Borrowed Plumes:

Kinnl, Klara; Möller, Jakob; Walter, Anna

DOI:
10.57938/2dfe25ed-1089-4454-929c-fdb54feda01b

Published: 01/08/2023

Document Version
Publisher's PDF, also known as Version of record

Department of Economics
Working Paper No. 345

Borrowed Plumes: The Gender Gap in Claiming Credit for Teamwork

Klara Kinnl
Jakob Moeller
Anna Walter

August 2023

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED
AACSB ACCREDITED
ASSOCIATION OF MBAS ACCREDITED

Borrowed Plumes: The Gender Gap in Claiming Credit for Teamwork

Klara Kinnl[1], Jakob Moeller[1], and Anna Walter[1,2]

[1]Vienna University of Economics and Business
[2]Institute for Advanced Studies Vienna

August 17, 2023

Abstract

We investigate gender differences in individual credit claiming for teamwork. In a large-scale online experiment, participants work on an interactive task in teams of two and subsequently report their subjective contribution to the teamwork. In three between-subject treatments, we incentivize participants to either i) state their beliefs about their contribution truthfully, ii) to exaggerate their contribution, or iii) to exaggerate and thereby harm the other team member. Our setup allows us to distinguish between overconfidence and exaggeration with and without negative externalities, and to test whether there is a gender gap in credit claiming. We find that men and women both equally overestimate their contributions, but men exaggerate more than women: As soon as there is an incentive to exaggerate, men claim to have contributed more than women, even when exaggeration harms the team member. This gender gap in credit claiming is particularly pronounced among very large claims and for high-contributors. Strategic misrepresentations of contributions to teamwork can thus have sizeable equity consequences on the labor market.

JEL codes: J16, C92, D9

Keywords: Experiment, Gender differences, Incentives, Team work, Overconfidence, Beliefs

We thank Mackenzie Alston, Roberto Asmat Belleza, Peter Brummund, Nadine Chlaß, Christine Exley, Susann Fiedler, Milos Fisar, Ben Greiner, Simone Haeckl-Schermer, Gergely Hajdu, Boon Han Koh, Pia Heckl, Stephanie Heger, Siri Isaksson, Melanie Koch, Martin G. Kocher, Ada Kovaliukaite, Marco Lambrecht, Eva Ranehill, Anna Sandberg, Rupert Sausgruber, Arthur Schram, Anna Schwarz, James Tremewan, Lise Vesterlund, Elisabeth Wurm and the participants at workshops and conferences – ESA World Meeting 2023, IAREP-SABE 2023, SPUDM 2023, ASFEE 2023, YEM 2023, WU Gender and Diversity Conference 2023, NOEG Winter Workshop 2022, seminars at WU Vienna and Insight Austria – for helpful comments and suggestions. This project received ethics approval at WU Vienna, approval number WU-HSRP-2021-021. Any errors are our own.

1 Introduction

A large share of modern production involves some form of teamwork (Deming, 2017), where individual contributions to the output are often ambiguous (Sarsons et al., 2021). This induces asymmetric information and gives particular weight to self-evaluations in wage and promotion decisions. Two potential problems with relying on self-evaluations are inaccurate beliefs and incentives to exaggerate. Over-reporting of own performance is well-documented and common (Ross and Sicoly, 1979), especially in strategic settings, i.e., when it affects decisions of others (Charness et al., 2018; Solda et al., 2020). Further, there are gender differences in self-evaluations, both for individual (Exley and Kessler, 2022), and for team work (Isaksson, 2018). In competitive environments, men pretend to be more self-confident than women (Charness et al., 2018; Brilon et al., 2023). Cavalan et al. (2022) find that men and women are equally overconfident when asked for their contribution to a jointly produced outcome, and women react to the way the question is framed. These findings suggest that men over-report their contributions to teamwork more than women, and there exist contextual effects. When labor market outcomes depend on subjective claims about contributions, inaccurate claims may contribute to existing gender inequalities and lead to inefficient market outcomes. While the majority of the research focuses on gender differences in beliefs, the labor market often introduces incentives to exaggerate.

In this paper, we investigate the drivers of gender differences in credit claiming for teamwork, distinguishing beliefs from exaggeration. We conduct an online experiment where participants work on an interactive task in teams of two and then claim credit for their contribution. We systematically vary participants' incentives for credit claiming in three between-subject treatments. In the True Beliefs Treatment, participants are incentivized to report beliefs about their contribution honestly, while in the Individual Bonus Treatment, we financially incentivize participants to exaggerate their contribution without negative consequences for themselves or their team member. These two treatments allow us to firstly measure *overconfidence* – the difference between beliefs about contribution and actual contribution to teamwork – and secondly distinguish overconfidence from *intentional exaggeration* – where participants intentionally claim more than they believe to have contributed. On the labor market, intentional exaggeration can have negative externalities for other team members, for example if they miss out on promotions because of incorrect credit attribution. To mirror this, we include the Split Bonus Treatment, where over-reporting one's own contribution harms the team member. This allows us to test the effect of negative externalities for the team member on intentional exaggeration.

We find the following: Firstly, we observe no gender differences in beliefs but in willingness to exaggerate. Both men and women equally overestimate their contribution to the team task under incentives to be honest. As soon as exaggeration pays off, men claim to have contributed more than women. Men also claim more when exaggeration has a negative externality. Secondly, the gender gap is most pronounced at the upper tail of the distribution, that is, among very large claims. The gender gap is also most pronounced among those participants who contribute the majority to the teamwork. In general, we observe that when there is an incentive to exaggerate, people on average claim more than they believe to have contributed, but less so when it harms the team member.

Our study contributes to the experimental literature on gender differences in the labor market, and adds to the discourse on gender differences in credit attribution and self-evaluation. Specifically, we explore mechanisms behind individual estimations of contributions to team tasks and isolate strategic behavior from individual beliefs. Studies show not only that womens' achievements are evaluated differently than mens', but also that women evaluate themselves less favorably – both to their disadvantage. While gender differences have been documented extensively in the literature, we add to this by looking at gender differences in incentivized beliefs about own contributions to teamwork. Looking at credit attribution, Sarsons et al. (2021) find that in the field of economics, men's probability of being tenured is not affected by co-authoring, but women are less likely to be tenured the more they co-author. They explain this finding with gender-biased credit attribution for teamwork, and provide experimental evidence for this bias. While Sarsons et al. (2021) consider credit attribution from the supervisor's side, Isaksson (2018) provides experimental evidence that women believe to have contributed less than equally contributing men when the gender of the team member is known. Related to this, Exley and Kessler (2022) find a large gender gap in subjective self-evaluations for a math and science task, both with and without an incentive to exaggerate. This gap is persistent and present already at a young age. Yet, the subjective nature of their outcome variable means that their design does not allow for a direct comparison of beliefs with willingness to exaggerate. We add to this literature by implementing a design that allows us to distinguish between beliefs and intentionally exaggerated claims about contributions to teamwork. Conducting a well-powered experiment, we find no evidence for a gender gap in incentivized beliefs about contributions. In our setting, both men and women equally overestimate their contribution to teamwork. However, when there is an incentive to exaggerate, men claim to have contributed more than women. Our findings indicate that this gender gap is driven by differences in willingness to exaggerate, rather than differences in beliefs.

Our results imply that people may have difficulties to correctly estimate their own contribution to teamwork and react to incentives to exaggerate. This requires treating self-reports cautiously, and improving governance towards more objective and transparent performance measures. Men and women respond to incentives to exaggerate differently, which may contribute to the gender pay gap. Therefore, incentives for exaggeration should be minimized, for example by decoupling self-reported performance from hiring and promotion decisions.

We further contribute to the emerging strand of literature documenting "distributional" gender differences, by showing that gender differences in credit-claiming are most pronounced at the top tail of the distribution, i.e., very large claims. Differences in distributions have also been documented in young boys' and girls' mathematical skills (Penner, 2008; Penner and Paret, 2008). Focusing only on gender differences at the means often leads to underestimation of gender differences (Thöni and Volk, 2021). Our results highlight the importance of considering gender differences along the entire distribution of behavior. We add new evidence by showing that there are distributional gender differences in claiming credit in a team context.

The remainder of the paper is structured as follows: We first introduce the experimental design, the team task and treatments in Section 2. Section 3 describes our hypotheses before we report our results in Section 4. Section 5 discusses our results, and Section 6 concludes.

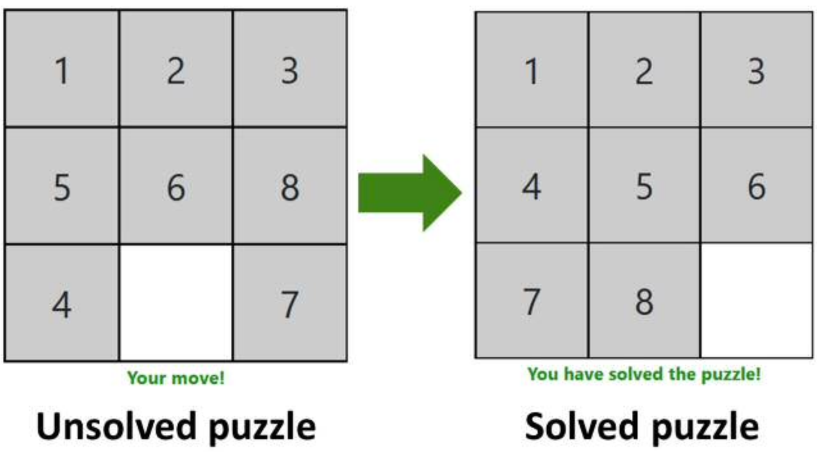
2 Experimental Design and Data

We conducted an online experiment in which participants were randomly assigned into teams of two, performed a team task together, and then estimated ("claimed") their individual contribution to the teamwork. We used a between-subjects treatment design and varied the incentives for claiming contributions to the team task. The experiment was preregistered (https://doi.org/10.17605/OSF.IO/gpdny). The following section describes the team task before explaining all steps of the experiment in more detail.

2.1 The Team Task

Participants worked on an incentivized, real-effort *sliding-puzzle* based on Isaksson (2018). The goal of the puzzle was to arrange eight numbered tiles in a three by three grid in ascending order. The two team members took turns to move a tile by clicking on it. See Figure 1 for an example of such a sliding-puzzle. Because there was time pressure to solve the puzzle in less than two minutes, it was in the team's best interest to solve the puzzle using the minimal number of moves (Isaksson, 2018). Participants had no information about their team member and no means of communicating.

Figure 1: Example of a sliding puzzle



The sliding-puzzle task has two useful characteristics. First, the task is gender-neutral: men and women are, on average, equally good at solving the puzzle (Isaksson, 2018) and also hold gender-neutral beliefs about ability in the puzzle, i.e., both men and women believe there are no gender differences in solving the puzzle (Takahashi, 2021). The mean number of individually solved puzzles in our sample is 5.13 for men and 5.15 for women (p=0.77 for a two-sided two-sample t-test; see Appendix B). Second, participants' actual contribution is observed by the researchers but not by the participants. While participants knew how their actual contribution was computed (see Formula 1 below), it was difficult to do so while solving the puzzle. Hence, there was uncertainty regarding both parties' contributions. Participants thus had to rely on subjective judgement for their claim. The correlation between actual contribution and contribution claim is as low as r=0.12 across all treatments. This suggests that, as intended, calculating the actual contribution while solving the puzzle is indeed hard and introduces subjectivity to claims. With regards to external validity, the sliding-puzzle task mimics some features of real-world teamwork. Team members have to work together interactively to solve the task, requiring adaption to each others' actions. In the extant literature, participants often work on independent tasks in isolation the sum of which represents joint production (Brilon et al., 2023; Cavalan et al., 2022; Solda et al., 2020). As often outside the laboratory, there is time pressure and an incentive to achieve the common goal of solving the puzzle, but the optimal strategy might not be obvious to all team members and there is no (time for) feedback. While everyones' actions could be perfectly monitored, team members focus on the task at hand rather than on tracking contributions. Think of remote teamwork as an example, where employees are assigned a team and do not necessarily see each other with no/limited means of communication (e.g., different time zones, no shared language/jargon...). The interdependence could be interpreted such that each team member has an expertise that they can only bring in if the team members do their job, otherwise they have to invest their resources on improving the team member's errors. A related example for this kind of teamwork, however with non-anonymous team members, could be co-authoring a paper: actions of every author in the shared document may be tracked, the goal of publishing is clear, and there exist interdependences between the co-authors and time pressure. Even though communication is possible, some coauthors might not read or respond to their mail. After several rounds of of re-writing, and re-organizing, often without external feedback, it becomes somewhat ambiguous which actions were good and which were not, and hence, who contributed how much to the final output.

2.2 Defining Actual Contribution

Before the team task, we informed participants about the definition of their *actual contribution*, which is as follows: For every solvable permutation of the puzzle, there exists a minimal number of moves required to reach the solution. We defined a *good move* as one which *reduces* the minimum number of moves needed to solve the puzzle, and a *bad move* as one which *increases* it. This way, there are no neutral moves. The categorization of moves into good and bad serves as a measure for each team member's individual contribution to solving the team task: Following Isaksson (2018), team member A's *actual contribution* is defined as follows.

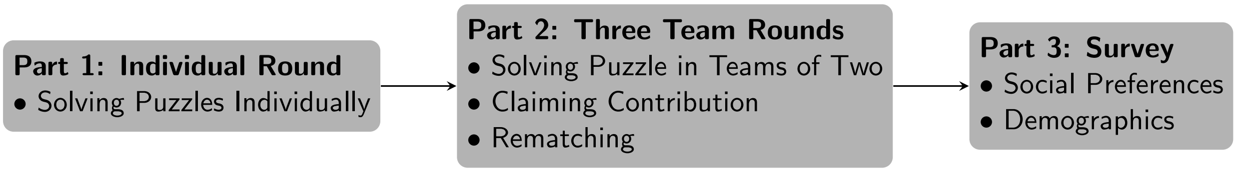
$$\frac{\text{A's good moves} - \text{A's bad moves}}{(\text{Team's good moves} - \text{Team's bad moves})} * 100 \tag{1}$$

For any solved puzzle, the sum of both team members' contributions is 100 and each team member's contribution can be interpreted as the percentage they have contributed to solving the puzzle. For border cases, such as when one participant had made zero good moves, we truncated the range of actual contributions at 0 and 100. We restrict our main analysis to solved puzzles (roughly 70 % of all puzzles, see Section2.4), in alignment with the methodology proposed by Isaksson (2018). Two key factors drove this decision. First, assessing contributions to unsolved puzzles requires making additional auxiliary assumptions, which may complicate our analysis. Actual contribution is undefined, for instance, for an unsolved puzzle where the total sum of good and bad moves is 0. Second, our research question is primarily concerned with successful teamwork, because successful teamwork is most relevant in the labor market, e.g., when determining promotions. These render solved puzzles a more suitable subject for our analysis. Excluding observations where the puzzle was not solved may introduce a bias by giving more weight to more skilled participants in the sample. However, this concern is mitigated because we do not find a significant difference in individual skill between participants who solved the puzzle in all three rounds and those who did not (p=0.25 for a two-sample t-test). Further, there is no gender difference in the number of successfully solved group puzzles, that is, no differential attrition by gender (p=0.16). A short exploratory analysis concerning unsolved puzzles is shown in Section 4.5.

2.3 Experiment Procedure

The following paragraphs explain the procedure of the experiment in detail, as visually represented in Figure 2.

Figure 2: Experimental sequence



Part 1: Individual Round

Participants first worked on the sliding-puzzle individually to become familiar with the task. They had one minute to solve as many puzzles as they could. For each solved puzzle, participants earned 2 points. We use the number of solved puzzles during the individual round as a proxy for individual skill. Afterward, participants were introduced to the contribution formula and were asked to calculate the actual contribution in an example as a comprehension check.

Part 2: Three Team Rounds

Next, participants attempted to solve the puzzle in randomly assigned teams of two and subsequently claimed their contribution. Participants were matched with a team member by arrival time. Teams are anonymous, i.e., team members know they are playing with another person, but they do not know each other's gender.

Part 2 was repeated over three rounds, with rematching between every round. To limit learning effects, the difficulty of the team task increased with every round, from six to seven to ten moves required to solve the puzzle. This should not matter for contribution claiming because the contribution depends on team member's actions in relation to each other and not on previous rounds. Collecting repeated observations over three rounds provides the following advantages beyond the mere gain of observations: First, we are able to reduce measurement noise on the individual level by observing each individual thrice rather than just once. This helps level out outlier observations that could occur, for example when a participant was randomly matched with a "bad team member", or the team just had "a bad round". Second, this allows us to still include individuals who did not solve the puzzle in all rounds when we exclude observations from non-solved puzzles.

Teams had two minutes to solve the sliding-puzzle. In every round, after having worked on the sliding-puzzle together, participants were asked to claim their individual contribution to solving the team task. Participants were instructed as follows: “Please make your claim about your contribution to the puzzle task on a scale from 0% to 100%”. Incentives to claim contributions differed between treatment groups, as described below. Participants did not receive feedback about their own or their team member’s actual contribution, which they knew through an earlier comprehension check (“The other participant will not see your claim.”). The computer randomly selected one of the three rounds for payoff. Regardless of solving the puzzle, participants earned points depending on their claim in the selected round, according to their treatment. In addition, each team member got 25 points if the team had solved the puzzle in the selected round.

Treatments

In all treatments, participants could earn a bonus payment for their claim. We varied the incentives participants had to claim their contribution in three between-subjects treatments.

In the *True Beliefs Treatment*, participants earned a bonus payment that was greater in expected value the closer their claim was to their actual contribution. The bonus was determined according to the binarized scoring rule (BSR) (Hossain and Okui, 2013). Participants’ payoff was determined based on the following function

$$\text{bonus} = \begin{cases} 100 & \text{if } K \geq L \\ 0 & \text{if } K < L \end{cases}$$

where K is a random number from a uniform distribution $\epsilon \; \{0, 0.5\}$ and $L =$ (actual contribution $-$ contribution claim)^2.

The BSR is a “proper scoring rule” independent of risk attitudes, but may facilitate distorted beliefs if participants are provided elaborate information on the incentives (Danz et al., 2022). To limit that, we used simple, verbal instructions stating that “the more accurate your claim, the higher your bonus, so you should claim what you truly think your contribution was”. On demand, participants could press a button for detailed instructions, but few did so (128 out of 845 participants). In the True Beliefs Treatment, it was in the participants’ best interest to report their true beliefs of how much they contributed.

In the *Individual Bonus Treatment*, the expected value of a participant’s bonus payment increased with their claim. Participants earned a bonus payment that was equal to their claim times 100 points. That is, $p_i = c_i$, where p_i is i’s payoff, and $c_i \in [0\%, 100\%]$ is i’s claim. Participants in the Individual Bonus Treatment thus had a clear incentive to exaggerate. Unlike in previous experiments, exaggeration directly increased expected payoff and did not depend, for example, on persuading another person (e.g., Brilon et al., 2023; Solda et al., 2020). As each team member received a separate bonus from an individual budget, the claim did not affect the team member.

In the *Split Bonus Treatment*, the expected value of a participant’s bonus payment also increased with the claim but the bonus was split between team members. After both team members submitted their claim, one team member was randomly selected as random dictator, whose claim determined the bonus payments for both team members. If participant i was selected, then i's bonus payment was $p_i = c_i$ and their team member j's bonus payment was $p_j = 100 - p_i$. Unlike in the Individual Bonus Treatment, in the Split Bonus Treatment, there was a negative externality of exaggerating for the other team member.

In the Split Bonus Treatment, the participant's claim is only payoff-relevant with a 50% chance. This allows to record relevant claims from both participants. Therefore, in order to keep the three treatments as comparable as possible, in the True Beliefs and the Individual Bonus Treatments there was a 50% chance that both team members received a random bonus payment drawn from a uniform distribution between 0 and 100. Otherwise, the payment was determined as described above.

Randomization into treatment groups happened at the session level. In total, we ran 29 sessions with an average number of 96 participants per session. The treatment was randomly drawn before the session and the respective oTree (Chen et al., 2016) script was run. We employed a between-subject design, so participants stayed within their experimental treatment group across all rounds. In Table 1, we show that the treatment groups are generally comparable and we control for factors on which they are not.

Part 3: Post-Experiment Survey

Finally, participants filled out a survey eliciting age, education, income, and gender. While we recruited participants who identify with a binary gender based on their Prolific profile, 33 participants selected "other" as their gender in the post-experiment survey. For these 33 participants, we use the binary gender they have selected on their Prolific profile instead. The survey contained further exploratory variables, such as their belief about what their team member had claimed in the last round. We incentivized this question to avoid responses that merely justify their prior claims. All other survey items were not incentivized.

In order to measure social preferences, we asked participants whether they agree or disagree with statements capturing preferences, on a five-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). In this way, we elicited measures for: downwards-/upwards-inequality aversion ("I dislike if I get a lower/higher bonus than other participants"), trust ("In general, one can trust other people", Dohmen et al., 2008), altruism ("I am generally willing to share with other people without expecting anything in return", Falk et al., 2018), utilitarianism ("Those participants who contributed more should get a greater bonus"), egalitarianism ("All participants should get the same bonus regardless of their contribution"), and team identification ("I felt connected with the other participants"). A detailed description of all collected variables can be found in Table C1 in Appendix C. After answering the survey, participants were informed about their payoff from the experiment, keeping team member's claims ambiguous in the split bonus treatment. While showing the payoff allows some participants in the split bonus treatment to induce the team members' claim, at the time of claiming, they were aware that they would not be shown the team members' *claim*, so we do not expect this indirect observability to influence their decisions.

Data Collection

The study was conducted on the survey platform Prolific in the spring of 2022. The experiment was neutrally described as a "decision-making study" in order to avoid sample self-selection. We chose to conduct the experiment online because this gave us access to a large and diverse sample.

We recruited participants from the US and UK with working age (18 to 66 years) to maximize the eligible sample. Other pre-screeners included speaking English fluently, identification with a binary gender (man/woman), and a high Prolific completion score (>90% of begun studies successfully completed). Moreover, to maximize the power to detect gender differences, we employed a gender balance of 50:50 across treatments.

Throughout the experiment, participants could collect points. At the end of the experiment, points were converted into a bonus payment at the exchange rate 1 point = 0.5 British Pence. Independently of the bonus payment, participants earned a completion fee of £1.20. The average total earnings were £1.62 and the median duration of the experiment was 11.83 minutes. Before the online experiment, we ran pilots to ensure comprehensibility of the instructions and a functioning technical setup. Observations from the pilot are not included in this paper.

2.4 Descriptive Statistics

In our main sample we recruited 2747 participants who each played three rounds, giving us a total of 8241 observations. According to power analyses, this allows us to estimate difference-in-difference effects as small as 4.5 percentage points, which corresponds to the magnitude of the effect found by Isaksson (2018) (see Appendix D.1). For the analysis, we exclude 219 participants who failed one main comprehension check two times in a row. For the main analysis, we focus on the sample of solved puzzles. Across all three rounds, this leaves us with 5302 observations of credit claims made by 2456 unique participants. 80.34% of the teams solved the puzzle in round 1, 81.47% in round 2 and 47.65% in round 3, where puzzle difficulty increased with each round. The average claim is around 60% and remains steady across all three rounds. Table 1 shows the mean age, percentage of women and UK residents, and the mean number of individual puzzles solved in one minute, which we interpret as a measure of individual skill, across the three treatment groups. The percentage of women is close to 50% in all three groups. The mean age is significantly different between groups, but the difference is negligibly small. We control for age in Appendix D.2. We randomly assigned treatments on the session level, and we increased the size of sessions over time. Because we unexpectedly depleted the UK subject pool on Prolific and had to resort to recruiting US residents, the share of UK residents is highly and significantly uneven between treatment groups. To account for this, we control for country of residence in all following regressions.

As a proxy for individual skill solving the sliding-puzzle, we use the number of puzzles solved individually in one minute (see Section 2.3). In line with Isaksson (2018), we find no difference in individual skill between men and women. Men solve 5.13 puzzles, and women solve 5.15 puzzles on average (p=0.77).

Table 1: Treatment group comparisons

	True Beliefs	Individual Bonus	Split Bonus	
Participants	845	789	822	
Age	31.02	30.93	32.12	$p<0.01^{***}$
Women	50.77%	50.57%	50.49%	p=0.99
UK	55.27%	69.20%	88.44%	$p<0.01^{***}$
Individual Skill	5.05	5.16	5.12	p=0.28

Note: Mean age, percentage of women, percentage of UK residents, and individual skill in the three treatment groups, for the sample excluding unsolved puzzles and participants who failed the attention check. p-values of an ANOVA test for differences in means of *age* and *individual skill*, and of χ^2-tests for differences in distributions for *women* and *UK*.

3 Hypotheses

The first set of hypotheses pertains to incentives to exaggerate. A purely payoff-maximizing individual will respond to this financial incentive by claiming to have contributed 100%, regardless of their beliefs about their actual contribution. Previous behavioral research shows that strategic or financial incentives to exaggerate can lead people to increase self-evaluations and reported confidence (Exley and Kessler, 2022; Schwardmann and Van der Weele, 2019; Solda et al., 2020), and that people are willing to (partially) lie for financial gain (Abeler et al., 2019; Fischbacher and Föllmi-Heusi, 2013). We therefore expect an increase in claims when it pays to exaggerate compared to when it pays to be honest. When there are negative externalities of exaggeration for the other team member, social preferences such as inequality aversion and fairness preferences ought to reduce willingness to exaggerate (Gneezy, 2005; Fehr and Schmidt, 1999; Eckel and Grossman, 1996). Further, individuals tend to cheat less when it hurts another participant compared to the experimenter (Meub et al., 2016). Therefore, we expect a decline in claims when exaggerating has a negative externality for the team member compared to when it does not.

Hypothesis 1: Incentives to exaggerate shift credit claims.

- **Hypothesis 1a:** When there is an incentive to exaggerate, individuals claim more than when there is an incentive to be honest.
- **Hypothesis 1b:** When there is an incentive to exaggerate and the team member is affected by one's claim, individuals claim less than when the team member is not affected.

The next set of hypotheses concerns gender differences in credit claiming and is based on behavioral findings in three specific domains. Firstly, a large body of literature finds that men are more overconfident than women (Bordalo et al., 2016, 2019; Exley and Kessler, 2022; Isaksson, 2018), with some notable exceptions (e.g., Cavalan et al., 2022; Brilon et al., 2023). Secondly, there are gender differences in willingness to lie (c.f. Abeler et al., 2019; Conrads et al., 2013; Dreber and Johannesson, 2008; Friesen and Gangadharan, 2012). Thirdly, studies have reported gender differences in social preferences such as inequality aversion (Croson and Gneezy, 2009) or altruism. For instance, women allocate themselves less than men in an anonymous dictator game (Eckel and Grossman, 1998). More recent evidence confirms this finding in a setting where a real-effort task precedes the allocation decision (Chowdhury et al., 2017; Halladay and Landsman, 2020), which is hence similar to our experiment. Based on these behavioral findings, we formulate hypotheses with respect to gender differences in credit claiming.

Hypothesis 2: Men claim greater contributions to team tasks than women after controlling for actual contribution.

Hypothesis 2a: Men overestimate their own contributions more than women.

Hypothesis 2b: When there is an incentive to exaggerate, men claim more than women after controlling for actual contribution.

Hypothesis 2c: When there is an incentive to exaggerate and the team member is affected by the claim, men claim more than women after controlling for actual contribution.

The final set of hypotheses is about relative differences in treatment effects between genders, i.e., differences-in-differences between treatments and genders. These hypotheses are motivated by behavioral findings about gender differences in response to strategic incentives. For example, men, more than women, strategically over-report their confidence in competitive settings (Charness et al., 2018; Brilon et al., 2023). On the other hand, Exley and Kessler (2022) find no difference in the response to the incentive to inflate self-promotions between men and women. While previous literature on gender differences in response to strategic or financial incentives is inconclusive, we provide new evidence by testing the following hypotheses.

Hypothesis 3: Men and women respond differently to incentives to exaggerate.

Hypothesis 3a: Men increase their claims more than women when there is an incentive to exaggerate, compared to when there is an incentive to be honest.

Hypothesis 3b: Men reduce their claims less than women when there is an incentive to exaggerate **and** the team member is affected by the claim, compared to when the team member is not affected.

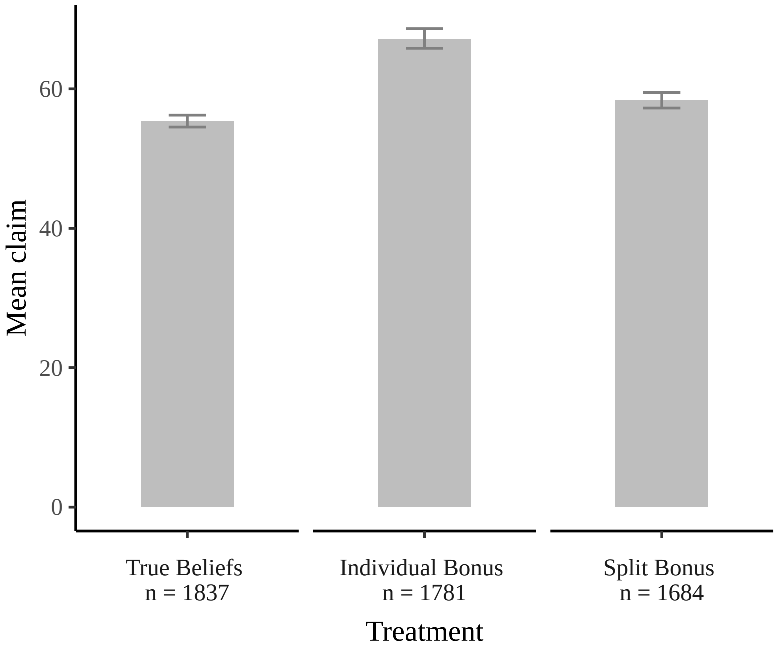
4 Results

We first report our main results before looking into mechanisms and exploratory analyses. Unless indicated otherwise, all regressions are panel regressions over the three rounds and include round fixed effects, participant random effects and clustered standard errors at the participant level. More details are listed in Appendix A.

4.1 Overconfidence and Exaggeration

Figure 3 shows the mean contribution claimed for the three treatment groups across all three rounds. The average contribution claim is 55.46 (SD=14.61) in the True Beliefs Treatment, 67.20 (SD=21.69) in the Individual Bonus Treatment, and 58.39 (SD=17.62) in the Split Bonus Treatment.

Figure 3: Mean claim in each treatment across all three rounds



Note: Whiskers mark participant-clustered bootstrapped 95% confidence intervals. Dashed horizontal lines show the average actual contribution. The sample includes only solved puzzles.

We begin by estimating overconfidence, which we define as the difference between claimed contribution and actual contribution in the True Beliefs Treatment. On average, across all rounds, participants overestimate their contributions by 6.13 percentage points. The average overconfidence is significantly different from 0 with p<0.01 for a one-sample t-test using participant clustered bootstrapped standard errors. For participants who did not solve the puzzle in all rounds, we use only the rounds in which the puzzle was solved.

Next, we estimate the average magnitude of exaggeration. Note that our design does not allow us to disentangle beliefs about own performance from willingness to exaggerate on an individual level, but on average. Table 2 shows panel regressions with participant random effects of claims on treatment and gender. In all regressions, we control for actual contribution, allowing us to compare equally contributing participants.[2]

The comparison of the Individual Bonus Treatment with the True Beliefs Treatment allows us to identify the average difference between participants' actual beliefs about their contribution and their intentional exaggeration given the incentive. Participants claim 11.37 percentage points more in the Individual Bonus Treatment than in the True Beliefs Treatment ($p<0.01$). This result is in support of Hypothesis 1a.

The presence of a negative externality of exaggerating significantly reduces claims, as participants claim 8.70 percentage points less in the Split Bonus Treatment than in the Individual Bonus Treatment ($p<0.01$, see also Appendix D.10 for pairwise treatment comparisons). This is in support of Hypothesis 1b. Nonetheless, even with a negative externality, when it pays to

[2]The divergence between contribution claims and actual contribution mainly occurs at actual contributions below 50% (see Figure 5). It is thus the participants who contribute little (low-contributors), who tend to overestimate their contribution.

Table 2: Regressions of claim on treatment and gender

	Dependent Variable:				
	Claim				
	(1)	(2)	(3)	(4)	(5)
Women	1.07	−2.95**	−2.96***		−1.24
	(0.87)	(1.49)	(1.12)		(1.23)
Individual Bonus				11.37***	13.53***
				(0.85)	(1.24)
Split Bonus				2.67***	5.11***
				(0.76)	(1.07)
Actual contribution	0.13***	0.15***	0.08***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
UK	2.28***	0.24	−0.90	1.11	−1.13
	(0.87)	(1.61)	(1.84)	(0.77)	(1.21)
Woman × Indi. Bonus					−4.27**
					(1.69)
Woman × Split Bonus					−5.04***
					(1.52)
Woman × UK					3.73**
					(1.59)
Constant	46.75***	60.84***	55.81***	48.41***	49.46***
	(1.71)	(2.26)	(2.40)	(1.04)	(1.34)
Wald Test p-values:					
Indi. = Split				<0.01***	<0.01***
Woman × Indi. = Woman × Split					0.94
Sample	True	Indi.	Split	Full	Full
Observations	1,837	1,781	1,684	5,302	5,302
R^2	0.107	0.103	0.127	0.139	0.141
Adjusted R^2	0.105	0.101	0.124	0.138	0.139

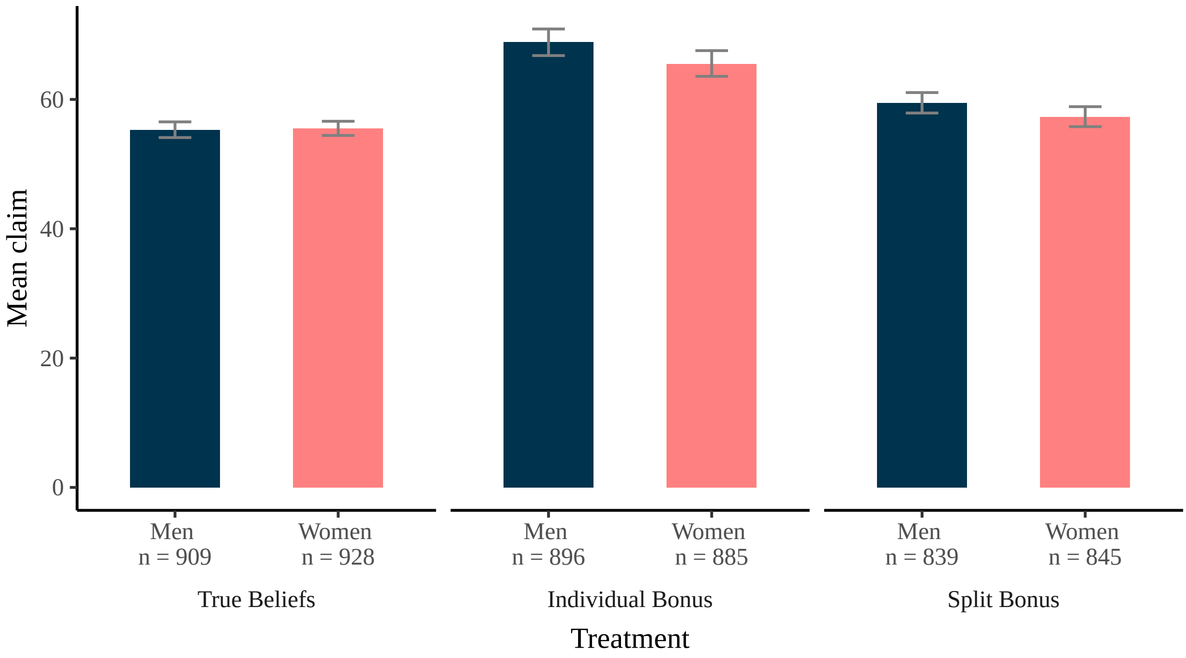
Note: Regressions of the amount claimed on gender and treatment across all three rounds. Regressions include participant random effects, round fixed effects (not reported) and clustered standard errors at the participant level. Columns 1, 2 and 3 are estimated on the samples of the True Beliefs, Individual Bonus and Split Bonus Treatments separately. Columns 4 and 5 are estimated on the full sample. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

exaggerate, participants claim 2.67 percentage points more in the Split Bonus Treatment than in the True Beliefs Treatment ($p<0.01$).

4.2 Gender Differences in Claiming Credit

Are there gender differences in claiming credit for teamwork? Figure 4 shows the mean amount claimed by men and women in each treatment across all three rounds.

Figure 4: Mean claim of men and women in each treatment across all three rounds



Note: Whiskers mark participant-clustered bootstrapped 95% confidence intervals. The sample includes only solved puzzles.

Overconfidence

Plot 1 in Figure 4 shows that in the True Beliefs Treatment, women claim 55.47% and men claim 55.24% on average. The difference is close to zero and not significant (p=0.22). We do not find evidence for a gender gap in overconfidence in a teamwork context; that is, we find no evidence in support of Hypothesis 2a. Both men and women claim to have contributed more than they actually did, and thus equally overestimate their contributions to teamwork. In Appendix E we perform a Bayesian estimation of the gender gap in the True Beliefs Treatment. The Bayesian approach allows us to include results from previous research as a prior when estimating the gender gap. Further, in contrast to frequentist methods, the Bayes factor can be interpreted as evidence in favor of the null hypothesis. Even when including the gender gap estimated by Isaksson (2018) as a prior, we find strong evidence in favor of the null hypothesis.

Exaggeration

We measure gender differences in intentional exaggeration with the Individual Bonus Treatment. Plot 2 in Figure 4 shows that in the Individual Bonus Treatment, men claim more than women. Given incentives to exaggerate, women claim to have contributed 2.95 percentage points less than men (p<0.05). This is evidence in support of Hypothesis 2b.

We further estimate the difference-in-differences between treatments and gender. The coefficient for the interaction between *women* and the Individual Bonus Treatment in column 5 in Table 2 is −4.27 and significantly different from zero (p=0.02). When comparing the Individual Bonus Treatment with the True Beliefs Treatment, both men and women increase average claims relative to what they believe to have contributed, but the average increase by men is larger than the average increase by women. This is in line with Hypothesis 3a.

Exaggeration with Externalities

The Split Bonus Treatment introduces a negative externality of exaggerating for the team member. Plot 3 in Figure 4 shows average claims of men and women when there is a negative externality. When there is an incentive to exaggerate but a negative externality for the team member, women claim to have contributed 2.96 percentage points less than men (p<0.01). This is in support of Hypothesis 2c.

The decline in claims from the Individual Bonus to the Split Bonus Treatment is nearly identical for men and women; see the Wald test for equality of the interaction terms in column 5 in Table 2. We thus do not find evidence in support of Hypothesis 3b.

4.3 The Gender Gap at the Upper Tail of the Distribution

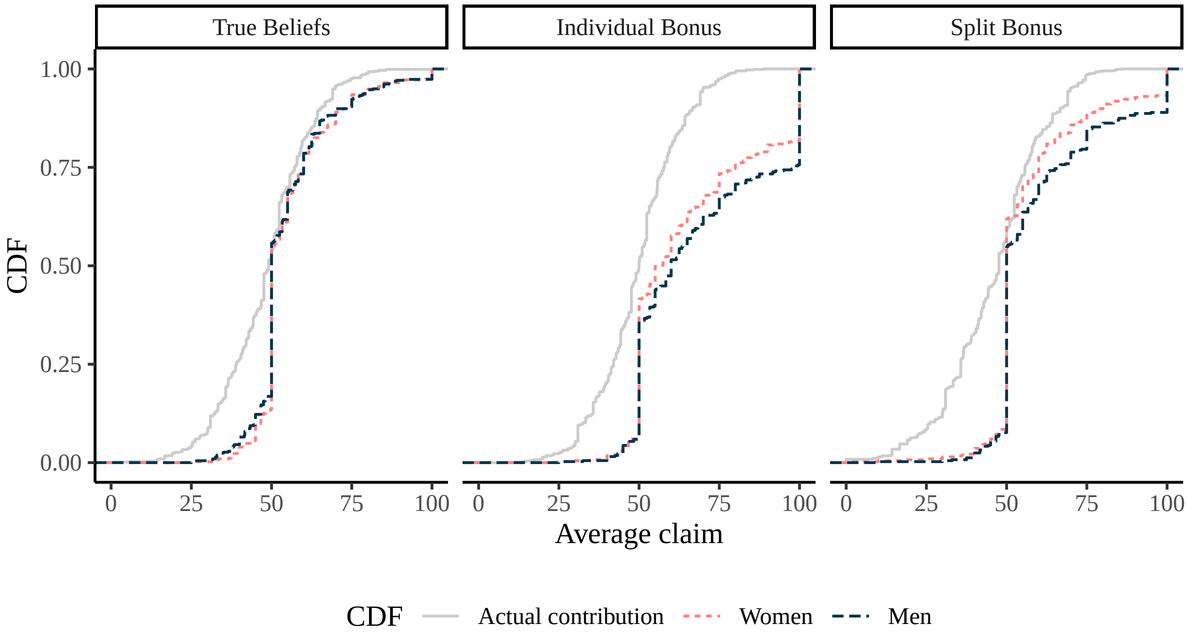
The previous two subsections document differences in the mean claims between treatments and genders. In this subsection, we present exploratory evidence that the gender gap is most pronounced among very large claims, and that high-contributing men claim more than high-contributing women.

Gender Differences in Extreme Claims

The gender gap in credit-claiming is relatively small at the means (Cohen's d=0.14 in the Individual Bonus and d=0.17 in the Split Bonus Treatment). However, at the upper tail of the distribution, the gap is much larger.

Figure 5 shows the cumulative density functions (CDFs) of the average claim per individual across the three rounds for each treatment. In the Individual Bonus Treatment and Split Bonus Treatment, the gap between men's and women's claims grows as claims increase. Table 3 shows the gender gap in each treatment estimated separately for the top quartile and the bottom three quartiles of the claim distribution. The gender gap is notably larger in the top quartile than in the bottom three quartiles for both the Individual and Split Bonus Treatments. In the True Beliefs treatment, there is no gender gap in claims in the top quartile or the bottom three quartiles.

Figure 5: CDFs of average claims across all three (solved) rounds per individual.



Further, we test for gender differences in the propensity to claim 100%. In round 1, 26.84% of men and 20.23% of women claim 100% in the Individual Bonus Treatment, and 11.28% of men and 8.33% of women in the Split Bonus Treatment. Only three percent of both men and women claim 100% in the True Beliefs Treatment. In Appendix D.5, we estimate logit regressions on the propensity to claim 100%, controlling for actual contribution and country of residence. Out of those participants who claim 100% in any round, 71.08% always claim 100%. That is: most participants either always or never claim 100%, which is in line with previous research by Fischbacher and Föllmi-Heusi (2013). Men are significantly more likely to claim 100% in the Individual Bonus Treatment, but not in the Split Bonus or the True Beliefs Treatment. After removing participants who claim 100% from the sample, there is no gender gap in any of the three treatments (see Appendix D.6). This suggests that the gender gap in credit claiming is driven by participants who claim 100%.

We also find treatment differences in the probability to claim 100%. Participants in the Individual Bonus Treatment are significantly more likely to claim 100% than in the other two treatments (see Appendix D.5 for the details). This is in line with the effects documented in Section 4.1.

The Gender Gap among High-Contributors

In Table 3 we estimate the gender gap in claims within each treatment separately for high-contributors (contribution >50%) and low-contributors (contribution ≤50%). While there is no significant gender gap among low or high-contributors in the True Beliefs Treatment, there are large and significant gender differences among the high-contributors in the Individual and Split Bonus Treatments: When there is an incentive to exaggerate, high-contributing men claim more credit than high-contributing women. This suggests that the gender gap in credit-claiming

Table 3: The gender gap at the top tails of the distributions

	Dependent Variable:		
		Claim	
	True Beliefs	Individual Bonus	Split Bonus
Top Quartile: Women	-0.07	-4.79***	-8.55***
	(1.38)	(0.64)	(1.65)
Bottom Three Quartiles: Women	0.51	-2.87***	-1.06***
	(0.37)	(0.82)	(0.33)
Z-Test for Equality of Coefficients:	p=0.68	p=0.06*	p<0.01***
High Contributors: Women	-0.23	-4.53***	-3.29***
	(1.31)	(1.96)	(1.66)
Low Contributors: Women	1.62	-1.74	-1.79
	(0.96)	(1.65)	(1.23)
Z-Test for Equality of Coefficients:	p=0.26	p=0.28	p=0.48

Note: Coefficient estimates for *women* from regressions of *claim* on gender, actual contribution and country of residence. Regressions include participant random effects, round fixed effects and clustered standard errors at the participant level. Row 1 shows coefficient estimates from the top quartile of claims, row 2 shows estimates from the bottom three quartiles of claims. The quartiles are constructed separately for each treatment and for men and women. Row 5 shows coefficient estimates for high contributors (>50%) and row 6 for low contributors (≤50%). The number of observations in the top quartile are 531 in the True Beliefs, 496 in the Individual Bonus, and 447 in the Split Bonus Treatment. The number of observations in the bottom three quartiles are 1306 in the True Beliefs, 1285 in the Individual Bonus and 1237 in the Split Bonus Treatment. The number of observations of high contributors are 584 in the True Beliefs, 583 in the Individual Bonus, and 534 in the Split Bonus Treatment. The number of observations of low contributors are 1253 in the True Beliefs, 1198 in the Individual Bonus, and 1150 in the Split Bonus Treatment. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

is driven by high-contributors.

4.4 What Drives Credit Claiming?

The previous subsections document that men respond more strongly to the incentive to exaggerate, even when it hurts the team member. One possible explanation for our results would be that men and women do not understand the incentives equally well. However, we find no evidence of gender differences in understanding, as the proportion of failed attention checks is the same among men and women (p=0.81). Therefore, we attribute the gender gap in claims we find in the Individual Bonus and the Split Bonus treatments to differences in willingness to exaggerate. In this subsection, we explore further factors that influence credit claims.

Social Preferences

Since the team setting involves interaction with another person, social preferences may play a role both in perceptions about contribution and in willingness to exaggerate. Previous literature is inconclusive, but there is a strand of research documenting gender differences in social preferences (Chowdhury et al., 2017; Croson and Gneezy, 2009; Eckel and Grossman, 1998; Halladay and Landsman, 2020, e.g.,). We find that women rate themselves significantly higher on altruism, egalitarianism, and downwards-inequality aversion. There are no gender differences in upwards-inequality aversion, team identification, trust, or utilitarianism (see Table D8 in Appendix D.7).

Do gender differences in social preferences explain gender differences in credit claiming? We observe significant correlations between social preferences and claims. In the Individual Bonus Treatment, the gender gap is no longer significant when we control for social preferences. This is surprising because, in this treatment, claims do not affect the team member. Social preferences thus should not influence claims. However, what we are picking up here may also be a correlation between our measures of social preferences and lying aversion. The gender gap in claims in the Split Bonus Treatment remains after controlling for social preferences, as can be seen in column 6 in Table 4. There is a significant negative effect on claims of downwards-inequality aversion, altruism, and team identification, while upwards-inequality aversion and egalitarianism have a significantly positive coefficient. The tendencies we observe are intuitively reasonable: Those who dislike getting more than others and who are altruistic or identify with the team (member) claim less. Those who dislike getting less than others claim more. Surprisingly, people who think everyone should get the same irrespectively of their contribution (egalitarians) claim more.

The differences-in-differences in claims between treatments and gender remain significant even when controlling for social preferences (see Table 4 column 8). Our evidence suggests that gender differences in credit claiming are not explained entirely by gender differences in social preferences. Even though our measures merely elicit stated social preferences and should thus be interpreted with caution, they should not be disregarded, as recent work finds substantial and strong correlations between stated preferences and experimentally elicited choices (e.g., for altruism, see Falk et al., 2023). Our results suggest that social preferences matter for credit claiming. This is worth exploring more thoroughly in future research.

The 50-50 Norm

Andreoni and Bernheim (2009) document a strong norm to share payoffs along a 50-50 split, motivated by the intention to appear fair. We find a large percentage of participants claiming to have contributed 50%: The percentage of participants who on average across all three rounds claim to have contributed 50% is 39.55% in the True Beliefs Treatment, 32.99% in the Individual Bonus Treatment and 48.06% in the Split Bonus Treatment. The proportion of participants claiming half is significantly larger in the Split Bonus Treatment than in the other two treatments (see Appendix D.8). When the claim directly affects the team member's payoff, the norm to appear fair seems to grow in importance. This might be explained by conditional reciprocity (e.g., Balafoutas et al., 2020): Participants who expect their team member to share the bonus equally reciprocate by also sharing the bonus equally. In line with this, in the Split Bonus Treatment 75.00% of participants who claim 50% in round 3, but only 31.74% of participants who do not claim 50%, expect their team member to claim 50%. There are no significant differences between men and women in the probability to claim 50% (see Appendix D.8).

4.5 Claiming Credit for Unsuccessful Teamwork

So far, the analysis has focused on the sample of successfully solved puzzles. However, exploratively analyzing claims for unsolved puzzles is also interesting, because it sheds some light on credit claiming for failed teamwork. Note the following caveat: The contribution formula is not always defined for unsolved puzzles (see Section 2). Further, we use the wording "contribution

Table 4: Regressions controlling for social preferences

	Dependent Variable:			
	Claim			
	(1)	(2)	(3)	(4)
Women	1.32	−1.50	−2.54**	−0.65
	(0.86)	(1.40)	(1.06)	(1.21)
Individual				13.05***
				(1.18)
Split				4.83***
				(1.06)
Actual contribution	0.13***	0.15***	0.08**	0.12***
	(0.03)	(0.03)	(0.03)	(0.02)
UK	2.59***	1.03	−0.32	−0.24
	(0.85)	(1.49)	(1.80)	(1.17)
Downw. Ineq. Av.	−1.43***	−5.67***	−4.27***	−3.75***
	(0.49)	(0.76)	(0.57)	(0.36)
Upw. Ineq. Av.	0.04	2.79***	1.40**	1.39***
	(0.50)	(0.72)	(0.60)	(0.36)
Trust	0.57	−0.29	−0.23	−0.09
	(0.43)	(0.71)	(0.59)	(0.35)
Altruism	−0.48	−2.20**	−1.28*	−1.25***
	(0.56)	(0.94)	(0.75)	(0.45)
Utilitarian	1.23	−2.39***	1.00	−0.25
	(0.93)	(0.87)	(0.87)	(0.54)
Egalitarian	0.66	0.10	2.20***	0.93**
	(0.66)	(0.79)	(0.69)	(0.42)
Team ident.	−0.83*	−1.89***	−1.72***	−1.39***
	(0.44)	(0.70)	(0.56)	(0.33)
Women × Indi. Bonus				−3.59**
				(1.62)
Women × Split Bonus				−4.44***
				(1.48)
Women × UK				2.96*
				(1.52)
Constant	45.48***	85.09***	60.70***	59.62***
	(6.22)	(6.44)	(6.05)	(3.81)
Sample	True	Indi.	Split	Full
Observations	1,837	1,781	1,684	5,302
R^2	0.12	0.17	0.18	0.18
Adjusted R^2	0.11	0.17	0.17	0.18

Note: Regressions of the amount claimed on gender across all three rounds. Regressions include participant random effects, round fixed effects (not reported) and clustered standard errors at the participant level. Column 1 is estimated of the sample of the True Beliefs Treatment, column 2 on the sample of the Individual Bonus Treatment, column 3 on the sample of the Split Bonus Treatment, and column 4 on the full sample. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

to *solving* the puzzle" to prompt claims. It is therefore not entirely clear how participants interpret the claim for unsolved puzzles. For the following analyses, we assume that claims still reflect participants' subjective perceptions of relative contributions. Because our measure of actual contribution is not always defined for unsolved puzzles, we do not control for actual contribution in the following regressions.

In Table 5 columns 1, 2, and 3 show regression estimates of the claim on not solving the puzzle, *unsolved*, using the full sample of solved and unsolved puzzles. In the True Beliefs and Individual Bonus Treatments, not solving the puzzle results in a significantly lower average claim. This is in line with earlier research that found people claim less contribution for themselves after unsuccessful teamwork compared to successful teamwork (Ross and Sicoly, 1979). However, in the Split Bonus Treatment, average contribution claims of participants who did not solve the puzzle are significantly higher. Contrary to the credit claiming behavior for solved puzzles, for unsolved puzzles, participants exaggerate more when there are externalities for their team member than when there are not.

In columns 4, 5 and 6 in Table 5 we look at gender differences among unsuccessful teams. In the True Beliefs Treatment, women claim to have contributed significantly more than men, while in the Individual Bonus and Split Bonus Treatments, women claim significantly less than men. This suggests that for unsuccessful teamwork, women believe to have contributed more than men; however, when it pays to exaggerate, men exaggerate more than women. These tendencies reflect the results for successful teamwork, but the gender gap is considerably more pronounced for unsuccessful teamwork.

Further, we find evidence for gender-specific reactions to team failure. In columns 1, 2, and 3 in Table 5 we find a significant interaction between failing to solve the puzzle and gender in the Individual Bonus and the Split Bonus Treatments, but not the True Beliefs Treatment. In both treatments where it pays to exaggerate, unsuccessful men claim significantly more than unsuccessful women. Interestingly, the interaction effect in the Split Bonus Treatment reveals that women claim the about same amount regardless of teamwork success. Men, on the contrary, increase their average claims when the puzzle remains unsolved. In the True Beliefs and Individual Bonus Treatments, where the claim does not affect the team member's payoff, both men and women seem to take some responsibility for unsuccessful teamwork. However, in the Split Bonus Treatment, where claiming less means that the team member gets more, neither men nor women take responsibility for unsuccessful teamwork. While women's claims are not different in solved or unsolved team work in the Split Bonus Treatment, our results indicate that men claim to have contributed significantly more when the puzzle is not solved. These results suggest men shifting blame for the unsuccessful teamwork to their team member when they can. These exploratory insights suggest that it may be worthwhile to systematically explore gender differences in taking responsibility for unsuccessful teamwork.

Table 5: The effect of failing to solve the puzzle

	Dependent Variable:					
	Claim					
	(1)	(2)	(3)	(4)	(5)	(6)
Unsolved	-11.16^{***}	-3.63^{**}	6.76^{***}			
	(1.49)	(1.59)	(1.52)			
Women	1.38	-3.29^{**}	-2.97^{***}	4.15^{**}	-8.45^{***}	-8.90^{***}
	(0.87)	(1.49)	(1.12)	(2.05)	(2.58)	(2.10)
UK	3.74^{***}	-0.32	0.15	7.59^{***}	-4.15	3.47
	(0.93)	(1.64)	(1.94)	(2.06)	(2.70)	(4.45)
Unsolved x Woman	1.76	-4.95^{**}	-5.42^{***}			
	(1.98)	(2.17)	(2.06)			
Constant	52.26^{***}	69.19^{***}	59.97^{***}	39.29^{***}	69.46^{***}	66.72^{***}
	(0.93)	(1.74)	(1.96)	(2.75)	(3.31)	(4.67)
Sample	True all	Indi. all	Split all	True unsolved	Indi. unsolved	Split unsolved
Observations	2,592	2,430	2,561	755	649	877
R^2	0.09	0.04	0.02	0.04	0.09	0.09
Adjusted R^2	0.08	0.04	0.02	0.03	0.08	0.08

Note: Regressions of the amount claimed on whether or not the puzzle was solved and gender across all three rounds. Regressions include participant random effects, round fixed effects (not reported) and clustered standard errors at the participant level. Columns 1, 2, and 3 are estimated on the samples of the True Beliefs, Individual Bonus and Split Bonus Treatments separately, including both solved and unsolved puzzles. Columns 4, 5, and 6 are estimated on the sample of only unsolved puzzles for each treatment respectively. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

5 Discussion

Why Do We not Find a Gender Gap in Beliefs about Contribution?

Previous research has found that men and women hold different beliefs about their subjective ability (Exley and Kessler, 2022) and about their contributions to teamwork (Isaksson, 2018). In this experiment, we find no evidence for a gender gap in beliefs about contributions to successful teamwork: on average, both men and women equally overestimate their contribution. Our True Beliefs Treatment is very similar to the experiment in Isaksson (2018), yet, unlike her, we find no gender gap in beliefs. There are several nuanced differences between the experiments that could account for the contrasting findings. For instance, Isaksson's experiment was conduced in the laboratory, not online, and several years before this experiment. We are not aware of any research documenting a decline in gender differences over the past few years. Another difference is that Isaksson has an all-US-American sample, while our sample consists of both US and UK residents. We can rule out this explanation, however, since we also find no evidence of a gender gap in claims in the True Beliefs Treatment in our US sample. The most promising explanation is that in Isaksson's experiment, participants know the gender of their team member, while in our experiment they do not. Possibly, knowing the gender of the team member leads people to project gender stereotypical beliefs about skills and contribution both onto their team member and onto themselves (Bordalo et al., 2016, 2019). There is evidence that knowing the gender of the other player matters in the context of an ultimatum game (Solnick, 2001), and that it changes giving behavior of women towards other female recipients (Ben-Ner et al., 2004).

Bandiera et al. (2022) use Bayesian analyses on data from 39 experiments that measure confidence. In line with our results, they find evidence for both genders being slightly overconfident. Several studies find that gender gaps in beliefs about own performance are contextual. Ludwig et al. (2017) find no gender differences in self-assessments under incentives to state the truth when real performance is imperfectly observable to others, but they do find that women, and not men, downgrade their self-assessment in response to observability. Haeckl (2022) finds that women increase their self-assessment when it is made public, but not when their actual performance is also publicly announced. The results for men are less conclusive. In a recent paper, Cavalan et al. (2022) investigate biases in bargaining over a joint outcome under uncertain individual contributions. Unlike in our setup, participants do not interact during the real-effort task. The authors find that men and women are equally overconfident (i.e., their belief about their contribution exceeds their actual performance, even after learning about the size of the jointly produced outcome). Yet, women, but not men, react to different framings of the instructions asking for their contribution, suggesting contextual gender differences.

We find no evidence that men believe to have contributed more than women, yet we find that men claim to have contributed more than women when it pays to exaggerate. This finding suggests that gender differences in claiming credit for contributions to teamwork are driven by differences in willingness to exaggerate, not differences in beliefs.

Gender Differences at the Extremes of the Distribution

The gender gap in credit claiming is particularly pronounced at the upper tail of the claim-distribution. A substantial body of literature documents gender differences at the extremes of distributions (see e.g. Andreoni and Vesterlund, 2001; Penner, 2008; Penner and Paret, 2008; Baye and Monseur, 2016; Liu et al., 2020; Thöni and Volk, 2021). While much of the gender literature focuses on differences in the means, differences at the extremes of distributions could have important implications for career advancements in highly competitive domains. In addition, we find that gender differences are most pronounced among high-contributors. If high-contributing women claim less credit for their contributions than high-contributing men, this may also exacerbate differences in labor market outcomes at the upper tail of the distribution.

One potential driver of the gap could be that men respond more strongly to financial incentives. In line with this reasoning, we find that, in particular in the Split Bonus Treatment, gender differences are strongly pronounced at the upper tails of the distributions. This also relates to findings that men are more likely to be either perfectly selfish or perfectly selfless (Andreoni and Vesterlund, 2001), and men react more strongly to competitive incentives (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Niederle et al., 2011). A second possible explanation is that women are less willing than men to lie for financial gain (Capraro, 2018; Erat and Gneezy, 2012). Finally, we cannot exclude the possibility that women are simply more averse to extreme claims. However, we are not aware of any literature documenting gender differences in preferences for extremes.

6 Conclusion

In this study, we explore the drivers of gender gaps in claiming credit for contributions to teamwork, distinguishing between honest beliefs and intentional exaggeration. This paper adds to the literature on gender differences in teamwork and self-evaluation. We address whether over-claiming is driven by overconfidence or intentional distortion of claims, and to what extent it matters whether the team member is adversely affected by exaggeration. To do so, we conducted a large-scale interactive online experiment recruiting 2747 participants. In the experiment, participants solved an interactive team task in teams of two and subsequently estimate how much they have contributed. We use a between-subjects design where we vary incentives for contribution claiming.

Our experiment shows that men and women both equally overestimate their contribution to teamwork. We find no evidence that men believe to have contributed more than women. Further, people intentionally exaggerate when it pays to do so, but less when it harms their team member. We find that men respond to the incentive to exaggerate more strongly than women. When it pays to exaggerate, men claim significantly more than women — irrespective of whether it does or does not harm the team member. In fact, men correct their claims relatively less when facing a negative externality of exaggerating than women. While we provide evidence for a gender gap in credit claiming at the means, it is even more pronounced at the upper tail of the distribution. The majority of previous research on gender differences in behavior has focused on differences at the means, possibly overlooking large differences at the extremes.

More research is needed on heterogeneous gender effects across the distribution of behavior.

Our results provide evidence for a gender gap in claiming credit for teamwork, and that this gap is driven by differences in willingness to exaggerate rather than differences in beliefs. This has important implications for the design of reward schemes, as it may exacerbate the gender gap in recognition for teamwork. Specifically, supervisors should treat self-reports with caution, as, in the setting of our experiment, people are generally bad at estimating their own contribution to teamwork. This suggests that to improve accurate contribution assessments and compensation equity, transparent, objective performance measures need to be the basis for reward schemes rather than self-reports. In order to reach truthful self-evaluations, the absence of incentives to exaggerate is desirable, for example by separating self-evaluations from financial incentives or introducing team rewards.

References

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., and Wagenmakers, E.-J. (2015). An introduction to bayesian hypothesis testing for management research. *Journal of Management*, 41(2):521–543.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J. and Vesterlund, L. (2001). Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics*, 116(1):293–312.

Balafoutas, L., Czermak, S., Eulerich, M., and Fornwagner, H. (2020). Incentives for dishonesty: An experimental study with internal auditors. *Economic Inquiry*, 58(2):764–779.

Bandiera, O., Parekh, N., Petrongolo, B., and Rao, M. (2022). Men are from mars, and women too: A bayesian meta-analysis of overconfidence experiments. *Economica*.

Baye, A. and Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-scale Assessments in Education*, 4(1):1–16.

Ben-Ner, A., Kong, F., and Putterman, L. (2004). Share and share alike? gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, 25(5):581–589.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–73.

Brilon, S., Grassi, S., Grieder, M., and Schulz, J. F. (2023). Strategic competition and self-confidence. *Management Science*.

Capraro, V. (2018). Gender differences in lying in sender-receiver games: A meta-analysis. *Judgment and Decision Making*, 13(4):345–355.

Cavalan, Q., de Gardelle, V., and Vergnaud, J.-C. (2022). I did most of the work! three sources of bias in bargaining with joint production. *Journal of Economic Psychology*, 93:102566.

Charness, G., Rustichini, A., and Van de Ven, J. (2018). Self-confidence and strategic behavior. *Experimental Economics*, 21(1):72–98.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chowdhury, S. M., Jeon, J. Y., and Saha, B. (2017). Gender differences in the giving and taking variants of the dictator game. *Southern Economic Journal*, 84(2):474–483.

Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34:1–7.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2008). Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry*, 46(1):84–90.

Dreber, A. and Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, 99(1):197–199.

Eckel, C. C. and Grossman, P. J. (1996). The relative price of fairness: Gender differences in a punishment game. *Journal of Economic Behavior & Organization*, 30(2):143–158.

Eckel, C. C. and Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, 108(448):726–735.

Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58(4):723–733.

Exley, C. L. and Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3):1345–1381.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, 69(4):1935–1950.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

Friesen, L. and Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters*, 117(3):624–626.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

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

Haeckl, S. (2022). Image concerns in ex-ante self-assessments–gender differences and behavioral consequences. *Labour Economics*, 76:102166.

Halladay, B. and Landsman, R. (2020). Shame on me: Emotions and gender differences in taking with earned endowments. *Available at SSRN 3729536*.

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

Isaksson, S. (2018). It takes two: Gender differences in group work. Technical report, Working Paper.

Liu, R., Alvarado-Urbina, A., and Hannum, E. (2020). Differences at the extremes? gender, national contexts, and math performance in latin america. *American Educational Research Journal*, 57(3):1290–1322.

Ludwig, S., Fellner-Röhling, G., and Thoma, C. (2017). Do women have more shame than men? an experiment on self-assessment and the shame of overestimating oneself. *European Economic Review*, 92:31–46.

Meub, L., Proeger, T., Schneider, T., and Bizer, K. (2016). The victim matters–experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters*, 23(16):1162–1167.

Niederle, M., Vesterlund, L., et al. (2011). Gender and competition. *Annual Review of Economics*, 3(1):601–630.

Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology*, 114(S1):S138–S170.

Penner, A. M. and Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37(1):239–253.

Ross, M. and Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, 37(3):322.

Sarsons, H., Gërxhani, K., Reuben, E., and Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, 129(1):101–147.

Schwardmann, P. and Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, 3(10):1055–1061.

Solda, A., Ke, C., Page, L., and Von Hippel, W. (2020). Strategically delusional. *Experimental Economics*, 23:604–631.

Solnick, S. J. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, 39(2):189–200.

Takahashi, Y. (2021). Gender differences in the cost of corrections in group work. *arXiv preprint arXiv:2108.10109*.

Thöni, C. and Volk, S. (2021). Converging evidence for greater male variability in time, risk, and social preferences. *Proceedings of the National Academy of Sciences*, 118(23):e2026112118.

Appendix

The supplementary online appendix (replication package) is available on request.

A Estimation Methods

In Section 4, we report the results of panel regressions on all rounds. This gives us the greatest possible statistical power to detect effects. In the panel regressions, we account for non-independence of observations in the data by including participant random effects, round fixed effects and clustered standard errors at the participant level. The round fixed effects absorb round-specific characteristics that are constant across individuals. Using participant random effects, we estimate an overall effect of each participant across all rounds, while we allow the effect to be different between rounds. We cluster standard errors at the participant level because we observe the same individual across three rounds.

The most conservative way to estimate the effects while making use of the information from all rounds is to take the average claim across rounds for each participant. This gives us one observation per participant, which includes information from all rounds. We show in Appendix D.4 that the results do not change when we use this approach. The downside to this approach is that we can not control for actual contribution as precisely, and we therefore can not compare individuals conditional on having contributed the same, but only conditional on having contributed the same on average.

In Appendix D.3 we also estimate all main results using only round 1. The estimates are of similar magnitude and direction as when using all rounds, but the gender differences are not significant because of the reduced statistical power.

B Gender-Neutrality of the Task

Table B1: Individual skill in puzzle-solving

	Dependent Variable:
	Individual Skill
Women	0.0189
	(0.0632)
Constant	5.128***
	(0.0449)
Observations	2,036
R-squared	0.000

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

C Variables

Table C1: **Overview of variables** used in analyses and their definitions.

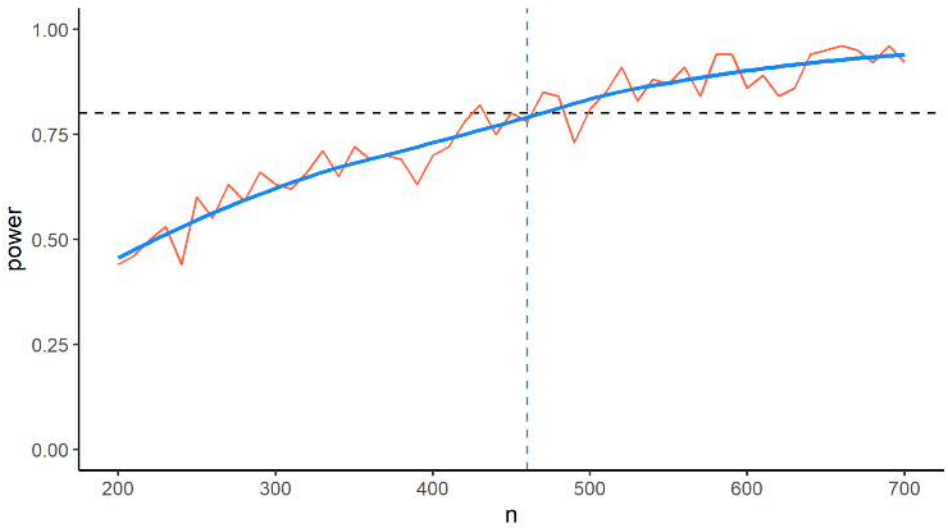
Variable	Range	Based on / definition
Outcome variables		
Claim	0–100	Isaksson (2018)
Average Contribution Claim	0–100	individual avg. claim across all rounds
Over-Claim	(-100) –100	claim - actual contribution
Explanatory variables		
Individual Bonus	0–1	dummy for Individual Bonus Treatment
Split Bonus	0–1	dummy for Split Bonus Treatment
Actual Contribution	0–100	Isaksson (2018)
		We truncate at 0 (for ac <0) and 100 (for ac >100)
Women	*Women*	0–1
Average Actual Contribution	0–100	individual avg. contribution across all rounds
Control variables		
Individual Skill	0–14	number of solved puzzles in the individual round
Round	1–3	current team task round
Age	18–66	age in years
UK	0–1	dummy for participant residence (UK)
US	0–1	dummy for participant residence (US)
Income	1–5	annual income before taxes:
		1 = less than 10 000 USD
		2 = 10 001 USD - 50 000 USD
		3 = 50 001 USD - 100 000 USD
		4 = 100 001 USD - 200 000 USD
		5 = 200 001 USD and above
High Education	1–5	1 = less than high school
		2 = high school
		3 = vocational school
		4 = undergraduate/Bachelor's degree
		5 = graduate/Master's degree
Participated on Phone	0–1	dummy for participation on mobile device
Exploratory variables		
Unsolved	0–1	indicator for team task being unsolved in that round
Social Preferences	1–5	agreement with statements that address potential mechanisms*
		1 = strongly disagree
		2 = disagree
		3 = neither agree nor disagree
		4 = agree
		5 = strongly agree

*Statements: "I dislike if I get a greater bonus than other participants" (Upwards inequality aversion), "I dislike if I get a smaller bonus than other participants" (Downwards inequality aversion), "In general, one can trust other people" (trust),"I am generally willing to share with other people without expecting anything in return" (Altruism), "Those participants who contributed more should get a greater bonus" (Utilitarian), "All participants should get the same bonus regardless of their contribution." (Egalitarian), "I felt connected with the other participants" (Team identification)

D Additional Analyses, Tables and Figures

D.1 Power Analysis

Figure D1: Power simulation: probability of detecting the difference-in-difference effects found in the current sample in a sample with n independent observations per experimental group.



Based on power simulations assuming standard deviations found in Isaksson (2018), we need to recruit n=2760 participants, allocated evenly across the three treatment groups, in order to be able to detect a difference-in-difference effect of 4.5 percentage points for Hypotheses 3a and 3b with 80% power. This assumes that we use the observations of round 1 only.

D.2 Regressions Including Control Variables

Table D1: Regressions including control variables

	Dependent Variable:					
	Claim					
	(1)	(2)	(3)	(4)	(5)	(6)
Women	1.04	−2.47	−2.83**	1.13	−2.60*	1.02
	(0.88)	(1.53)	(1.15)	(0.89)	(1.49)	(0.88)
Individual Bonus				13.13***		13.19***
				(1.24)		(1.23)
Split Bonus					−8.29***	4.72***
					(1.36)	(1.05)
Actual Contribution	0.13***	0.14***	0.08***	0.14***	0.11***	0.12***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Round 2	1.24**	0.09	1.19*	0.62	0.63	0.81**
	(0.57)	(0.52)	(0.61)	(0.39)	(0.40)	(0.33)
Round 3	0.53	0.61	2.30***	0.54	1.47***	1.10**
	(0.81)	(0.70)	(0.78)	(0.54)	(0.53)	(0.45)
UK	2.34***	0.70	−0.07	1.71*	0.42	1.34*
	(0.88)	(1.65)	(1.90)	(0.90)	(1.27)	(0.81)
Individual Skill	−0.10	1.12**	−0.30	0.43	0.44	0.22
	(0.27)	(0.52)	(0.42)	(0.28)	(0.33)	(0.23)
High Education	0.35	0.32	−0.09	0.35	0.16	0.21
	(0.38)	(0.69)	(0.54)	(0.39)	(0.44)	(0.32)
Income	−0.04	1.05	1.52*	0.42	1.27*	0.77*
	(0.56)	(1.03)	(0.81)	(0.56)	(0.65)	(0.46)
Age	−0.04	−0.12	−0.06	−0.07	−0.07	−0.06
	(0.05)	(0.10)	(0.08)	(0.05)	(0.06)	(0.04)
Participated on Phone	0.14	−4.49***	−0.23	−2.27**	−2.42**	−1.63**
	(0.99)	(1.51)	(1.26)	(0.91)	(0.99)	(0.74)
Women x Indi. Bonus				−3.71**		−3.63**
				(1.68)		(1.68)
Women x Split Bonus					−0.48	−4.07***
					(1.86)	(1.42)
Constant	47.13***	56.34***	55.33***	45.28***	59.42***	46.51***
	(2.87)	(4.93)	(4.27)	(2.74)	(3.45)	(2.31)
Sample	True	Indi.	Split	Full	Full	Full
Observations	1,837	1,781	1,684	3,618	3,465	5,302
R^2	0.11	0.11	0.13	0.15	0.13	0.14
Adjusted R^2	0.10	0.11	0.12	0.15	0.12	0.14

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

D.3 Round 1

Table D2: Regressions on round 1

	Dependent Variable:				
	Claim				
	(1)	(2)	(3)	(4)	(5)
Women	0.48	-1.97	-2.54^{*}		-2.00
	(1.04)	(1.65)	(1.33)		(1.69)
Individual Bonus				12.10^{***}	13.34^{***}
				(0.94)	(1.33)
Split Bonus				2.79^{***}	4.60^{***}
				(0.99)	(1.37)
Actual Contribution	0.14^{***}	0.14^{**}	0.11^{**}	0.13^{***}	0.13^{***}
	(0.03)	(0.06)	(0.05)	(0.03)	(0.03)
UK	1.80^{*}	-0.46	-0.61	0.64	-1.59
	(1.04)	(1.82)	(2.01)	(0.88)	(1.36)
Women $\times$ Indi. Bonus					-2.51
					(1.88)
Women $\times$ Split Bonus					-3.87^{*}
					(1.99)
Women $\times$ UK					3.77^{**}
					(1.82)
Constant	46.52^{***}	61.12^{***}	54.24^{***}	47.91^{***}	49.34^{***}
	(1.85)	(3.53)	(3.01)	(1.53)	(1.85)
Sample	True	Indi.	Split	Full	Full
Observations	708	685	643	2,036	2,036
R^2	0.03	0.01	0.01	0.09	0.10
Adjusted R^2	0.03	0.01	0.01	0.09	0.09

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

D.4 Results with the Average Claim across All Rounds

Table D3: Treatment level regressions on the average claim across all rounds

	Dependent Variable:		
	Average Claim		
	(1)	(2)	(3)
Women	1.31	-2.96^{**}	-3.18^{***}
	(0.89)	(1.49)	(1.14)
Average Actual Contribution	0.07^{**}	0.09	0.03
	(0.03)	(0.06)	(0.04)
UK	2.12^{**}	0.08	-1.08
	(0.89)	(1.61)	(1.79)
Constant	50.49^{***}	64.34^{***}	59.55^{***}
	(1.81)	(3.31)	(2.55)
Sample	True	Indi.	Split
Observations	845	789	822
R^2	0.01	0.01	0.01
Adjusted R^2	0.01	0.004	0.01

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table D4: Regressions on the average claim across all rounds comparing treatments

	Dependent Variable:	
	Average Claim	
	(1)	(2)
Women	1.10	-3.17^{**}
	(1.18)	(1.29)
Individual Bonus	13.30^{***}	8.36^{***}
	(1.19)	(1.32)
Average Actual Contribution	0.07^{**}	0.05
	(0.03)	(0.03)
UK	1.21	-0.29
	(0.89)	(1.18)
Women x Individual Bonus	-3.84^{**}	0.12
	(1.66)	(1.85)
Constant	50.75^{***}	57.86^{***}
	(1.87)	(2.07)
Sample	True & Indi.	Indi. & Split
Observations	1,634	1,611
R^2	0.11	0.06
Adjusted R^2	0.11	0.06

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

D.5 Probability to claim 100%

Table D5: Propensity to claim 100% in round 1

	Dependent Variable:		
	Claim = 100		
	(1)	(2)	(3)
Women	0.079	-0.370^{**}	-0.330
	(0.449)	(0.185)	(0.268)
Actual Contribution	-0.009	0.002	-0.004
	(0.013)	(0.007)	(0.009)
UK	0.367	-0.011	-0.020
	(0.466)	(0.205)	(0.399)
Constant	-3.275^{***}	-1.092^{***}	-1.870^{***}
	(0.776)	(0.393)	(0.588)
Sample	True	Indi.	Split
Observations	708	685	643
Log Likelihood	-97.475	-371.391	-205.281
Akaike Inf. Crit.	202.950	750.782	418.563

Note: Logit regressions of the propensity to claim 100% in round 1. The sample includes solved only solved puzzles. Column 1 is estimated on the True Beliefs Treatment, column 2 on the Individual Bonus Treatment, and column 3 on the Split Bonus Treatment. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table D6: Logit regressions for the propensity to claim 100%

	Dependent Variable:				
	Claim = 100				
	(1)	(2)	(3)	(4)	(5)
Women	-0.03	-0.37^{**}	-0.31		-0.43
	(0.35)	(0.17)	(0.22)		(0.35)
Individual Bonus				2.01^{***}	2.18^{***}
				(0.18)	(0.25)
Split Bonus				0.99^{***}	1.16^{***}
				(0.20)	(0.27)
Actual Contribution	0.003	0.004	0.01	0.005	0.004
	(0.01)	(0.004)	(0.01)	(0.003)	(0.003)
UK	0.41	0.12	-0.06	0.20	-0.16
	(0.35)	(0.19)	(0.33)	(0.15)	(0.20)
Round 2	0.44^{**}	0.03	0.15	0.11^{**}	0.11^{**}
	(0.19)	(0.06)	(0.12)	(0.05)	(0.05)
Round 3	0.37	0.15	0.16	0.19^{**}	0.18^{**}
	(0.29)	(0.10)	(0.17)	(0.08)	(0.08)
Women x Indi. Bonus					-0.36
					(0.37)
Women x Split Bonus					-0.39
					(0.41)
Women x UK					0.58^{*}
					(0.30)
Constant	-3.82^{***}	-1.27^{***}	-2.39^{***}	-3.61^{***}	-3.31^{***}
Sample	True	Indi.	Split	Full	Full
Observations	1,837	1,781	1,684	5,302	5,302

Note: Logit regressions of the propensity to claim 100% across all three rounds. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

D.6 The gender gap disappears when removing those who claim 100%

Table D7: Within and between treatment regressions without participants who claim 100%

	Dependent Variable:					
	Claim					
	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.98	−0.20	−1.02	0.84	−0.23	0.77
	(0.65)	(0.92)	(0.65)	(0.66)	(0.91)	(0.65)
Individual Bonus				3.38***		3.42***
				(0.79)		(0.79)
Split Bonus					−2.72***	0.66
					(0.80)	(0.67)
Actual Contribution	0.13***	0.18***	0.11***	0.15***	0.14***	0.13***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Round 2	0.75	−0.09	0.62	0.38	0.28	0.46
	(0.51)	(0.53)	(0.50)	(0.37)	(0.37)	(0.30)
Round 3	0.45	−0.13	1.14*	0.22	0.56	0.53
	(0.68)	(0.69)	(0.67)	(0.49)	(0.49)	(0.40)
UK	1.57**	−0.05	−0.37	0.96*	−0.12	0.70
	(0.66)	(1.00)	(1.12)	(0.56)	(0.75)	(0.50)
Women x Indi. Bonus				−0.83		−0.81
				(1.10)		(1.10)
Women x Split Bonus					−0.83	−1.81**
					(1.11)	(0.91)
Constant	45.34***	48.18***	48.58***	45.11***	49.98***	45.89***
	(1.45)	(1.85)	(1.73)	(1.15)	(1.35)	(0.97)
Sample	True	Indi.	Split	True & Indi.	Indi. & Split	Full,
Observations	1,763	1,347	1,504	3,110	2,851	4,614
R^2	0.12	0.20	0.15	0.17	0.18	0.16
Adjusted R^2	0.12	0.20	0.15	0.17	0.18	0.16

Note: Regressions of the amount claimed on gender and treatment across all three rounds. Regressions include participant random effects, round fixed effects and clustered standard errors at the participant level. All regressions are estimated on the sample excluding participants who claim 100%. Columns 1, 2 and 3 are estimated on the samples of the True Beliefs, Individual Bonus and Split Bonus Treatments separately. Column 4 is estimated for the sample of the True Beliefs and the Individual Bonus Treatments. Column 5 is estimated on the sample of the Individual Bonus and the Split Bonus Treatments. Column 6 is estimated on the full sample of all three treatments pooled. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

D.7 Social Preferences

Table D8: Gender differences in social preferences

	Dependent Variable:						
	Altruism	Egalitarianism	Utilitarianism	Downwards-Inequality Aversion	Upwards-Inequality Aversion	Team identification	Trust
Women	0.087***	0.194***	-0.076***	0.119***	0.001	0.005	-0.00703
	(0.019)	(0.025)	(0.020)	(0.022)	(0.022)	(0.024)	(0.023)
Constant	3.592***	2.558***	3.822***	2.259***	3.679***	2.714***	3.067***
	(0.014)	(0.018)	(0.014)	(0.016)	(0.016)	(0.017)	(0.016)
Observations	7,875	7,875	7,875	7,875	7,875	7,875	7,875
R-squared	0.003	0.008	0.002	0.004	0.000	0.000	0.000

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

D.8 Probability to claim 50%

Table D9: Treatment differences in propensity to claim 50% on average across all three rounds

	Dependent Variable:
	Average Claim = 50
Individual Bonus	-0.286^{***}
	(0.104)
Split Bonus	0.451^{***}
	(0.105)
Actual Contribution	-0.007^{***}
	(0.002)
UK	-0.185^{*}
	(0.096)
Constant	0.069
	(0.133)
Wald Tests:	
Individual Bonus = Split Bonus	$p<0.001^{***}$
Observations	2,455
Log Likelihood	−1,630.715
Akaike Inf. Crit.	3,271.431

Note: Logit regression on the propensity to claim 50% on average across all three rounds. The outcome variable is whether the average claim per individual across all three rounds (for solved puzzles only) is equal to 50. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table D10: Propensity to claim 50% on average across all three rounds

	Dependent Variable:	
	Average Claim = 50	
	(1)	(2)
Women	0.062	0.193
	(0.143)	(0.155)
Individual Bonus	-0.359^{**}	
	(0.150)	
Split Bonus		0.698^{***}
		(0.150)
Average Actual Contribution	-0.004	-0.014^{***}
	(0.004)	(0.004)
UK	-0.195^{*}	-0.222^{*}
	(0.110)	(0.133)
Women x Indi. Bonus	0.138	
	(0.207)	
Women x Split Bonus		0.058
		(0.209)
Constant	-0.098	0.015
	(0.231)	(0.238)
Sample	True & Indi.	Indi. & Split
Observations	1,634	1,611
Log Likelihood	$-1{,}066.225$	$-1{,}058.757$
Akaike Inf. Crit.	2,144.451	2,129.514

Note: Logit regressions of the propensity to claim 50% on average across all three rounds. The outcome variable is whether the average claim per individual across all three rounds (for solved puzzles only) is equal to 50. Column 1 is estimated on the True Beliefs and the Individual Bonus Treatments, column 2 is estimated on the Individual Bonus and the Split Bonus Treatments. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

D.9 The US-UK Difference

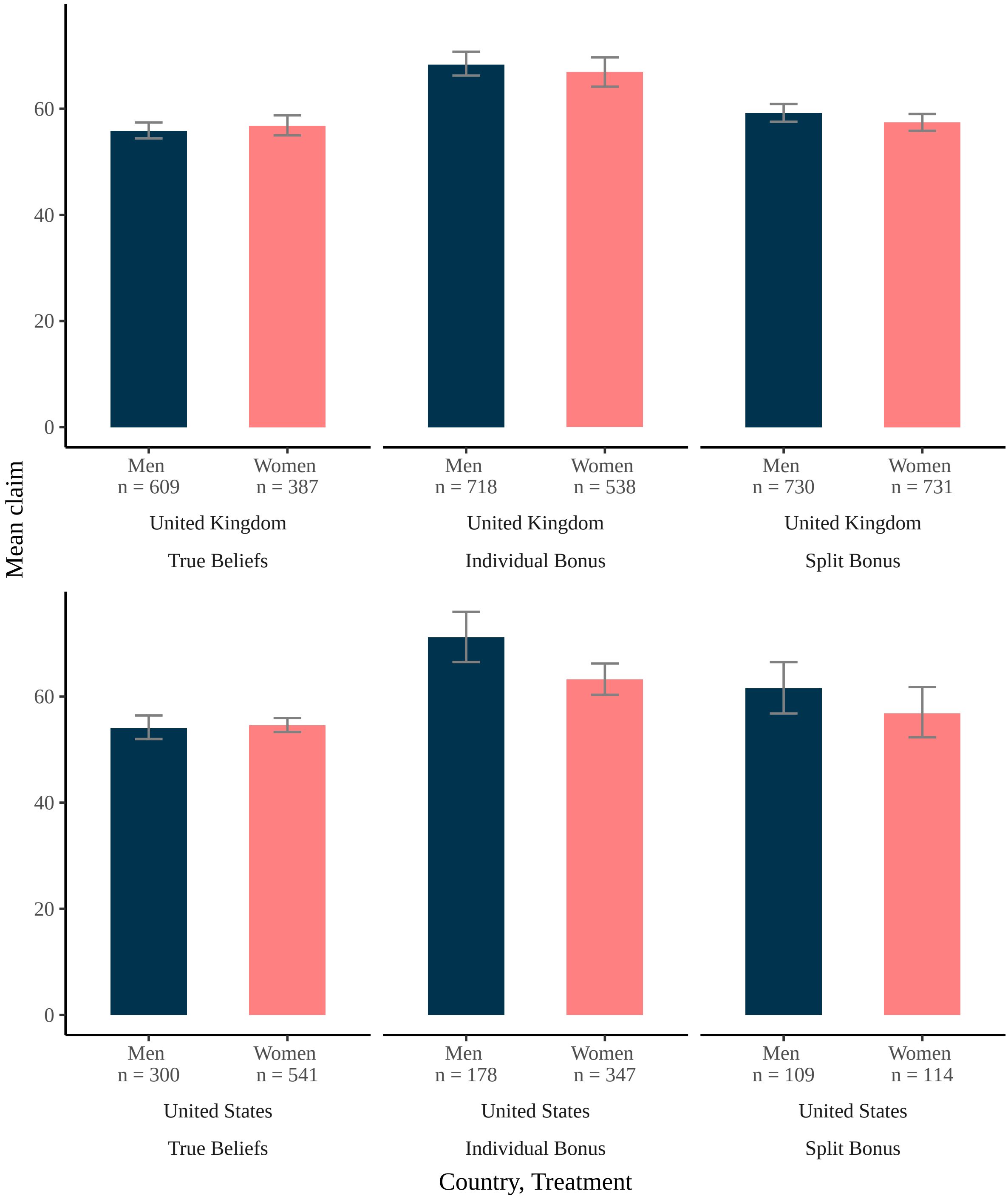
Table D11: Regressions of claim on gender estimated separately by country and treatment

	Dependent Variable:					
	Claim					
	(1)	(2)	(3)	(4)	(5)	(6)
Women	1.33	0.60	-0.77	-8.43^{***}	-2.53^{**}	-6.37^{*}
	(1.21)	(1.23)	(1.75)	(2.83)	(1.18)	(3.51)
Actual contribution	0.07^{*}	0.23^{***}	0.14^{***}	0.17^{***}	0.07^{**}	0.17^{**}
	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.07)
Round 2	0.68	1.74^{**}	0.45	-0.61	0.95	2.98^{*}
	(0.72)	(0.87)	(0.67)	(0.72)	(0.66)	(1.54)
Round 3	1.81^{*}	-1.11	0.90	0.16	2.33^{***}	2.10
	(1.09)	(1.21)	(0.88)	(1.07)	(0.87)	(1.68)
Constant	51.91^{***}	42.12^{***}	60.40^{***}	63.67^{***}	55.23^{***}	52.80^{***}
	(2.05)	(2.44)	(2.18)	(3.31)	(1.84)	(4.18)
Treatment	True	True	Indi.	Indi.	Split	Split
Country	UK	US	UK	US	UK	US
Observations	996	841	1,256	525	1,461	223
R^2	0.10	0.13	0.07	0.20	0.12	0.16
Adjusted R^2	0.10	0.12	0.07	0.19	0.12	0.15

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Although we did not initially intend to study country differences, we find suggestive evidence that there are differences in contribution beliefs and in the size of the gender gap between the two countries. We estimate the gender gap in claims in each treatment separately for the US and the UK. In the True Beliefs Treatment, there is no gender gap in either country. In the Individual Bonus Treatment, we find no gender gap in the UK, but a large and significant gender gap in the US. In the Split Bonus Treatment, we find a gender gap in both countries. Note that because we did not intend to study country-differences ex ante, the US sample for the Split Bonus Treatment is small.

Figure D2: Mean claim by country, treatment and gender



Note: Whiskers mark participant-clustered bootstrapped 95% confidence intervals. The sample includes only solved puzzles.

D.10 Pairwise Treatment Comparisons

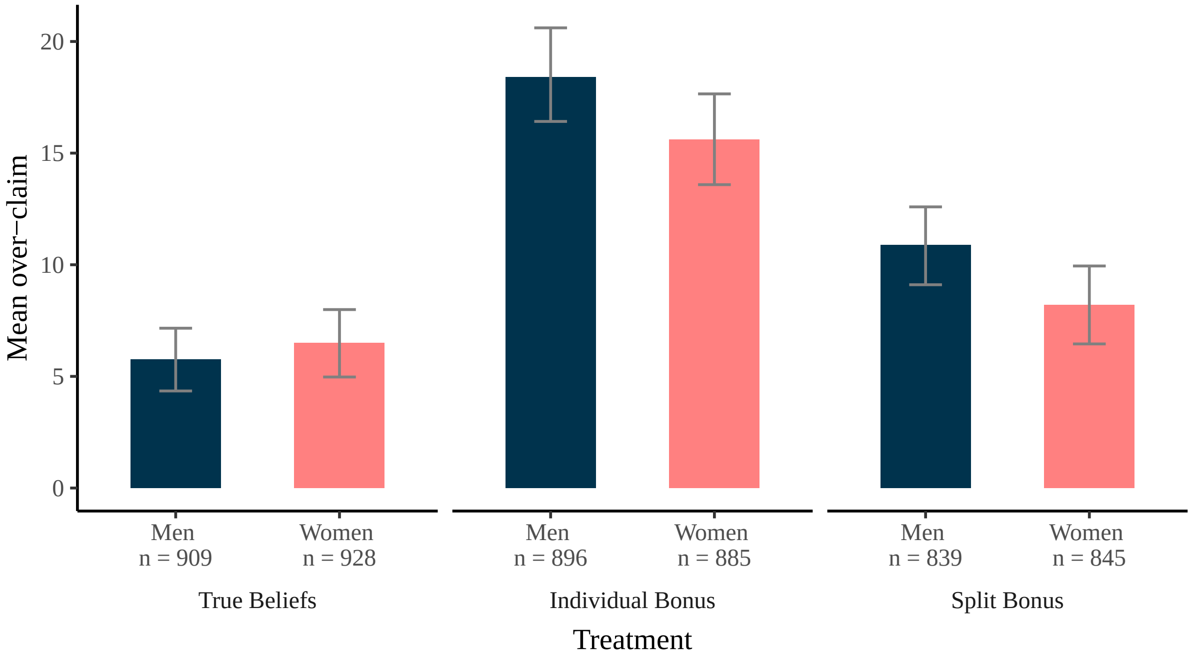
Table D12: Pairwise treatment comparisons

	Dependent Variable:			
	Claim			
	(1)	(2)	(3)	(4)
Women			1.00	-3.06^{**}
			(0.87)	(1.48)
Individual Bonus	11.28^{***}		13.21^{***}	
	(0.85)		(1.23)	
Split Bonus		-8.58^{***}		-8.48^{***}
		(0.95)		(1.36)
Actual Contribution	0.14^{***}	0.11^{***}	0.14^{***}	0.11^{***}
	(0.02)	(0.02)	(0.02)	(0.02)
Round 2	0.63	0.62	0.63	0.62
	(0.39)	(0.40)	(0.39)	(0.40)
Round 3	0.53	1.47^{***}	0.53	1.47^{***}
	(0.54)	(0.53)	(0.54)	(0.53)
UK	1.56^{*}	0.39	1.41	-0.12
	(0.84)	(1.22)	(0.87)	(1.23)
Women × Indi. Bonus			-3.77^{**}	
			(1.67)	
Women × Split Bonus				-0.004
				(1.85)
Constant	47.41^{***}	60.77^{***}	47.00^{***}	62.66^{***}
	(1.20)	(1.56)	(1.35)	(1.85)
Sample	True & Indi.	Indi & True	True & Indi.	Indi. & True
Observations	3,618	3,465	3,618	3,465
R^2	0.15	0.12	0.15	0.12
Adjusted R^2	0.15	0.12	0.15	0.12

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

D.11 Mean over-claim by treatment and gender

Figure D3: Mean over-claim of men and women in each treatment across all three rounds



Note: Over-claim is the difference between claim and actual contribution. Whiskers mark participant-clustered bootstrapped 95% confidence intervals. The sample includes only solved puzzles.

E Bayesian Estimation of the Gender Gap in Beliefs

The True Beliefs Treatment is very similar in design to the experiment in Isaksson (2018). The main difference is that in Isaksson, participants know the gender of their team member, while in the present experiment they do not. In this section, we will follow a Bayesian approach to update our posterior estimate of the gender gap in beliefs about contributions to teamwork with the evidence from this experiment, while using the estimate from Isaksson as a prior. This part of the analysis was not preregistered.

Isaksson estimates a gender gap of 4.40 percentage points, with a standard error of 1.79.[3] We estimate the following Bayesian regression with a burn-in of 1000 MCMC iterations and 40,000 iterations after burn-in:

$$claim = \beta_0 + \beta_1 women + \beta_2 ac + \beta_3 uk$$

As a prior for the coefficient for *women* we use a normal distribution with mean -4.40 and standard deviation 1.79. For all other coefficients we use an uninformed prior with mean 0.

Figure E1: Prior and posterior estimate of the gender gap in beliefs

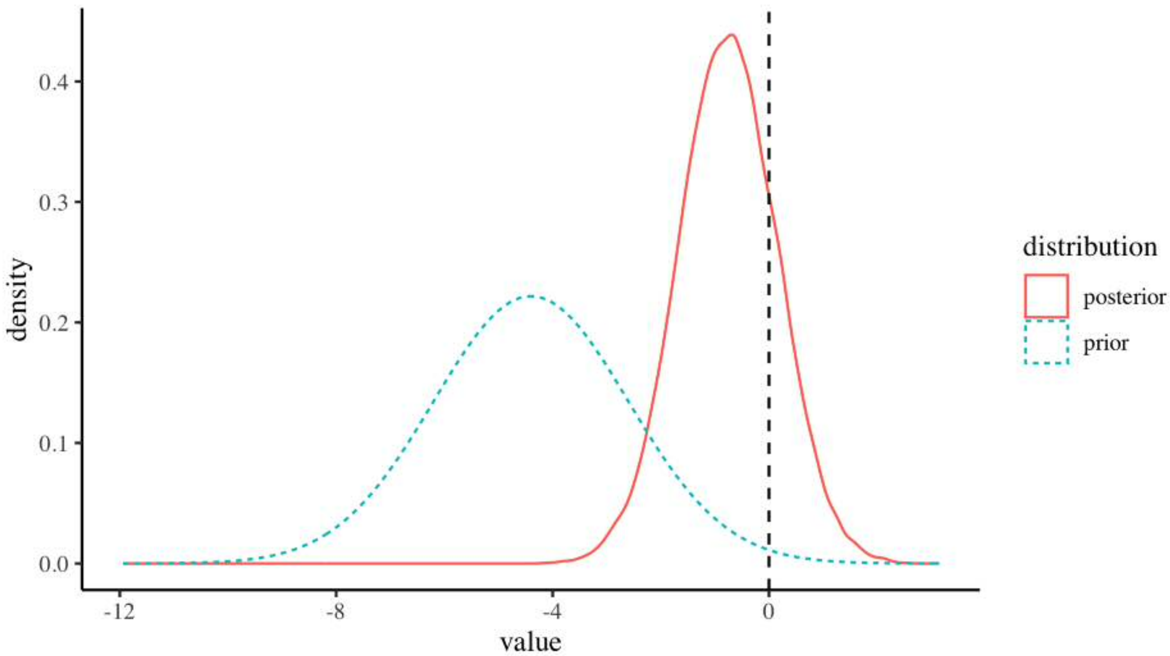


Figure E1 shows the prior and posterior distributions for the coefficient for *women*. The posterior distribution has a mean of -0.76 and a standard deviation of 0.90.

We can use the prior and posterior distributions to provide evidence for or against the hypothesis that the mean gender gap is 0. For this, we compute the Bayes factor against the null hypothesis $\beta_1 = 0$ based on the prior and posterior. The Bayes factor describes the change in the posterior density relative to the prior density at the null value, and can thus be interpreted as evidence in favor of or against the null hypothesis. We obtain a Bayes factor of 0.035 for the alternative hypothesis (or 28.57 for the null hypothesis). This is strong evidence in favor of the

[3] See Isaksson (2018) Table 4.

null hypothesis (see Andraszewicz et al., 2015).

Under the prior assumption that the gender gap in claiming credit when the gender of the team member is unknown is equal to the gender gap when the gender of the team member is known, estimated by Isaksson (2018), we use a Bayesian approach to update our estimate of the gender gap with the evidence from this experiment. The posterior is close to 0, and 0 is well within the Highest Density Interval (HDI) for a 95% credible interval. Further, the Bayes factor indicates that the null hypothesis is 28.57 times more likely at the posterior than at the prior. Our results thus provides strong evidence in favor of the null hypothesis.