6 vs $26.0, p=0.138). Perhaps most importantly, we observe that the safeguard choice treatment creates a significant gender wage gap of $4.2 (wage for women = $26.0, men = $30.2, p=0.02), which is insignificant in the other two treatments (p>0.38).

INSERT FIGURE 4 ABOUT HERE

Figure 5 provides insights as to why the safeguard choice backfires for women. This figure illustrates the ranking and prizes in the different treatments depending on gender. The middle panel shows that women are much more likely to score in the bottom 50% in the safeguard choice treatment than men and that this is not paralleled in the compulsory and no safeguard treatment. In fact, such a gender difference is significant for the choice treatment (p=0.027), but not for the baseline (p=0.28) and compulsory treatment (p=1.0).

INSERT FIGURE 5 ABOUT HERE

Figure 6 provides a more fine-grained illustration of effort levels depending on gender and treatment. In this figure, we compare the effort level of each quantile in the no safeguard treatment against the same quantile in the safeguard choice treatment. While this quantile plot does not show the change in the effort level caused by our intervention at each quantile as we use a between subject design, it allows for causal interpretation at the aggregate level. For men, the choice of safeguard shifts the overall distribution of correct answers lower than the baseline, except for the quantiles near the two prize cut-offs. Women, in contrast, appear to differently respond to the safeguard choice. All quantiles in the safeguard choice treatment are higher in the no safeguard treatment, except for few between the cut-offs. Thus, the

safeguard choice appears to cause most men and women to provide less effort as compared to the no safeguard, but this reduction is more pronounced for women whose performance is close to move up the ranks.

INSERT FIGURE 6 ABOUT HERE

Result 3: *The availability of a complimentary safeguard disadvantages women more than men and creates a gender wage gap.*

3.2 Survey findings

The experimental findings are consistent with our alternative conjecture based on the assumption that the safeguard increases temptation to rest and decreases stress. To provide insights on these potential underlying mechanisms and their relationships to gender, we conduct a survey with the participants after the tournament but before revealing information about their performance. In this survey, we ask participants to report their stress and temptation to rest levels during the real-effort task and how a removal of the safeguard would change these.[6]

Figure 7 illustrates several corresponding survey insights. First, while we find that only a minority of workers experiences high levels of stress and temptation, many report to experience some stress and temptation suggesting the presence of mental costs. Second, we observe that a significant proportion of participants report that stress (45%) and temptations (35%) are more pronounced when there is no safeguard. Third, and perhaps most importantly, the psychological impact of the safeguard appears to be gender independent. The reported stress and temptation levels are similar for men and women in the presence and absence of a safeguard ($p=0.91$ for change in stress and $p=0.465$ for change in temptations), suggesting that explanations based on mental costs cannot explain the observed gender differences in choice of the safeguard, performance and wage levels.

INSERT FIGURE 7 ABOUT HERE

4. Discussion

There is substantial evidence that uncertainty in labor relations is more detrimental to women than to men (e.g., Frederiksen 2008; Hirsch & Schnabel 2012; García 2017). There is also evidence that giving workers choice over their employment conditions can improve labor

[6] We only asked participants in the compulsory treatment to ensure that everyone has experienced the safeguard.

outcomes (e.g., Bloom et al. 2014; Leslie et al. 2012; Beckmann, Cornelissen & Kräkel 2017). We investigate a workplace tournament that reduces uncertainty by providing workers with the choice of a safeguard that increases the minimum wage. Our findings suggest that giving workers the autonomy to select the incentive scheme can disadvantage women. This is because women are more tempted to choose safeguards, even though they weaken the incentive to exert effort. Our findings provide novel evidence on the limitations of tournaments to create gender neutral outcomes.

Our experiment also provides insights for the literature on incentive contracts beyond the economics of genders. First, providing safeguards to low-performing workers appears to be not only costly but also counter-productive as they lower effort and performance. At the same time, such safeguards can even have a detrimental impact on high-performing workers and thus decrease effort and performance throughout the whole distribution of workers. More generally, our study contributes to the discussion of optimal incentives for policies that target low-performing individuals (e.g., Rosen 1986; Heckman 2006; Mario et al. 2020). For example, poverty alleviation programs are more likely to affect the wage of low-skilled workers and it is of key importance to understand whether and under which circumstances they cause lower effort and thus might lead to larger, not smaller wage gaps.

Our study is a novel attempt to study the role of endogenous choice and safeguards in tournaments. We envision several extensions for future research. First, it seems important to further investigate in which environments workers benefit from having the choice over their compensation scheme and whether it is a general property that giving this choice is less beneficial for women than men. Second, our data suggests that opting against the safeguard serves as a self-control mechanism. It may be interesting to study ways to make individuals "burn the boat" to achieve greater success.

Reference

Allen, R. D., Hitt, M. A., & Greer, C. R. (1982). Occupational stress and perceived organizational effectiveness in formal groups: An examination of stress level and stress type. *Personnel Psychology*, *35*(2), 359-370.

Beckmann, M., Cornelissen, T., & Kräkel, M. (2017). Self-managed working time and employee effort: Theory and evidence. *Journal of Economic Behavior & Organization*, *133*, 285-302.

Bloom N, Garicano L, Sadun R, et al. The distinct effects of information technology and communication technology on firm organization[J]. Management Science, 2014, 60(12): 2859-2885.

Carpenter, J., Matthews, P. H., & Schirm, J. (2010). Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review*, *100*(1), 504-17.

Cason, T. N., Masters, W. A., & Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, *94*(9-10), 604-611.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50-58.

Compte, O., & Postlewaite, A. (2004). Confidence-enhanced performance. *American Economic Review*, *94*(5), 1536-1557.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, *47*(2), 448-74.

Dargnies, M. P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, *58*(11), 1982-2000.

Dechenaux, E., Kovenock, D., & Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, *18*(4), 609-669.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, *31*(2), 305-326.

Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American economic review*, *101*(2), 556-90.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, *23*(4), 281-295.

Eckel, C. C., & Grossman, P. J. (2008). Differences in the economic decisions of men and women: Experimental evidence. *Handbook of experimental economics results*, *1*, 509-519.

Erkal, N., Gangadharan, L., & Nikiforakis, N. (2011). Relative wage and giving in a real-effort experiment. *American Economic Review*, *101*(7), 3330-48.

Erkal, N., Gangadharan, L., & Xiao, E. (2019). Leadership Selection: Can Changing the Default Break the Glass Ceiling?. *Available at SSRN 3022386*.

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

Flory, J. A., Leibbrandt, A., & List, J. A. (2015). Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, *82*(1), 122-155.

Frederiksen, A. (2008). Gender differences in job separation rates and employment stability: New evidence from employer-employee data. *Labour Economics*, *15*(5), 915-937.

García, A. N. (2017). Gender differences in unemployment dynamics and initial wages over the business cycle. *Journal of Labor Research*, *38*(2), 228-260.

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469-503.

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

Gneezy, U., Leonard, K. L., & List, J. A. (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. *The Quarterly Journal of Economics*, *118*(3), 1049-1074.

Hall, D. T., & Lawler, E. E. (1971). Job Pressures and Research Performance: Contrary to popular opinion, job pressures are not necessarily undesirable; in fact, certain pressures seem to enhance researchers' job attitudes and performance. *American Scientist*, *59*(1), 64-73.

Harbring, C., & Irlenbusch, B. (2003). An experimental study on tournament design. *Labour Economics*, *10*(4), 443-464.

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

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902.

Hirsch, B., & Schnabel, C. (2012). Women move differently: Job separations and gender. *Journal of Labor Research*, *33*(4), 417-442.

Hong, F., Hossain, T., & List, J. A. (2015). Framing manipulations in contests: a natural field experiment. *Journal of Economic Behavior & Organization*, *118*, 372-382.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, *80*(3), 984-1001.

Kalra, A., & Shi, M. (2001). Designing optimal sales contests: A theoretical perspective. *Marketing Science*, *20*(2), 170-193.

Leibbrandt, A., & List, J. A. (2018). Do Equal Employment Opportunity Statements Backfire? Evidence from A Natural Field Experiment On Job-Entry Decisions. *National Bureau of Economic Research*.

Leibbrandt, A., Wang, L. C., & Foo, C. (2018). Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science*, *64*(8), 3501-3516.

Leslie, L. M., Manchester, C. F., Park, T. Y., & Mehng, S. A. (2012). Flexible work practices: A source of career premiums or penalties? *Academy of Management Journal*, *55*(6), 1407-1428.

Mario, B., Nicole, G., Ute, B., Lisa, F., & Benjamin, L. (2020). The German Minimum Wage: Effects on Productivity, Profitability, and Investments. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, *240*(2-3), 321-350.

Moldovanu, B., & Sela, A. (2001). The optimal allocation of prizes in contests. *American Economic Review*, *91*(3), 542-558.

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

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

Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, *59*(1), 1-16.

Orrison, A., Schotter, A., & Weigelt, K. (2004). Multiperson tournaments: An experimental examination. *Management Science*, *50*(2), 268-279.

Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics*, *1*, 641-692.

Sheremeta, R. M. (2011). Contest design: An experimental investigation. *Economic Inquiry*, *49*(2), 573-590.

Sheremeta, R. M. (2015). Behavioral dimensions of contests. In *Companion to the political economy of rent seeking*. Edward Elgar Publishing.

Van Dijk, F., Sonnemans, J., & Van Winden, F. (2001). Incentive systems in a real effort experiment. *European Economic Review*, *45*(2), 187-214.

Vandegrift, D., Yavas, A., & Brown, P. M. (2007). Incentive effects and overcrowding in tournaments: An experimental analysis. *Experimental Economics*, *10*(4), 345-368.

Appendix

A1 Instructions

Here, we provide the instructions for the choice treatment. The instructions for the other treatments only differ in the availability of the safeguard and available upon request.

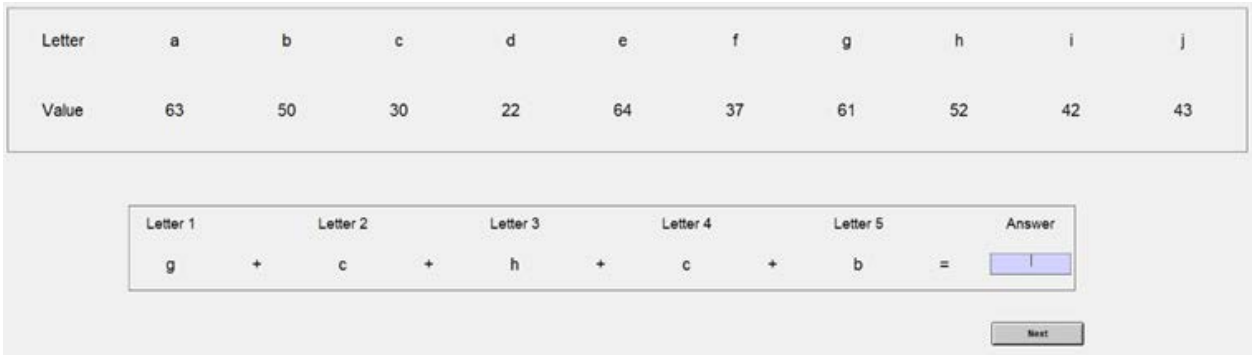
1. General Instructions

Thank you for participating in this experiment. Please read the following instructions carefully. After reading the instructions, there will be some questions to check that all participants have understood the experiment. Thereafter, the main experiment will start. Note that you will be paid in private and in cash at the end of the experiment.

If you have any questions during the experiment please raise your hand and we will come to you. Please do not ask your questions out loud, or attempt to communicate with other participants, or look at other participants' computer screens at any time during the experiment. Please turn your phone to silent mode and put it in your bag. Please do not use any calculator during the experiment, including the one available with the Windows system.

2. The Task

You will have 40 minutes (2400 seconds) to solve puzzles similar to the one in the below graph.



To solve this puzzle, you first need to decipher the code in the box below. For example, letter 1 (g) has value 61, letter 2 (c) 30, letter 3 (h) 52, letter 4 (c) 30 and letter 5 (b) 50. The correct answer is the summation of all 5 values: correct answer=61+30+52+30+50=223. After typing your answer in the answer box, you have to press the Next button below to access the next puzzle.

3. Payment

Your payment depends on your performance relative to the other participants in this session and whether you choose to have safeguard.

If you <u>do not choose the safeguard</u>, only your relative performance determines your wage:

- If your number of correct answers is among the top 10%, then you will receive **$60**.
- If your number is among the top 10%-50%, then you will receive **$30**.
- If your number is among the bottom 50%, then you will receive **$15**.

If you choose to have the safeguard, then you earn at least $20. However, it costs you $1 if your number of correct answers are among the top 50%:

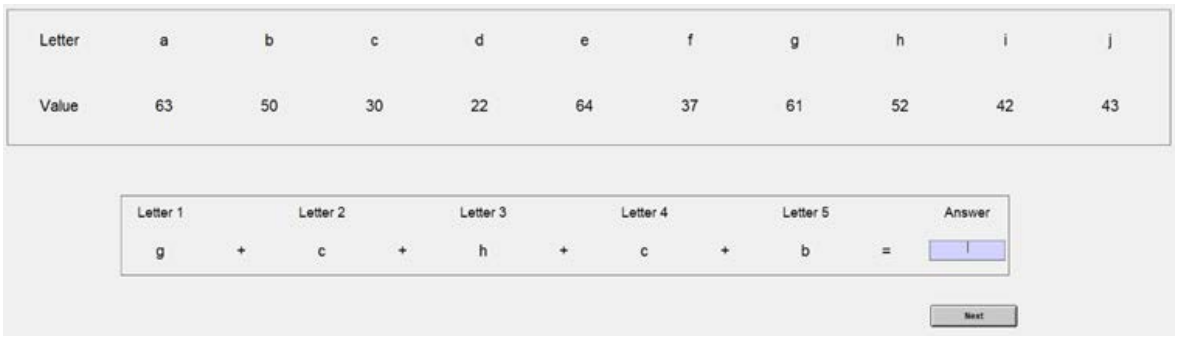
- If your number of correct answers is among the top 10%, then you will receive **$59**.
- If your number is among the top 10%-50%, then you will receive **$29**.
- If your number is among the bottom 50%, then you will receive **$20**.

A2 Framing of the safeguard

We investigate the role of the framing of the safeguard here. We use two frames in the compulsory safeguard treatment, which vary whether participants are made aware of the presence of a safeguard. More precisely, in the first frame we only mention the payment rule without any mentioning of a safeguard. In contrast, in the second frame, we present participants with the concept of the safeguard and inform them that their minimum payment ($20) is higher than in some of the other groups ($15), because of the presence of a safeguard. We observe no differences between frames in terms of quizzes attempted and quizzes correct ($p=0.8$ for attempt $p=0.9$ for correct). This also holds true if we analyse males and females separately. For males, the p-value testing no difference caused by the framing is 0.37 for questions attempted and 0.30 for questions correct; for females, the numbers are 0.62 and 0.38, respectively.

Figures

Figure 1: Illustration of Task

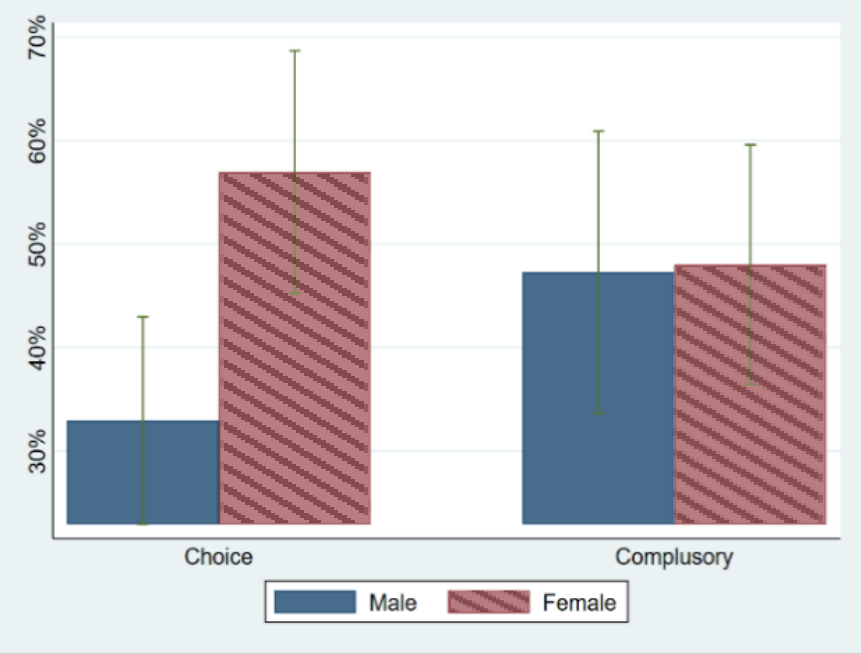


Notes: the figure provides a screenshot of an example puzzle that workers have to solve during the experiment. They first need to decipher the code in the lower box using the upper table and then sum up all the values.

Figure 2: Effort and performance levels depending on treatment

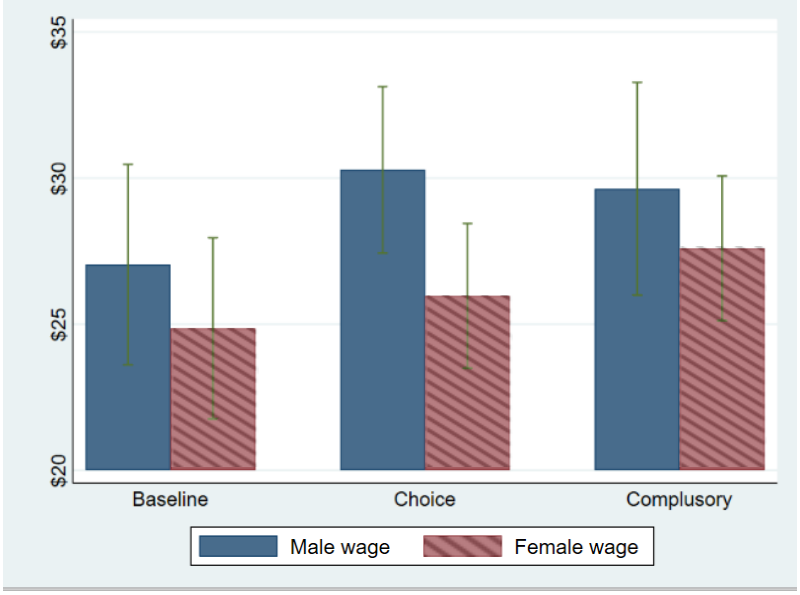
Notes: This figure shows the average attempts and performance depending on treatment. Participants had 40 minutes to attempt as many quizzes as possible. Error bars indicate 95% confidence intervals.

Figure 3: Materialization of safeguard depending on treatment and gender



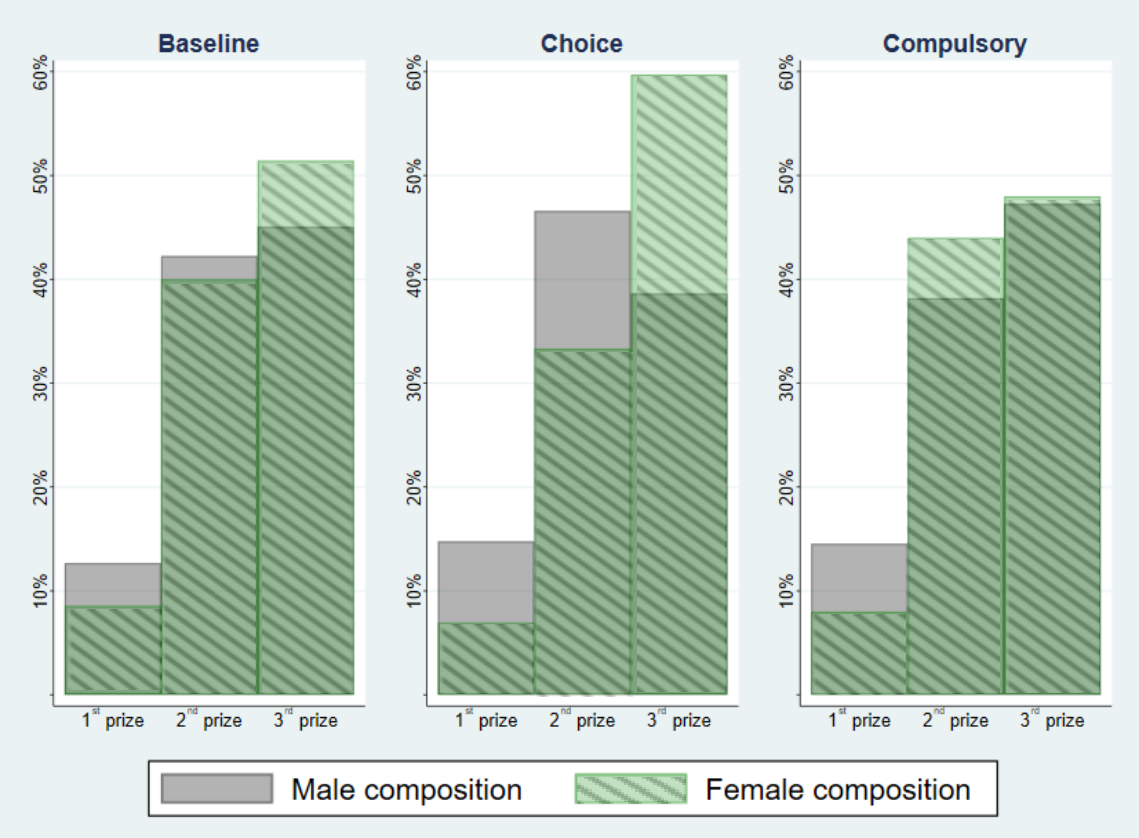
Notes: This figure illustrates the percentage of workers whose wage increased because of the safeguard.

Figure 4: Average wage by gender in each treatment



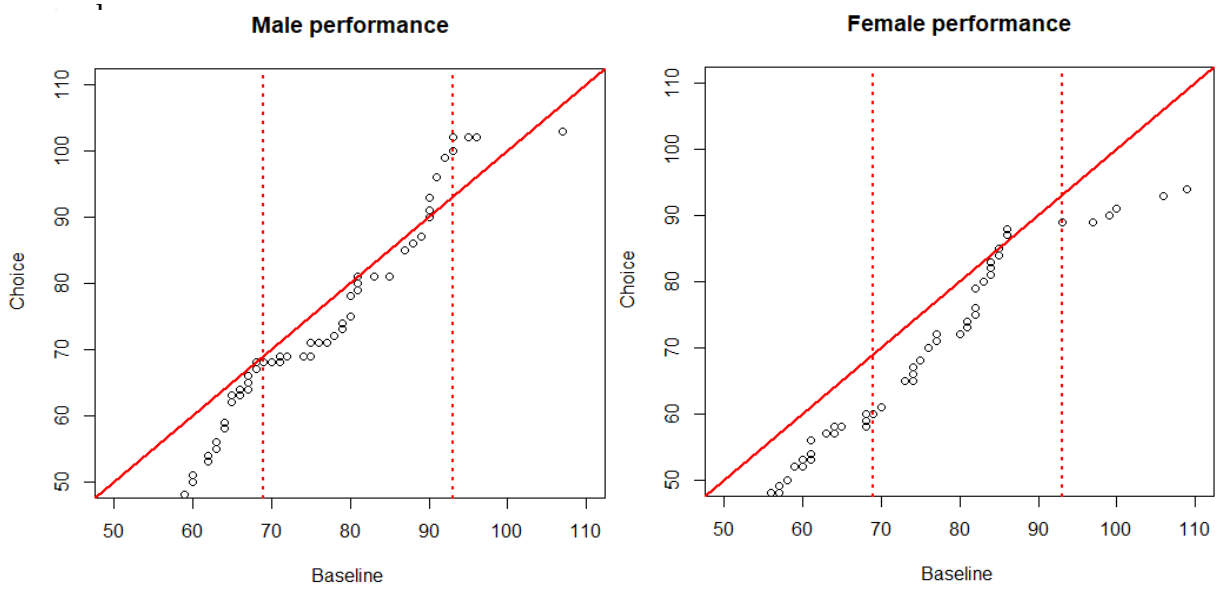
Notes: this figure shows the average wage for each gender across treatments. The error bars indicate 95% confidence intervals.

Figure 5: The gender composition of each prize level for each treatment



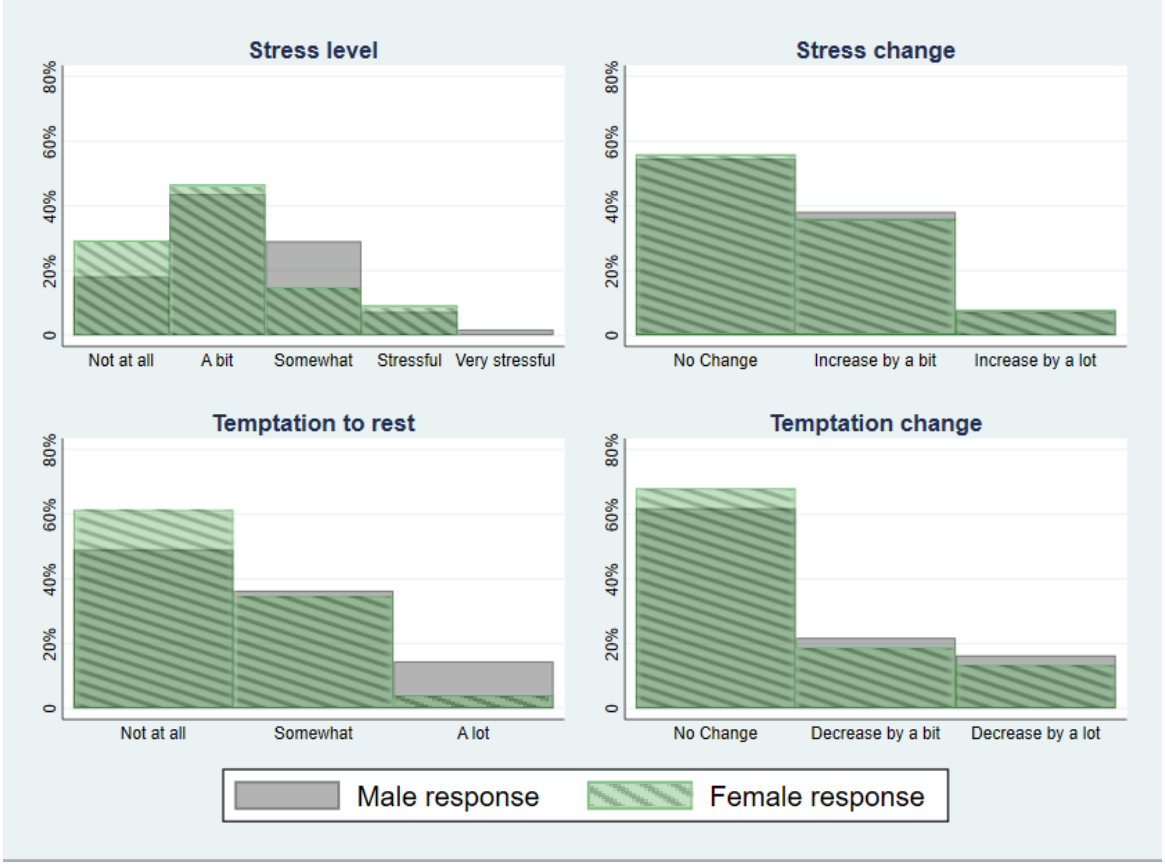
Notes: this figure plots the gender composition of each prize level in all treatments. The first price is awarded to the top 10% performers, the second is awarded to the 10%-50% performers, and the third is awarded to the bottom 50%.

Figure 6: Performance differences between choice and baseline treatment depending on



Notes: this figure compares the performance in the baseline with the choice treatment for both genders. Each quantile of the baseline is matched with that of the choice treatment and plotted against the 45-degree line. Any point above the line indicates that the choice treatment has a higher performance for that particular quantile and vice versa. The two dashed vertical lines indicate the 50% and 10% cut-offs for the different prizes in the baseline.

Figure 7: Treatment impact on stress and temptation depending on gender



Notes: this figure shows the survey response regarding the stress level and temptation to rest during the tournament and how do they expect a removal of the safeguard would change these between male and female workers. All survey questions are implemented in the form of multiple-choice questions and the possible choices map to the labels in the subfigures. The first row of subfigures plots the reported stress level and the expected change in the stress level from using the safeguard to not using it. The second row shows the reported temptation to take a rest during the tournament and the expected change in the temptation from opting for the safeguard to opt against it. We collect data from the compulsory treatment to ensure that all participants have experienced the safeguard.

Tables

Table 1: Treatment overview

	Safeguard choice (n=160)		No safeguard (n=141)	Compulsory (n=130)
	No safeguard	Safeguard		
Top 10%	$60	$60	$60	$60
Top 10%-50%	$30	$30	$30	$30
Bottom 50%	$15	$20	$15	$20

Notes: this table shows the tournament payment structure for each of the three treatments.

Table 2: Effort and performance depending on treatment

	Performance	Effort
Compulsory	-8.985***	-10.962***
	(2.628)	(2.753)
Choice × not use safeguard	-4.374 (4.837)	-2.627 (5.136)
Choice × use safeguard	-5.480** (2.626)	-6.283** (2.738)
Male	4.961**	5.259**
	(2.077)	(2.390)
Constant	67.594***	75.430***
	(2.264)	(2.309)

Notes: This table shows the OLS estimation results of the regression comparing the outcomes across treatments and the choice of the safeguard. Performance is defined by the number of correctly solved quizzes. Effort defines the number of attempted quizzes. The constant correspond to the outcomes of female workers in the baseline and all other independent variables are dummy variables. There are 431 observations for each regression. We report the estimate result of the regression models and include the robust standard error in parentheses under each point estimate. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Determinants of safeguard choice

Variable	Use safeguard	Do not use safeguard	Difference	p-value
Male (dummy)	51.8%	76.2%	-24.4%	0.04
Risk>medium (dummy)	81.0%	62.6%	18.4%	0.10
Good at math (dummy)	64.7%	66.7%	-2.0%	0.86
Confidence (guessed own performance/ guessed group average)	1.20	1.38	-0.18	0.30

Notes: this table compares key variables between the groups who choose to use the safeguard against those who opt out within the choice treatment. All these variables are collected by the post-experiment survey, where we ask the participants questions on their gender, risk preference (Eckel & Grossman 2002), subjective belief about their mathematical abilities, and guess about their own performance relative to the group average. Male is a dummy variable that takes value 1 if and only if the participant is a male; Risk>medium is a dummy variable that takes value 1 if and only if the participant is more risk-loving than the medium in the risk eliciting task; Good at math is a dummy variable that takes value 1 if and only if the participant's answer to this question is yes; confidence is a continuous variable calculated as the ratio of the participant's guess about her own performance and her guess about the average performance in the respective experimental session. The numbers in the second and third columns in the table are the average values of these variables (Male, Risk>medium, Good at math, Confidence) of the group who choose to use the safeguard and the one who choose not to use the safeguard, respectively. Each p-value reports the Mann-Whitney test result testing the null hypothesis that there is no difference between the group use the safeguard and the one does not use the safeguard.

Table 4: Individual determinants of the safeguard choice

	Choice of safeguard
Male (dummy)	-0.106**
	(0.052)
Good at Math (dummy)	0.023
	(0.058)
Risk>medium (dummy)	-0.071
	(0.049)
Confidence (guessed own performance	-0.107
/ guessed group average)	(0.090)
Constant	1.092***
	(0.118)

Notes: This table shows the OLS estimation the linear provability model studying how individual characteristics affect the choice of safeguard. Male is a dummy variable that takes value 1 if and only if the participant is a male; Risk>medium is a dummy variable that takes value 1 if and only if the participant is more risk-loving than the medium in the risk eliciting task; Good at math is a dummy variable that takes value 1 if and only if the participant's answer to this question is yes; confidence is a continuous variable calculated as the ratio of the participant's guess about her own performance and her guess about the average performance in the respective experimental session. There are 160 observations for the regression. We include the robust standard error in parentheses under each point estimate. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.