Gender Differences in Private and Public Goal Setting[†]

Jordi Brandts*, Sabrine El Baroudi**, Stefanie J. Huber***, Christina Rott**[‡]

January 2021

* Instituto de Análisis Económico (CSIC) and Barcelona GSE.

** Department of Management and Organization, School of Business and Economics, Vrije Universiteit Amsterdam.

*** Department of Economics, School of Economics, University of Amsterdam.

Abstract

We conduct a field and an online classroom experiment to study gender differences in self-set performance goals and their effects on performance in a real-effort task. We distinguish between public and private goals, performance being public and identifiable in both cases. Participants set significantly more ambitious goals when these are public. Women choose lower goals than men in both treatments, but in particular when goals are private information. Men perform better than women under private and public goals as well as in the absence of goal setting, consistent with the identifiability of performance causing gender differences, as found in other studies. Compared to private goal setting, public goal setting does not affect men's performance at all but it leads to women's performance being significantly lower. Comparing self-set goals with actual performance we find that under private goal setting women's performance is on average 67% of goals, whereas for men it is 57%. Under public goal setting the corresponding percentages are 43% and 39%, respectively.

Keywords: gender gap, goal setting, public observability, experiment.

JEL Classifications: C91, J01, J16, J82

[†] Acknowledgement: Brandts thanks the Spanish Ministry of Economics and Competitiveness through Grant: ECO2017-88130 and through the Severo Ochoa Program for Centers of Excellence in R&D (CEX2019-000915-S) and the *Generalitat de Catalunya* (Grant: 2017 SGR 1136) for financial support.

[‡] Corresponding author. Email: c.e.rott@vu.nl. Postal address: Department of Management and Organization, School of Business and Economics, Vrije Universiteit Amsterdam, Room 4A-64; Boelelaan 1105; 1081 HV Amsterdam; The Netherlands.

1. Introduction

Gender equality is one of the 17 Sustainable Development Goals (SDG 5) elaborated by the United Nations Development Programme in the 2030 Agenda for Sustainable Development.[1] A more specific goal is to increase women's participation and leadership in all forms of decision-making in the public, judiciary, and private sector. As pointed out by the World Economic Forum (2018): "[...] while there are still relevant gender-biased labour market outcomes, the presence of women in management roles is today one of the main barriers to overcome, both in the public and private sector, in order to achieve full economic gender parity [...]."

Why should we care about the underrepresentation of women in managerial positions? The empirical literature provides initial evidence that macroeconomic benefits of gender equality exist in the workplace because of complementarities between the two genders. These benefits are especially important in organizations requiring high-skill workers. García-Meca et al. (2015) show that board-level gender diversity improves the performance of firms. The positive effects on firm performance are especially large for those whose strategy is based on innovation (Dezsö and Ross, 2012) and for firms in high-tech manufacturing and knowledge-intensive services (Christiansen et al., 2016). Gender diversity on the boards of banking-supervision agencies has also been associated with greater financial stability (Sahay and Cihak, 2018).

There are many reasons for the imbalance between women and men in leading positions. An important distinction is that between demand-side and supply-side factors (Gino et al., 2015). Demand-side factors are those that women face because of the different ways in which women are judged and treated on the labor market and in society at large (prejudice, discrimination, etc.). Supply-side factors are related to differential beliefs and behavior of women and men that are relevant for access to high-level positions.

Research in experimental economics has particularly contributed to a better understanding of supply-side factors. In particular, gender differences in the reaction to various aspects of competition effects have been studied in detail in a large experimental literature. Niederle (2016) presents a recent survey of relevant studies, distinguishing between those that deal with gender differences in performance under competitive incentives and those that analyze gender differences in choices between competitive and non-competitive incentive schemes. More recently, research has also addressed the role of public observability for gender differences in public speaking.[2]

In this paper, we study gender differences in a novel, potentially important supply-side dimension of behavior: goal setting. The specific question that motivates our work is whether gender differences in public goal setting could be one reason for the female underrepresentation in high-level positions. Leaders in the public and private sector must typically announce their goals for the firm/society publicly. The public typically judges their leaders by which goals are set and

[1] https://www.undp.org/content/undp/en/home/2030-agenda-for-sustainable-development/people/gender-equality.html; accessed on Dec 15th, 2020.

[2] Survey-based and observational data show that women fear more to speak up in public (Stein et al., 1996; Turk et al., 1998; Behnke and Sawyer, 2001; Marinho et al., 2017), feel more stressed about it (Buser and Yuan, 2020), and also do speak up less frequently in public settings (Hinsley et al., 2017; Carter et al., 2018; Parthasarathy et al., 2019).

to what extent they are reached. If females have difficulties with this part of the job, that might explain why they apply less for leading positions (among other factors).

We study how women and men set goals for themselves and perform in a particular real-effort task under observability and identifiability of performance. We study behavior under two distinct goal-setting conditions: Public goals – that is, the self-set goal is observable and identifiable by the public - and private goals – that is, the self-set goal is only observable by oneself. We run a field experiment and an online classroom experiment, in which participants perform a real-effort task. Depending on the randomly assigned treatment, participants perform the task without goal setting (control condition), after setting a goal privately, or after setting a goal publicly. Our primary outcome variables are participants' self-set goals and their effects on performance.

Our results show that participants set significantly more ambitious goals when these are public. Women choose lower goals than men in both treatments, but more so when goals are private information. Men perform better than women under private and public goals as well as in the absence of goal setting, consistent with the identifiability of performance causing gender differences, as found in other studies. Compared to private goal setting, public goal-setting does not affect men's performance at all but it leads to women's performance going significantly down. In terms of the ratio between performance and goals participants are more realistic under private than under public goal setting, with women being more realistic than men in both cases.

The remainder of the paper is structured as follows. In section 2, we present a review of the relevant literature in psychology and economics. Section 3 discusses the experimental procedures, the design, the hypotheses, and the research questions. Section 4 contains the results and section 5 the conclusions.

2. Review of the Literature on Goal Setting

There is a rich literature in psychology on goal setting and performance which finds that setting goals, whether self-set, assigned by others or set jointly through participation, is better for performance than not setting any goals (Latham and Locke, 2007). These goal-setting effects have been shown to be salient in the realm of sports, academic performance, managerial and professional jobs, and teamwork, to mention a few examples (Locke, 1996; Locke and Latham, 2002; Locke and Latham, 2006, for literature reviews). Support for goal-setting effects on performance has also been found worldwide in experimental and non-experimental research with samples consisting of participants from Asia, Australia, Europe, and North America (Locke and Latham, 1990). The psychology literature further shows that relevant mediators in goal-setting and performance are individual choice, effort, persistence, and goal-achievement strategy. Potential moderators are the ability to achieve a goal, goal commitment, feedback concerning goal pursuit, the complexity of a goal, and other situational factors (e.g., presence of needed resources to achieve a goal) (see Locke and Latham, 2006; Latham and Locke, 2007; Locke and Latham, 2019 for an overview of mediators and moderators).

2.1 Self-set goals and gender differences

Regarding task-based goal setting as a specific type of goal setting, the psychology literature offers potential explanations for it and posits that there are gender differences in goal-setting behavior and offers potential explanations. However, this research is relatively old and likely not fully generalizable to contemporary times, because societal events, these days, continuously put pressure on individuals to change or adapt their attitudes towards gender differences (Szekeres et al., 2020) This could also influence men and women's goal-setting behavior (Latham and Locke, 2007; Locke and Latham, 2019). According to the older stream of goal-setting research in the psychology literature, men chose higher task-based goals than women (e.g., Kurman, 2001; Levy and Baumgardner, 1991; de Pater et al., 2009). The psychology literature offers a likely explanation: men are generally perceived by both genders as more competent, leading to superior male performance (McCarty, 1986; Wood and Karten, 1986). Thus, men are more confident about their competences (e.g., Beyer, 1990; Lundeberg et al., 1994; Beyer and Bowden, 1997) and opt therefore for more challenging goals than their less confident female peers (McCarty, 1986; Wood and Karten, 1986).

More recently, economists started investigating gender differences in goal-setting behavior. For instance, Clark et al. (2017) examine whether self-set goals, task-based goals, or performance-based goals improve student performance. The authors find that the task-based goal setting increases task completion (i.e., practice exams) and course performance, but only for men. Women completed more practice exams in the control group without goal setting. In another study, Dalton et al. (2016) provide a simple model of self-chosen goals that predicts that (i) the self-chosen goal contract is more cost-effective than a piece-rate contract for an employer aiming for a specific output level, and that (ii) workers set goals that they systematically outperform. The authors test these predictions in the laboratory and find that the self-chosen goal contract increases men's performance compared to the piece-rate contract. However, this is not the case for women. Concerning the self-set goals, women set lower goals than men but outperform their self-set goals to a greater extent than men.[3]

The experimental economics literature offers potential explanations for such a gender difference in goal setting. First, there is clear evidence that women take less risk than men. In a series of ten experimental studies, Croson and Gneezy (2009) find evidence that women are indeed more risk-averse compared to men. Dohmen et al. (2011) measure and validate self-assessed risk aversion and show that women are much less likely to take risks in general. This finding applies to several domains; car driving, finance, sports and leisure, health, and career. A similar conclusion on gender differences in risk aversion is echoed in more recent studies (Buser et al., 2020; Falk et

[3]There is some agreement that self-set goals likely have a stronger positive effect on performance than goals assigned or set in cooperation due to a higher locus of control (Latham and Marshall, 1982; Hollenbeck et al., 1989).

al., 2018). In addition to being more risk-averse, women report a higher intensity of nervousness and fear than men---in anticipation of negative outcomes (Brody, 1993; Fujita et al., 1991). Therefore women might want to avoid negative outcomes more than men (i.e., lower performance than the self-set goal) by taking less risk in not meeting the self-set goal. Given that higher goals are more challenging to achieve, females might reduce the probability of this negative outcome by setting a lower goal than men.

Another individual characteristic that influences which goal a person sets is self-efficacy, that is, self-confidence that the goal for a specific task is attainable (Bandura, 1997; Latham and Locke, 2007; Locke and Latham, 2006). There is consensus in the literature that men are more confident than women. For example, Niederle and Vesterlund (2007) asked participants in a laboratory experiment to solve a real task, first under a non-competitive piece-rate and afterwards under a competitive tournament incentive scheme. After solving the task in the competitive tournament incentive, participants were asked to select which of these two compensation schemes they wanted to apply to their next performance. They found that 73 percent of the men and only 35 percent of the women selected the competitive tournament. The authors conclude that this gender gap difference is to a large extent driven by men being more confident about their performance than women. Möbius et al. (2018) implement an experimental test with a sample of 656 undergraduate students. The authors track the evolution of students' beliefs about their own relative performance on an IQ test and find that women are less confident about their performance than men. Buser et al. (2020) use data from lab experiments on preferences for redistribution conducted in the U.S. and several European countries to investigate gender differences and their causes. Across all sampled locations, they found that men are more confident about their ability compared to women.

In the psychology literature goal-setting theory (Bandura, 1997; Latham and Locke, 2007; Locke and Latham, 2006) offers an explanation for why women, being less confident about their competences and performance, set lower goals compared to men (Dalton et al., 2016). The goal-setting theory suggests that higher goals lead to higher performance than easy lower goals because the former motivates individuals to put more effort into achieving the challenging goal such as looking for new knowledge and developing new skills (Locke and Latham, 1990, 2002; Locke and Latham, 2006). If women are likely to set lower goals, it is also reasonable to assume that their actual performance will be lower than their male peers' performances.

2.2 Self-set goals and gender differences in private versus public environments

We are not aware of any experimental work that examines gender differences between publicly and privately self-set goals. However, as part of the literature on gender differences in reacting to competition, Schram et al. (2019) study the difference between providing public ranking (referred to as status ranking) and private ranking information about performance. They found no gender differences in performance or attempted summations when there is no status ranking. By contrast, inducing status ranking leads to gender differences. Men significantly increased the number of

attempted summations and performance, while women significantly decreased the number of attempted summations and performance.

In another related study, Ariely et al. (2009) examine the impact on performance when an audience watches the subject working on a cognitive task that involves performance-contingent payment. Across the two conditions (public and private), there was no evidence of any gender difference in the ability to solve anagrams, nor any evidence for the two genders to be differentially influenced by social pressure.

Moreover, research suggests that in a competitive environment, a "desire to win" can emerge within individuals, which motivates them to beat the other side, rather than focusing solely on maximizing their payoffs (Cooper and Fang, 2008; Delgado et al., 2008). In environments in which self-set goals and performance are revealed publicly, competition is triggered which could motivate individuals to opt for higher goals and performance just to beat others. Whether this is indeed the case is not yet studied in goal-setting research. Moreover, given that research on gender differences in performance and attitudes in such competitive environments shows mixed results, it is yet to be explored how men and women will set goals and achieve their performance in public versus private environments.

3. Experimental design and procedures

The experiment is composed of three parts: the goal setting, a real-effort task, and a questionnaire, that includes socio-economic background questions.[4] The participants were not allowed to communicate with anyone during the whole experiment.

It is a crucial design feature that the audience can identify participants. Therefore, we reveal the students' names in a particular way. Before the task takes place, the students indicate their first and last names. In all treatments, we display the following information on a shared public screen at the end of the experiment: Students' first and last names together with their performance in the task. Our experiment was pre-registered and approved by the Research Ethics Review Board (School of Business and Economics, Vrije Universiteit Amsterdam).[5]

Section 3.1. explains the 15-minute real effort task in detail. Section 3.2 describes the three treatments that allow us to investigate potential gender differences in private and public goal setting and whether women and men perform differently under private and public goal setting as well as in the absence of any goal setting. Section 3.3. presents the procedure and the subject pool.

[4] Appendix B provides the instructions as displayed on the screen.

[5] The study was pre-registered at AsPredicted.org (#48703, https://aspredicted.org/blind.php?x=4t5by6) and the experimental design of the online classroom experiment approved by the Research Ethics Review Board, School of Business and Economics, Vrije Universiteit Amsterdam (20200828.1.xxx where xxx stands for one of the author's employee ID). Ethics approval is not required, but we still opted to apply for it.

3.1 Task

The task is identical to the one used in Schram et al. (2019) and before that in Weber and Schram (2017).[6] Participants are presented with a sequence of pairs of 10x10 matrices filled with random two-digit numbers (Figure 1).[7]

For each matrix pair, the participants' task is to search for the highest number in each matrix and then calculate the sum of these numbers. Participants have to enter this sum at the center-bottom of the computer screen. After entering the sum, the participant immediately learns if she/he has entered a correct answer or not. Regardless of whether the sum was correct or not, the next new pair of matrices appear. This task stops after 15 minutes and participants can see the remaining time on the screen at the top left of screen. We measure a participant's performance by the number of accurate summations within the time limit of 15 minutes.

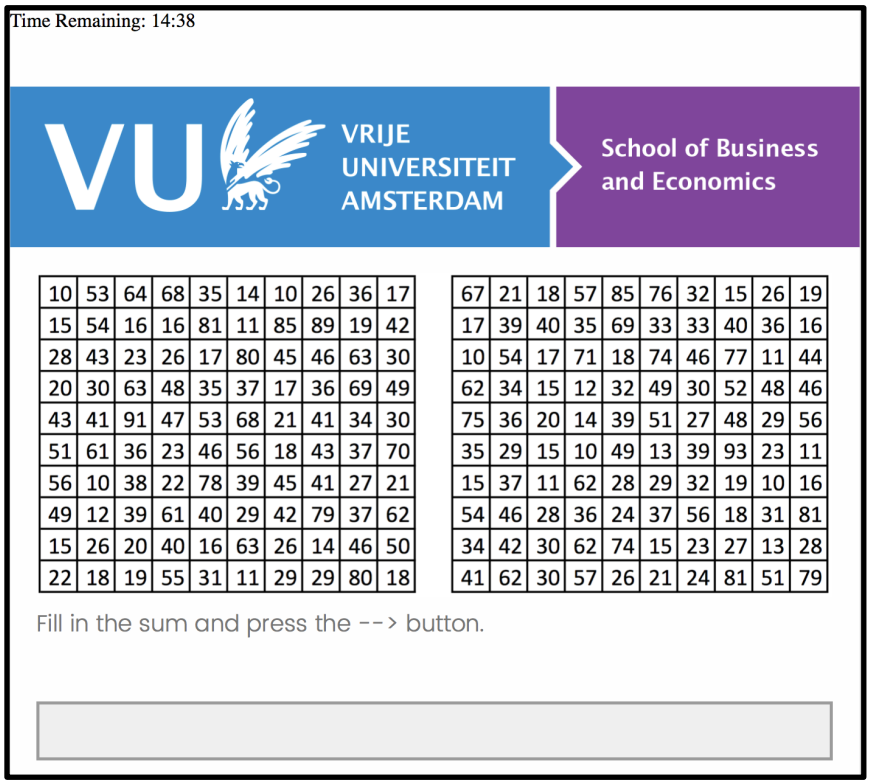


Figure 1. *Screenshot of the Task*

All participants perform this task individually without interacting with other participants. The instructions highlight the importance of doing well in the task by informing the participants

[6] The choice of task is an important issue, which we do not study here. See Flory et al. (2015) and Günther et al. (2010).

[7] A possible alternative would have been to use the summation task of Niederle and Vesterlund (2017). As discussed in Schram et al. (2019), this task involves a risk of a stereotype threat (Shurchkov, 2012), where females feel that men have an advantage in this task. Therefore, we use the summation task of Weber and Schram (2017) and Schram et al. (2019), as these previous studies have found no gender performance differences.

that doing well in such a task is positively correlated with professional life success. In addition, we give the participants information about the performance distribution of similar participants doing this task in previous studies.[8] The performance of each participant is public information to all participants. We inform the participants that their performance (i.e., the total number of correct answers within the 15 minutes), together with their name, will be displayed on a shared screen after the study is finished.

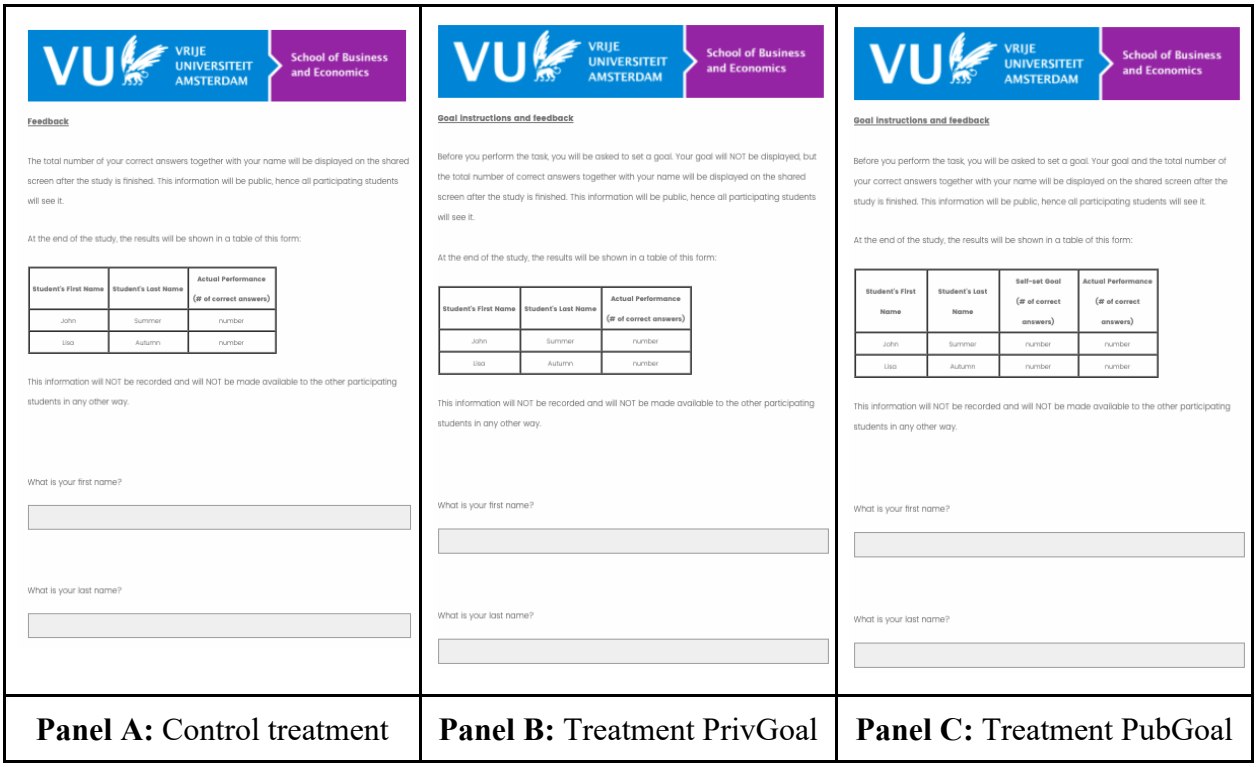
We decided not to incentivize participants. The literature in psychology on goal setting theory suggests that self-set goals induce intrinsic motivation---in contrast to externally-set goals. Intrinsically motivated behavior is commonly referred to as a behavior that is engaged for its own sake without any external inducement (Pinder, 1984; Cerasoli et al., 2014), whereas extrinsically motivated behaviors are guided towards achieving some instrumental outcomes such as money or financial rewards (Erez et al., 1990). Self-set goals allow for personal control in setting a goal that is attainable with one's ability (Erez et al., 1990). A similar argument has been recently echoed by Welsh et al. (2020) that self-set goals induce positive feelings such as enthusiasm, because they are perceived as beneficial and achievable. Whether financial rewards motivate individuals to perform is deemed to depend on individual values and personal dispositions. A failure to consider these individual differences could decrease one's motivation and even result in lower performances (Malik et al., 2015). Since intrinsic motivations are considered to mainly trigger self-set goals and drive performance, we decided to not incentivize our participants financially.

3.2 Treatments

We implement two goal-setting treatments, private and public, next to a control treatment without goal setting. We randomly assigned participants to one of the treatments (between-subject design). The only difference between the two goal-setting treatments is that the self-set goal is private or public information at the end of the study. Figure 2 highlights the implementation of the treatment variation. Panel A shows the implementation of the control treatment, Panel B the implementation of the private goal setting, and Panel C the implementation of the public goal setting.

Control treatment (NoGoal):
Participants do not set a goal, but they are informed that the total number of correct answers (i.e., performance) together with the participant's name will be displayed on the shared screen at the end of the experiment.

[8] The instructions state: "This is an important task that is often used to measure people's talents. Many scientific studies have found that people who do well in a task like this are more successful in professional life than people who do less well. In a previous session, students like you performed the same task. Most of them gave between 9 and 17 correct answers." We refer to the participants' performance in Schram et al. (2019). We mention the prior performance in the instructions to give the participants a broad idea of what their performance could be in a task with which they have no experience.



Panel A: Control treatment	**Panel B:** Treatment PrivGoal	**Panel C:** Treatment PubGoal

Figure 2. *Screenshots of Treatment Implementation. Panel A: Screenshot for control treatment NoGoal. Panel B: Screenshot for treatment PrivGoal. Panel C: Screenshot for treatment PubGoal.*

Private Goal Setting Treatment (PrivGoal):
Before performing the 15 minutes task, the participants are asked to set a goal (the number of correct answers). The precise wording of the goal question is: "What is your self-set goal - How many questions do you WANT to answer correctly in the 15 minutes available?" Next to this question, the instructions remind the participant that this self-set goal will NOT be displayed, but the total number of correct answers (i.e., performance), together with the participant's name, will be displayed on the shared screen at the end of the experiment. Hence, the goal setting is private but the performance is public.

Public Goal Setting Treatment (PubGoal):
Before performing the 15 minutes task, the participants are asked to set a goal (the number of correct answers). The precise wording of the goal question is: "What is your self-set goal - How many questions do you WANT to answer correctly in the 15 minutes available?" Next to this question, the instructions remind the participant that this self-set goal will be displayed, together with the total number of correct answers (i.e., performance) and with the participant's name, on the shared screen at the end of the experiment. Hence, both the goal setting and the performance are public.

Before performing the matrix-task, all participants – irrespective of the treatment assignment – are asked how many questions they *expect* to answer correctly in the 15 minutes available. Note that this is different from their goal, which refers to how many questions participants *want* to answer correctly.

3.3 Experimental Sessions, Procedure and Participant Pool

The experiments were conducted at the School of Business and Economics of the Vrije Universiteit Amsterdam in September 2019 and October 2020. Participants were first-year Bachelor students from the International Business Administration program. We used the software Qualtrics to program the experiment, and the duration of the experiment was on average less than 30 minutes. The experiments in 2019 and 2020 differed in several key aspects.[9] Next, we describe the details of the implementation of both experiments.

Field experiment in September 2019: The experiment was conducted on location at the School of Business and Economics of the Vrije Universiteit Amsterdam. The field experiment was integrated in the first lecture of the course as a quiz. Participation was absolutely voluntary and students were informed that it would not have any impact on their assessment in the course and that they could leave the quiz at any moment in time. The students were randomly assigned to different treatments on the online course platform and a different Qualtrics link was sent to each treatment group. In total, 302 students participated in our experiment, out of which 124 were female and 178 male.[10] Sixty-nine students were assigned to the control treatment (NoGoal), 97 to the Private Goal Setting Treatment (PrivGoal), and 136 to the Public Goal Setting Treatment (PubGoal).[11]

Online classroom experiment in October 2020: The experiment was conducted during the last online-lecture of the first-year Bachelor course 'People in Business and Society'. The Qualtrics link was sent to the attending students during the lecture. The participation was voluntary. Students could earn a fixed number of course participation credits, which were given independently of

[9] The experiments in 2019 and 2020 also differ in terms of the ordering of the matrix task and the questionnaire. In 2019, the questionnaire was completed before the task, while in 2020, the task was completed before the questionnaire. The reversed order in 2020 was possible because of resolved technical issues.

[10] We dropped the following observations leading to a final sample of 302 observations: 5 who did not consent, 130 who did not start the matrix task, 5 double entries, 32 who finished the survey in less than 1000 seconds (the matrix task alone takes 15 minutes = 900 seconds), 8 who worked on the survey for more than 2100 seconds (= 35 minutes; the average duration was 1499 seconds. Less than 2% of participants took more than 2100 seconds. Since the experiment took place in a less controlled environment than in the lab, we tried to make the conditions of the participants as comparable as possible.), and 7 who attempted to solve more than 49 matrix tasks (= mean + one standard deviation in the top 10 percentile of attempts) and were considered as not working seriously on the task.

[11] Since it was the first lecture of the course, some students were not able to access the notification with the Qualtrics link and they were given one of the treatment links during the lecture.

whether they consented to participate in the study, finished, or left the study, and these conditions were announced one week before the experiment. The students were provided with one Qualtrics link and randomly assigned to one treatment within the survey. In total, 333 students participated in the experiment, out of which 144 were female and 189 male. 113 students were randomly assigned to the control treatment (NoGoal), 112 to the Private Goal Setting Treatment (PrivGoal), and 108 to the Public Goal Setting Treatment (PubGoal).[12]

3.4 Hypotheses and Research Questions

Our hypotheses and research questions are motivated by two research streams as discussed in detail in the literature review in section 2: First, the large body of mainly psychological studies analyzing goal setting and, in particular, self-set goals and their impact on behavior and individuals' performance. And second, the broad stream of experimental economics literature addressing gender differences in different contexts involving various elements of competition.

Since this is, to our best knowledge, the first study addressing gender differences in private versus public goal-setting environments, we developed our hypotheses based on the findings in the two motivating research fields on self-set goals and gender differences. We expect several factors to play a role in the way women and men set their own goals: self-confidence (in own performance skills), beliefs about learning capabilities, optimism, as well as social norms on the gendered appropriateness of modest versus ambitious goals that are likely to influence private and public goal setting and performance. Second, concerns about the public image also play a role in public goal setting and performance, but cannot affect private goal setting and performance by design.

Because of the first-mentioned factors, we expect to find a gender gap in self-set goals and consequently in performance in the private goal setting treatment PrivGoal, whereby men set higher goals and perform better than women. In addition, we expect that the public image concerns in the public goal-setting condition PubGoal increase the gender differences. For the control treatment NoGoal we do not have a clear guideline to go by. Using the same task and without goals, Schram et al. (2019) find no gender differences under private ranking and that women perform worse than men under public ranking. In our case we do not provide any type of explicit ranking, but individual results are public at the end of the experiment and, hence the environment may be closer to one with public ranking. Given these opposing influences, we do not have a clear basis for formulating a hypothesis for the NoGoal condition.

[12] We dropped the following observations leading to a final sample of 333 observations: 5 who did not consent, 39 who did not start filling in the questionnaire (no self-identified gender available), 6 who preferred to not self-identify their gender, 4 who worked on the survey for more than 2100 seconds (= 35 minutes; the average duration was 1499 seconds. Less than 2% of participants took more than 2100 seconds. Since the experiment took place in a less controlled environment than in the lab, we tried to make the conditions of the participants as comparable as possible.), and 2 who attempted to solve more than 49 matrix tasks (= mean + one standard deviation in the top 10 percentile of attempts) and were considered as not working seriously on the task.

The following hypotheses summarize our expected behavior (as formulated in the pre-registration[13]):

Hypothesis 1: Men set significantly higher goals than women when self-set goals are private information (treatment PrivGoal), and this difference becomes larger when goals are set publicly (PubGoal).

Hypothesis 2: While women and men do not perform differently without goals (control treatment NoGoal), a significant gender gap in performance emerges with privately self-set goals (treatment PrivGoal), and it becomes larger when goals are set publicly (treatment PubGoal).

One of the critical research questions that follow the hypotheses is whether changes in women's and/or men's behavior drive gender differences across treatments. Public image concerns combined with social norms on the appropriateness of ambitious goals are likely to affect men's self-set goals and performance positively. For women, the impact of public image concerns is more unambiguous. While women are often expected to be more modest and less competitive than men, the observability and identifiability of the goal can also lead to a boost in ambition, which can affect performance positively (higher aspiration) or negatively (higher performance pressure).

A separate issue is the comparison between the two goal-setting treatments and the treatment without goals. Referring to previous findings on the effect of self-set goals on performance (e.g., van Lent and Souverijn, 2020), we also expect a positive impact of private goals on women's and men's performance compared to no goals.

4. Results

Before turning to the analysis of participants' self-set goals and their performance, we present descriptive statistics. The distribution of socio-demographics does not differ across treatments and experiments overall. In treatments NoGoal, PrivGoal, and PubGoal, the respective share of women is 47%, 43%, and 38% (chi2 test, $p = 0.136$), the respective average age is 19.0, 19.2, and 18.9 (Kruskal-Wallis test, $p = 0.3605$), and 58%, 63%, and 64% of the participants indicate that they feel attached to the Dutch culture (chi2 test, $p = 0.346$).

Comparing the field and the online classroom experiments, respectively 41% and 43% of the participants are women (chi2 test, $p = 0.578$) while 62% indicate affinity with the Dutch culture in either cohort (chi2 test, $p = 0.919$). There is a small, yet significant age difference across cohorts (18.9 in the field vs. 19.2 in the online classroom experiment; Mann-Whitney U test---hereinafter MWU test, $p = 0.0387$).

In the following, we present the results on participants' goal setting and performance across gender and treatments. We focus on the pooled analysis of both experiments, but also present the

[13] The study was pre-registered at AsPredicted.org (#48703, https://aspredicted.org/blind.php?x=4t5by6).

results for the field and the online classroom experiment separately. We show non-parametric tests and regression results from Ordinary Least Square regressions with robust standard errors.[14]

4.1 Goal Setting

In the treatments PrivGoal and PubGoal, participants choose a goal for the number of correct answers that they want/aim to give. Participants are free to choose any goal between 0 and 99, and it does not have any monetary or assessment consequences for them. The average self-set goal in PrivGoal is 18.0, and the corresponding goal of 23.3 is significantly higher in PubGoal (MWU test, $p = 0.0682$). This difference is driven by the behavior in the field experiment (MWU test, $p = 0.0906$; online classroom experiment: MWU test, $p = 0.9058$). We expected to find a larger treatment effect in the field experiment, given that observability and identifiability of goals are arguably higher in an in-person situation than in an online environment.

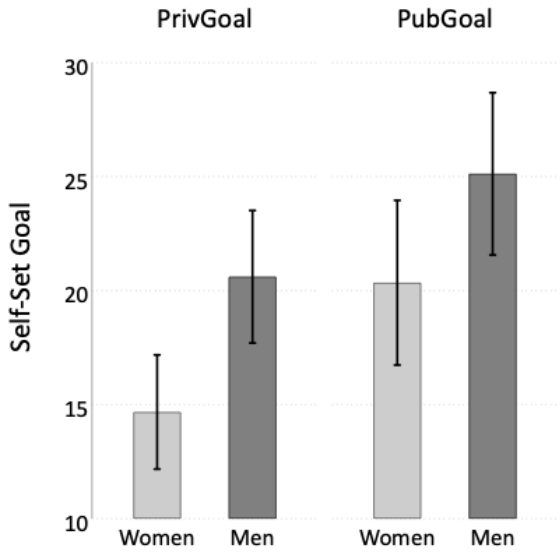


Figure 3. *Self-Set Goal. Average goal set by women and men in the treatments PrivGoal and PubGoal. 90% confidence intervals are calculated with robust standard errors.*

Figure 3 shows the average goal set by women and men in the treatments PrivGoal and PubGoal and highlights a key finding. The corresponding Ordinary Least Square regression

[14] The main text and the tables refer to uncorrected p-values. For our main analysis (gender differences), we run a total of 14 tests with two outcome variables (goal setting and performance) testing for gender gaps across treatments and treatment effects on women, men, and the gender gap. This is reflected by the regression post-estimation tests in table 1 for goal setting and table 2 for performance. The chance of at least one false positive result with 14 (independent) tests and a significance level of 10% is 0.77. We apply the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) for 14 multiple comparisons with an acceptable false discovery rate of 0.20 and apply the correction to the OLS post-estimation F-tests as well as the MWU tests. With this multiple testing correction, all uncorrected significant results remain significant. In Appendix A we show additional results for the mediator variables 'expected performance' and 'attempts'. They do not pertain to our main hypotheses.

analysis with robust standard errors and post-estimation F-tests are presented in Table 1. Four observations emerge, where, for Table 1 we refer to the results of the F-tests shown in the bottom part of the table. First, starting with the gender gap results, in treatment PrivGoal, men are significantly more ambitious than women---as revealed by the male self-set average goal of 20.6 compared to the average goal of 14.7 set by women (MWU test, $p = 0.0000$; Table 1, model 1, $p = 0.0113$). This is robust across the experimental settings and cohorts (field experiment: MWU test, $p = 0.0018$, Table 1, model 3, $p = 0.0617$; online classroom experiment: MWU test, $p = 0.0104$, Table 1, model 5, $p = 0.0951$).[15] This gender difference in goal setting is large. Male set a 40% larger goal than women.

Second, in the treatment PubGoal, men are more ambitious and set a higher goal than women. This gender difference in goal setting is substantial. Male set a 24% larger goal than women. The gender difference is significant with non-parametric tests, but insignificant with parametric tests (female goal of 20.3 versus male goal of 25.1; MWU test, $p = 0.0299$; Table 1, model 1, $p = 0.122$). A closer look reveals that the weaker results in treatment PubGoal stem from different responses in the field and in the online classroom experiment. In the field experiment, men set a 40% larger goal than women and in the online classroom experiment men set only a 4% larger goal than women. To be precise, while women and men aim publically for a better performance than privately in the field experiment (15.0 versus 21.9 for women: MWU test, $p = 0.0803$, Table 1, model 3, $p = 0.0649$; 20.9 versus 30.5 for men: MWU test, $p = 0.2840$; Table 1, model 3, $p = 0.0183$), this is not the case in the online classroom experiment (14.4 versus 18.1 for women: MWU test, $p = 0.8487$, Table 1, model 5, $p = 0.349$; 20.3 versus 18.9 for men; MWU test, $p = 0.6128$; Table 1, model 5, $p = 0.698$). We can only speculate about the reasons for this difference across experiments. The different setups (in-person versus online lecture; first versus last lecture of the course) might explain part of these differences.[16]

[15] Self-set goals are significantly correlated with an individual's risk attitudes (correlation coefficient = 0.1568, $p = 0.0008$). Therefore, the gender gap in both treatments becomes insignificant when adding controls to the regression models (Table 1, models 2, 4, and 6).

[16] For example, the observability and identifiability of goals might be higher in an in-person situation (field experiment) than in an online environment (online classroom experiment). The treatment effect is weaker for both genders in the online classroom experiment. Women set a 46% (25.7%) larger goal in the PubGoal treatment in the field (online classroom) experiment. And men set a 46% larger goal in the PubGoal treatment in the field experiment and a 1.4% lower goal in the online classroom experiment compared to the PrivGoal treatment. In addition, one of the topics covered in the course "People in Business and Society" is gender diversity on the labor market. This might have impacted the more conscious choice of the self-set goal. Differences are less pronounced for the less conscious performance in the matrix-task as discussed in section 4.2.

Table 1. Self-Set Goal: Gender Gap and Treatment Effect on Women, Men, and Gender Gap

	Self-Set Goal					
	(1) Pooled	(2) Pooled	(3) Field	(4) Field	(5) Classroom	(6) Classroom
Female	-5.932**	-3.929	-5.922*	-4.450	-5.837*	-4.008
	(2.331)	(2.473)	(3.154)	(3.563)	(3.482)	(3.487)
Public Goal	4.516	5.336*	9.594**	9.950**	-1.414	0.0277
	(2.795)	(2.788)	(4.037)	(4.105)	(3.643)	(3.398)
Female*Public Goal	1.154	-0.357	-2.677	-2.851	5.059	3.408
	(3.864)	(4.065)	(5.496)	(5.844)	(5.326)	(5.206)
Constant	20.61***	47.91	20.95***	123.3	20.27***	-103.7
	(1.766)	(54.97)	(2.139)	(81.75)	(2.831)	(72.35)
Controls	NO	YES	NO	YES	NO	YES
Observations	451	450	231	230	220	220
R-squared	0.029	0.049	0.055	0.076	0.013	0.034
Gender Gap						
Private Goal	-5.932**	-3.929	-5.922*	-4.450	-5.837*	-4.008
[F-test p-value]	*[0.0113]*	*[0.113]*	*[0.0617]*	*[0.213]*	*[0.0951]*	*[0.252]*
Public Goal	-4.778	-4.286	-8.599*	-7.301	-0.778	-0.600
[F-test p-value]	*[0.122]*	*[0.166]*	*[0.0573]*	*[0.103]*	*[0.847]*	*[0.885]*
Effect of Public (vs. Private) Goal						
Women	5.670**	4.979*	6.917*	7.099*	3.645	3.436
[F-test p-value]	*[0.0342]*	*[0.0747]*	*[0.0649]*	*[0.0767]*	*[0.349]*	*[0.372]*
Men	4.516	5.336*	9.594**	9.950**	-1.414	0.0277
[F-test p-value]	*[0.107]*	*[0.0563]*	*[0.0183]*	*[0.0162]*	*[0.698]*	*[0.994]*
Gender Gap	1.154	-0.357	-2.677	-2.851	5.059	3.408
[F-test p-value]	*[0.765]*	*[0.930]*	*[0.627]*	*[0.626]*	*[0.343]*	*[0.513]*

Notes. *** p<0.01, ** p<0.05, * p<0.1. Ordinary Least Square Regression results with robust standard errors (in parentheses). The table shows post-estimation F-tests with corresponding p-values [in parentheses]. The dependent variable is the participant's self-set goal for the matrix-task and the explanatory variables are a gender dummy (taking value 1 if female and 0 if male), a treatment dummy (taking value 1 if PubGoal and 0 if PrivGoal), and their interaction term. Controls (Risk Attitudes, Age, Age^2, and a dummy for feeling attached to the Dutch culture) are included in models (2), (4), and (6). The data come from treatments PrivGoal and PubGoal and the samples are both experiments pooled in models (1) and (2), the field experiment in models (3) and (4), and the classroom experiment in models (5) and (6).

Third, focusing now on the effects of public vs. private goals, we find that both genders are more ambitious and set higher goals in the PubGoal treatment compared to the PrivGoal treatment. For the overall data this effect is significant for women (14.7 versus 20.3, MWU test, p = 0.0583; Table 1, model 1, p = 0.0342), but not for men (20.6 versus 25.1, MWU test, p = 0.5482; Table 1, model 1, p=0.107). For both genders the positive effects stem from the field experiment data. Men increase their goals in the field experiment (20.9 versus 30.5, MWU test, p = 0.2840; Table 1, model 3, p = 0.0183), but in the classroom experiment the effect is not significant and

negative. Women increase their goals significantly at the 10% level in the field and not significantly in the online classroom experiment.

Fourth, the gender gap in the difference between public and private goals is minor. Men's goals are 29 percent higher than women's self-set goals in the treatment PrivGoal, but they are only 19 percent higher in the treatment PubGoal. Moving from private to public goal setting reduces the gap by ten percentage points, an insignificant change (Table 1, model 1, $p = 0.765$); this result also holds separately for the field and the classroom.

We can summarize our findings with respect to private and public goal setting in the following results:

Result 1a: Men set significantly higher goals compared to women in both treatments. This difference is larger when goals are private information.

Result 1b: The gender gap in goal-setting is not larger when the goal is public versus private information.

Result 1c: Women set significantly higher goals when goals are public compared to private information in both experiments and mainly in the field environment.

Result 1d: Men set significantly higher goals when goals are public compared to private information, but only in the field environment.

Results 1a and 1b speak directly to Hypothesis 1 (pre-registered). Our data are consistent with the first part of it but not with the second part.

4.2 Performance

Before we turn to the second outcome variable---women's and men's performance in the different goal setting conditions---we shortly discuss the impact of the goal treatments on the (potential) mediating factors 'expected performance' and 'attempts'. Recall that before performing the matrix task, participants are asked to indicate how many problems they *expect* to solve correctly.[17] Participants' expected performance and their actual performance are positively correlated (correlation coefficient = 0.0867, $p = 0.0299$). The correlation of attempts and performance is very strong (correlation coefficient = 0.6671, $p = 0.000$). In Appendix A, gender differences across treatments are shown visually and with regression analysis (Figure A1 and Table A1 for expected performance; Figure A2 and Table A2 for attempts).

While women's performance expectations go slightly up when the self-set goal is publicly visible (13.5 in NoGoal, 13.6 in PrivGoal, 17.5 in PubGoal; NoGoal versus PubGoal: MWU test,

[17] Participants receive feedback about the correctness of their answer after each summation problem. We therefore elicit beliefs before the real effort task.

p = 0.0964; otherwise p > 0.2844), the effect is strongly pronounced among men (13.5 in NoGoal, 17.8 in PrivGoal, 21.5 in PubGoal; NoGoal versus PrivGoal: MWU test, p = 0.0083; NoGoal versus PubGoal: MWU test, p = 0.0001). The post-estimation F-tests of Ordinary Least Square regressions in Table A1 draw a similar picture.

Foreseeably, expected performance and self-set goals are strongly positively correlated (correlation coefficient = 0.8708, p = 0.0000). The vast majority of participants (91.5%) believes to achieve maximally their self-set goal and some interesting patterns emerge: While half of the participants (50.6%) are confident to meet their self-set goal precisely, 40.9% expect to perform worse than their self-set goal. Among the 91.5% of participants, the goal setting environment does not significantly affect the distribution of participants confident to meet their goal (chi3 test, p = 0.605), also not separately for women and men (chi2 tests, p > 0.244). However, while roughly half of women and men expect to meet their self-set goal in PrivGoal (51.9% of women and 55.5% of men; chi2 test, p = 0.622), a gender gap emerges in PubGoal: 45.6% of women versus 62.8% of men indicate that they are confident to meet their self-set goal (chi2 test, p = 0.014). While the public goal-setting environment seems to boost men's goal-compliance confidence, the opposite tendency can be observed for women.

With respect to the number of attempted summations we find a large and highly significant gender gap in attempts across treatments (2.4 in NoGoal, 3.7 in PrivGoal, 4.3 in PubGoal; MWU tests, p < 0.0023; Table A2, model 1, p < 0.00242), which is consistent across experiments (Table A2, models 3 - 6). While men attempt to solve more matrix summations after setting a goal for themselves (17.4 in NoGoal, 18.9 in PrivGoal, 19.2 in PubGoal; MWU tests, p < 0.0622) women's attempts are literally unchanged (15.0 in NoGoal, 15.3 in PrivGoal, 15.0 in PubGoal; MWU tests, p > 0.9420). The different reaction to goal setting affects the gender gap in attempts across treatment however only slightly (Table A2, model 5, p = 0.0799; non-reported difference in gender gap in NoGoal versus PubGoal, Table A2, model 1, p = 0.0942).

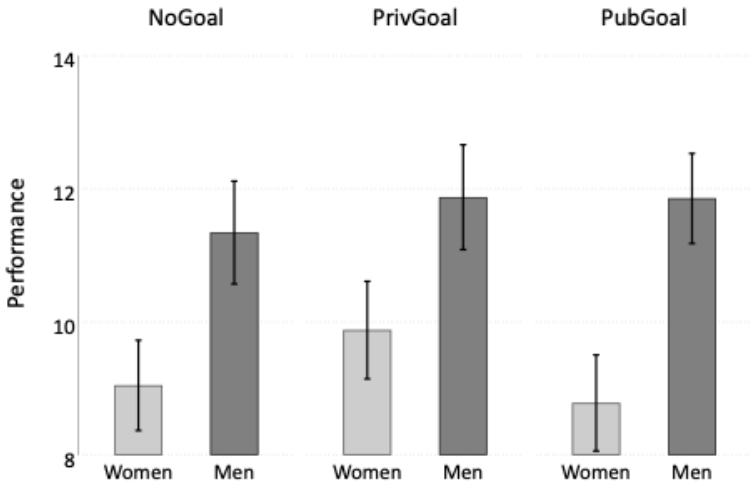


Figure 4. *Performance. Average performance by women and men in the treatments NoGoal, PrivGoal, and PubGoal. 90% confidence intervals are calculated with robust standard errors.*

Table 2. Performance: Gender Gap and Treatment Effects on Women, Men, and Gender Gap

	Performance					
	(1) Pooled	(2) Pooled	(3) Field	(4) Field	(5) Classroom	(6) Classroom
Female	-1.996***	-1.983***	-1.550	-1.906	-2.092***	-1.802***
	(0.654)	(0.658)	(1.197)	(1.181)	(0.656)	(0.655)
No Goal	-0.530	-0.602	-0.439	-1.223	-0.303	-0.0814
	(0.671)	(0.664)	(1.246)	(1.205)	(0.707)	(0.702)
Public Goal	-0.0187	-0.174	-0.660	-1.057	0.628	0.630
	(0.632)	(0.620)	(1.016)	(0.994)	(0.730)	(0.717)
Female*No Goal	-0.301	0.0320	-0.773	0.120	-0.206	-0.247
	(0.904)	(0.896)	(1.669)	(1.608)	(0.964)	(0.963)
Female*Public Goal	-1.076	-0.932	-0.951	-0.683	-1.801*	-1.869*
	(0.888)	(0.896)	(1.501)	(1.509)	(0.972)	(0.976)
Constant	11.87***	1.534	12.55***	-19.37	11.19***	26.68**
	(0.479)	(11.84)	(0.797)	(16.86)	(0.519)	(11.85)
Controls	NO	YES	NO	YES	NO	YES
Observations	635	626	302	294	333	332
R-squared	0.071	0.103	0.046	0.094	0.128	0.164
Gender Gap						
No Goal	-2.297***	-1.951***	-2.323**	-1.786	-2.298***	-2.049***
[F-test p-value]	*[0.000253]*	*[0.00154]*	*[0.0468]*	*[0.113]*	*[0.00127]*	*[0.00321]*
Private Goal	-1.996***	-1.983***	-1.550	-1.906	-2.092***	-1.802***
[F-test p-value]	*[0.00237]*	*[0.00268]*	*[0.196]*	*[0.108]*	*[0.00156]*	*[0.00628]*
Public Goal	-3.073***	-2.915***	-2.501***	-2.589***	-3.893***	-3.671***
[F-test p-value]	*[0.000]*	*[0.000]*	*[0.00605]*	*[0.00656]*	*[0.000]*	*[0.000]*
Effect of No (vs. Private) Goal						
Women	-0.831	-0.570	-1.212	-1.102	-0.509	-0.329
[F-test p-value]	*[0.171]*	*[0.348]*	*[0.276]*	*[0.313]*	*[0.438]*	*[0.617]*
Men	-0.530	-0.602	-0.439	-1.223	-0.303	-0.0814
[F-test p-value]	*[0.429]*	*[0.365]*	*[0.725]*	*[0.311]*	*[0.668]*	*[0.908]*
Gender Gap	-0.301	0.0320	-0.773	0.120	-0.206	-0.247
[F-test p-value]	*[0.739]*	*[0.972]*	*[0.644]*	*[0.940]*	*[0.831]*	*[0.798]*
Effect of Public (vs. Private) Goal						
Women	-1.095*	-1.107*	-1.611	-1.740	-1.173*	-1.239*
[F-test p-value]	*[0.0801]*	*[0.0867]*	*[0.146]*	*[0.130]*	*[0.0685]*	*[0.0623]*
Men	-0.0187	-0.174	-0.660	-1.057	0.628	0.630
[F-test p-value]	*[0.976]*	*[0.779]*	*[0.517]*	*[0.288]*	*[0.390]*	*[0.380]*
Gender Gap	-1.076	-0.932	-0.951	-0.683	-1.801*	-1.869*
[F-test p-value]	*[0.226]*	*[0.298]*	*[0.527]*	*[0.651]*	*[0.0647]*	*[0.0563]*

Notes. *** p<0.01, ** p<0.05, * p<0.1. Ordinary Least Square Regression results with robust standard errors (in parentheses). The table shows post-estimation F-tests with corresponding p-values [in parentheses]. The dependent variable is the participant's number of attempts in the matrix-task and the explanatory variables are a gender dummy (taking value 1 if female and 0 if male), a treatment dummy for NoGoal and PubGoal (taking value 1 if applies and 0 otherwise), and the interaction terms of the gender dummy with each treatment dummy. Controls (Expected Performance, Risk Attitudes, Age, Age^2, and a dummy for feeling attached to the Dutch culture) are included in models (2), (4), and (6). The data come from treatments NoGoal, PrivGoal, and PubGoal and the samples are both experiments pooled in models (1) and (2), the field experiment in models (3) and (4), and the classroom experiment in models (5) and (6).

With respect to performance, we make three observations. First, a preliminary result pertains to the case of NoGoal. While women solve on average 9.0 summations correctly in the NoGoal treatment, men give 11.3 correct answers (MWU test, $p = 0.0016$; Table 2, model 1, $p = 0.000253$). Considering that performance is made public after the study, we can give an ex-post explanation of why men outperform women in NoGoal. The control condition NoGoal is more comparable to the Status Ranking treatment than to the Control or Conformity treatments in Schram et al. (2019). In the Status Ranking treatment, a third party can compare participants' performance with each other, which is not the case in the Control or Conformity treatments. With public performance, this is indeed the case in our control treatment NoGoal.

Second, we find a robust and strong gender gap in performance across treatments, illustrated in Figure 4 and analyzed with post-estimation F-tests in Table 2.[18] In both treatments, when setting a goal (privately or publicly), men perform significantly better than women (9.9 versus 11.9 in PrivGoal: MWU test, $p = 0.0015$, Table 2, model 1, $p = 0.00237$; 8.8 versus 11.9 in PubGoal: MWU test, $p = 0.000$, Table 2, model 1, $p = 0.000$), with the effects being stronger in the online classroom than in the field experiment. The gender gap is hardly affected by the goal-setting environment (only one significant change from PrivGoal to PubGoal in the online classroom experiment: Table 2, model 5, $p = 0.0647$).

Third, we observe an interesting pattern in women's performance: a privately set goal improves female performance slightly and insignificantly compared to no goal. However, women's performance worsens significantly when they set a goal publicly compared to privately (MWU test, $p = 0.0983$; Table 2, model 1, $p = 0.0801$). This is an interesting and novel observation that is worth attention and further research. Men's performance is not affected at all by the goal setting environment (MWU test, $p = 0.8960$; Table 2, model 1, $p = 0.976$).

Our results can be summarized as follows:

Result 2a: A gender gap in performance exists in all treatments with and without goal setting.

Result 2b: The gender gap in performance is larger, but not significantly, when the goal is public compared to private information.

Result 2c: Women perform significantly worse when goals are public compared to private information.

Result 2d: Men's performance is not affected by the goal setting conditions.

[18] Though performance is not incentivized in our experiments, participants' performance is strikingly similar to the incentivized performance in Schram et al. (2019) where women's and men's average performance range between 10 and 14 correct answers across treatments. We are thus confident that participants in our experiments take the study and the real effort task seriously. The same is true for the number of attempts, see Schram et al. (2019).

Results 2a and 2b speak directly to Hypothesis 2 (pre-registered). Our data are not consistent with the hypothesis, since we do find a gender gap in performance in NoGoal and the gender gap does not vary significantly between private and public goals.

Finally, we can now compare set goals with actual performance. Under private goal setting women's performance is on average 67% of goals, whereas for men it is 57%. Under public goal setting the corresponding percentages are 43% and 39% respectively. Seen ex-post participants are more realistic under private than under public goal setting, with women being more realistic than men in both cases.

5. Discussion and Conclusion

We conduct a field and an online classroom experiment to test the hypothesis that men and women set different goals and perform differently when information about their goals is private versus publicly revealed. In general, participants set higher goals when they are publicly observable and identifiable. Overall across treatments, we find that men set higher goals than women. For goal setting in specific, results in the private setting show that men set more ambitious goals than women, but this effect weakens in the public setting in which especially women set higher goals and thus the gender gap becomes weaker. For performance, we find that men perform better than women across treatments (i.e., without goals and with privately or publicly self-set goals). This result might not be surprising given that performance is publicly revealed in all treatments. Both genders perform slightly, but insignificantly better in the private goal-setting environment compared to the treatment without goals. Men perform very similarly across the two goal-setting environments (private and public). However, the public visibility of goals harms women's performance significantly, but it does not affect the overall gender gap. Since the evidence on how women and men set their goals and perform is limited in the experimental economics and in the psychology literature (Welsh et al., 2020), we will use related research in both streams of literature to discuss our results.

For the gender gap in private goal setting it is interesting to view our results in the light of various psychological theories. A possible explanation may be derived from the self-concordance theory; a theory in the psychology literature formulated by Sheldon and Elliot (1998, 1999). According to this theory self-concordant goals emanate directly from the integrated self and are thus in line with an individual's preferences and interests. Such goals are free from any pressure to please others or to respond to one's external environment (Sheldon and Elliot, 1999) and are therefore more likely salient to private settings in which individuals do not have to take their social environment into account when setting their own goals.[19] This is also in line with the goal setting theory which posits that self-efficacy, self-confidence that the goal for a specific task is attainable, is an important individual characteristic for self-set goals (Bandura, 1997; Latham and Locke, 2007; Locke and Latham, 2006).

[19] Sheldon et al (2020) demonstrate recently that self-concordant goals foster positive personal resources such as optimism and hence do reflect an individual's belief that they can achieve goals.

In our experiment both men and women were free to set a goal that matches their personal interests and abilities in the private setting. It could thus be that men set higher goals in the private setting, because they are likely more confident about their competences (e.g., Beyer, 1990; Lundeberg et al., 1994; Beyer and Bowden, 1997) and opt therefore for more difficult goals than their less confident female peers (McCarty, 1986; Wood and Karten, 1986) as suggested in the old stream of psychology literature and also more recently in research in the experimental economics literature showing that men are more confident about their abilities than women (Buser et al., 2020).

Several studies reveal substantial gender differences in the attribution of success and failure to internal factors (personal abilities and skills) and external factors (for instance, luck). While some evidence suggests that boys show a stronger self-serving attributional bias than girls (e.g., Stipek and Gralinski, 1991), the findings seem to be stronger for adolescents (Hankin and Abramson, 2001) and adults (Boggiano and Barrett, 1991; Mezulis et al., 2004). If not achieving a self-set goal can be seen as failure (increasing in the size of the mismatch), men might be more likely to attribute such 'failure' to external factors whereas women possibly tend to internalize it. These attribution differences could explain why men set higher goals than women, especially when goals are private information.

In the public setting, both men and women increased their goals in the field experiment, with men setting more ambitious goals than women. There may be several potential explanations for this finding. For instance, social conformity is a phenomenon that has received much attention in the psychology literature since its introduction to the literature (Asch, 1951). It is considered a powerful social phenomenon that encourages individuals to adapt their opinions and behaviours to conform to the majority in the group, especially to fit in the group and to be "liked" by others (Asch, 1951; Deutsch and Gerard, 1955). It has been widely observed in face-to-face groups, but more recently the psychological mechanism was also found in online environments (Wijenayake et al., 2020). Both men and women were aware that their self-set goals were observable and identifiable by their peers; they could have therefore increased their goals knowing that the majority would do so. In the field experiment this effect was likely more salient due to the physical presence; participants were much more visible and identifiable in the field experiment than in the online classroom experiment.[20]

The findings regarding performance suggest that without and with (privately or publicly) self-set goals, men perform significantly better than women when performance is public. According to the goal setting theory this makes sense as men set higher goals than women in private and public settings (Locke and Latham, 1990, 2002; Locke and Latham, 2006). Men's performance is also better when they set goals in the private and public setting compared to no goal setting. This finding supports the research of Clark et al. (2017) who find that task-based goal setting increases task completion (i.e., practice exams) and course performance especially for

[20] Another potential explanation is offered by psychology researchers suggesting that in competitive environments, a "desire to win" can emerge within individuals which motivates them to beat the other side, rather than focusing solely on maximizing their individual payoffs (Cooper and Fang, 2008; Delgado et al., 2008) which could thus be true for both men and women.

men. In the experiment in Schram et al. (2019) men significantly increased the number of attempted summations as well as the number of correct summations when there was status ranking. Participants' performance is publicly available and identifiable across treatments in our experimental study. This design feature is in fact in line with the status ranking treatment in Schram et al. (2019), which might explain gender differences even in the control condition. With our design, we are however not able to distinguish between baseline gender differences in performance and the role of public performance.

Women's performance drops significantly when they set a goal publically compared to privately. This finding is in line with previous experimental studies which show that women underperform in competitive environments (Gneezy et al., 2003; Gneezy and Rustichini, 2004). We do add a novel insight to this extant work, by showing that women set higher goals than they could achieve and others have thus observed this phenomenon.

Several important and interesting research questions emerge from our study that open the way for a new stream of research. To which factors do women and men attribute (un-) successfully met self-set goals? Can women and men be coached in setting better goals and how does that affect performance? Advice institutions have been shown to improve decisions and outcomes in a variety of settings (for instance, Schotter, 2003; Brandts et al., 2015; Brandts and Rott, 2021). How does the public perceive women's and men's goal setting and performance? And how do women and men process the public perception? We do not know whether failing to achieve a publically set goal results in gender differences in negative consequences such as reputational damages and stress or future goal setting and performance in competitive environments. It is interesting to examine whether such consequences take place as they might bear important implications for practice. Especially, since women tend to be more sensitive to negative outcomes than men (Buser and Yuan, 2019; Brody, 1993; Fujita et al., 1991). Additionally, public observability and public perception might affect self-selection into public positions. Our experimental setup was entirely non-strategic. However, the strategic context in which individuals set goals might have an important impact on goal setting (and possibly performance).

References

Alesina, A., and Giuliano, P. (2010). The Power of the Family. Journal of Economic Growth, 15, 93–125.

Ariely, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009). Large stakes and big mistakes. The Review of Economic Studies, 76(2), 451–469.

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), Groups, leadership, and men. Pittsburgh: Carnegie.

Bandura, A. (1997). Self-efficacy: The exercise of control. Stanford: W.H. Freeman.

Behnke, R.R., and Sawyer, C.R. (2001). Patterns of psychological state anxiety in public speaking as a function of anxiety sensitivity. Communication Quarterly 49 (1):84–94.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, 57(1), 289–300.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. Journal of Personality and Social Psychology, 59, 960-970. https://doi.org/10.1037/0022-3514.59.5.960

Beyer, S., and Bowden, E.M. (1997). Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias. Personality and Social Psychology Bulletin, 23(2), 157–172. https://doi.org/10.1177/0146167297232005

Boggiano, A.K., and Barrett, M. (1991). Gender differences in depression in college students. Sex Roles, 25, 595–605.

Brandts, J., Groenert, V., and Rott, C. (2015). The impact of advice on women's and men's selection into competition. Management Science, 61(5), 1018–1035.

Brandts, J., and Rott, C. (2021). Advice from women and men and selection into competition. Journal of Economic Psychology, 82, article 102333.

Brody, L.R. (1993). On understanding gender differences in the expression of emotion." In Human Feelings: Explorations in Affect Development and Meaning, ed. Steven L. Ablon,

Daniel Brown, Edward J. Khantzian, and John E. Mack, 87-121. Hillsdale, N.J.: Analytic Press.

Buser, T. and Yuan, H. (2020). Public Speaking Aversion, working paper.

Buser, T. and Yuan, H. (2019). Do Women give up Competing more easily? Evidence from the Lab and the Dutch Math Olympiad. American Economic Journal: Applied Economics, 11(3), 225-52.

Buser, T., Niederle, M. and Oosterbeek, H. (2020). Can Competitiveness Predict Education and Labor Market Outcomes? Evidence from Incentivized Choice and Survey Measures. Working paper available on SSRN. http://dx.doi.org/10.2139/ssrn.3549354

Buser, T., Grimalda, G., Putterman, L. and van der Weele, J. (2020). Overconfidence and gender gaps in redistributive preferences: Cross-country experimental evidence. Journal of Economic Behavior & Organization, 187, 267-286.https://doi.org/10.1016/j.jebo.2020.07.005 0167-2681/

Carter, A.J., Croft, A., Lukas, D. and Sandstrom, G.M. (2018). Women's visibility in academic seminars: Women ask fewer questions than men. PloS one 13 (9):e0202743.

Cerasoli, C.P., Nicklin, J.M., and Ford, M.T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. Psychological Bulletin, 140(4),980–1008. doi: 10.1037/a0035661

Clark, M., Gill, D., Prowse, V., and Rush, M. (2017). Using goals to motivate college students: Theory and evidence from field experiments. NBER Working Paper, No. 10283. https://doi-org.vu-nl.idm.oclc.org/10.1162/rest_a_00864

Christiansen, L., Lin, H., Pereira, J., Topalova, P.B. and Turk-Ariss, R. (2016): Gender Diversity in Senior Positions and Firm Performance: Evidence from Europe. IMF Working Paper.

Cooper, D. J., and Fang, H. (2008). Understanding overbidding in second price auctions: An experimental study. Economic Journal, 18(532), 1572–1595. https://doi-org.vu-nl.idm.oclc.org/10.1111/j.1468-0297.2008.02181.x

Croson, R., Gneezy, U., (2009). Gender differences in preferences. Journal of Economic Literature, 47 (2), 448–474. doi:10.1257/jel.47.2.448.

Dalton, P. S., V. Gonzalez and C. N. Noussair (2016). Self-Chosen Goals: Incentives and Gender Differences, CentER Discussion Paper Series No. 2016-036. http://dx.doi.org/10.2139/ssrn.2583872

Delgado, M.R., Schotter, A., Ozbay, E.Y. and Phelps, E.A. (2008). Understanding overbidding: Using the neural circuitry of reward to design economic auctions. Science, 321, 1849–1852. doi.10.1126/science.1158860

De Pater, I.E., Van Vianen, A.E.M., Fischer, A.H., and Van Ginkel, W.P. (2009). Challenging experiences: Gender differences in task choice. Journal of Managerial Psychology, 24, 4-28. doi. 10.1108/02683940910922519

Deutsch, M. and Gerard, H.B.(1955). A study of normative and informational social influences upon individual judgment. Journal of Abnormal and Social Psychology, 51, 629–636. https://doi.org/10.1037/h0046408

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G.G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. Journal of the European Economic Association, 9 (3), 522–550. https://doi-org.vu-nl.idm.oclc.org/10.1111/j.1542-4774.2011.01015.x

Dezsö, C. L. and D.G. Ross (2012). Does Female Representation in Top Management Improve Firm Performance? a Panel Data Investigation. Strategic Management Journal, 33, 1072–1089.

Erez, M., Gopher, D. and Arzi, N. (1990). Effects of goal difficulty, self-set goals, and monetary rewards on dual task performance. Organizational Behavior and Human Decision Processes, 47, 247–269. https://doi.org/10.1016/0749-5978(90)90038-B

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018). Global evidence on economic preferences. The Quarterly Journal of Economics. 133 (4), 1645–1692. https://doi-org.vu-nl.idm.oclc.org/10.1093/qje/qjy013

Fernández, R. (2007): "Alfred Marshall Lecture Women, Work, and Culture," Journal of the European Economic Association, 5, 305–332.

Flory, J.A., Leibbrandt, A. and List., J.A. (2015). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. Review of Economic Studies, 82(1), 122–155. https://doi-org.vu-nl.idm.oclc.org/10.1093/restud/rdu030

Fujita, F., Diener, E. and Sandvik, E. (1991). Gender Differences in Negative Affect and Weil-Being: The Case for Emotional Intensity. Journal of Personality and Social Psychology, 61(3), 427-434. https://doi-org.vu-nl.idm.oclc.org/10.1037/0022-3514.61.3.427

García-Meca, E., García-Sánchez, I.M. and Martínez-Ferrero, J. (2015). Board diversity and its effects on bank performance: An international analysis. Journal of Banking & Finance, 53, 202–214.

Gino, F., Wilmuth, C.A. and Brooks, A.W. (2015). Compared to men, women view professional advancement as equally attainable, but less desirable. Proceedings of the National Academy of Sciences of the United States of America, 112, 12354 –12359. http://dx.doi.org/10.1073/ pnas.1502567112

Gneezy, U., Niederle, M. and Rustichini, A. (2003). Performance in competitive environments: Gender differences. Quarterly Journal of Economics, 118, 1049–1074. https://doi.org/10.1162/00335530360698496

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. American Economic Review, 4, 377–381. doi: 10.1257/0002828041301821

Günther, C., Ekinci, N.A., Schwieren, C. and Strobel, M. (2010). Women can't jump? An experiment on competitive attitudes and stereotype threat. Journal of Economic Behavior & Organization, 75, 395-401. doi:10.1016/j.jebo.2010.05.003

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008): Culture, Gender, and Math. Science, 320, 1164–1165.

Hankin, B.L. and Abramson, L.Y. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability-transactional stress theory. Psychological Bulletin, 127, 773–796.

Hinsley, A., Sutherland, W.J. and Johnston, A. (2017). Men ask more questions than women at a scientific conference. PloS one 12 (10).

Hofstede, G. (1980): Culture´s consequences: International differences in work-related values. Beverly Hills, C.A: Sage.

Hollenbeck, J. R., Williams, C.R. and Klein, H.J. (1989). An empirical examination of the antecedents of commitment to difficult goals, Journal of Applied Psychology, 74. https://doi-org.vu-nl.idm.oclc.org/10.1037/0021-9010.74.1.18

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W. and Gupta, V. (2004). Culture, leadership, and organizations. The GLOBE study of 62 societies. Thousand Oaks, CA: Sage Publications.

Kurman, J. (2001). Self-regulation strategies in achievement settings: Culture and gender differences. Journal of Cross-Cultural Psychology, 32(4), 491-503. https://doi.org/10.1177/0022022101032004008

Latham, G. P. and Locke, E.A. (2007). New developments in and directions for goal-setting research. European Psychologist, 12, 290-300. https://doi-org.vu-nl.idm.oclc.org/10.1027/1016-9040.12.4.290.

Latham, G.P. and Marshall, H.A. (1982). The effects of self-set, participatively set, and assigned goals on the performance of government employees. Personnel Psychology, 35, 399–404. https://doi-org.vu-nl.idm.oclc.org/10.1111/j.1744-6570.1982.tb02204.x

Levy, P. and Baumgardner, A. (1991). Effects of self-esteem and gender on goal choice. Journal of Organizational Behavior, 12, 529-541. https://doi-org.vu-nl.idm.oclc.org/10.1002/job.4030120606

Locke, E.A. (1996). Motivation through conscious goal setting. Applied and Preventive Psychology, 5(2), 117-124. https://doi.org/10.1016/S0962-1849(96)80005-9

Locke, E.A. and Latham, G.P. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice-Hall.

Locke, E.A. and Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705-717. https://doi.org/10.1037/0003-066X.57.9.705

Locke, E. A. and Latham, G.P. (2006). New directions in goal-setting theory. Current directions in psychology science, 15(5), 265-268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

Locke, E.A. and Latham, G.P. (2015). Breaking the rules: A historical overview of goal-setting theory. In A. J. Elliot (Ed.), Advances in motivation science (pp. 99 –126). Cambridge, MA: Elsevier.

Locke, E.A. and Latham, G.P. (2019). The development of goal setting theory: a half century retrospective. Motivation Science. https://doi.org/10.1037/mot0000127.

Lundeberg, M.A., Fox, P.W. and Punccohar, J. (1994). Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments. Journal of Educational Psychology, 86(1), 114–121. https://doi.org/10.1037/0022-0663.86.1.114

Marinho, A.C.F., de Medeiros, A.M., Cortes Gama, A.C., and Caldas Teixeira, L. (2017). Fear of public speaking: Perception of college students and correlates. Journal of Voice 31 (1):127–134.

Malik, M.A.R., Butt, A.N. and Choi, J.N. (2015). Rewards and employee creative performance: Moderating effects of creative self-efficacy, reward importance, and locus of control. Journal of Organizational Behavior, 36, 59–74. doi: 10.1002/job.1943. https://doi.org/10.2753/IJS0020-7659380103

McCarty, P.A. (1986). Effects of feedback on the self-confidence of men and women. Academy of Management Journal, 29, 840-847. https://doi.org/10.5465/255950

McDaniel, A.E. (2008). Measuring gender egalitarianism: The attitudinal difference between men and women. International Journal of Sociology, 38, 58–80.

Mezulis, A.H., Abramson, L.Y., Hyde, J.S. and Hankin, B.L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. Psychological bulletin, 130(5), 711.

Möbius, M., Niederle, M., Niehaus, P. and Rosenblat, T. (2011). Managing self-confidence: theory and experimental evidence. Working paper.

Niederle, M. (2016). Gender. In: Handbook of Experimental Economics, second edition, by John Kagel and Alvin E. Roth (eds). Princeton University Press: 481-553.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? Quarterly Journal of Economics, 122(3), 1067–1101. https://doi-org.vu-nl.idm.oclc.org/10.1162/qjec.122.3.1067

Niederle, M. and Vesterlund, L. (2011). Gender and competition. Annual Review in Economics, 3, 601–630. doi-org.vy-nl.idm.oclc.org/10.1146/annurev-economics-111809-125122.

Parthasarathy, R., Rao, V., and Palaniswamy, N. (2019). Deliberative Democracy in an Unequal World: A Text-As-Data Study of South India's Village Assemblies. American Political Science Review 113 (3):623–640.

Pinder, C.C. (1984). Work motivation: theory, issues and applications. Glenview, IL: Scott, Foresman.

Sahay, R. and M. Cihak (2018). Women in Finance: A Case for Closing Gaps. IMF Staff Discussion Notes.

Schotter, A. (2003). Decision making with naive advice. American Economic Review, 93(2), 196–201.

Schram, A., Brandts, J. and Gërxhani, K. (2019). Social-status ranking: a hidden channel to gender inequality under competition. Experimental Economics, 1-23. https://doi.org/10.1007/s10683-018-9563-6.

Sheldon, K.M. and Elliot, A.J. (1998). Not all personal goals are personal: Comparing autonomous and controlled reasons for goals as predictors of effort and attainment. Personality and Social Psychology Bulletin, 24, 546-557.

Sheldon, K.M., and Elliot, A.J. (1999). Goal striving, need satisfaction, and longitudinal well-being: The self-concordance model. Journal of Personality and Social Psychology, 76, 482-497.

Sheldon, K., Gordeeva, T., Sychev, O., Osin, E. and Titova, L. (2020). Self-concordant goals breed goal-optimism and thus well-being. Current Psychology. https://doi.org/10.1007/s12144-020-01156-7

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. Journal of the European Economic Association, 10, 1189–1213.

Stein, M.B., Walker, J.R., and Forde, D.R. (1996). Public-speaking fears in a community sample: Prevalence, impact on functioning, and diagnostic classification. Archives of general psychiatry 53 (2):169–174.

Stipek, D.J., and Gralinski, J.H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. Journal of educational psychology, 83(3), 361.

Szekeres, H., Shuman, E. and Saguy, T. (2020). Views of sexual assault following #MeToo: the role of gender and individual differences. Personality and Individual Differences, 166(1), 1-6. https://doi.org/10.1016/j.paid.2020.110203

van Lent, M. and Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. Journal of Behavioral and Experimental Economics, 101570.

Turk, C.L., Heimberg, R.G., Orsillo, S.M, Holt, C.S., Gitow, A., Street, L.L., Schneier, F.R., and Liebowitz, M.R. (1998). An investigation of gender differences in social phobia. Journal of anxiety disorders 12 (3):209–223.

Weber, M. and Schram, A. (2017). The non-equivalence of labor market taxes: A real-effort experiment. The Economic Journal, 127, 2187–2215.

World Economic Forum (2018). The Global Gender Gap Report 2018.

Wood, W. and Karten, S.J. (1986). Perceived sex differences in competence as a determinant of sex differences in interaction, Journal of Personality and Social Psychology, 50, 341-347.

Wijenayake, S., van Berkel, N., Kostakos, V. and Goncalves, J. (2020). Impact of contextual and personal determinants on online social conformity, Computers in Human Behavior, 108, 2-11 https://doi.org/10.1016/j.chb.2020.106302

Appendix A. Additional Results

A1. Figures: Expected Performance and Attempts

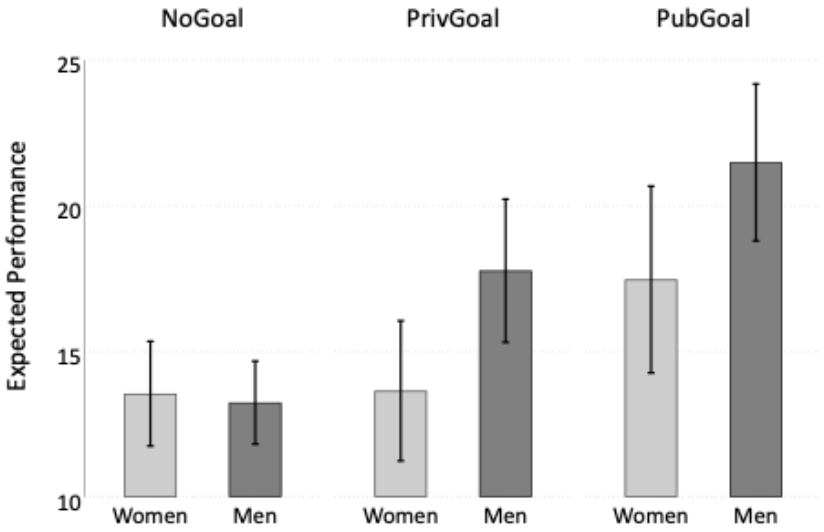


Figure A1. *Expected performance. Average expected performance by women and men in the treatments NoGoal, PrivGoal, and PubGoal. 90% confidence intervals are calculated with robust standard errors.*

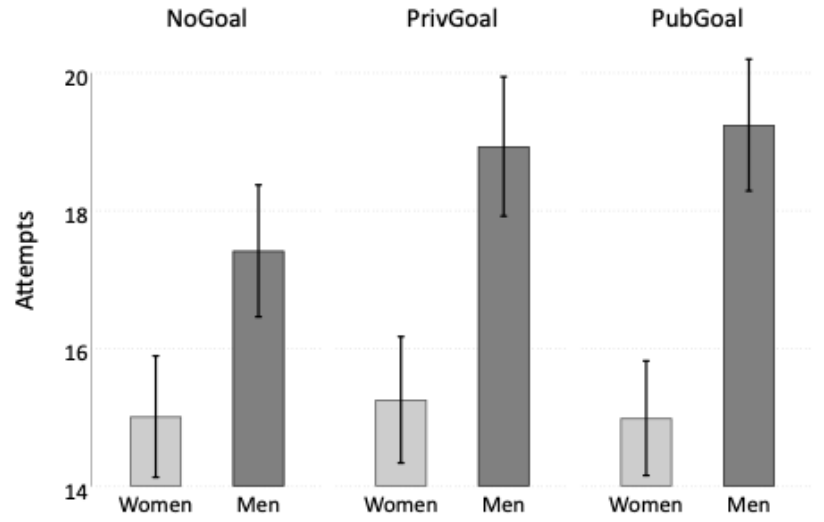


Figure A2. *Expected performance. Average expected performance by women and men in the treatments NoGoal, PrivGoal, and PubGoal. 90% confidence intervals are calculated with robust standard errors.*

A2. Tables: Expected Performance and Attempts

Table A1. Expected Performance: Gender Gap and Treatment Effects on Women, Men, and Gender Gap

	Expected Performance					
	(1) Pooled	(2) Pooled	(3) Field	(4) Field	(5) Classroom	(6) Classroom
Female	-4.125**	-2.784	-2.944	-2.098	-4.869	-3.454
	(2.091)	(2.142)	(3.021)	(3.337)	(3.054)	(2.874)
No Goal	-4.530***	-3.712**	-4.635**	-2.955	-4.358	-3.917
	(1.728)	(1.692)	(2.010)	(2.282)	(2.784)	(2.709)
Public Goal	3.721*	4.271*	6.887**	6.973**	0.00847	1.178
	(2.220)	(2.194)	(3.062)	(3.094)	(3.085)	(2.876)
Female*No Goal	4.432*	2.875	5.133	2.838	4.000	2.994
	(2.513)	(2.535)	(4.010)	(4.328)	(3.420)	(3.229)
Female*Public Goal	0.101	-0.961	-3.375	-3.207	3.316	1.767
	(3.295)	(3.416)	(4.790)	(5.015)	(4.550)	(4.437)
Constant	17.77***	46.81	18.05***	116.3**	17.49***	-31.85
	(1.495)	(33.58)	(1.624)	(54.06)	(2.514)	(40.86)
Controls	NO	YES	NO	YES	NO	YES
Observations	628	626	295	294	333	332
R-squared	0.041	0.063	0.055	0.082	0.026	0.049
Gender Gap						
No Goal	0.307	0.0917	2.189	0.740	-0.869	-0.459
[F-test p-value]	*[0.826]*	*[0.948]*	*[0.407]*	*[0.779]*	*[0.573]*	*[0.772]*
Private Goal	-4.125**	-2.784	-2.944	-2.098	-4.869	-3.454
[F-test p-value]	*[0.0489]*	*[0.194]*	*[0.331]*	*[0.530]*	*[0.112]*	*[0.230]*
Public Goal	-4.024	-3.745	-6.318*	-5.304	-1.553	-1.687
[F-test p-value]	*[0.115]*	*[0.138]*	*[0.0903]*	*[0.144]*	*[0.646]*	*[0.616]*
Effect of No (vs. Private) Goal						
Women	-0.0979	-0.837	0.498	-0.117	-0.358	-0.923
[F-test p-value]	*[0.957]*	*[0.654]*	*[0.886]*	*[0.973]*	*[0.857]*	*[0.638]*
Men	-4.530***	-3.712**	-4.635**	-2.955	-4.358	-3.917
[F-test p-value]	*[0.00897]*	*[0.0286]*	*[0.0218]*	*[0.196]*	*[0.118]*	*[0.149]*
Gender Gap	4.432*	2.875	5.133	2.838	4.000	2.994
[F-test p-value]	*[0.0783]*	*[0.257]*	*[0.202]*	*[0.513]*	*[0.243]*	*[0.354]*
Effect of Public (vs. Private) Goal						
Women	3.821	3.310	3.512	3.767	3.325	2.944
[F-test p-value]	*[0.117]*	*[0.188]*	*[0.341]*	*[0.328]*	*[0.321]*	*[0.379]*
Men	3.721*	4.271*	6.887**	6.973**	0.00847	1.178
[F-test p-value]	*[0.0942]*	*[0.0520]*	*[0.0253]*	*[0.0250]*	*[0.998]*	*[0.682]*
Gender Gap	0.101	-0.961	-3.375	-3.207	3.316	1.767
[F-test p-value]	*[0.976]*	*[0.779]*	*[0.482]*	*[0.523]*	*[0.467]*	*[0.691]*

Notes. *** p<0.01, ** p<0.05, * p<0.1. Ordinary Least Square Regression results with robust standard errors (in parentheses). The table shows post-estimation F-tests with corresponding p-values [in parentheses]. The dependent variable is the participant's expected performance in the matrix-task and the explanatory variables are a gender dummy (taking value 1 if female and 0 if male), a treatment dummy for NoGoal and PubGoal (taking value 1 if applies and 0 otherwise), and the interaction terms of the gender dummy with each treatment dummy. Controls (Risk Attitudes, Age, Age^2, and a dummy for feeling attached to the Dutch culture) are included in models (2), (4), and (6). The data come from treatments NoGoal, PrivGoal, and PubGoal and the samples are both experiments pooled in models (1) and (2), the field experiment in models (3) and (4), and the classroom experiment in models (5) and (6).

Table A2. Attempts: Gender Gap and Treatment Effects on Women, Men, and Gender Gap

	Attempts					
	(1) Pooled	(2) Pooled	(3) Field	(4) Field	(5) Classroom	(6) Classroom
Female	-3.677***	-3.526***	-2.545*	-2.735**	-4.144***	-3.916***
	(0.830)	(0.842)	(1.347)	(1.357)	(1.006)	(1.017)
No Goal	-1.516*	-1.515*	-0.422	-1.090	-1.748	-1.556
	(0.846)	(0.833)	(1.362)	(1.293)	(1.065)	(1.075)
Public Goal	0.311	0.181	0.550	0.194	-0.127	-0.173
	(0.847)	(0.836)	(1.264)	(1.247)	(1.066)	(1.070)
Female*No Goal	1.272	1.533	-1.013	-0.152	2.409*	2.449*
	(1.145)	(1.137)	(1.906)	(1.870)	(1.371)	(1.377)
Female*Public Goal	-0.577	-0.456	-2.270	-2.017	0.372	0.416
	(1.132)	(1.142)	(1.769)	(1.788)	(1.383)	(1.407)
Constant	18.93***	10.04	19.95***	-19.29	17.90***	32.40**
	(0.615)	(14.17)	(0.875)	(21.48)	(0.852)	(14.26)
Controls	NO	YES	NO	YES	NO	YES
Observations	635	626	302	294	333	332
R-squared	0.088	0.112	0.082	0.114	0.102	0.123
Gender Gap						
No Goal	-2.405***	-1.993**	-3.558***	-2.886**	-1.735*	-1.467
[F-test p-value]	*[0.00242]*	*[0.0114]*	*[0.00876]*	*[0.0317]*	*[0.0634]*	*[0.120]*
Private Goal	-3.677***	-3.526***	-2.545*	-2.735**	-4.144***	-3.916***
[F-test p-value]	*[0.000]*	*[0.000]*	*[0.0598]*	*[0.0448]*	*[0.000]*	*[0.000142]*
Public Goal	-4.254***	-3.982***	-4.815***	-4.752***	-3.771***	-3.500***
[F-test p-value]	*[0.000]*	*[0.000]*	*[0.000]*	*[0.000]*	*[0.000]*	*[0.000247]*
Effect of No (vs. Private) Goal						
Women	-0.244	0.0184	-1.436	-1.242	0.660	0.893
[F-test p-value]	*[0.752]*	*[0.981]*	*[0.282]*	*[0.353]*	*[0.445]*	*[0.301]*
Men	-1.516*	-1.515*	-0.422	-1.090	-1.748	-1.556
[F-test p-value]	*[0.0737]*	*[0.0696]*	*[0.757]*	*[0.400]*	*[0.102]*	*[0.149]*
Gender Gap	1.272	1.533	-1.013	-0.152	2.409*	2.449*
[F-test p-value]	*[0.267]*	*[0.178]*	*[0.595]*	*[0.935]*	*[0.0799]*	*[0.0762]*
Effect of Public (vs. Private) Goal						
Women	-0.266	-0.275	-1.720	-1.823	0.245	0.243
[F-test p-value]	*[0.723]*	*[0.720]*	*[0.166]*	*[0.160]*	*[0.781]*	*[0.785]*
Men	0.311	0.181	0.550	0.194	-0.127	-0.173
[F-test p-value]	*[0.714]*	*[0.828]*	*[0.664]*	*[0.876]*	*[0.905]*	*[0.872]*
Gender Gap	-0.577	-0.456	-2.270	-2.017	0.372	0.416
[F-test p-value]	*[0.610]*	*[0.689]*	*[0.200]*	*[0.260]*	*[0.788]*	*[0.768]*

Notes. *** p<0.01, ** p<0.05, * p<0.1. Ordinary Least Square Regression results with robust standard errors (in parentheses). The table shows post-estimation F-tests with corresponding p-values [in parentheses]. The dependent variable is the participant's number of attempts in the matrix-task and the explanatory variables are a gender dummy (taking value 1 if female and 0 if male), a treatment dummy for NoGoal and PubGoal (taking value 1 if applies and 0 otherwise), and the interaction terms of the gender dummy with each treatment dummy. Controls (Expected Performance, Risk Attitudes, Age, Age^2, and a dummy for feeling attached to the Dutch culture) are included in models (2), (4), and (6). The data come from treatments NoGoal, PrivGoal, and PubGoal and the samples are both experiments pooled in models (1) and (2), the field experiment in models (3) and (4), and the classroom experiment in models (5) and (6).

Appendix B. Instructions and Screenshots

Welcome to this study!

Your answers in this study will not be used in this course in any way. They will not affect your assessment in the course at all.

The study will take approximately 30 minutes. This study is part of a research project in social sciences that we - a group of professors from different universities - are conducting. Your engagement and attention when responding all parts of the study are very valuable for the success of this study, which will contribute to a better understanding of our society.

The study is divided into a task and a questionnaire. You are not allowed to communicate with anyone else until the study is over. For each part, you will receive instructions. We guarantee that everything we tell you in these instructions will happen exactly as described.

The answers to the questionnaire are entirely voluntary. The file with personal information will be password protected and saved on a secure university drive. It will be deleted after publication. We will immediately create two datasets: one dataset for feedback and one anonymized dataset. We will use only the anonymized and non-identifiable dataset for the analysis. We have no interest whatsoever in identifying an individual's decisions and answers or in sharing that information with a third party.

Your participation in this study is voluntary. You have the right to withdraw at any point during the study, for any reason, and without any prejudice.

By clicking the button below, you acknowledge that your participation in the study is voluntary and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

The research team:
Dr. Sabrine El Baroudi, Vrije Universiteit Amsterdam (s.elbaroudi@vu.nl)
Prof. Jordi Brandts, Instituto de Análisis Económico (CSIC) and Barcelona GSE (jordi.brandts@iae.csic.es)
Dr. Stefanie Huber, Universiteit van Amsterdam (s.j.huber@uva.nl)
Dr. Christina Rott, Vrije Universiteit Amsterdam (c.e.rott@vu.nl)

- ◯ I consent, begin the study
- ◯ I do not consent, I do not wish to participate

VU VRIJE UNIVERSITEIT AMSTERDAM | School of Business and Economics

Task instructions

You will now independently perform a task during 15 minutes.

This is an important task that is often used to measure people's talents. Many scientific studies have found that people who do well in a task like this are more successful in professional life than people who do less well. In a previous session, students like you performed the same task. Most of them gave between 9 and 17 correct answers.

The task is as follows. You will see two matrices on the computer screen. Each matrix has 10 rows and 10 columns and is filled with randomly generated numbers. Your job is to find the largest number in each of the two matrices and then to add them up. You are not allowed to use calculators, but you can use paper and pencil.

Example:

74	20	35	43	59	72	42	44	29	26
40	20	11	43	43	21	91	35	36	18
17	21	92	74	71	14	70	47	22	32
16	28	15	61	67	38	20	34	17	63
14	26	45	41	44	40	43	52	28	40
24	37	17	60	37	27	46	12	26	42
61	27	65	23	15	35	35	26	16	79
18	34	23	24	58	36	36	19	11	28
67	10	33	53	28	36	32	48	68	31
19	25	43	78	29	69	57	27	19	14

91	40	28	10	42	22	56	33	48	65
41	55	18	30	32	36	64	25	68	57
45	20	72	50	74	65	24	28	45	21
28	34	37	94	27	19	11	17	10	63
88	54	15	33	69	39	14	19	12	53
60	53	48	37	32	35	17	32	61	46
16	28	26	65	20	62	17	33	14	49
10	18	62	28	27	28	10	32	19	33
19	56	24	16	53	15	35	27	37	10
29	16	36	27	15	28	53	30	45	36

The largest number in the left matrix above is 92. The largest number in the right

matrix above is 94. The sum of the two numbers is 186 (= 92 + 94).

After entering the sum, the computer will tell you whether your answer is correct or incorrect (please note that the time will continue to run while you see this result). Subsequently, irrespective of whether your answer is correct or incorrect, a new pair of matrices will appear. This means that for each pair, you have only one attempt to provide the correct answer. However, there will always be a new pair of matrices as long as you are within the 15 minutes limit. The remaining time will be displayed on the shared screen. The objective is to provide as many correct answers as possible within the 15 minutes available.

Remember that studies have found that people who do well in a task like this are more successful in professional life. As mentioned earlier, in a previous session, students like you performed the same task. Most of them gave between 9 and 17 correct answers within the time limit of 15 minutes.

Feedback

The total number of your correct answers together with your name will be displayed on the shared screen after the study is finished. This information will be public, hence all participating students will see it.

At the end of the study, the results will be shown in a table of this form:

Student's First Name	Student's Last Name	Actual Performance (# of correct answers)
John	Summer	number
Lisa	Autumn	number

This information will NOT be recorded and will NOT be made available to the other participating students in any other way.

What is your first name?

What is your last name?

VU VRIJE UNIVERSITEIT AMSTERDAM | School of Business and Economics

How many questions do you EXPECT to answer correctly in the 15 minutes available?