



Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Gender differences in private and public goal setting

Jordi Brandts^a, Sabrine El Baroudi^b, Stefanie J. Huber^c, Christina Rott^{b,*}^a Instituto de Análisis Económico (CSIC) and Barcelona School of Economics, Spain^b Department of Management and Organisation, School of Business and Economics, Vrije Universiteit Amsterdam, Room 4A-64; Boelelaan 1105, Amsterdam 1081 HV, the Netherlands^c CeNDEF, University of Amsterdam and Tinbergen Institute, the Netherlands

ARTICLE INFO

Article history:

Received 14 January 2021

Revised 28 June 2021

Accepted 11 September 2021

Available online 29 October 2021

JEL classifications:

C91

J01

J16

J82

Keywords:

Gender differences

Goal setting

Public observability

Experiment

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. 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.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

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.² Why should we care about the underrepresentation of women in leading positions? The empirical literature provides initial evidence that benefits of gender equality in the workplace exist because of complementarities between the two genders.³

* Corresponding author.

E-mail address: c.e.rott@vu.nl (C. Rott).¹ <https://www.undp.org/content/undp/en/home/2030-agenda-for-sustainable-development/people/gender-equality.html> (link accessed on Dec 15, 2020).² As pointed out by the World Economic Forum (2018): "[...] while there are still relevant gender-biased labor 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 [...]"³ These benefits are especially important in organizations requiring high-skill workers. García-Meca et al. (2015) shows 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 (Dezso 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. On the end of supply-side factors, gender differences in the reaction to various aspects of competition have been studied in detail in a large experimental literature (see Niederle, 2016, for a survey of relevant studies).⁴

In this paper, we study gender differences in a novel, potentially important supply-side dimension of behavior: performance 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 often announce their goals for the society or the firm publicly; for instance, governments setting public targets for debts, unemployment, vaccination, or emission levels or companies declaring self-set targets for a more diverse workforce.⁵ Leaders are arguably more visible than lower-ranked employees. If females have (more) difficulties with this part of the job, this may be part of the explanation of why they apply less for leading positions (among other factors). With respect to the specific issue of gender-balance in organizations, public self-set goals are, compared to more disruptive interventions like quotas, a ‘softer’, less invasive, and potentially cheaper intervention. Public institutions like the European Central Bank and large private corporations opt ever more often for publicly announcing self-set goals for a more diverse workforce (and for committing to publish the corresponding data of whether the self-set goals are reached).⁶

Not only leaders, but individuals in general face many situations on the labor market and in private life, in which their performance is observable and identifiable and they may use self-set goals – private or public – as a commitment device. An example from the labor market is a researcher’s performance in terms of number and quality of publications. This information is publicly observable. To overcome self-control problems, s/he may set a goal upfront – either privately (that is, without telling anyone) or publicly (for instance, in a yearly assessment with the head of department or in front of colleagues). An example from the private environment is an individual’s intention to lose weight. If the intended weight loss is large enough it will be publicly observable. Also here, the individual may want to set a goal for him-/herself – either privately (that is, without telling anyone) or publicly (for instance, by sharing it with friends or family).

In this paper, we focus on the goal-setter in a non-strategic setup. A feature that all previously mentioned examples and many real-life situations have in common to a varying degree is the presence of strategic concerns. In a strategic situation, the goal-setter’s performance (and goal) and at least one other party’s decision(s) determine the outcome. The goal-setter is influenced by individual factors like confidence and also considers how others react to the own performance (and goal). To the best of our knowledge, we are the first to investigate the difference between the private versus public goal setting explicitly and its interaction with gender differences. As a first step, we implement a non-strategic setup in this paper to rule out beliefs about how others react to the own goal and performance.

In this paper, we shed light on this question, and investigate experimentally how women and men set goals for themselves and perform in a particular real-effort task, where performance is publicly observable and identifiable. We analyze behavior under two distinct goal-setting conditions: Private goals – that is, the self-set goal is only observable by oneself – and public goals – that is, the self-set goal is observable and identifiable by the public. 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. The primary outcome variables are participants’ self-set goals and performance.

Our results show that participants set significantly more ambitious goals when these are public. 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 Schram et al. (2019). 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 discussion and conclusion.

⁴ More recently, research has also addressed the role of public observability for gender differences in public speaking. 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).

⁵ Since 2016, German firms not falling under the gender quota for supervisory boards (that is, firms that are publicly listed or that are third codetermined and have at least 500 employees) need to set their own targets for the proportion of women in boards and the leading management levels below the executive board.

⁶ See for instance, <https://www.ft.com/content/0d1d2d4d-8bb8-42ce-b263-9863a1f377ed> for the goal set by the European Central Bank, <https://www.bloomberg.com/news/articles/2021-04-14/amazon-pledges-to-promote-more-women-black-employees> for Amazon’s diversity goals, or <https://www.cnbc.com/2020/06/09/gm-ceo-its-my-responsibility-to-drive-change-after-floyds-death.html> for GM’s diversity goals (links accessed on May 06, 2021).

2. Review of the literature on goal setting

In this section we briefly refer to some related research, starting with the general issue of goal-setting and moving then to gender differences and then to the distinction between private and public goals. 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, 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).⁷

Economists have studied possible mechanisms behind the effects of goal setting in several theoretical studies. Heath et al. (1999) propose the “goals as reference points” approach to explain the motivational process behind the achievement of goals. Their idea is that a reference point systematically alters the value of outcomes as described by the psychological principles of Prospect Theory (Kahneman and Tversky, 1979). The reference point divides outcomes into regions of gains and losses and the value function includes loss aversion in the loss domain. Because loss aversion implies that losses are more painful than gains of the same size are pleasurable, individuals will be more motivated to perform better after moving their reference point up through the goal. Wu et al. (2008) provide a formal model for Heath et al.’s (1999) assumptions and Koch and Nafziger (2011) show how self-set goals act as a commitment device to help overcome self-control problems. The key idea is that loss aversion makes low performance/non-compliance of goals psychologically painful and thereby motivates individuals to commit to their goal.

2.1. Self-set goals and gender differences

The psychology literature 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, 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).⁸ 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. (2020) examine whether self-set goals that are task-based or performance-based improve student performance.⁹ 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 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.¹⁰

The experimental economics literature offers potential explanations for such a gender difference in goal setting. First, there is 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., 2020a; Falk et 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

⁷ 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, 2007, 2019 for an overview of mediators and moderators).

⁸ A distinction is made between task-based goals and performance-based goals where in the former the level of task difficulty guides goal setting behavior; participants are asked to set goals to complete easy or challenging tasks. In the latter, the level of performance guides goal setting behavior; participants are asked to set a goal to achieve a certain performance which can be considered to be high or low (e.g. Kurman, 2001)

⁹ For performance-based goals, students were asked to set goals for their performance in the course. For task-based goals, students were asked to set goals for the number of online practice exams that they would complete in advance of each midterm exam and the final exam.

¹⁰ 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).

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.¹¹

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, 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% of the men and only 35% of the women selected the competitive tournament. The authors conclude that this gender difference is to a large extent driven by men being more confident about their performance than women. Möbius et al. (2011) 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. (2020b) 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, 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). 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, 2006).

2.2. 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, there are some studies on how men and women are differentially affected by various dimensions of the publicness of the environment in which they perform. 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 was only private ranking (as well as in the absence of any ranking). By contrast, inducing status ranking leads to gender differences in performance. 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.¹² The participants are 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).¹³

¹¹ In a repeated competitive experimental environment, Gill and Prowse (2014) shows that being “unlucky” in the past, reduces the work effort of women but not of men.

¹² Appendix B provides the instructions as displayed on the screen.

¹³ 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 was 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.

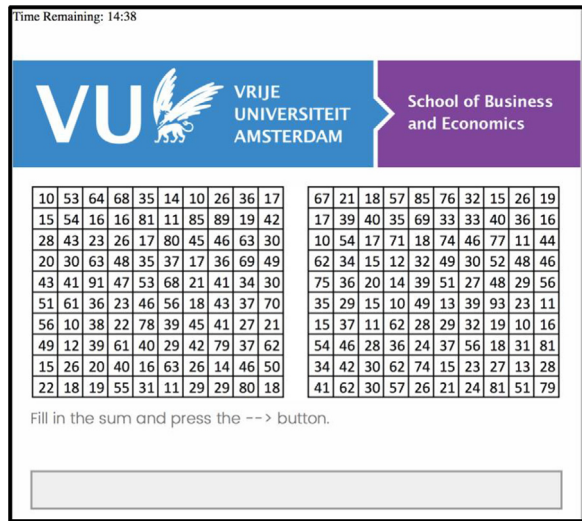


Fig. 1. Screenshot of the task.

Section 3.1 explains the 15 min 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.

3.1. Task

The task is identical to the one used in Schram et al. (2019) and before that in Weber and Schram (2017).¹⁴ Participants are presented with a sequence of pairs of 10 × 10 matrices filled with random two-digit numbers (Fig. 1).¹⁵

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, a new pair of matrices appears. This task stops after 15 min and participants can see the remaining time on the screen at the top left of the screen. We measure a participant’s performance by the number of accurate summations within the time limit of 15 min. The instructions inform the participants that there is no ceiling on their possible performance (and hence on their performance goals they can set for themselves). The instructions state: “(...) there will always be a new pair of matrices as long as you are within the 15 minutes limit”. We programmed the task such that we could be sure that nobody (even not the participant with the highest ability) could do the task and reach the limit.

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 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.¹⁶ 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 min), together with their name, will be displayed on a shared screen after the study is finished.

As in Clark et al. (2020), 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 it’s own sake without any external inducement (Pinder, 1984; Cerasoli et al., 2014), whereas extrinsically motivated behaviors are guided towards achieving some instrumental outcomes

¹⁴ The task choice is an important issue, which we do not study here. See Flory et al. (2015) and Günther et al. (2010).
¹⁵ A possible alternative would have been to use the summation task of Niederle and Vesterlund (2007). 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 studies have found no gender performance differences.
¹⁶ 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.” This statement is identical to the statement in Schram et al. (2019). The instructions continue: “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 included the statement about success in the task being correlated with professional success to keep our environment close to the one in Schram et al. (2019), where the same real-effort task and this statement was used. More importantly, given that we do not have monetary incentives, we thought that the statement would help participants to take the task seriously. We mention the prior performance to give the participants a broad idea of what their performance could be in a task with which they have no experience.

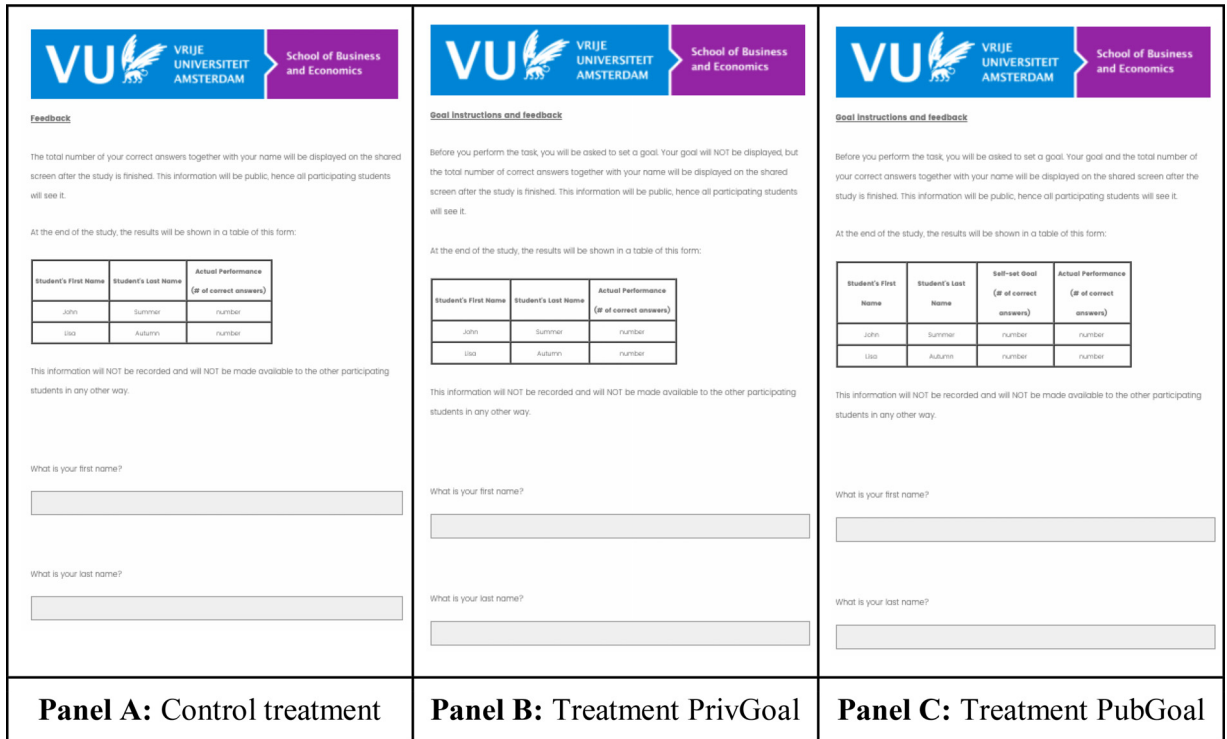


Fig. 2. Screenshots of treatment implementation. Panel A: screenshot for control treatment NoGoal. Panel B: screenshot for treatment PrivGoal. Panel C: screenshot for treatment PubGoal.

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.¹⁷

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). Performance is public information in all (control) treatments. In the control treatment, participants are not asked to set a goal, but in the two goal-setting treatments they are. 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. We thus implement a clean design in which only one feature changes at a time. Fig. 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.¹⁸

¹⁷ One could argue that we should incentivize performance (independently of the goal-setting). However, the orthogonality of goals and incentives has not been tested either in the economic or in the psychological literature, to the best of our knowledge. Hence, not incentivizing performance seems to us the appropriate design choice here to avoid possible interaction effects of goals and incentives.

¹⁸ Koch and Nafziger (2011) investigates men's use of self-set, non-binding goals as self-regulation tools and their effectiveness. The authors find that men use self-set goals as self-regulation tools. In our control treatment NoGoal, we cannot identify whether subjects have set "implicitly" a goal for themselves. Kaiser et al. (2021) highlight that this issue can be a concern "(...) the existing experimental studies that consider whether self-set goals are effective motivators suffer from a problem of treatment migration that arises from unobserved goal setting." Note that however our key research question is not about understanding whether self-set goals improve performance. Instead, we are mainly interested in (i) understanding the differences between the two goal-setting treatments "private goal (PrivGoal)" and "public goal (PubGoal)", and (ii) whether differential effects exist for men versus women.

Private Goal Setting Treatment (PrivGoal): Before performing the 15 min task, the participants are asked to set a goal (the number of correct answers). Participants know that their goal will be private before they are asked to set a goal. They are informed about the setup in the feedback (and goal-setting) instructions (see Fig. 2). On the subsequent screen, participants indicate their goal. 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?" On the decision screen next to this question, the instructions remind the participant that the 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 min task, the participants are asked to set a goal (the number of correct answers). Participants know that their goal will be public before they are asked to set a goal. They are informed about the setup in the feedback (and goal-setting) instructions (see Fig. 2). On the subsequent screen, participants indicate their goal. 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?" On the decision screen next to this question, the instructions remind the participant that the 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 min available. Note that this is different from their goal, which refers to how many questions participants *want* to answer correctly. We think that eliciting expected performance is important because it can point to a potential explanation of why women and men possibly react differently to the goal-setting environment.

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 enrolled in the course 'People in Business and Society' from the International Business Administration program. We used the Qualtrics software to program the experiment and the duration of the experiment was on average less than 30 min. The experiments in 2019 and 2020 differ in one key dimension. The experiment in 2019 was conducted during an in-person lecture, while the experiment in 2020 was conducted in an online-lecture.¹⁹ 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 one lecture of the first-year Bachelor course 'People in Business and Society' 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 (41%) were female and 178 (59%) male.²⁰ 69 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).²¹

Online classroom experiment in October 2020: The experiment was conducted during one 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 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 (43%) were female

¹⁹ The reason for this significant design difference (in-person lecture versus online lecture) was the COVID-19 outbreak. For transparency, we present our results for the pooled data and for each experiment separately. A minor additional difference between the two experiments concerns 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.

²⁰ 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 s (the matrix task alone takes 15 min = 900 s), 8 who worked on the survey for more than 2100 s (= 35 min; the average duration was 1499 s. Less than 2% of participants took more than 2100 s. 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) were considered as not working seriously on the task.

²¹ Since it was the first lecture of the course, some students were not yet enrolled in the course and not able to access the email with the Qualtrics link. They were given one of the treatment links during the lecture—which led to the unbalanced distribution across treatments. All participants started at the same time and we cannot distinguish who received the link by email or manually. Ex-ante, there is no reason to believe that these two groups differ systematically. Ex-post, we can compare socio-economic statistics across treatments and do not find significant differences in gender, age, and cultural identity across treatment (see tests in the beginning of Section 4). This gives us comfort in believing that there are not significant differences in characteristics of participants across experiments and treatments.

and 189 (57%) 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).²²

3.4. Research question and hypotheses

How do women and men set goals and perform on a real-effort task when the self-set goal is observable by the public (versus private)?²³ This primary research question and our main hypotheses 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-self-setting environments, we developed our hypotheses based on the findings in the two motivating research fields on self-set goals and gender differences.

The focus in this paper is on the comparison of public versus private goal setting. Starting with private goals, the main difference between treatments NoGoal and PrivGoal may be related to the goal setter's usage of the goal as commitment device and his/her perception of compliance with the goal. A way to conceptualize this is in terms of additional benefits and costs of setting goals. A private goal may entail a benefit (higher motivation and satisfaction) and a cost (disappointment in case of private failure). No (empirical or theoretical) studies exist that investigate whether private goal-settings differ for women compared to men. We conjecture that the net benefit might be lower for women—as several studies find that women attribute private failure mainly to their own ability while men tend to attribute it to bad luck (Hankin and Abramson, 2001; Boggiano and Barrett, 1991; Mezulis et al., 2004).

The main difference between treatments PrivGoal and PubGoal may be related to the goal setter's belief about and preference for how others *perceive* the goal and compliance with it.²⁴ When goal-setting is public, the individual may receive an additional benefit (for instance, looking confident, skillful, or ambitious) and an additional cost (for instance, beliefs about and utility from others judging failure to meet the self-set goal) related to public image concerns. Individuals will set higher public than private goals if the additional benefits are larger than the additional costs. Such benefits and costs may differ across genders. Public image concerns related to the appropriateness of ambitious goals are likely to affect men positively: The benefits from high public goals are likely to outweigh the costs of high public goals. For women, the impact of public image concerns is more ambiguous, since women are often expected to be more modest and less competitive than men. Hence, women's public image concerns can lead to lower goals in the public setting.

How does goal setting affect performance across treatments? The only difference between treatments NoGoal and PrivGoal are private self-set goals. As discussed in the literature review section, prospect theory provides a mechanism through which goals can improve performance (loss aversion transforms goals into commitment devices). In the private goal setting, goals can be seen as the expression of commitment devices. Combined with previous findings on the effect of self-set goals on performance (e.g., Van Lent and Souverijn, 2020), this leads us to expect a positive impact of private goals on women's and men's performance compared to no goals. Public goals can function as commitment devices just as private goals, with the additional effect of the visibility of goals referred to above.

Taking together these arguments, we can now formulate our hypotheses on the gender gap in goal-setting and performance across treatments. We define a gender gap as the absolute mean difference of a variable between women and men. 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 additional public image concerns in the public goal-setting condition PubGoal increase the gender differences as argued above. For the control treatment NoGoal we relate to the previous literature employing this task. Using the same task in a setting without goals, Schram et al. (2019) find no gender differences in performance under private ranking and without ranking and a gender gap under public ranking (women perform worse than men). In our case we do not provide any type of explicit public ranking. Also Weber and Schram (2017) do not find gender differences in performance in this task. We therefore conjecture no gender differences in the NoGoal condition. This expectation seemed to us a strong one and it is the one we use in Hypothesis 2. But one has to keep in mind that individual performance is public at the end of the experiment. This could create a feeling of public comparability, which may create an environment that is closer to one with public ranking in Schram et al. (2019).

²² 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 s (= 35 min; the average duration was 1499 s. Less than 2% of participants took more than 2100 s. 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.

²³ The study was pre-registered at AsPredicted.org (#48703, <https://aspredicted.org/blind.php?x=4t5by6>). The preregistration included this main research question and the main hypotheses presented at the end of this section.

²⁴ Note that, given that in our setting strategic considerations do not apply, the observer's perception has no consequence for the goal-setter other than these public image concerns (Bénabou and Tirole, 2006).

The following hypotheses summarize our conjectures about participants' behavior (as formulated in the pre-registration²⁵):

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).

4. Results

Before turning to the analysis of participants' self-set goals and their performance, we present some 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, $p = 0.0387$).

One might be concerned that compliance of the study is selective and differs across treatments and/or gender. In the field experiment, the dropout percentages are 39% (44 of 113) in NoGoal, 36% (54 of 151) in PrivGoal, and 36% (78 of 214) in PubGoal. In the online classroom experiment, the dropout percentages are 10% (12 of 125) in NoGoal, 11% (14 of 126) in PrivGoal, and 14% (18 of 126) in PubGoal.²⁶ The differences across treatments are not significant - neither in the field experiment (chi2 test; $p = 0.859$) nor in the online classroom experiment (chi2 test; $p = 0.498$). The gender distribution in the study is comparable to the gender distribution of the students enrolled in the course in 2019/20 and 2020/21: Both academic years pooled, 39% (394) of the students enrolled in the course were female and 61% (604) students were male - compared to 42% (268) female and 58% (367) male participants in the study. Also for each study year and experiment separately, the distributions are comparable.²⁷ This leads us to believe that there was no selective dropout by gender.

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 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.²⁸

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, $p = 0.9058$). This larger treatment effect in the field experiment can be explained by the observability and identifiability of goals being arguably higher in an in-person situation than in an online environment.

Fig. 3 shows the average goal set by women and men in the treatments PrivGoal and PubGoal. The corresponding Ordinary Least Square regression 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 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

²⁵ The study was pre-registered at AsPredicted.org (#48703, <https://aspredicted.org/blind.php?x=4t5by6>).

²⁶ Note that some students dropped out before seeing the treatment instructions, whom we do not consider here. Hence, we define a dropout as a student that started reading the instructions, but decided to not participate in the experiment.

²⁷ In the academic year 2019/20 and the field experiment, the corresponding distributions are 36% (165) female and 64% (292) male enrolled students versus 41% (124) female and 59% (178) male participants. In the academic year 2020/21 and the online classroom experiment, the corresponding distributions are 42% (229) female and 58% (312) male students versus 43% (144) female and 57% (189) male participants.

²⁸ 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.

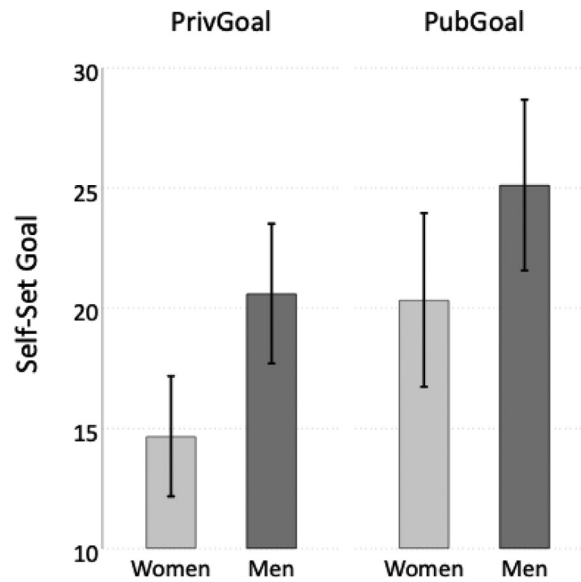


Fig. 3. Self-set goal. Average goal set by women and men in the treatments PrivGoal and PubGoal. 90% confidence intervals are calculated with robust standard errors.

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** (2.331)	-3.929 (2.473)	-5.922* (3.154)	-4.450 (3.563)	-5.837* (3.482)	-4.008 (3.487)
Public Goal	4.516 (2.795)	5.336* (2.788)	9.594** (4.037)	9.950** (4.105)	-1.414 (3.643)	0.0277 (3.398)
Female*Public Goal	1.154 (3.864)	-0.357 (4.065)	-2.677 (5.496)	-2.851 (5.844)	5.059 (5.326)	3.408 (5.206)
Constant	20.61*** (1.766)	47.91 (54.97)	20.95*** (2.139)	123.3 (81.75)	20.27*** (2.831)	-103.7 (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** [0.0113]	-3.929 [0.113]	-5.922* [0.0617]	-4.450 [0.213]	-5.837* [0.0951]	-4.008 [0.252]
Public Goal	-4.778 [0.122]	-4.286 [0.166]	-8.599* [0.0573]	-7.301 [0.103]	-0.778 [0.847]	-0.600 [0.885]
Effect of Public (vs. Private) Goal						
Women	5.670** [0.0342]	4.979* [0.0747]	6.917* [0.0649]	7.099* [0.0767]	3.645 [0.349]	3.436 [0.372]
Men	4.516 [0.107]	5.336* [0.0563]	9.594** [0.0183]	9.950** [0.0162]	-1.414 [0.698]	0.0277 [0.994]
Gender Gap	1.154 [0.765]	-0.357 [0.930]	-2.677 [0.627]	-2.851 [0.626]	5.059 [0.343]	3.408 [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, Age², 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).

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 mainly 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 (20.3 versus 18.9, MWU test, $p = 0.6128$; Table 1, model 5, $p = 0.6983$). Women increase their goals significantly at the 10% level in the field and insignificantly in the online classroom experiment.

Second, moving to the gender gap results pertinent to Hypothesis 1, 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, $p = 0.0951$).²⁹ This gender difference in goal setting is large: Men set a 40% higher goal than women.

Third, in the treatment PubGoal, men are more ambitious and set a higher goal than women. This gender difference in goal setting is substantial. Men 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 publicly 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. As already mentioned above, perhaps the higher degree of observability and identifiability in an in-person versus online environment can explain part of these differences.³⁰

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

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

Result 1a: Women set significantly higher goals when goals are public compared to private information and in particular in the field environment.

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

Result 1c: Men set significantly higher goals than women, both when goals are private or public.

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

Results 1a and 1b address the goal-setting part of the pre-registered primary research question and Results 1c and 1d speak directly to the pre-registered Hypothesis 1. Our data are consistent with the first prediction of Hypothesis 1 (Result 1c) but not with the second prediction (Result 1d). The above analysis of the goal-setting part of the primary research question and of the corresponding hypothesis was laid out in the pre-analysis plan in the pre-registration.

4.2. Performance

Before we turn to the second main 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'. The analysis of these factors was not explicitly pre-registered as part of the *main* research question, hypotheses, and analysis, but we believe that it is of interest. Recall that before performing the matrix task, participants are asked to indicate how many problems they *expect* to solve correctly.³¹ 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 (Fig. A1 and Table A1 for expected performance; Fig. A2 and Table A2 for attempts).

²⁹ 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).

³⁰ 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 basically the same (1.4% lower) goal in the online classroom experiment compared to the PrivGoal treatment. The choice of the self-set goal is more conscious than performance in the matrix-task, which might explain why differences across experiments are less pronounced for performance as discussed in Section 4.2.

³¹ Participants receive feedback about the correctness of their answer after each summation problem. We therefore elicit beliefs before the real effort task.

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, $p = 0.0964$; 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%) believe that they will at most achieve their self-set goal. 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 these 91.5% of participants, the goal setting environment does not significantly affect the distribution of participants confidence to meet their goal (chi2 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, $p < 0.0622$; [Table A2](#), model 1, $p = 0.0737$; non-reported difference in men's attempts in NoGoal versus PubGoal, [Table A2](#), model 1, $p = 0.0267$) women's attempts are literally unchanged (15.0 in NoGoal, 15.3 in PrivGoal, 15.0 in PubGoal; MWU tests, $p > 0.9420$; [Table A2](#), model 1, $p > 0.7521$). The different reaction to goal setting affects the gender gap in attempts across treatment slightly ([Table A2](#), model 5, $p = 0.0799$; non-reported difference in gender gap in NoGoal versus PubGoal, [Table A2](#), model 1, $p = 0.0942$).

With respect to performance, we make three observations. First, 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. In contrast, men's performance is not affected at all by the goal setting environment (MWU test, $p = 0.8960$; [Table 2](#), model 1, $p = 0.976$).

Second, 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, $p = 0.000253$). This result is at odds with our pre-registered Hypothesis 2. However, considering that performance is made public after the study, we can give an ex-post explanation of why men outperform women in NoGoal, which we have already mentioned above. The control condition NoGoal could be more comparable to the public ranking treatment than to the private or no ranking treatments in [Schram et al. \(2019\)](#). There, a third person can observe and compare participants' performance in public ranking. Though we do not provide any explicit (public) ranking, performance is publicly observable in our control treatment NoGoal. This might create a feeling of comparability and be closer to the setting with public than private or no ranking in [Schram et al. \(2019\)](#).³²

Third, we find a robust and strong gender gap in performance across treatments, illustrated in [Fig. 4](#) and analyzed with post-estimation F -tests in [Table 2](#).³³ 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, $p = 0.0647$). Our results can be summarized as follows:

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

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

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

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

³² Given our design, we cannot rule out alternative explanations for finding gender differences in performance in NoGoal. We have to take into account that the no-difference result stems from two laboratory studies, [Weber and Schram \(2017\)](#) and [Schram et al. \(2019\)](#), and that we are using a different subject pool in a field experiment and not a laboratory experiment. Some results in the literature suggest that men could outperform women because of (minor) differences in math-related skills ([Halpern et al., 2007](#)) or because of gender differences in visual perception performance ([Shaqiri et al., 2018](#)) and these tendencies may come through in our experiment, although they did not in [Weber and Schram \(2017\)](#) and [Schram et al. \(2019\)](#). Implicit goals that participants set for themselves without being asked to do so could also play a role (for instance, [Kaiser et al., 2021](#)).

³³ 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. \(2019\)](#).

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*** (0.654)	-1.983*** (0.658)	-1.550 (1.197)	-1.906 (1.181)	-2.092*** (0.656)	-1.802*** (0.655)
No Goal	-0.530 (0.671)	-0.602 (0.664)	-0.439 (1.246)	-1.223 (1.205)	-0.303 (0.707)	-0.0814 (0.702)
Public Goal	-0.0187 (0.632)	-0.174 (0.620)	-0.660 (1.016)	-1.057 (0.994)	0.628 (0.730)	0.630 (0.717)
Female*No Goal	-0.301 (0.904)	0.0320 (0.896)	-0.773 (1.669)	0.120 (1.608)	-0.206 (0.964)	-0.247 (0.963)
Female*Public Goal	-1.076 (0.888)	-0.932 (0.896)	-0.951 (1.501)	-0.683 (1.509)	-1.801* (0.972)	-1.869* (0.976)
Constant	11.87*** (0.479)	1.534 (11.84)	12.55*** (0.797)	-19.37 (16.86)	11.19*** (0.519)	26.68*** (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*** [0.000253]	-1.951*** [0.00154]	-2.323** [0.0468]	-1.786 [0.113]	-2.298*** [0.00127]	-2.049*** [0.00321]
Private Goal	-1.996*** [0.00237]	-1.983*** [0.00268]	-1.550 [0.196]	-1.906 [0.108]	-2.092*** [0.00156]	-1.802*** [0.00628]
Public Goal	-3.073*** [0.000]	-2.915*** [0.000]	-2.501*** [0.00605]	-2.589*** [0.00656]	-3.893*** [0.000]	-3.671*** [0.000]
Effect of No (vs. Private) Goal						
Women	-0.831 [0.171]	-0.570 [0.348]	-1.212 [0.276]	-1.102 [0.313]	-0.509 [0.438]	-0.329 [0.617]
Men	-0.530 [0.429]	-0.602 [0.365]	-0.439 [0.725]	-1.223 [0.311]	-0.303 [0.668]	-0.0814 [0.908]
Gender Gap	-0.301 [0.739]	0.0320 [0.972]	-0.773 [0.644]	0.120 [0.940]	-0.206 [0.831]	-0.247 [0.798]
Effect of Public (vs. Private) Goal						
Women	-1.095* [0.0801]	-1.107* [0.0867]	-1.611 [0.146]	-1.740 [0.130]	-1.173* [0.0685]	-1.239* [0.0623]
Men	-0.0187 [0.976]	-0.174 [0.779]	-0.660 [0.517]	-1.057 [0.288]	0.628 [0.390]	0.630 [0.380]
Gender Gap	-1.076 [0.226]	-0.932 [0.298]	-0.951 [0.527]	-0.683 [0.651]	-1.801* [0.0647]	-1.869* [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, Age², 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).

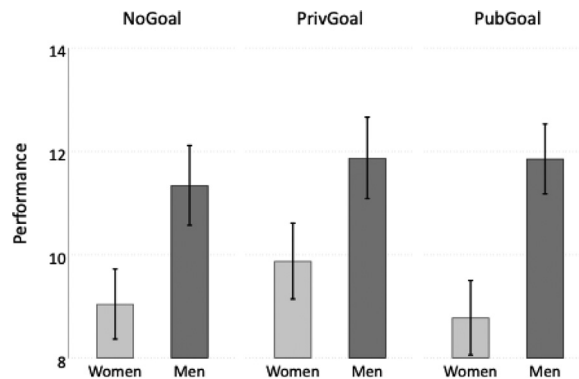


Fig. 4. Performance. Average performance by women and men in the treatments NoGoal, PrivGoal, and PubGoal. 90% confidence intervals are calculated with robust standard errors.

Results 2a and 2b address the performance part of the pre-registered primary research question and Results 2c and 2d speak directly to the pre-registered Hypothesis 2. Our data are overall not consistent with Hypothesis 2. Though we find a significant gender gap in performance in PrivGoal and PubGoal as hypothesized, we also find a gender gap in performance in NoGoal (Result 2c). Though the gender gap changes in the expected direction when moving from PrivGoal to PubGoal, the gender gap does not vary statistically significantly across treatments overall (Result 2d). The above analysis of performance was laid out in the pre-analysis plan in the pre-registration.

A potential explanation for the absence of a significant effect of goals on performance and on the gender performance gap could be that the marginal effect of the accumulated commitment devices is too small to affect performance significantly as we move from NoGoal (public performance) to PrivGoal (public performance and private *self-set goal*) to PubGoal (public performance and *public self-set goal*). Note however, that men do try to solve significantly more matrices after setting a goal, which affects the gender gap in attempts in the expected direction. Also an emerging gender gap in goal-compliance confidence in PubGoal (compared to PrivGoal) provides supporting evidence for our hypothesis. Attempts and expected performance/ goal-compliance confidence had not been pre-registered as part of the *main* research questions, hypotheses, and analysis.

To link the findings of goal setting and performance, we 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. Or put differently, one may want to take publicly announced goals with a grain of salt. This result speaks directly to our pre-registered primary research question and indicates that the key difference between the public versus private goal-setting environment (i.e., social image concerns) leads to larger differences between goals and actual performance – for both, men and women. This result provides preliminary evidence that the perceived benefit of social image concerns outweighs the perceived associated costs (i.e., belief about the perception of failure when not meeting the self-set goal). And second, the difference between these perceived benefits and costs are larger for men than women. A more in-depth analysis of these differences is beyond the scope of this paper.

5. Discussion and conclusion

We conduct a field and an online classroom experiment to study whether men and women set different goals and perform differently in the absence of goals, when information about goals is private and when it is public. Our main findings point to the following: (1) Women and men set higher goals when they are publicly observable and identifiable than when they are private. This effect is particularly observable in the field experiment. (2) Women choose both lower private and public goals than men. (3) Men perform better than women under private and public goals as well as in the absence of goal setting. (4) Women's performance does not change between no goals and private goals, but worsens between public and private goals, whereas men's performance is unaffected by the goal-setting condition. As a consequence of differential goal setting and performance we find that, 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. In what follows we connect our results to some relevant literature.

The difference in behavior between the public and the private goal-setting environments is consistent with several explanations. One is that it is the result of public image concerns (Bénabou and Tirole, 2006). Such social image concerns can be thought of in terms of a benefit (belief of being *perceived* as ambitious/high ability) and a cost (beliefs of how failure is *perceived*). Our result shows that both men and women set more ambitious goals when these are public which suggests that the benefits resulting from social image concerns are larger than the costs.

Social conformity is a related, but distinct concept that could also explain our result that men and women set more ambitious goals in the public treatment. Social conformity is considered a powerful social phenomenon that encourages individuals to adapt their opinions and behaviors 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 in the public condition; they could have increased their goals because they expected that the majority would do so.³⁴

We find that women choose lower goals than men in both treatments. In terms of a benefit-cost explanation, women may perceive the net benefits (benefits minus costs) to be lower than men. Goal setting theory posits that self-efficacy and self-confidence that the goal for a specific task is attainable are important individual characteristic for self-set goals (Bandura, 1997; Latham and Locke, 2007, 2006). It could thus be that women set lower goals than men, because they are likely less confident about their competences (e.g., Beyer, 1990; Lundeberg et al., 1994; Beyer and Bowden, 1997) than their confident male peers (McCarty, 1986; Wood and Karten, 1986) as suggested in the old stream of psychology literature and also more recently in the experimental economics literature showing that men are more confident about their abilities than women (Buser et al., 2020b). Our finding that women set lower goals than men could also be explained by gender

³⁴ 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.

differences in the attribution of success and failure to internal factors (personal abilities and skills) and external factors (for instance, luck). Some evidence suggests that young boys show a stronger self-serving attributional bias than young 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 women set lower goals than men in both treatments.

Men perform better across all treatments. Surprisingly, for both men and women, performance does not increase significantly with the introduction of privately self-set goals. Hence, we do not find necessarily evidence that participants are using goals as successful commitment devices. Further, this result contradicts findings from experimental and non-experimental studies testing the goal setting theory that have shown that setting goals is better for performance than not setting any goals (for literature reviews see Locke, 1996; Locke and Latham 2002, 2006, 2007). Our environment differs from the aforementioned studies by having performance being public and identifiable in all treatments. Hence, a potential ex-post explanation of our result could be that, even in the no-goal condition, individuals feel committed to performing well because their performance is publicly observable. The marginal effect of an additional commitment device in the form of a private or public self-set goal could be too small to affect performance significantly. An interesting observation in this context is however that men attempt to solve significantly more problems when they set a goal privately or publicly compared to no goal setting while women’s attempts are literally unchanged across treatments.

Finally, whereas both women’s and men’s performance is unaffected by private self-set goals, public self-set goals negatively affect women’s performance and do not affect men’s performance. Our finding that women perform worse with public than with private goals, while men are unaffected is in line with other experimental studies which show that women underperform in competitive environments (Gneezy et al., 2003; Gneezy and Rustichini, 2004). We do not have enough information to identify the mechanisms behind this result. However, we find that women’s goal-compliance confidence declines, whereas men’s increases resulting in a significant gender gap.

Several important and interesting research questions emerge from our study. Can an interpretation of gender goal-setting differences in terms of gendered (perceived) benefits and costs be backed up by data? To which factors do women and men attribute (un-) successfully met self-set goals? Does failing to achieve a publicly (self-)set goal result in gender differences in negative consequences such as reputational damages and stress in competitive environments? It would be 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). In this context, it would be interesting to analyze how individuals set goals and perform when they are held accountable. Additionally, public observability and the public perception might affect self-selection into public versus private goal setting or (leading) positions that entail public goal announcements. 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).

Acknowledgment

Brandts thanks the Spanish *Ministerio de Ciencia e Innovación* (Grant number PID2020-114251GB-I00), the Spanish *Agencia Estatal de Investigación* (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S) and the *Generalitat de Catalunya* (Grant: 2017 SGR 1136) for financial support.

Appendix A. Additional results

A1. Figures: Expected Performance and Attempts

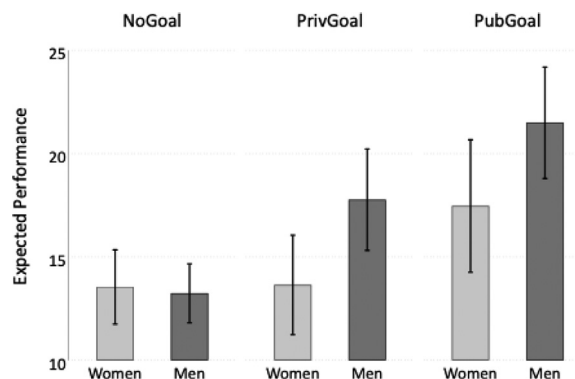


Fig. A1. Expected performance. Average expected performance by women and men in the treatments NoGoal, PrivGoal, and PubGoal. 90% confidence intervals are calculated with robust standard errors.

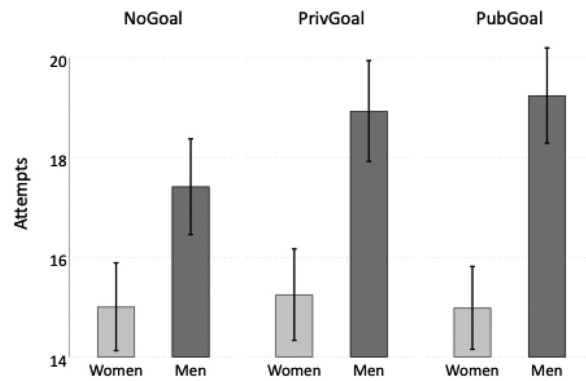


Fig. A2. Attempts. Average attempts by women and men in the treatments NoGoal, PrivGoal, and PubGoal. 90% confidence intervals are calculated with robust standard errors.

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** (2.091)	-2.784 (2.142)	-2.944 (3.021)	-2.098 (3.337)	-4.869 (3.054)	-3.454 (2.874)
No Goal	-4.530*** (1.728)	-3.712** (1.692)	-4.635** (2.010)	-2.955 (2.282)	-4.358 (2.784)	-3.917 (2.709)
Public Goal	3.721* (2.220)	4.271* (2.194)	6.887** (3.062)	6.973** (3.094)	0.00847 (3.085)	1.178 (2.876)
Female*No Goal	4.432* (2.513)	2.875 (2.535)	5.133 (4.010)	2.838 (4.328)	4.000 (3.420)	2.994 (3.229)
Female*Public Goal	0.101 (3.295)	-0.961 (3.416)	-3.375 (4.790)	-3.207 (5.015)	3.316 (4.550)	1.767 (4.437)
Constant	17.77*** (1.495)	46.81 (33.58)	18.05*** (1.624)	116.3** (54.06)	17.49*** (2.514)	-31.85 (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, Age², 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*** (0.830)	-3.526*** (0.842)	-2.545* (1.347)	-2.735** (1.357)	-4.144*** (1.006)	-3.916*** (1.017)
No Goal	-1.516* (0.846)	-1.515* (0.833)	-0.422 (1.362)	-1.090 (1.293)	-1.748 (1.065)	-1.556 (1.075)
Public Goal	0.311 (0.847)	0.181 (0.836)	0.550 (1.264)	0.194 (1.247)	-0.127 (1.066)	-0.173 (1.070)
Female*No Goal	1.272 (1.145)	1.533 (1.137)	-1.013 (1.906)	-0.152 (1.870)	2.409* (1.371)	2.449* (1.377)
Female*Public Goal	-0.577 (1.132)	-0.456 (1.142)	-2.270 (1.769)	-2.017 (1.788)	0.372 (1.383)	0.416 (1.407)
Constant	18.93*** (0.615)	10.04 (14.17)	19.95*** (0.875)	-19.29 (21.48)	17.90*** (0.852)	32.40** (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*** [0.00242]	-1.993** [0.0114]	-3.558*** [0.00876]	-2.886** [0.0317]	-1.735* [0.0634]	-1.467 [0.120]
Private Goal	-3.677*** [0.000]	-3.526*** [0.000]	-2.545* [0.0598]	-2.735** [0.0448]	-4.144*** [0.000]	-3.916*** [0.000142]
Public Goal	-4.254*** [0.000]	-3.982*** [0.000]	-4.815*** [0.000]	-4.752*** [0.000]	-3.771*** [0.000]	-3.500*** [0.000247]
Effect of No (vs. Private) Goal						
Women	-0.244 [0.752]	0.0184 [0.981]	-1.436 [0.282]	-1.242 [0.353]	0.660 [0.445]	0.893 [0.301]
Men	-1.516* [0.0737]	-1.515* [0.0696]	-0.422 [0.757]	-1.090 [0.400]	-1.748 [0.102]	-1.556 [0.149]
Gender Gap	1.272 [0.267]	1.533 [0.178]	-1.013 [0.595]	-0.152 [0.935]	2.409* [0.0799]	2.449* [0.0762]
Effect of Public (vs. Private) Goal						
Women	-0.266 [0.723]	-0.275 [0.720]	-1.720 [0.166]	-1.823 [0.160]	0.245 [0.781]	0.243 [0.785]
Men	0.311 [0.714]	0.181 [0.828]	0.550 [0.664]	0.194 [0.876]	-0.127 [0.905]	-0.173 [0.872]
Gender Gap	-0.577 [0.610]	-0.456 [0.689]	-2.270 [0.200]	-2.017 [0.260]	0.372 [0.788]	0.416 [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, Age², 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 Screen: Identical across treatments



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. Sabine 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

Task Instruction Screen: Identical across treatments



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 Instruction Screen: NoGoal Treatment



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
Isla	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?

Powered by Qualtrics X

Feedback Instruction Screen: PubGoal Treatment



Goal instructions and feedback

Before you perform the task, you will be asked to set a goal. Your goal and 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	Self-set Goal (# of correct answers)	Actual Performance (# of correct answers)
John	Summer	number	number
lisa	Autumn	number	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?

Feedback Instruction Screen: PrivGoal Treatment



Goal instructions and feedback

Before you perform the task, you will be asked to set a goal. Your goal will NOT be displayed, but the total number of 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?

Goal-setting Decision Screen: PubGoal Treatment



What is your self-set goal – How many questions do you WANT to answer correctly in the 15 minutes available?

Recall: Your goal and the number of correct answers together with your name will be displayed on the shared screen after the study is finished (= will be shared with everyone participating).

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	Self-set Goal (# of correct answers)	Actual Performance (# of correct answers)
John	Summer	number	number
Lisa	Autumn	number	number



Goal-setting Decision Screen: PrivGoal Treatment



What is your self-set goal – How many questions do you WANT to answer correctly in the 15 minutes available?

Recall: Your goal will NOT be displayed on the shared screen after the study is finished.

Only the number of correct answers together with your name will be displayed (= will be shared with everyone participating).

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



Expectation Decision Screen: Identical across all treatments



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

Powered by Qualtrics A

References

- Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N., 2009. Large stakes and big mistakes. *Rev. Econ. Stud.* 76 (2), 451–469.
- Asch, S., Guetzkow, H., 1951. Effects of group pressure upon the modification and distortion of judgment. *Groups, Leadership, and Men*. Carnegie, Pittsburgh.
- Bandura, A., 1997. *Self-Efficacy: The Exercise of Control*. W.H. Freeman, Stanford.
- Behnke, R.R., Sawyer, C.R., 2001. Patterns of psychological state anxiety in public speaking as a function of anxiety sensitivity. *Commun. Q.* 49 (1), 84–94.
- Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. *Am. Econ. Behav.* 96 (5), 1652–1678.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57 (1), 289–300.
- Beyer, S., 1990. Gender differences in the accuracy of self-evaluations of performance. *J. Personal. Soc. Psychol.* 59, 960–970. doi:10.1037/0022-3514.59.5.960.
- Beyer, S., Bowden, E.M., 1997. Gender differences in self-perceptions: convergent evidence from three measures of accuracy and bias. *Personal. Soc. Psychol. Bull.* 23 (2), 157–172. doi:10.1177/0146167297232005.
- Boggiano, A.K., Barrett, M., 1991. Gender differences in depression in college students. *Sex Roles* 25, 595–605.
- Brody, L.R., 1993. On understanding gender differences in the expression of emotion. In: Ablon, S.L., Brown, D., Khantzian, E.J., Mack, J.E. (Eds.), *Human Feelings: Explorations in Affect Development and Meaning*. Analytic Press, Hillsdale, N.J., pp. 87–121.
- Buser, T. and Yuan, H. (2020). Public speaking aversion, working paper. Tinbergen Institute Discussion Paper No. TI 2020-074/I
- Buser, T., Yuan, H., 2019. Do women give up competing more easily? Evidence from the lab and the dutch math olympiad. *Am. Econ. J. Appl. Econ.* 11 (3), 225–252.
- Buser, T., Niederle, M., Oosterbeek, H., 2020a. Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures. NBER Working Paper No. 28916 doi:10.3386/w28916.
- Buser, T., Grimalda, G., Putterman, L., Van Der Weele, J., 2020b. Overconfidence and gender gaps in redistributive preferences: cross-country experimental evidence. *J. Econ. Behav. Organ.* 187, 267–286. doi:10.1016/j.jebo.2020.07.005 0167-2681/.
- Carter, A.J., Croft, A., Lukas, D., 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., Ford, M.T., 2014. Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. *Psychol. Bull.* 140 (4), 980–1008. doi:10.1037/a0035661.
- 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 No. 16/50, Available at SSRN: <https://ssrn.com/abstract=2759759>
- Clark, D., Gill, D., Prowse, V., Rush, M., 2020. Using goals to motivate college students: theory and evidence from field experiments. *Review of Economics and Statistics* 102 (4), 648–663.
- Cooper, D.J., Fang, H., 2008. Understanding overbidding in second price auctions: an experimental study. *Econ. J.* 118 (532), 1572–1595. doi:10.1111/j.1468-0297.2008.02181.x.
- Crosan, R., Gneezy, U., 2009. Gender differences in preferences. *J. Econ. Lit.* 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. 10.2139/ssrn.2583872
- Delgado, M.R., Schotter, A., Ozbay, E.Y., 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., Van Ginkel, W.P., 2009. Challenging experiences: gender differences in task choice. *J. Manag. Psychol.* 24, 4–28. doi:10.1108/02683940910922519.
- Deutsch, M., Gerard, H.B., 1955. A study of normative and informational social influences upon individual judgment. *J. Abnorm. Soc. Psychol.* 51, 629–636. doi:10.1037/h0046408.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* 9 (3), 522–550. doi:10.1111/j.1542-4774.2011.01015.x.
- Dezső, C.L., Ross, D.G., 2012. Does female representation in top management improve firm performance? A panel data investigation. *Strateg. Manag. J.* 33, 1072–1089.
- Erez, M., Gopher, D., Arzi, N., 1990. Effects of goal difficulty, self-set goals, and monetary rewards on dual task performance. *Organ. Behav. Hum. Decis. Process.* 47, 247–269. doi:10.1016/0749-5978(90)90038-B.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018. Global evidence on economic preferences. *Q. J. Econ.* 133 (4), 1645–1692. doi:10.1093/qje/qjy013.
- Flory, J.A., Leibbrandt, A., List, J.A., 2015. Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions. *Rev. Econ. Stud.* 82 (1), 122–155. doi:10.1093/restud/rdu030.
- Fujita, F., Diener, E., Sandvik, E., 1991. Gender differences in negative affect and well-being: the case for emotional intensity. *J. Personal. Soc. Psychol.* 61 (3), 427–434. doi:10.1037/0022-3514.61.3.427.
- García-Meca, E., García-Sánchez, I.M., Martínez-Ferrero, J., 2015. Board diversity and its effects on bank performance: an international analysis. *J. Bank. Financ.* 53, 202–214.
- Gill, D., Prowse, V., 2014. Gender differences and dynamics in competition: the role of luck. *Quant. Econ.* 5, 351–376.

- Gino, F., Wilmoth, C.A., Brooks, A.W., 2015. Compared to men, women view professional advancement as equally attainable, but less desirable. *Proc. Natl. Acad. Sci. U. S. A.* 112, 12354–12359. doi:[10.1073/pnas.1502567112](https://doi.org/10.1073/pnas.1502567112).
- Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. *Q. J. Econ.* 118, 1049–1074. doi:[10.1162/00335530360698496](https://doi.org/10.1162/00335530360698496).
- Gneezy, U., Rustichini, A., 2004. Gender and competition at a young age. *Am. Econ. Rev.* 4, 377–381. doi:[10.1257/0002828041301821](https://doi.org/10.1257/0002828041301821).
- Günther, C., Ekinçi, N.A., Schwieren, C., Strobel, M., 2010. Women can't jump? An experiment on competitive attitudes and stereotype threat. *J. Econ. Behav. Organ.* 75, 395–401. doi:[10.1016/j.jebo.2010.05.003](https://doi.org/10.1016/j.jebo.2010.05.003).
- Hankin, B.L., Abramson, L.Y., 2001. Development of gender differences in depression: an elaborated cognitive vulnerability-transactional stress theory. *Psychol. Bull.* 127, 773–796.
- Halpern, D.F., Benbow, C.P., Geary, D.C., Gur, R.C., Hyde, J.S., Gernsbacher, M.A., 2007. The science of sex differences in science and mathematics. *Psychol. Sci. Public Interest* 8 (1), 1–51.
- Heath, C., Larrick, R.P., Wu, G., 1999. Goals as reference points. *Cogn. Psychol.* 38, 79–109.
- Hinsley, A., Sutherland, W.J., Johnston, A., 2017. Men ask more questions than women at a scientific conference. *PLoS One* 12 (10), 1–14.
- Hollenbeck, J.R., Williams, C.R., Klein, H.J., 1989. An empirical examination of the antecedents of commitment to difficult goals. *J. Appl. Psychol.* 74. doi:[10.1037/0021-9010.74.1.18](https://doi.org/10.1037/0021-9010.74.1.18).
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi:[10.1142/9789814417358_0006](https://doi.org/10.1142/9789814417358_0006).
- Kaiser, J.P., Koch, A. and Nafziger, J. (2021). Self-set goals are effective regulation tools – despite goal revision. CEPR Discussion Paper No. DP15716, Available at SSRN: <https://ssrn.com/abstract=3783942>.
- Koch, A.K., Nafziger, J., 2011. Self-regulation through goal setting. *Scand. J. Econ.* 113 (1), 212–227. doi:[10.1111/j.1467-9442.2010.01641.x](https://doi.org/10.1111/j.1467-9442.2010.01641.x).
- Kurman, J., 2001. Self-regulation strategies in achievement settings: culture and gender differences. *J. Cross Cult. Psychol.* 32 (4), 491–503. doi:[10.1177/0022022101032004008](https://doi.org/10.1177/0022022101032004008).
- Latham, G.P., Locke, E.A., 2007. New developments in and directions for goal-setting research. *Eur. Psychol.* 12, 290–300. doi:[10.1027/1016-9040.12.4.290](https://doi.org/10.1027/1016-9040.12.4.290).
- Latham, G.P., Marshall, H.A., 1982. The effects of self-set, participatively set, and assigned goals on the performance of government employees. *Pers. Psychol.* 35, 399–404. doi:[10.1111/j.1744-6570.1982.tb02204.x](https://doi.org/10.1111/j.1744-6570.1982.tb02204.x).
- Levy, P., Baumgardner, A., 1991. Effects of self-esteem and gender on goal choice. *J. Organ. Behav.* 12, 529–541. doi:[10.1002/job.4030120606](https://doi.org/10.1002/job.4030120606).
- Locke, E.A., 1996. Motivation through conscious goal setting. *Appl. Prev. Psychol.* 5 (2), 117–124. doi:[10.1016/S0962-1849\(96\)80005-9](https://doi.org/10.1016/S0962-1849(96)80005-9).
- Locke, E.A., Latham, G.P., 1990. *A Theory of Goal Setting and Task Performance*. Prentice-Hall, Englewood Cliffs, NJ.
- Locke, E.A., Latham, G.P., 2002. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *Am. Psychol.* 57 (9), 705–717. doi:[10.1037/0003-066X.57.9.705](https://doi.org/10.1037/0003-066X.57.9.705).
- Locke, E.A., Latham, G.P., 2006. New directions in goal-setting theory. *Curr. Dir. Psychol. Sci.* 15 (5), 265–268. doi:[10.1111/j.1467-8721.2006.00449.x](https://doi.org/10.1111/j.1467-8721.2006.00449.x).
- Locke, E.A., Latham, G.P., 2019. The development of goal setting theory: a half century retrospective. *Motiv. Sci.* doi:[10.1037/mot0000127](https://doi.org/10.1037/mot0000127).
- Lundeberg, M.A., Fox, P.W., Puncochar, J., 1994. Highly confident but wrong: gender differences and similarities in confidence judgments. *J. Educ. Psychol.* 86 (1), 114–121. doi:[10.1037/0022-0663.86.1.114](https://doi.org/10.1037/0022-0663.86.1.114).
- Marinho, A.C.F., De Medeiros, A.M., Cortes Gama, A.C., Caldas Teixeira, L., 2017. Fear of public speaking: perception of college students and correlates. *J. Voice* 31 (1), 127–134.
- Malik, M.A.R., Butt, A.N., Choi, J.N., 2015. Rewards and employee creative performance: moderating effects of creative self-efficacy, reward importance, and locus of control. *J. Organ. Behav.* 36, 59–74. doi:[10.1002/job.1943](https://doi.org/10.1002/job.1943), 10.2753/IJS0020-7659380103.
- McCarty, P.A., 1986. Effects of feedback on the self-confidence of men and women. *Acad. Manag. J.* 29, 840–847. doi:[10.5465/255950](https://doi.org/10.5465/255950).
- Mezulis, A.H., Abramson, L.Y., Hyde, J.S., 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. *Psychol. Bull.* 130 (5), 711.
- Möbius, M., Niederle, M., Niehaus, P., Rosenblat, T., 2011. Managing self-confidence: theory and experimental evidence. NBER Working paper, 17014 doi:[10.3386/w17014](https://doi.org/10.3386/w17014).
- Niederle, M., 2016. Gender. In: Kagel, J., Roth, A.E. (Eds.), *Handbook of Experimental Economics*. Princeton University Press, pp. 481–553.
- Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? *Q. J. Econ.* 122 (3), 1067–1101. doi:[10.1162/qjec.122.3.1067](https://doi.org/10.1162/qjec.122.3.1067).
- Parthasarathy, R., Rao, V., Palaniswamy, N., 2019. Deliberative democracy in an unequal world: a text-as-data study of South India's village assemblies. *Am. Political Sci. Rev.* 113 (3), 623–640.
- Pinder, C.C., 1984. *Work Motivation: Theory, Issues and Applications*. Scott, Foresman, Glenview, IL.
- Sahay, M.R., Cihak, M.M., 2018. Women in finance: a case for closing gaps. *International Monetary Fund* 2018 (005) ISBN 9781484375907.
- Schram, A., Brandts, J., Gërçhani, K., 2019. Social-status ranking: a hidden channel to gender inequality under competition. *Exp. Econ.* 1–23. doi:[10.1007/s10683-018-9563-6](https://doi.org/10.1007/s10683-018-9563-6).
- Shaqiri, A., Roinishvili, M., Grzeczowski, L., Chkonia, E., Pilz, K., Mohr, C., Brand, A., Kunchulia, M., Herzog, M.H., 2018. Sex-related differences in vision are heterogeneous. *Sci. Rep.* 8 (1), 1–10.
- Shurchkov, O., 2012. Under pressure: gender differences in output quality and quantity under competition and time constraints. *J. Eur. Econ. Assoc.* 10, 1189–1213.
- Stein, M.B., Walker, J.R., Forde, D.R., 1996. Public-speaking fears in a community sample: prevalence, impact on functioning, and diagnostic classification. *Arch. Gen. Psychiatry* 53 (2), 169–174.
- Stipek, D.J., Gralinski, J.H., 1991. Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *J. Educ. Psychol.* 83 (3), 361.
- Szekeress, H., Shuman, E., Saguy, T., 2020. Views of sexual assault following #MeToo: the role of gender and individual differences. *Personal. Individ. Differ.* 166 (1), 1–6. doi:[10.1016/j.paid.2020.110203](https://doi.org/10.1016/j.paid.2020.110203).
- Van Lent, M., Souverein, M., 2020. Goal setting and raising the bar: a field experiment. *J. Behav. Exp. Econ.* 87, 101570.
- Turk, C.L., Heimberg, R.G., Orsillo, S.M., Holt, C.S., Gitow, A., Street, L.L., Schneier, F.R., Liebowitz, M.R., 1998. An investigation of gender differences in social phobia. *J. Anxiety Disord.* 12 (3), 209–223.
- Weber, M., Schram, A., 2017. The non-equivalence of labor market taxes: a real-effort experiment. *Econ. J.* 127, 2187–2215.
- Welsh, D.T., Baer, M.D., Sessions, H., 2020. Hot pursuit: the affective consequences of organization-set versus self-set goals for emotional exhaustion and citizenship behavior. *J. Appl. Psychol.* 105 (2), 166–185. doi:[10.1037/apl0000429](https://doi.org/10.1037/apl0000429).
- World Economic Forum, 2018. *The Global Gender Gap Report*. World Economic Forum.
- Wood, W., Karten, S.J., 1986. Perceived sex differences in competence as a determinant of sex differences in interaction. *J. Personal. Soc. Psychol.* 50, 341–347.
- Wijenayake, S., Van Berkel, N., Kostakos, V., Goncalves, J., 2020. Impact of contextual and personal determinants on online social conformity. *Comput. Hum. Behav.* 108, 2–11. doi:[10.1016/j.chb.2020.106302](https://doi.org/10.1016/j.chb.2020.106302).
- Wu, G., Heath, C., Larrick, R., 2008. A prospect theory model of goal behavior. Working paper.