Department of Economics
ISSN number 1441-5429
Discussion number 2/20

Tournaments with Safeguards: A Blessing or a Curse for Women? *

Zhengyang Bao[a†] and Andreas Leibbrandt[a]

[a] Department of Economics, Monash University, Clayton, VIC 3800, Australia.

This version: March 19, 2020

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 experiment, participants take part in a tournament with a real-effort task and choose whether they want to have a complementary safeguard that guarantees higher wages for the low ranked. As expected, we find that women are more likely than men to choose 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.

Key words: Workplace tournaments, gender differences, tournament safeguards, incentive contracts

* We would like to thank Lata Gangadharan, Philip Grossman for their comments. Zhengyang Bao is grateful to the faculty and staff at the University of Chicago for their hospitality during his visit and the time when this paper was written.
† Corresponding author.
E-mail addresses: zhengyang.bao@monash.edu (Zhengyang Bao), Andreas.Leibbrandt@monash.edu (Andreas Leibbrandt).

1. 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 compose the 2018 Fortune 500 list, less than 5% are women. Such low representation of women at the CEO level 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). A few mechanisms that are studied in the literature encourage women to join tournaments ranging from team competitions to individual competitions (Dargnies 2012; Flory, Leibbrandt, & List 2015; Healy & Pate 2011) to gender quotas (Leibbrandt, Wang, & Foo 2018; Niederle, Segal, & Vesterlund 2013) and making tournaments the default choice (Erkal, Gangadharan, & Xiao 2019). However, luring women into tournaments may not be sufficient to narrow gender gaps. Once they have participated in a tournament, women need a competitive environment where they can thrive.

Many organizations employ tournaments, but they vary the risk exposure for low-ranked employees. At one extreme, some use up or out 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 difficult to identify whether such safeguards are useful in tackling 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 that mitigates the consequences of being low ranked. More precisely, workers are informed of 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 their relative performance falls into the lowest ranking category. Our conjecture is that the safeguard is particularly popular among women, improves women's outcomes relative to men's outcomes, and alleviates psychological competitive pressure to perform.

[1] For example, in many countries, government jobs are more secure than jobs that require similar skills in the private sector because the government sector usually has a lower dismissal rate.

Our findings show that women are indeed more likely than men to select a complementary safeguard and that risk-aversion is an important predictor for the choice of safeguard. However, we also find that this safeguard increases the gender wage gap compared to more standard tournaments. Further, we observe that the safeguard reduces the performance of both women and men regardless of whether it was voluntarily selected or automatically implemented. Our survey findings suggest that a safeguard does not alleviate pressure more for women than for men, and the safeguard encourages genders to slack off.

Our study is closely related to the literature on gender differences in the selection of incentive schemes (Dohmen & Falk 2011; Eckel & Grossman 2008; Flory, Leibbrandt, & List 2015; Gneezy, Leonard, & List 2009). This literature provides evidence that women are more likely than men to be disadvantaged by competitive environments such as workplace tournaments (Gneezy, Niederle, & Rustichini 2003; Gneezy & Rustichini 2004; Niederle & Vesterlund 2007). 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 (Dargnies 2012; Healy & Pate 2011; Niederle, Segal, & Vesterlund 2013). While there is some evidence that gender quota and policies can address gender gaps (Erkal, Gangadharan, & Xiao 2019), there is also evidence that they might backfire (Gangadharan, Jain, Maitra & Vecci 2016; Leibbrandt, Wang, & Foo 2018; 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 subject in this literature is how individuals react to exogenously imposed variations in the prize structure (see e.g., Harbring & Irlenbusch 2003; Moldovanu & Sela 2001; Orrison, Schotter, & Weigelt 2004; Sheremeta 2011) and how this relates to behavioral aspects (e.g., Delfgaauw, Dur, Sol, & Verbeke 2013; Hong, Hossain, & List 2015; Sheremeta 2015). We complement this literature

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

by investigating reactions to safeguards that allow for endogenous selection of the lowest prize and how this relates to gender differences in tournament outcomes. Finally, our work is related to the literature studying the link between incentive schemes and labor productivity (see e.g., Anderhub, Gächter & Königstein 2002; Corgnet & Hernán-González 2019; Corgneta, Fehr & Goette 2007; Gómez-Miñambresbc & Hernán-Gonzálezd 2018). In a typical setup in this literature, workers cannot change or choose their incentive schemes. We fill this gap by giving workers the freedom to choose between two incentive schemes.

2. Experimental Design

2.1 The real-effort task

In this experiment, we used 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; Erkal, Gangadharan, & Nikiforakis 2011; Gill & Prowse 2012; Leibbrandt, Wang, & Foo 2018;). In this task, participants were asked to solve as many puzzles as possible in a time span of 40 minutes. Fig. 1 provides one example of such a puzzle. Participants first deciphered the value of five letters and then summed up the corresponding values. In this example, letters 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 pressing the *Next* button, the computer displayed a different set of numbers until the participants ran 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 took part in a rank-order tournament and their performance (number of correctly solved puzzles) relative to other participants determined their monetary compensation. There were three compensation levels: (*i*) the top 10% of participants in a given session receive \$60, (*ii*) the top 10 to 50% of participants receive \$30, and (*iii*) the remaining participants (bottom 50%) receive \$15 or \$20, depending on treatment and choice. Thus, the

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

tournaments used a multiple-prize payment rule common to many workplaces (Cason, Masters, & Sheremeta 2010; Kalra & Shi 2001; Vandegrift, Yavas, & Brown 2007).[4]

2.3 The treatments

This study had three main treatments. The *choice* treatment allowed participants to select their own tournament incentives. To properly identify the impact of this choice, there were two other treatments: *compulsory* and *baseline*. All participants took part in one treatment only. The experiment instructions are in the Appendix.

2.3.1 Choice treatment

In this treatment, workers chose whether to have a complementary safeguard before the task started. The safeguard guaranteed a higher minimum payment if performance was in the lower half but did not affect payment if performance was in the upper half. More precisely, as can be seen in Table 1, if a participant chose the safeguard, they received $20 if their performance was in the bottom 50% instead of only $15 if they did not choose the safeguard. If their performance was in the top 50%, their payment was not affected.

INSERT TABLE 1 ABOUT HERE

2.3.2 Compulsory treatment

In this treatment, the safeguard was already embedded in the incentive structure for all participants. That is, the minimum payment was $20, just as it was in the choice treatment if participants chose a safeguard. The safeguard guaranteed a minimum payment of $20 for all participants. We tested two versions of this treatment. In one frame, we made the safeguard explicit (i.e., we told participants that the minimum compensation was only $15 in some of the other sessions). In the other frame, the safeguard was implicit and unknown to the participants (i.e., we gave workers no reference concerning the other group's prize structures).

2.3.3 Baseline

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

2.4 Conjectures

[4] When the cut off between prize levels was not an integer, we rounded up the number of workers who received the better prize in favor of workers. For example, if there were 28 workers in a group, then the top three performers received $60.

The standard prediction for all three treatments was that all workers would be incentivized to provide effort and thus increase their likelihood of higher compensation. Thus, the first conjecture was that there are no treatment differences in effort across treatments. The second conjecture applied to the choice treatment, for which we conjectured that all participants chose the safeguard as it weakly dominated not choosing the safeguard.

These standard predictions did not take into account mental effort costs associated with fatigue and the temptation to rest. While tournaments provide a significant incentive to work hard (possible tripling to quadrupling compensation), it is possible that the cost of the mental effort for some participants is sufficiently high that it prevents them from providing (maximal) effort. Accordingly, the alternative prediction is that effort is lower for the compulsory treatment than it is for the baseline because the incentive is less pronounced. Further, it is possible that some participants rejected the safeguard if they believed that it might undermine their effort and opportunity to increase their payments. On the other hand, it is also possible that the safeguard relieves 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 in this regard (e.g., Allen, Hitt, & Greer 1982; Compte & Postlewaite 2004; Hall & Lawler 1971; Harbring & Irlenbusch 2003; Van Dijk, Sonnemans, & Van Winden 2001).

So far, these predictions do not take gender into account. However, there is evidence that gender plays a crucial role in tournaments. In particular, there is evidence that women are more risk averse than men (Charness & Gneezy 2012; Croson & Gneezy 2009), less likely to enter tournaments (Flory, Leibbrandt, & List 2015; Gneezy, Leonard, & List 2009; Niederle & Vesterlund 2007), and underperform in competitions (Gneezy, Niederle, & Rustichini 2003; Gneezy & Rustichini 2004). For the choice treatment, we allowed for endogenous selection in two different payment rules, which permitted 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 the safeguard increases the temptation to rest and has a gender-dependent impact on pressure and stress.

2.5 Experimental procedures

The experiment was programmed in z-Tree (Fischbacher 2007), and workers were recruited with the software SONA. In total, 431 workers (160 in choice, 130 in compulsory, and 141 in baseline) participated in the three treatments, and they earned, on average, $32 for

an approximate 70-minute experimental session. There was no mentioning of gender throughout the entire experiment to prevent potential experimenter demand effects.

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 that included identical questions to the real-effort task and quiz questions to ensure that workers understood the instructions and payment mechanism. After the participants answered these questions correctly, they started the real-effort task. When all workers had 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 incentivized them to reveal their beliefs about their own performance, the group average performance.[5]

In addition, we elicited in this questionnaire risk preferences. We used the risk task proposed by Eckel & Grossman (2002). Each worker must choose one gamble out of six gambles with different risk exposure. The realization of the gamble prize is added to the worker's final payment. The choice is recorded as an ordinal categorical variable can be any integers between 1 to 6, where a larger number indicates less risk aversion.

3. Findings

3.1 Experimental findings

We observe high effort levels in our experiment. Fig. 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 were 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 made a substantial effort to do well in the experiment.

INSERT FIGURE 2 ABOUT HERE

We found significant treatment differences in effort levels. In particular, we observed that workers attempted less quizzes in the compulsory treatment[6] than in the baseline (78.1 vs

[5] We used the binarized scoring rule proposed by Hossain & Okui (2013) to incentivize participants to report their best guess about their own performance and the group average performance. Each answer within 5 quizzes around the true value is awarded with $2.

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

66.7, p<0.001) and that the performance was also significantly lower (70.1 vs 60.7, p<0.001).[7] This finding is consistent 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 observed that participants attempted less quizzes in the choice treatment than in the baseline (72.5 vs 78.1, p=0.037) and that the performance was also lower (65 vs. 70.1, p=0.062). Additionally, there were significant differences between the compulsory and choice treatments for 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 observed the following. The compulsory safeguard significantly reduced 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 chose to opt against the safeguard performed similarly to the baseline (p=0.61 for effort and p=0.37 for performance); those who chose to use the safeguard performed similarly to those in the compulsory treatment, and their performance was 8.1% to 8.3% worse than the baseline in terms of effort and performance (p<0.05 for both cases). Finally, the female dummy was significant at the 5% level and shows that men made more effort and had a higher performance (p<0.05 for both effort and performance).

INSERT TABLE 2 ABOUT HERE

Result 1: *The safeguard reduced effort regardless of whether it was implemented as a choice or as a default. Individuals who decided against using a safeguard provided more effort than those who decided to use a safeguard. Only workers that decided against using a safeguard performed similarly well as workers who did not have a safeguard.*

The safeguard was popular despite the negative impact on effort and performance (87% choose the safeguard). Importantly, and as conjectured, we found that women were more likely to choose the safeguard than men (93.1% vs. 81.8%, p=0.0368).

INSERT TABLE 3 ABOUT HERE

Table 3 compares the characteristics of the participants who chose the safeguard with the characteristics of those who did not choose the safeguard. We can see that the fraction of women is much lower in the sample of participants who did not choose the safeguard (23.8%)

[7] Without further specification, all p-values reported in this paper are from two-tail Mann-Whitney tests.

than in the sample of participants who chose the safeguard (48.2%). In addition, risk preference is also a significant predictor of safeguard choice. Participants make more risk averse choices in the risk task are more likely to be in the sample of those who chose the safeguard ($p=0.02$). As expected, we also observe gender differences in the willingness to take risks with females being more risk-averse (3.52 vs. 4.12, $p<0.01$). Characteristics other than gender and risk preference play less important roles in the safeguard choice. Task ability (math skills) and confidence (the ratio of guessed own performance and guessed group performance) are not significant drivers of safeguard choice.

In Table 4, we regress the choice of safeguard on these characteristics (gender, risk preferences, math skills, confidence) sequentially and simultaneously. We observe that both gender and risk preference remain significant predictors in all models even if both are used simultaneously ($p=0.03$-0.06 for gender and $p=0.03$-0.07 for risk preference). In particular, in Model 1 and 2, we regress the choice of safeguard on gender and risk preference alone, respectively. We find females are 11.2% ($p=0.03$) more likely to use the safeguard compared to males, and those choose the gamble with the risk level one number higher are 3.5% ($p=0.03$) less likely to use the safeguard. These results are robust if we interact gender with risk in Model 3 and add additional control variables, task ability and confidence, in Model 4.

INSERT TABLE 4 ABOUT HERE

Result 2: *Gender is an important predictor of safeguard choice. Women are more likely to choose a complementary safeguard in tournaments.*

The safeguard benefits women only at first sight. Fig. 3 shows how likely the safeguard is to materialize and increase the compensation in the choice treatment for each gender. 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 treatment, we see that gender plays no role as an almost identical percentage of either gender receiving the safeguard payment of $20.

INSERT FIGURE 3 ABOUT HERE

A closer look at the data reveals that women fare worse than men when there is a choice for safeguard. Fig. 4 compares the compensation for male and female workers across

treatments. Compared to the baseline, we observed that compensation was 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 was gender specific. Men's average wages increased by $3.3 (12.2%) while women's wages only increased by $1.1 (4.4%). In addition, compared with the compulsory treatment, giving workers the choice of safeguard slightly increased men's payoff ($29.6 vs $30.3, p=0.60) but somewhat decreased women's payoff ($27.6 vs $26.0, p=0.138). Perhaps most importantly, we observed that the choice treatment created a significant gender wage gap of $4.2 (wage for women=$26, men=$30, p=0.02), which was insignificant in the other two treatments (p>0.38).

INSERT FIGURE 4 ABOUT HERE

Fig. 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 were much more likely to score in the bottom 50% in the choice treatment than men, and that this was not paralleled in the compulsory and baseline treatment. In fact, such a gender difference was 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

Fig. 6 provides a more fine-grained illustration of effort levels depending on gender and treatment. In this figure, we compared the effort level of each quantile in the no baseline against the same quantile in the 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 shifted the overall distribution of correct answers lower than the baseline, except for the quantiles near the two prize cut-offs. Women, in contrast, appeared to respond differently to the safeguard choice. All quantiles in the choice treatment were higher in the baseline, except for the few that were between the cut-offs. Thus, the choice of safeguard appeared to cause most men and women to exert less effort compared to the baseline, but this reduction was more pronounced for women whose performance was close and moved them up the ranks.

INSERT FIGURE 6 ABOUT HERE

Result 3: *The availability of a complementary safeguard disadvantaged women more than men and created 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 with gender, we conducted a survey with the participants after the tournament but before revealing information on their performance. In this survey, we asked participants to report their stress levels and temptation to rest during the real-effort task and how a removal of the safeguard would change these factors.[8]

Fig. 7 illustrates several corresponding survey insights. First, while we found that only a minority of workers experienced high levels of stress and temptation, many reported experiencing some stress and the temptation to rest suggesting the presence of mental costs. Second, we observed that a significant proportion of participants expected to experience higher stress level (45%) and less temptation to rest (35%) when there was no safeguard. Third, and perhaps most importantly, the psychological impact of the safeguard appeared to be gender independent. The reported stress and temptation levels were 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 mental costs cannot explain the observed gender differences in safeguard choice, 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 men (e.g., Frederiksen 2008; García 2017; Hirsch & Schnabel 2012). There is also evidence that giving workers choice over their employment conditions can improve labor outcomes (e.g., Beckmann, Cornelissen, & Kräkel 2017; Bloom, Garicano, Sadun, & Van Renee 2014; Leslie, Manchester, Park, & Mehng 2012). We investigate a workplace tournament that reduces uncertainty by providing workers with a safeguard choice 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

[8] We only asked participants in the compulsory treatment because all of them had experienced the safeguard. Experiencing the safeguard may facilitate the participants to understand the question and give more reliable answers.

choose safeguards, although they weaken the incentive to exert effort. Our findings provide novel evidence on the limitations of tournaments in creating 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 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., Heckman 2006; Mario, Nicole, Ute, Lisa, & Benjamin 2020; Rosen 1986). For example, poverty alleviation programs are more likely to affect the wages of low-skilled workers, and it is of key importance to understand whether and under what circumstances they cause less 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 a choice in their compensation scheme and whether it is a general property that giving this choice is less beneficial for women than men. Second, our data suggest 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.

References

Anderhub, V., Gächter, S., & Königstein, M. (2002). Efficient contracting and fair play in a simple principal-agent experiment. *Experimental Economics*, 5(1), pp.5-27.

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., & Van Reneen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 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), 504–517.

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.

Corgnet, B., & Hernan-Gonzalez, R. (2019). Revisiting the trade-off between risk and incentives: The shocking effect of random shocks? *Management Science*, 65(3), pp.1096-1114.

Corgnet, B., Gómez-Miñambres, J., & Hernán-Gonzalez, R. (2018). Goal setting in the principal–agent model: Weak incentives for strong performance. *Games and Economic Behavior*, 109, pp.311-326.

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), 3330–3348.

Erkal, N., Gangadharan, L., & Xiao, E. (2019). Leadership selection: Can changing the default break the glass ceiling? Available at SSRN 3022386.

Fehr, E., & Goette, L. (2007). Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review*, 97(1), pp.298-317.

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.

Gangadharan, L., Jain, T., Maitra, P. and Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders. *European Economic Review*, 90, 302-325.

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. & 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.

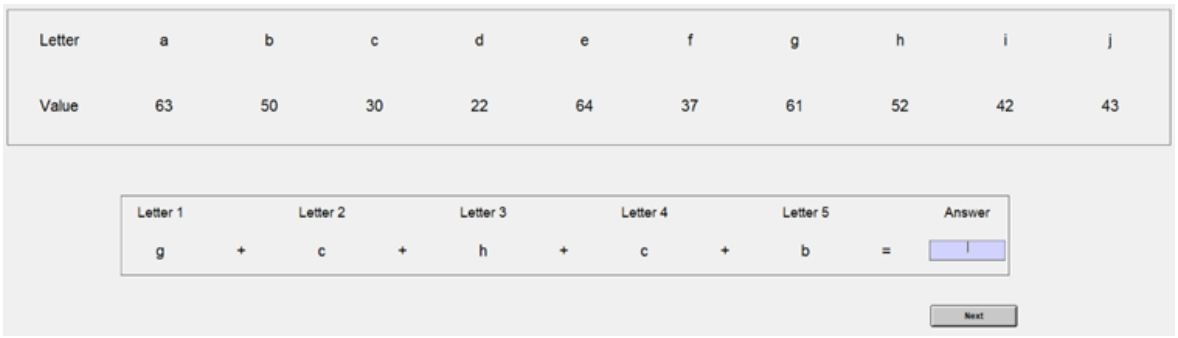
Sheremeta, R. M. (2015). Behavioral dimensions of contests. In R. D. Congleton & A. L. Hillman (Ed.), *Companion to the political economy of rent seeking* (pp. 150–164). Cheltenham: Edward Elgar Publishing.

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.

Figures

Fig. 1 Illustration of task

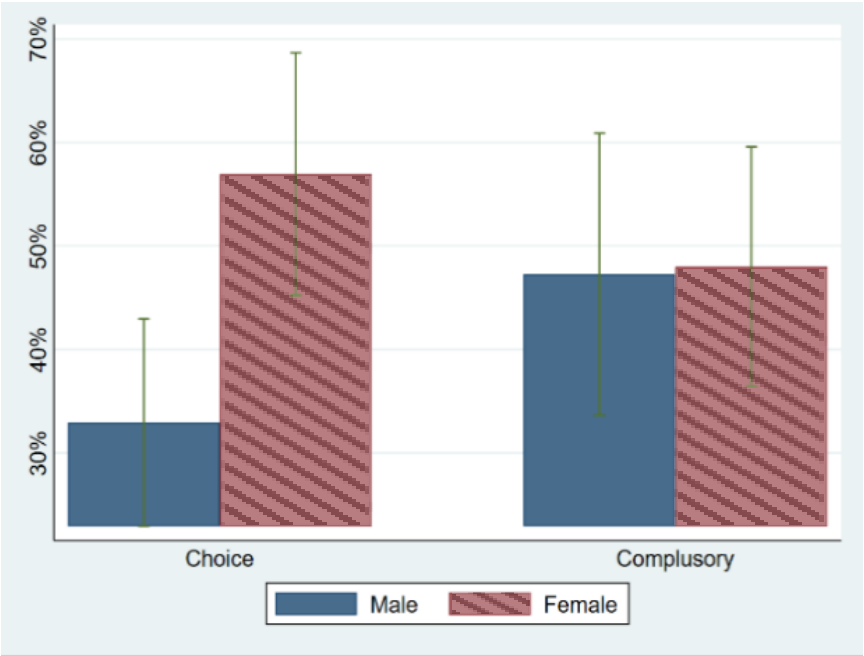


Notes: the figure provides a screenshot of an example puzzle that workers have to solve during the experiment. They first need to decipher the code in the lower box using the upper table and then sum up all the values.

Fig. 2 Effort and performance levels depending on treatment

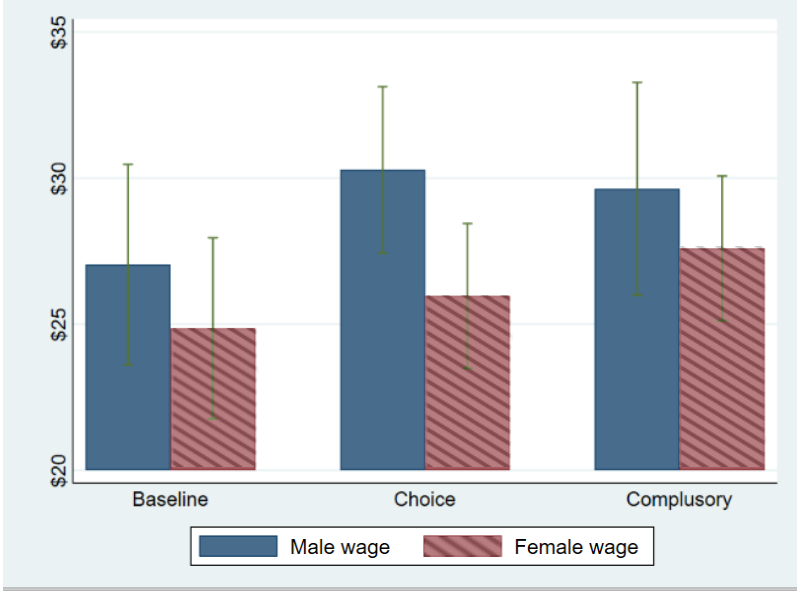
Notes: This figure shows the average attempts and performance depending on treatment. Participants had 40 minutes to attempt as many quizzes as possible. Error bars indicate 95% confidence intervals.

Fig. 3 Materialization of safeguard depending on treatment and gender



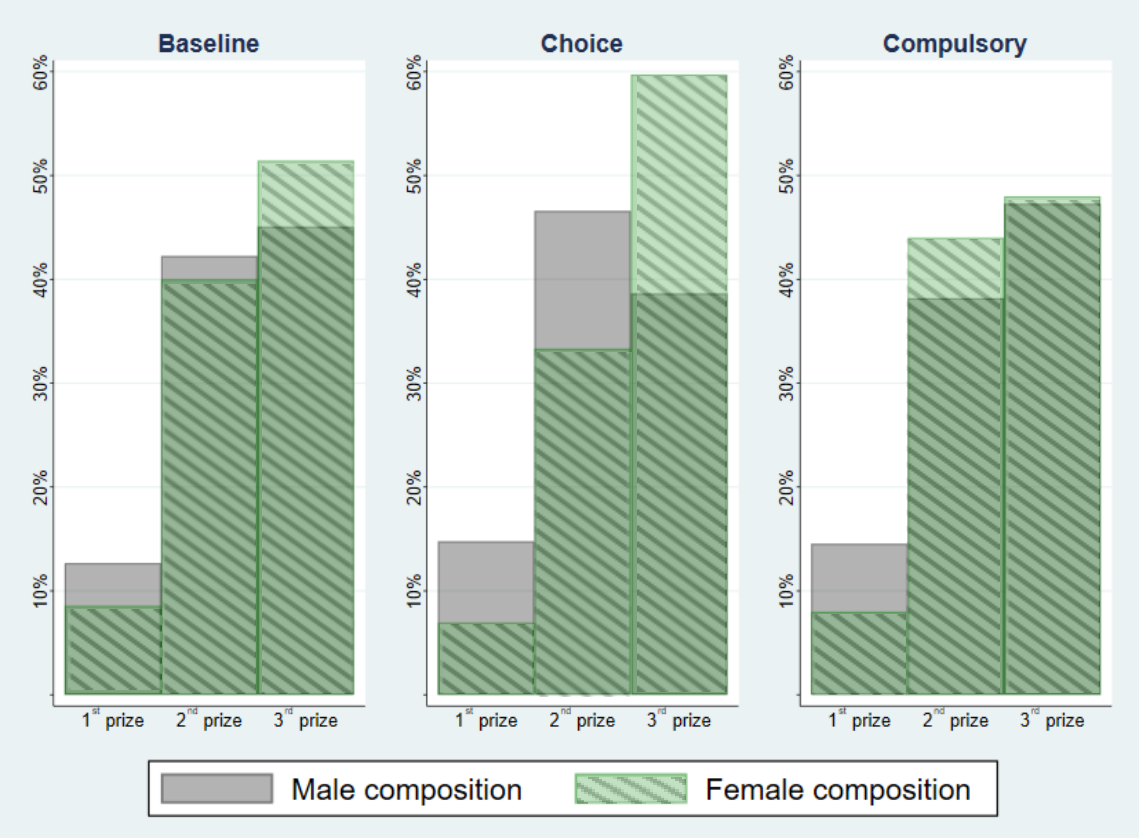
Notes: This figure illustrates the percentage of workers who received a higher minimum wage due to the materialization of the safeguard.

Fig. 4 Average wage by gender in each treatment



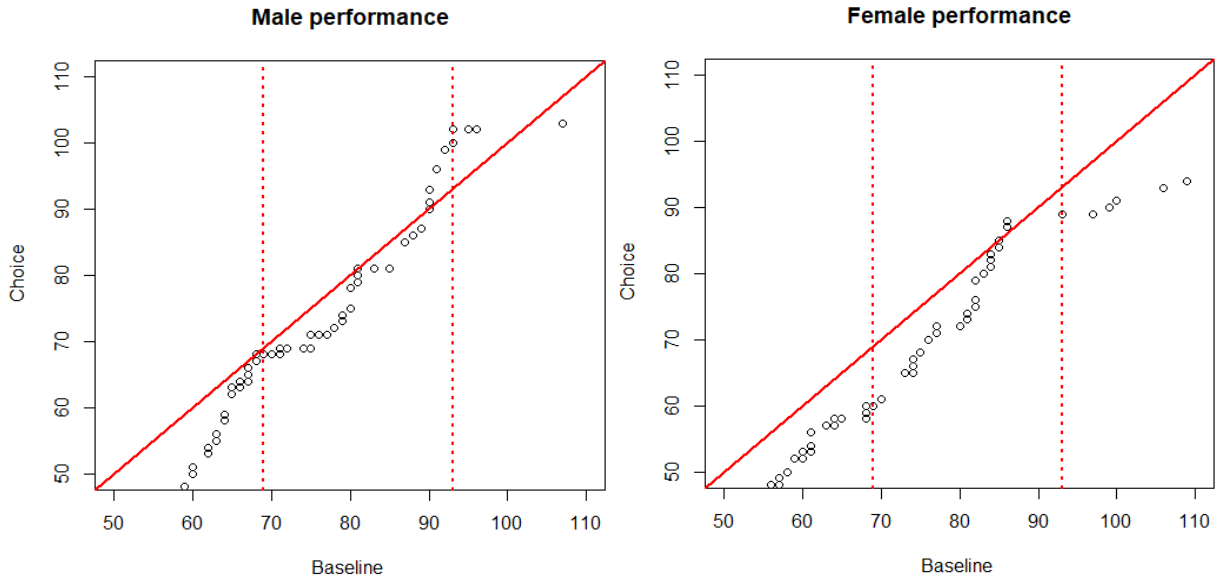
Notes: This figure shows the average wage for each gender across treatments. The error bars indicate 95% confidence intervals.

Fig. 5 The gender composition of each prize level for each treatment



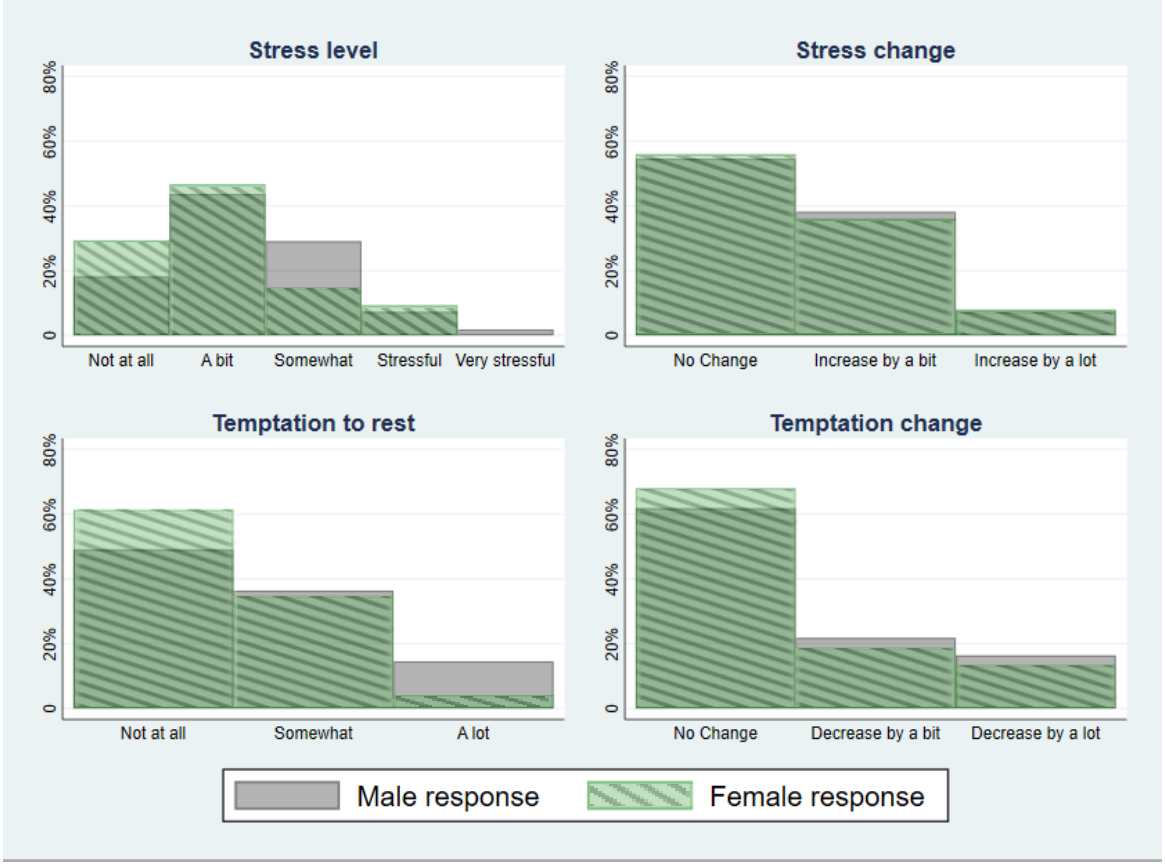
Notes: This figure plots the gender composition of each prize level in all treatments. The first price is awarded to the top 10% performers, the second is awarded to the top 10% to 50% performers, and the third is awarded to the bottom 50%.

Fig. 6 Performance differences between choice and baseline treatment depending on gender



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 is 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.

Fig. 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 participants expect a removal of the safeguard to 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 to not using the safeguard. 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 opting against it. We collect data from the compulsory treatment to ensure that all participants have experienced the safeguard.

Tables

Table 1: Treatment overview[a]

	Choice (n=160)		Baseline (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

[a] This table shows the tournament payment structure for each of the three treatments.

Table 2: Effort and performance depending on treatment[a]

	Performance	Effort
Compulsory	-8.985***	-10.962***
	(2.628)	(2.753)
Choice × not to use safeguard	-4.374 (4.837)	-2.627 (5.136)
Choice × use safeguard	-5.480** (2.626)	-6.283** (2.738)
Female	-4.961**	-5.259**
	(2.077)	(2.390)
Constant	72.555***	80.689***
	(2.264)	(2.309)

[a] This table shows the OLS estimation results of the regression comparing the outcomes across treatments and the safeguard choice. Performance is defined by the number of correctly solved quizzes. Effort defines the number of attempted quizzes. The constant corresponds to the outcomes of male 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[a]

Variable	Use safeguard (n=139)	Do not use safeguard (n=21)	Difference	p-value
Female (dummy)	48.2%	23.8%	24.4%	0.04
Risk	4.01	4.86	-0.85	0.02
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

[a] 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 asked the participants questions on their gender, risk preference (Eckel & Grossman 2002), subjective belief about their mathematical abilities, and their estimate of their own performance relative to the group average. Female is a dummy variable that takes the value 1 if and only if the participant is a female; Risk is an ordinal categorical variable that can be any integers between 1 to 6, where a larger number indicates less risk aversion; Good at math is a dummy variable that takes the 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 their own performance and their 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 the variables (Male, Risk > medium, Good at math, Confidence) of the group that choose to use the safeguard and the group that chooses 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 that uses the safeguard and the group that does not use the safeguard.

Table 4: Individual determinants of the safeguard choice[a]

	Choice of safeguard			
	Model 1	Model 2	Model 3	Model 4
Female (dummy)	0.112**		0.095**	0.097*
	(0.051)		(0.048)	(0.052)
Risk (categorical)		-0.035**	-0.030**	-0.027*
		(0.015)	(0.015)	(0.015)
Good at math (dummy)				0.019
				(0.058)
Confidence (guessed own performance				-0.104
/ guessed group average)				(0.089)
Constant	0.818***	1.012***	0.949***	1.054***
	(0.030)	(0.059)	(0.068)	(0.131)

[a] This table shows the OLS estimation of the linear probability model studying how individual characteristics affect the safeguard choice. Female is a dummy variable that takes the value 1 if and only if the participant is a female; Risk is an ordinal categorical variable that can be any integers between 1 to 6, where a larger number indicates less risk aversion. ; Good at math is a dummy variable that takes the 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 their own performance and their 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.

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 are available upon request.

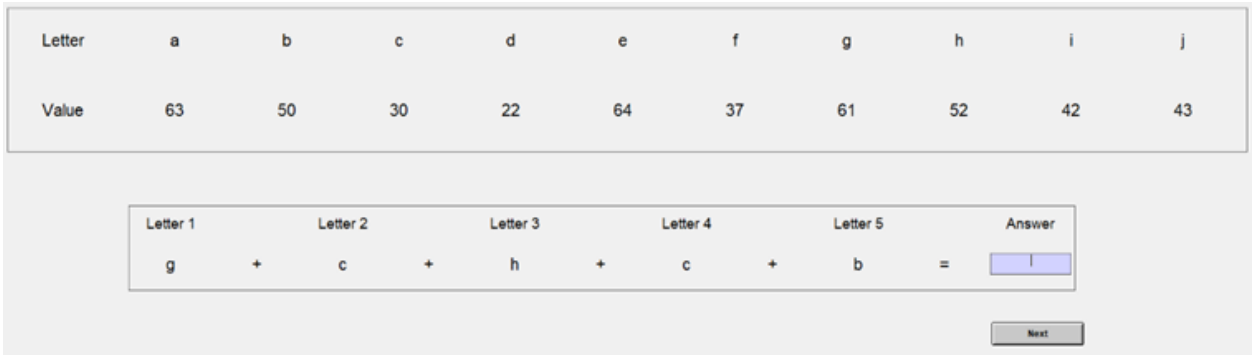
1. General Instructions

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, 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 (2,400 seconds) to solve puzzles similar to the one in the graph below.



To solve this puzzle, you first need to decipher the code in the box below. For example, letter 1 (g) has a value of 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 five 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 a safeguard.

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

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

- If your total number of correct answers is among the top 10% to 50%, then you will receive **$30**.
- If your total number of correct answers 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 total number of correct answers is among the top 50%:

- If your total number of correct answers is among the top 10%, then you will receive **$59**.
- If your total number of correct answers is among the top 10% to 50%, then you will receive **$29**.
- If your total number of correct answers 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 treatment, which vary depending on 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 ($15) because of the presence of a safeguard. We observe no differences between frames in terms of quizzes attempted and quizzes answered correctly ($p=0.8$ for attempt $p=0.9$ for correct). This also holds true if we analyze males and females separately. For males, the p-value testing for no difference caused by the framing is 0.37 for attempted questions and 0.30 for questions answered correctly; for females, the numbers are 0.62 and 0.38, respectively.