Experimental Economics (2026), 1–25
doi:10.1017/eec.2026.10063

economic science association

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL PAPER

Borrowed plumes: the gender gap in claiming credit for teamwork

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

[1]Vienna University of Economics and Business: Wirtschaftsuniversitat Wien, Vienna, Austria
[2]Institute for Advanced Studies Vienna, Vienna, Austria
Corresponding author: Anna Walter-Dockx; Email: anna.walter-dockx@wu.ac.at

(Received 5 December 2024 UTC; revised 7 October 2025 UTC; accepted 14 September 2026 UTC)

Abstract

Teamwork is prevalent in many forms of production, but individual contributions to team output are often ambiguous. We use a large-scale online experiment with an interactive team task to investigate gender differences in claiming credit for contributions to teamwork with misaligned incentives: Team members have an incentive to exaggerate their contribution, but this harms the other team member. In this setting, we observe that men claim to have contributed more than equally contributing women. Using two further between-subject treatments, we find no evidence that gender differences in altruism or in overconfidence explain the observed gender claim-gap. Further, the gender gap is present only among high contributors, and we show that the gap is particularly pronounced for large claims, which has potentially important consequences for gender equity in labour market outcomes.

Keywords: experiment; gender differences; incentives; teamwork; overconfidence

JEL codes: J16; C91; D91

1. Introduction

Teamwork is prevalent in a large share of modern production (Deming, 2017). Individual contributions to team output are often ambiguous, both to team members themselves and to outside observers, giving particular weight to self-evaluations. Credit attribution for teamwork can have substantial consequences for career advancement (Sarsons et al., 2021). At the same time, by the nature of teamwork, credit attributions are a zero-sum game. Thus, team members often face conflicting incentives when making claims about their contributions to teamwork: On the one hand, there is an incentive to exaggerate one's own contributions; on the other hand, exaggerating can harm other team members by diminishing their recognition. This creates a unique and complex setting of misaligned incentives, which so far has received only limited scholarly attention. Sarsons et al. (2021) find that women are attributed less credit for their contributions to teamwork when they work together with men. A possible source for this bias could be a gender difference in claiming credit. Recent research shows that men and women differ in self-evaluations, both for individual work (Exley & Kessler, 2022) and teamwork (Isaksson, 2018), and that men are more strategic in their self-evaluations (Brilon et al., 2024; Charness et al., 2018). If men and women differ in claiming credit for their contributions to teamwork, this could add to the gender pay gap.

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

In this paper, we study gender differences in claiming credit for contributions to teamwork under misaligned incentives. We conduct a large-scale online experiment where participants interactively solve a puzzle in teams of two and then claim credit for their contribution. Our design allows us to measure each team member's actual contribution and compare claims of equally performing men and women. In the Split Bonus Treatment, we mirror the misaligned incentive structure observed in many labour market settings: It is in participants' financial best interest to exaggerate their contribution, but exaggerating harms the other team member. In this setting, we find that men claim to have contributed more than equally contributing women.

Two plausible explanations for this gender gap are that women care more about the other team member's payoff than men (i.e., a difference in altruism) or that men and women differ in their self-assessment abilities (i.e., a difference in overconfidence). We provide a theoretical framework to model these potential mechanisms and employ two between-subject treatments to test them.

In the Individual Bonus Treatment, the participant's claim no longer affects the other team member's payoff. Thereby, we eliminate the impact of preferences over the other team member's payoff on claims. This manipulation directly isolates what Fehr and Fischbacher (2002) call "pure altruism", i.e., where "...a person values the material resources allocated to a relevant reference agent positively" (page 3). We discuss other social preferences, such as inequality aversion and how they might affect claims in Section 4.2.1. The comparison of the Individual Bonus with the Split Bonus Treatment identifies the influence of altruism on claims. The gender claim-gap persists; thus, we find no support for this mechanism being the primary driver for the gender gap in claims.

To test if the gender gap is caused by differences in self-assessment abilities, in the Accuracy Treatment, we incentivise participants to estimate their contribution accurately. Following previous literature, we measure overconfidence as the difference between incentivised beliefs and actual performance (e.g., Bordalo et al., 2019) (see also Section 3). We find no gender difference in claims in the Accuracy Treatment. Because of this, we conclude that differences in overconfidence are not the main driver of the gender claim-gap.

We acknowledge that several other mechanisms could explain the gender claim-gap: Gender differences in lying aversion, in the weight given to payoff relative to other preferences, and inequality aversion could also explain our findings. In this paper, we experimentally test the mechanisms of altruism and of overconfidence with high statistical power rather than testing more channels with insufficient power. We provide exploratory evidence suggesting that gender differences in lying aversion, payoff weights, and inequality aversion may indeed be important mechanisms.

This paper contributes to the literature on gender differences in self-assessments. Previous research has focused on settings with financial incentives to be accurate (Bordalo et al., 2019; Isaksson, 2018; Ring et al., 2016; Sutter & Glätzle-Rützler, 2015), no financial incentives at all (Buser et al., 2020; Haynes & Heilman, 2013; Saccardo et al., 2018), or compared both these schemes (Neyse et al., 2016). A number of studies also investigate the effect of making self-assessments public (Ewers & Zimmermann, 2015; Haeckl, 2022; Ludwig et al., 2017). However, there is limited research on self-assessments with financial incentives to exaggerate. A notable exception is Exley & Kessler (2022), who study gender differences in self-promotion. In their experiment, participants evaluate their own performance subjectively, and this evaluation is later shown to "potential employers". Subjects thus have an incentive to evaluate themselves favourably. Whereas Exley & Kessler (2022) study qualitative, and thus inherently subjective self-evaluations, our experimental design includes an objectively correct claim, which we term the participant's "actual contribution". This enables us to make comparisons that would not be possible otherwise: First, we can compare claims to actual contributions. Second, we can compare claims under the incentive to exaggerate with claims under the incentive to be accurate in the same task and setting. Exley & Kessler (2022) find that when there is an incentive to exaggerate, men evaluate themselves more favourably than women. Our results indicate that this pattern also extends to a teamwork setting with misaligned incentives.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

The paper also contributes to the literature on gender differences in social preferences. Isaksson (2018) studies gender differences in claiming credit for teamwork, absent the influence of social preferences. In a setting where participants know the gender of the other team member and there is an incentive to self-evaluate accurately, Isaksson (2018) finds that men believe they have contributed more than equally contributing women. In our experiment, we introduce misaligned incentives between team members such that the utility derived from the other team member's payoff influences claims. Participants can exaggerate their contribution and increase their own expected payoff by taking away from the other team member. This mirrors a key attribute of teamwork: credit for contributions is a zero-sum game—i.e., the more one team member gets, the less remains for others. Prior research on gender differences in social preferences suggests that men and women may take externalities on their team member into account differently (Croson and Gneezy, 2009). Exploiting our treatment variation, we show that the gender gap in credit claiming remains, irrespective of whether claims affect the other team member or not. This suggests that gender differences in altruism are not a primary driver behind the gender claim-gap.

We further contribute to the emerging strand of literature documenting "distributional" gender differences, that is, the distributions of behaviour of men and women differ more than just in the mean. With our work, we add to the growing recognition that focusing solely on gender differences in means may underestimate the true extent of these differences (Thöni & Volk, 2021). We find that gender differences in credit claiming are most pronounced at the upper tail of the claim-distribution, that is, for very large claims. Further, we only find a gender gap among high contributors. This has potentially important consequences for gender equity, because large contribution claims and claims by high contributors likely have the most substantial impact on labour market outcomes.

Our findings have important implications for the design of performance evaluation systems, especially in settings where individual contributions to teamwork are difficult to observe. First, misattributed credit in teams can lead to loss of allocative efficiency, discouraging honest effort while encouraging strategic exaggeration. We show that team members respond to incentives to exaggerate, leading to claims significantly above actual contributions. As accuracy incentives reduce but do not entirely eliminate overclaiming, evaluation systems should rely less on self-reports and instead place greater weight on objective, transparent measures. At the same time, alternative evaluators such as peers or supervisors may themselves be subject to attribution biases (e.g., Sarsons et al., 2021). Such evaluations should therefore be viewed as mitigating rather than eliminating potential sources of bias.

Second, gender differences in claiming credit may add to the gender pay gap. We find that when there is an incentive to exaggerate, men claim to have contributed more than equally contributing women. Thus, performance evaluations that rely on self-assessments may reinforce gender disparities. However, the disappearance of the gender gap under accuracy incentives suggests a practical remedy: Incentive systems that reward accuracy or rely on objective measures can help to promote fair and equitable outcomes.

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 the results in Section 4. Section 5 discusses the 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-subject treatment design to vary the incentives for claiming contributions to the team task. The experiment was preregistered at the OSF registry.[1] The following section describes the team task before explaining all steps of the experiment in more detail.

[1] https://doi.org/10.17605/OSF.IO/gpdny

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

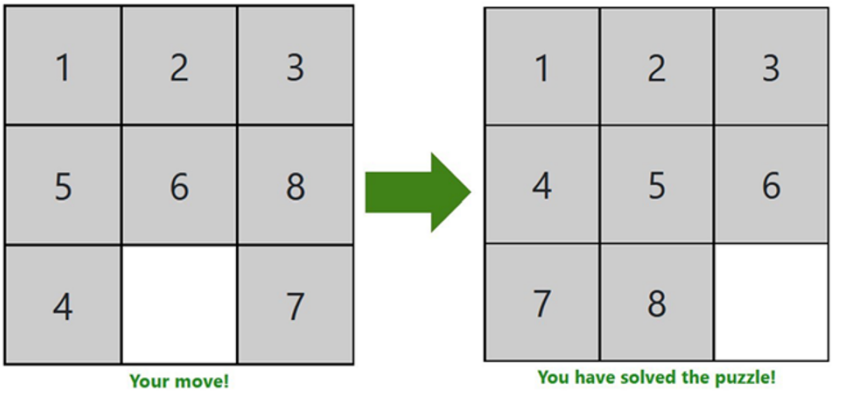


Fig. 1 Example of a sliding puzzle
Note: An unsolved puzzle (left), and a solved puzzle (right).

2.1. The team task

Participants worked on an incentivised, interactive, real-effort *sliding-puzzle*. This task is also used in other experiments (e.g., Bartoš et al., 2025; Isaksson, 2018; Takahashi, 2025). 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 use the minimal number of moves (Isaksson, 2018). Participants had no information about the other team member and no means of communicating.

The sliding-puzzle task has two useful characteristics. First, the task is gender-neutral: on average, men and women are equally good at solving the puzzle (Isaksson, 2018) and also hold gender-neutral beliefs about ability in the puzzle, that is, both men and women believe there are no gender differences in solving the puzzle (Takahashi, 2025). The mean number of individually solved puzzles in our sample is 5.10 for men and 5.11 for women ($p = 0.89$ for a two-sided two-sample t-test). Second, participants' actual contributions are directly observed by the researchers but not by the participants. While participants knew how their actual contribution was defined (see Equation (1) below), it was challenging to compute it while solving the puzzle. Hence, there was ambiguity regarding both team members' contributions. Participants thus also had to rely on subjective judgment for their claim. This is supported by the small but positive and statistically significant correlation between actual contribution and claims across all treatments ($r = 0.12, p < 0.01$). The correlation does not differ between genders ($p = 0.61$).[2]

2.2. Defining actual contribution

Given the large variety in conceptualisations of contributions, it is important to clearly define and communicate how contribution is measured in this specific setting. To ensure that participants had a common understanding, we gave participants a simple, coherent definition of *actual contribution* before the team task: For every solvable permutation of the puzzle, there exists a minimal number of moves required to reach the solution. Following Isaksson (2018), we define a *good move* as one which reduces, and a *bad move* as one which increases the minimal number of moves needed to solve the puzzle. There are no neutral moves. The categorisation of moves into good and bad moves allows

[2]P-value for an interaction term between gender and actual contribution in a panel regression of claims on gender and actual contribution.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

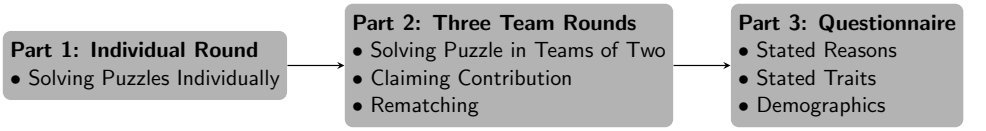


Fig. 2 Experimental sequence

for measuring each team member's individual contribution to solving the team task. Conditional on having solved the puzzle, team member A's *actual contribution* is defined as:

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

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, for example when one participant had made zero good moves, we truncated the range of actual contributions at 0 and 100. We ensured that participants knew and understood how contributions are measured using comprehension check questions and a practice example.

We restrict our main analysis to solved puzzles (roughly 70 % of all puzzles, see Section 2.4).[3] Two key factors drove this decision. First, our research question is primarily concerned with successful teamwork, because successful teamwork is most relevant in the labour market (Isaksson, 2018). Second, assessing contributions to unsolved puzzles requires making additional auxiliary assumptions, which may complicate our analysis. For instance, actual contribution is undefined for an unsolved puzzle where the total sum of good and bad moves is 0. In Section 4.3 we report exploratory analyses of unsolved puzzles.

2.3. Experiment procedure

The following paragraphs explain the procedure of the experiment in detail, as visually represented in Figure 2. The participant instructions can be found in "Online Resource 1".

2.3.1. 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 (Isaksson, 2018). Afterwards, participants were introduced to the contribution formula and asked to calculate the actual contribution in an example as a comprehension check.

2.3.2. 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

[3] We do not find evidence that this sample restriction introduces a bias by giving more weight to more skilled participants, as there is no statistically 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). In principle, the sample restriction could also bias gender comparisons if it leads to a different selection of men than women. We test whether this is the case in several ways, but do not find evidence suggesting a bias: There is no gender difference in individual skill in the sample of solved puzzles ($p = 0.89$ for a two-sample t-test) or in the unrestricted sample ($p = 0.90$), nor is there a gender difference in the probability of solving team puzzles (p= 0.27 for a χ^2-test. That is, we do not find differences in attrition by gender. Further, we test whether the relationship between claims and individual skill differs between men and women in Appendix B.11, Table 16, but do not find a difference. Lastly, in Table 4 we estimate regressions of claims on gender using the unrestricted sample of solved and unsolved puzzles. The gender differences are statistically significant and go in the same direction as for the sample of solved puzzles.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

are anonymous, i.e., team members know they are playing with another person, but they do not know each other's gender or other personal details.

Part 2 was repeated over three rounds, with rematching between all rounds. Collecting repeated observations over three rounds provides the following advantages beyond the gain in observations: First, we can 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 unsolved puzzles. With each of the rounds, puzzles got more difficult, as measured by the minimal number of moves required to solve the puzzle (from six to seven to ten moves). We did this because for more difficult puzzles, the optimal strategy is less salient, leaving more room for error and subjectivity in claims. Importantly, contribution claims remain comparable across rounds because they depend on the relative, not the absolute number of good and bad moves.

Teams had two minutes to solve a sliding-puzzle. After every round, participants were asked to claim their individual contribution. They were instructed as follows. The incentives for the respective treatments were explained before we asked participants to enter their claim: "Please make your claim about your contribution to the puzzle task on a scale from 0% to 100%". Participants did not receive feedback about their own or their team member's actual contribution. Claims were not revealed either, and this was common knowledge. We ensured understanding through a main comprehension check where participants had to confirm that their claim would not be seen by the other team member and demonstrate knowledge about their incentive scheme, that is, whether their claim could affect the other team member's bonus and whether accuracy of claims mattered as laid out in the section below. The computer randomly selected one of the three rounds for the payoff. Each team member got 25 points if the team had solved the puzzle in the selected round.

2.3.3. Treatments

Regardless of solving the puzzle, participants could earn a bonus payment for their claim in the selected round in all treatments. We varied the incentives participants had to claim their contribution in three between-subject treatments. Note that in any of the treatments, the financial incentive to exaggerate did not depend on what the other team member claimed.

In the *Split Bonus Treatment*, the expected value of participants' bonus payments increased with the claim, but at the cost of the other team member. After both team members had submitted their claims, one team member was randomly selected whose claim determined the bonus payments for both team members. If participant i's claim $c_i \in [0, 1]$ was selected, then i's bonus payment was $p_i = 100c_i$ and their team member j's bonus payment was $p_j = 100 - p_i$. We explained this to the participants in the instructions, saying that: "If your claim determines both bonuses, then you will get your own claim in points as a bonus, and the other participant will get the remainder (i.e., 100 points minus your claim). This means your claim affects the other participant's bonus. If the other participant's claim determines both bonuses, then the other participant will get her/his claim in points, and you will get the remainder (i.e., 100 points minus the other participant's claim)."

In the *Individual Bonus Treatment*, we kept the financial incentive to exaggerate, but we removed the link between the participant's claim and the other team member's payoff. That is, we excluded the effect of altruism on claims by design. Like in the Split Bonus Treatment, the expected value of participants' bonus payments increased with their claim. That is, $p_i = 100c_i$, where p_i is i's payoff, and $c_i \in [0, 1]$ is i's claim. Participants in the Individual Bonus Treatment thus had a financial incentive to exaggerate. However, each team member received a bonus from an individual budget, that is, the claims are not zero-sum. We instructed the participants, stating that: "[...] each of you will get your own claim in points as a bonus, no matter what the other participant stated. That is, if you claim to have contributed X%, you will get a bonus of X points, no matter what the other participant stated.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Similarly, if the other participant claims to have contributed Y%, she/he will get a bonus of Y points, no matter what you stated. Your claim does not affect the other participant's bonus."

With the *Accuracy Treatment*, we elicited participants' ability to estimate their own contribution under incentives to be accurate. Following the literature, we define overconfidence as an overestimation in incentivised self-assessments (e.g., Bordalo et al., 2019; Cavalan et al., 2022; Grosch & Fischer, 2024; Niederle & Vesterlund, 2007). Operationally, we measure overconfidence as the difference between a participant's claim in the Accuracy Treatment and their actual contribution. Participants earned a bonus payment that was greater in expected value the closer their claim was to their actual contribution. It was thus in participants' financial best interest to report their best estimate of their own contribution. To keep altruistic concerns towards the other team member eliminated, the claim did not affect the team member's payoff. The bonus was determined according to the binarised scoring rule (BSR) (Hossain & Okui, 2013).

The BSR is a "proper scoring rule" independent of risk attitudes, but may facilitate distortion in beliefs if participants receive elaborate information on the incentives (Danz et al., 2022). To avoid that, we used simple, verbal instructions stating that "[...] the closer your claim is to your own actual contribution, the higher your chances of receiving a bonus of 100 points. That is, the more accurate your claim, the higher your bonus, so you should claim what you truly think your contribution was." We also highlighted that "Your claim does not affect the other participant's bonus." Participants could read a more detailed description of the payoff function by clicking a button, but few did so (128 out of 845 participants).

Note that in the Split Bonus Treatment, a participant's claim is payoff-relevant with a 50% chance, which allows obtaining relevant claims from both participants. To keep the three treatments comparable, in the Individual Bonus and the Accuracy 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 indicated in the treatment descriptions above.

Randomisation 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) program was run. We employed a between-subject design to avoid experimenter-demand effects, 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 where they are not.

2.3.4. Part 3: post-experiment questionnaire

After the three team rounds, participants filled out a short questionnaire. We asked participants to "[p]lease explain in a few words how you decided how much to claim". The open responses to this question were sorted into seven pre-specified, non-exclusive categories by two independent research assistants who were blind to treatment and gender. The categories were developed by the authors based on a random subset of 30 responses, without knowing respondents' details. They capture the most relevant reasons for making claims and include whether the decision was based on perceived contribution, fairness concerns, splitting credit equally, taking blame for a mistake, blaming the other team member for a mistake, maximising one's payoff and deciding randomly. For a more detailed description of each category see Appendix D. The interrater-reliability is moderate (κ above 0.48% for all categories).[4] We use a conservative coding approach in that a statement falls into a category only if both research assistants selected it. Next, we collected an indicative measure of other preferences and personal traits. We refer to these measures as *stated traits*. We asked participants whether they agree or disagree with several statements on a five-point Likert scale from 1 (*strongly disagree*)

[4] κ for contribution is 0.51%, for fairness 0.87%, for splitting equally 0.48%, for taking blame 0.69%, for blaming the other 0.71%, for maximising payoff 0.70% and for random reason 0.75%

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

to 5 (*strongly agree*). In this way, we elicited stated advantageous/disadvantageous inequality aversion ("I dislike if I get a greater/smaller 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), meritocratic fairness principles ("Those participants who contributed more should get a greater bonus"), egalitarian fairness principles ("All participants should get the same bonus regardless of their contribution"), and team identification ("I felt connected with the other participants"). We also asked participants how much they believed the other team member had claimed in the final round. This question was incentivised. Participants also reported their age, level of education, income, and gender. While we recruited only participants who identify with a binary gender on Prolific, 33 participants selected "other" as their gender in the post-experiment questionnaire. In our analysis, for these participants, we use the binary gender they had indicated on their Prolific profile. A detailed description of all collected variables can be found in Appendix A. No feedback was given during the entire experiment. Only after answering the questionnaire, participants were informed about their earnings. We provided only vague information about the composition of total earnings to keep the other team member's claim in the Split Bonus Treatment ambiguous.

2.3.5. Generalisability

We study credit claiming for teamwork in an experiment because this gives us precise control over participants' decision context. The experimental design mimics the key features of the workplace setting we are interested in while abstracting from other aspects. At a basic level, a professional teamwork setting involves (1) some jointly and often interactively produced team output (Isaksson, 2018) to which (2) contribution is zero-sum and often ambiguous (Haynes & Heilman, 2013). Other common but not ubiquitous features are (3) some social context, including behavioural norms and (4) an ongoing relationship between team members. With (1) and (2), the basic credit-claiming decision is affected by individual-level factors such as payoff maximisation and altruism towards the other team member. With (3) and (4), extensions such as social norms and reputation concerns may play a role. It is crucial to understand the basic setting before extending to richer settings. Because of this, the experiment incorporates features (1) and (2), while deliberately removing (3) and (4). Team output is jointly and interactively produced, and contributions to the output are clearly defined and zero-sum and, at the same time, ambiguous to team members. The interactive nature of our team task distinguishes this paper from much previous experimental research on teamwork, which uses joint individual production tasks (Bruhin et al., 2024; Cavalan et al., 2022; Haynes & Heilman, 2013; Solda et al., 2020). We remove reputation concerns by keeping participants anonymous and minimising social context. This allows us to study the impact of altruism and overconfidence on credit claiming in a cleanly identified way. We discuss which mechanisms our experiment can identify in more detail in Section 3. It is plausible that (3) and (4) also affect credit-claiming behaviour. Whether and in what way this is the case remains an empirical question. We discuss possible extensions of our design to address these questions in Section 6.

2.3.6. 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" 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 of working age (18 to 66 years) to maximise the eligible sample. To ensure high data quality, other pre-screeners included speaking English fluently and a high Prolific completion score ($>$ 90% of begun studies successfully completed). To maximise 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. Independent

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Table 1. Treatment group comparisons

	Split Bonus	Individual Bonus	Accuracy	
Participants	822	789	845	
Age	32.12	30.93	31.02	$<0.01^{***}$
Women	50.49	50.57	50.77	0.99
UK	88.44	69.20	55.27	$<0.01^{***}$
Individual skill	5.12	5.16	5.05	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*.

of the bonus payment, participants earned a completion fee of £1.20. The median total earnings were £1.64, and the median duration of the experiment was 11.83 minutes. This corresponds to an hourly wage of £8.32, which exceeds the platform's allowed minimum pay (£5.00) and recommended pay (£7.50) per hour at the time of data collection. Before the online experiment, we ran pilot sessions to ensure comprehensibility of the instructions and a functioning technical setup. Observations from the pilot are not included in the analysis.

2.4. Descriptive statistics

In our main sample, we recruited 2747 participants who each participated in three rounds, giving us a total of 8241 observations. According to an ex-ante power analysis, 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 B.1). For all analyses, we exclude 219 participants who failed the main comprehension check twice, because we cannot assume that these participants have understood the incentives. For the main analysis, we restrict the sample to solved puzzles. 80.34% of all 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. Across all three rounds, this leaves us with 5302 observations of credit claims made by 2456 unique participants. 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 (individual skill) across the three treatment groups. The percentage of women is close to 50% in all three groups. The mean age is statistically significantly different between groups, but the difference is negligibly small. We control for age in Appendix B.2. We randomly assigned treatments on the session level, and we increased the size of sessions over time. Because we unexpectedly depleted the pool of active UK subjects on Prolific and had to resort to recruiting US residents, the share of UK residents is highly statistically significantly uneven between treatment groups. To account for this, we control for country of residence in all the following regressions. We report country-specific estimates in Appendix B.3. However, as country differences are not part of our original, pre-registered research question and because of the uneven distribution of countries between treatment groups, we refrain from interpreting any country-specific effects here and invite future research to study this.

The number of puzzles solved individually in one minute indicates individual skill (see Section 2.3.1). We find no difference in individual skill between men and women. Men solve 5.10 puzzles, and women solve 5.11 puzzles on average ($p= 0.89$).

3. Theoretical model and hypotheses

In this section, we introduce a simple model of claiming behaviour, which serves as a framework for analysing the factors that influence claims and for formulating our hypotheses. Individual i chooses her claim $c_i \in [0, 1]$ to maximise her utility:

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

$$U_i = \pi p_i^T(c_i) + \lambda p_j^T(c_i) - \kappa(c_i - b_i)^2$$

$$b_i = a_i + e; e \sim \mathcal{N}(\mu, \sigma^2)$$

where $\pi \geq 0$ is the weight given to payoff, $\lambda \geq 0$ is the weight given to altruism, and $\kappa \geq 0$ is the weight given to accurate claims. $b_i \in [0, 1]$ is i's true belief about her contribution, which depends on her actual contribution $a_i \in [0, 1]$ and an error term e. p_i^T is i's payoff, which depends on c_i according to treatment $T \in \{S, I, A\}$ (Split Bonus, Individual Bonus, and Accuracy):

$$p_i^S = 100c_i;\ p_j^S = 100 - 100c_i$$

$$p_i^I = 100c_i$$

$$p_i^A = \begin{cases} 100, & \text{if } l \leq n \\ 0, & \text{if } l > n \end{cases} \text{ where } l = (c_i - a_i)^2; n \sim \mathcal{U}[0, 1]$$

p_j^T is the other team member j's payoff, which depends on c_i only in the Split Bonus Treatment and is independent of c_i otherwise. λ thus captures the weight of utility derived from the other team member's payoff, that is, altruism. We abstract from other social preferences, such as fairness, inequality aversion (Fehr & Schmidt, 1999), or adherence to social norms, and so on at this point. $-\kappa(c_i - b_i)^2$ represents the intrinsic disutility from claims that deviate from beliefs. In sum, when i chooses her claim, she considers her own payoff, the other team member's payoff and her belief about her contribution. p_i^A is the binarised scoring rule (Hossain & Okui, 2013), such that $\frac{\partial \mathbb{E}(p_i^A)}{\partial |c_i - b_i|} < 0$. If we assume purely selfish preferences such that $\pi > 0, \lambda = \kappa = 0$, i.e., i only cares about her own payoff, then i claims 100% in Split Bonus and Individual Bonus and her belief in Accuracy $c_i^S = 1, c_i^I = 1, c_i^A = b_i$.

Next, let us assume there are social preferences and a preference for accurate claims, that is, $\lambda > 0,\ \kappa > 0$. We begin by considering the Split Bonus Treatment, where there are misaligned incentives between team members. Previous research has found that women behave more socially in allocation tasks. Andreoni and Vesterlund (2001) document that women behave more kindly than men in a modified dictator game, where a benevolent allocation is costly. Halladay and Landsman (2024) find that women take less than men in a taking-framed dictator game with earned endowments. See Croson and Gneezy (2009) for an overview of the literature. This suggests that women may have stronger altruistic preferences than men in our setting: $\lambda_W > \lambda_M$ (where W denotes women and M denotes men). All else equal, $\lambda_W > \lambda_M \implies c_W^S < c_M^S$. Other research finds gender differences in willingness to lie (Abeler et al., 2019; Conrads et al., 2013; Dreber & Johannesson, 2008; Friesen & Gangadharan, 2012; Grosch & Rau, 2017). Women are less willing to lie than men when lying has negative externalities (Capraro, 2018; Erat & Gneezy, 2012). This suggests that $\kappa_W > \kappa_M$, which implies $c_W^S - b_W < c_M^S - b_M$. If we assume that $b_W = b_M$, then $c_W^S < c_M^S$, c.p. Further, research on self-assessments has found men to be more overconfident than women (Bordalo et al., 2016; Bordalo et al., 2019; Exley & Kessler, 2022; Isaksson, 2018), with some notable exceptions which find no gender difference (Brilon et al., 2024) or find that gender differences are context dependent (Grosch & Fischer, 2024; Haeckl, 2022; Ludwig et al., 2017). If men are more overconfident than women, their belief will be biased upwards such that $\mu_M > \mu_W \implies b_M > b_W \implies c_M > c_W$, c.p. Based on the implications of both altruism and overconfidence, we derive the following hypothesis.

Hypothesis 1: When there is an incentive to exaggerate at the cost of the team member, men claim more than women after controlling for actual contribution.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

In the Individual Bonus Treatment, we remove the link between c_i and p_j. Thus, altruistic preferences λ no longer affect claims. Because of this, the comparison to the Split Bonus Treatment identifies the effect of λ on claims. In particular, if a gender claim-gap is caused by gender differences in altruism, then this will be captured by the difference-in-differences in claims between men and women in the Individual Bonus and Split Bonus treatments, i.e $\lambda_W > \lambda_M \iff (c_M^S - c_W^S) - (c_M^I - c_W^I) > 0$. With $\lambda > 0$, claims will also be greater if they do not negatively affect the other team member's payoff, i.e., $c^I > c^S$. We therefore test the following hypotheses.

Hypothesis 2a: When there is an incentive to exaggerate without affecting the other team member, men claim more than women after controlling for actual contribution.

Hypothesis 2b: Men increase their claims more than women when there is an incentive to exaggerate without affecting the other team member, compared to when exaggeration is at the cost of the team member.

Another possible explanation for a gender claim-gap could be gender differences in self-assessment abilities. In the Accuracy Treatment, we give participants the incentive to make accurate claims. Here, U_i is at a maximum where $c_i^A = b_i$. Following a large strand of literature on overconfidence (Bordalo et al., 2019; Cavalan et al., 2022; Grosch & Fischer, 2024; Niederle & Vesterlund, 2007), we define overconfidence as the difference between beliefs and actual performance $b_i - a_i$. With the Accuracy Treatment, we measure overconfidence because $c_i^A = b_i \implies c_i^A - a_i = b_i - a_i$. This means that if men are more overconfident than women, men will claim more in the Accuracy Treatment than women, i.e., $b_M > b_W \iff c_M^A > c_W^A$ (controlling for actual contribution). When comparing to the Split Bonus and Individual Bonus Treatments, $\frac{\partial p_i^S}{\partial c_i}$ and $\frac{\partial p_i^I}{\partial c_i}$ are strictly positive. Hence, optimizing p_i^S and p_i^I with respect to c_i requires maximising c_i. In contrast, optimising p_i^A with respect to c_i has an inner solution depending on b_i. Therefore, $c^A < c^S$ and $c^A < c^I$. The difference-in-differences between men and women in the Accuracy and Individual Bonus Treatments identifies a potential gender difference in how much participants care about their payoff, relative to making accurate claims. That is $\pi_M > \pi_W \iff (c_M^I - c_W^I) - (c_M^A - c_W^A) > 0$, c.p. We, therefore, test the following hypotheses.

Hypothesis 3a: When there is an incentive to be accurate, men claim more than women after controlling for actual contribution.

Hypothesis 3b: Men increase their claims more than women when there is an incentive to exaggerate at the cost of the team member, compared to when there is an incentive to be accurate.[5]

Hypothesis 3c: Men increase their claims more than women when there is an incentive to exaggerate without affecting the other team member, compared to when there is an incentive to be accurate.

In the final set of hypotheses, we test the direct treatment effects predicted by the model. Previous behavioural research shows that strategic or financial incentives to exaggerate can lead people to increase self-evaluations and reported confidence (Schwardmann & Van der Weele, 2019; Solda et al., 2020; Exley & Kessler, 2022), and that people are willing to lie for financial gain (Fischbacher & Föllmi-Heusi, 2013; Abeler et al., 2019). However, individuals tend to cheat less when it harms another participant compared to when it harms the experimenter (Meub et al., 2016). Moreover, Hermann

[5]This hypothesis was not preregistered, but we include it for completeness.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

and Muß hoff (2019) find that intrinsic costs for stealing are larger than for lying and exaggerating in the Split Bonus Treatment, which might be interpreted as stealing from the other team member. At the same time, a desire-to-win effect can dominate social preference concerns (Dato et al., 2024) and might thus facilitate more lying. Taken together, we expect smaller claims when exaggerating harms the other team member (Split Bonus Treatment) compared to when the other team member is not affected (Individual Bonus Treatment), i.e., $c^S < c^I$. We expect claims to be greater when it pays to exaggerate (Split Bonus and Individual Bonus Treatments) than when it pays to be accurate (Accuracy Treatment), that is, $c^A < c^I$, $c^A < c^S$.

Hypothesis 4a: When there is an incentive to exaggerate at the cost of the other team member, individuals claim less than when the team member is not affected.

Hypothesis 4b: When there is an incentive to exaggerate without affecting the other team member, individuals claim more than when there is an incentive to be accurate.

Hypothesis 4c: When there is an incentive to exaggerate at the cost of the other team member, individuals claim more than when there is an incentive to be accurate.[6]

We preregistered the hypotheses at OSF in March 2022.[7]

4. Results

In this section, we first report the main results of testing the hypotheses listed in Section 3, before looking into exploratory analyses in subsections 4.2 and 4.3.

Table 2 shows regressions of claims on gender and treatment. In the main body of the paper, we report the results of panel regressions over the three rounds using the sample of solved puzzles, including round fixed effects, participant random effects and clustered standard errors at the participant level. This specification gives us the greatest statistical power. A potential threat to this specification would be if claiming behaviour evolves differently across rounds by gender, with increasing puzzle difficulty. We estimate regressions with round-gender interactions in Appendix B.12 and do not find evidence for systematic differences in claims across rounds between men and women. In all regressions, we control for country of residence as described in Section 2.4 and actual contribution, allowing us to compare equally contributing participants. For more conservative estimates, we also report the results of regressions using only data from round 1 in Appendix B.5, as well as regressions using the participant's average claim across all solved rounds in Appendix B.11. The results are comparable, though effect sizes are slightly smaller in round 1, and gender differences are not significant, likely because of reduced statistical power.

4.1. Gender differences in claiming credit

Are there gender differences in claiming credit for teamwork? Figure 3 shows the mean difference between claim and actual contribution (to which we refer to as over-claim) for men and women in each treatment across all rounds, using only the rounds in which the puzzle was solved for each participant. In the Split Bonus Treatment, men over-claim more than women ($p<0.01$).[8] The gap persists in the Individual Bonus Treatment ($p=0.01$), but not in the Accuracy Treatment ($p=0.24$). Appendix B.4 shows the mean claims by treatment and gender.

[6]This hypothesis was not preregistered, but we include it for completeness.
[7]We have adapted the order and slightly rephrased some of the hypotheses here for better readability.
[8]P-values for two-sample t-tests with bootstrapped participant-clustered standard errors.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Table 2. Regressions of claim on treatment and gender

	Dependent variable:				
	Claim				
	(1)	(2)	(3)	(4)	(5)
Women	−2.96***	−2.95**	1.07		−6.28***
	(1.12)	(1.49)	(0.87)		(1.82)
Individual				8.70***	8.42***
				(0.93)	(1.35)
Accuracy				−2.67***	−5.11***
				(0.76)	(1.07)
Actual contribution	0.08***	0.15***	0.13***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
UK	−0.90	0.24	2.28***	1.11	−1.13
	(1.84)	(1.61)	(0.87)	(0.77)	(1.21)
Women × Individual					0.78
					(1.86)
Women × Accuracy					5.04***
					(1.52)
Women × UK					3.73**
					(1.59)
Constant	55.81***	60.84***	46.75***	51.08***	54.58***
	(2.40)	(2.26)	(1.71)	(1.24)	(1.61)
Wald test p-values:					
Individual = Accuracy				<0.01***	<0.01***
Women × Individual = Women × Accuracy					0.01**
Sample	Split	Indi.	Accuracy	Full	Full
Observations	1,684	1,781	1,837	5,302	5,302
R^2	0.13	0.10	0.11	0.14	0.14
Adjusted R^2	0.12	0.10	0.10	0.14	0.14

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. Columns 1, 2 and 3 are estimated on the samples of the Accuracy Treatment, 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.

4.1.1. The gender claim-gap under misaligned incentives

In Table 2, we report regression estimates controlling for actual contribution. We begin with the Split Bonus Treatment, where participants have an incentive to exaggerate their contribution, but this harms the other team member. In this setting, men claim to have contributed 2.96 percentage points more than equally contributing women ($p< 0.01$, see Table 2 column 1). This is in support of Hypothesis 1.

4.1.2. Altruism does not drive the gender claim-gap

A possible mechanism for the gender claim-gap could be that women care more about the other team member's payoff than men, and are thus more reluctant to exaggerate their contribution at the expense of the other team member. In the Individual Bonus Treatment, we test this mechanism by removing the link between claims and the other team member's payoff, while keeping the financial incentive to exaggerate. If the gender claim-gap is driven by a gender difference in altruism, we would expect the gap to disappear or at least diminish in the Individual Bonus Treatment. However, the size of the gender claim-gap remains unchanged. Men claim to have contributed 2.95 percentage points

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

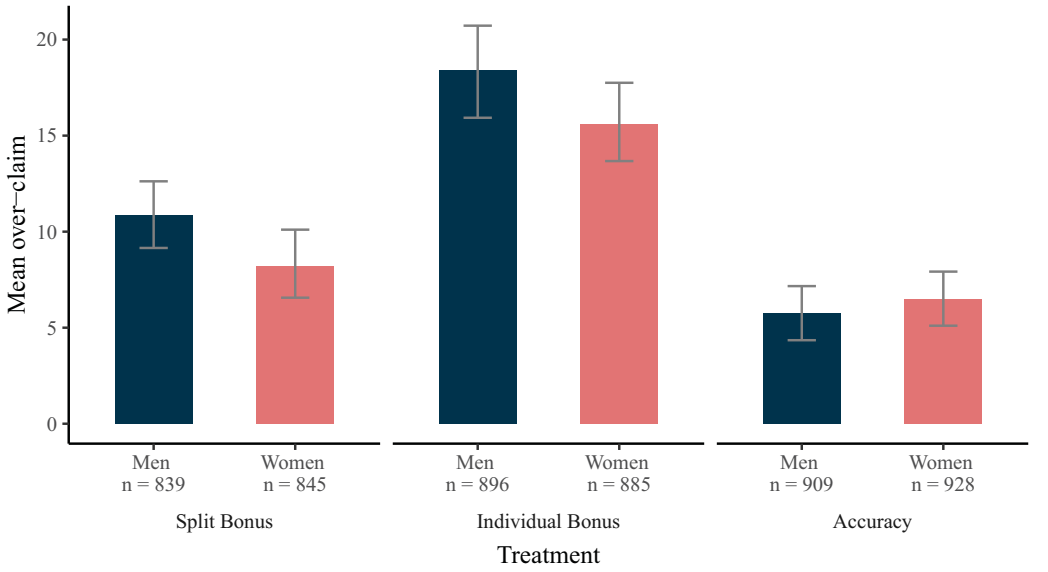


Fig. 3 Mean over-claim by gender and treatment
Note: Mean over-claim (difference between claim and actual contribution) over all three rounds (solved puzzles only), error-bars mark bootstrapped participant-clustered 95% confidence intervals.

more than equally contributing women ($p< 0.05$; see Table 2 column 2). This is evidence in support of Hypothesis 2a. The difference-in-difference between gender and treatment compared to the Split Bonus Treatment is close to 0 and not statistically significant: We cannot reject the null hypothesis for Hypothesis 2b. Since the gender claim-gap persists regardless of whether the claim affects the other team member or not, we conclude that gender differences in altruism are not the main driver of the gap.

4.1.3. No gender difference in overconfidence

Another possible mechanism for the gender claim-gap could be a gender difference in overconfidence. We measure overconfidence as the difference between claimed and actual contribution under incentives to be accurate (see Section 3). To test this mechanism, we give participants a financial incentive to make accurate claims in the Accuracy Treatment. On average, across all rounds, participants overestimate their contributions by 6.13 percentage points (statistically significantly different from 0 with $p< 0.01$ for a one-sample t-test using participant-clustered bootstrapped standard errors). However, we observe no difference between men and women in the Accuracy Treatment. Women claim 55.47% and men claim 55.24% on average ($p= 0.71$). That is, we find no evidence of a gender difference in overconfidence that would explain the gender claim-gap (no support of Hypothesis 3a). The difference-in-differences between gender and treatment is statistically significant both compared to the Split Bonus Treatment ($p< 0.01$) and compared to the Individual Bonus Treatment ($p= 0.01$), supporting Hypotheses 3b and 3c.[9]

In Appendix C we conduct a Bayesian estimation of the gender claim-gap in the Accuracy Treatment. Using the gender gap estimated by Isaksson (2018) as a conservative prior that favours the existence of a gender gap, we find strong evidence for the absence of such a gap. The Bayes factor can be interpreted as evidence in favour of the null hypothesis.

[9]The differences-in-differences in claims between treatments and gender remain statistically significant even when controlling for stated traits (see Appendix B.7).

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

4.1.4. Direct treatment effects

We also test the direct effects of the incentive schemes on claims. Compared to the Split Bonus Treatment, both men and women increase their claims in the Individual Bonus Treatment. Claims increase by 8.70 percentage points on average ($p< 0.01$; see Table 2 column 4; see also Appendix B.6 for pairwise treatment comparisons). This supports Hypothesis 4a. In the Accuracy Treatment, participants claim 11.37 percentage points less than in the Individual Bonus Treatment ($p< 0.01$) and 2.67 percentage points less than in the Split Bonus Treatment ($p< 0.01$), supporting Hypotheses 4b and 4c.

4.2. Other channels and heterogeneities (exploratory)

Above, we document that men claim to have contributed more to successful teamwork than women when incentives are misaligned between team members. As we find no evidence that gender differences in altruism or in overconfidence explain the gender claim-gap, the question of what drives the gap remains. In this subsection, we document exploratory analyses that suggest potential mechanisms and provide further insight into the heterogeneity of treatment effects.

4.2.1. Stated reasons for claims

There are several other mechanisms besides gender differences in altruism and in overconfidence that could explain the gender claim-gap. In the model in Section 3, a gender difference in lying aversion κ or in the relative importance of payoff π would also lead to a claim-gap. Previous research provides evidence for gender differences in willingness to lie (Abeler et al., 2019; Capraro, 2018; Conrads et al., 2013; Dreber & Johannesson, 2008; Erat & Gneezy, 2012; Friesen & Gangadharan, 2012; Grosch & Rau, 2017). Other possible mechanisms suggested by the literature that are not included in our model are inequality aversion (Croson and Gneezy, 2009), motivated reasoning (Thaler, 2021) or conforming to norms (Babcock et al., 2003; Coffman et al., 2024; Haeckl et al., 2025). Hereinafter, we explore data from our post-experiment questionnaire to gain some insights into which of these channels may be important.

After participants had made their claim in the third (and final) round, they were asked to state the reason for their claim in their own words. Two independent research assistants categorised these statements into seven categories (see Appendix D for category details). This methodology is commonly used to gain insights into behaviour within economic experiments and to generate new hypotheses (for a review see Haaland et al., 2025).

Figure 4 shows the proportion of men and women in each treatment who mention each reason. Across all treatments, the most common reason for claims is (perceived) contribution. Beyond this, several patterns emerge: Men more often than women base their claim on maximising their own payoff ($p< 0.10$ in Split Bonus and $p< 0.05$ in Individual Bonus) and blaming the other team member for a mistake ($p< 0.05$ in Split Bonus and $p< 0.10$ in Accuracy).[10] Women more often indicate splitting credit equally ($p< 0.05$ in Split Bonus) and allocating credit based on perceived contributions ($p< 0.10$ in Individual Bonus). Overall, these patterns are broadly in line with the potential mechanisms of (1) a gender difference in lying aversion ($\kappa_M < \kappa_W$), (2) a gender difference in preference weights for payoff ($\pi_M > \pi_W$), or (3) a gender difference in inequality aversion.

This evidence suffers from several caveats: Stated reasons were not incentivised and may be subject to social desirability bias or limited introspection ability. In addition, stated reasons are endogenous to claims and patterns are thus only correlational. Nonetheless, we consider it relevant descriptive evidence to guide further research on the channels behind the gender claim-gap.

[10] All p-values for comparisons between proportions reported in this subsection are for Fisher's exact tests.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

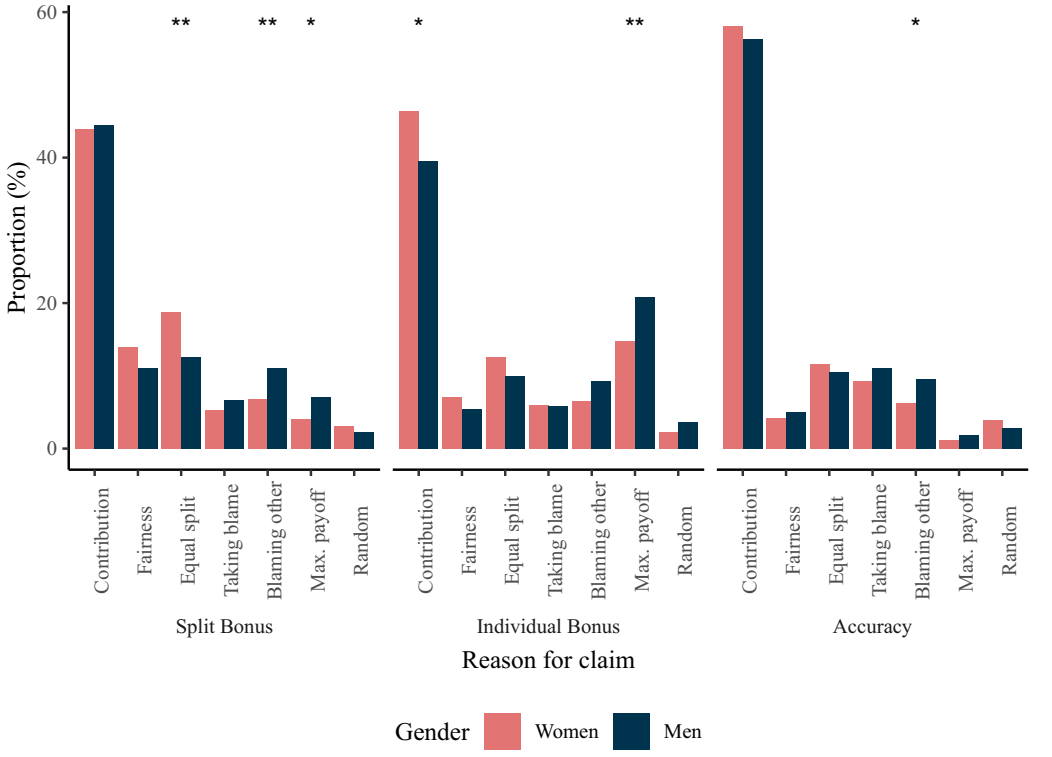


Fig. 4 Stated reasons for claims

Note: Percentage of participants whose stated reason for their claim falls into each category. Categories are non-exclusive, see Section 2 for the coding. *Contribution:* This participant based their claim on performance or effort on the puzzle. *Fairness:* This participant based their claim on what they perceive as fair. *Equal split:* This participant always wants credits to be split equally (50-50), regardless of the quality of contribution or effort. *Taking blame:* This participant based their claim on taking the blame for or acknowledging a mistake they made. *Blaming other:* This participant based their claim on a mistake the other team member made. *Max. payoff:* This participant based their claim on maximising their payoff. *Random:* This participant mentions that they did not understand the instructions and/or decided what to claim randomly. Stars indicate p-values of Fisher's exact tests for differences in proportions between men and women within the respective category. *p<0.1; **p<0.05; ***p<0.01.

4.2.2. The gender gap is more pronounced for large claims

In the following paragraphs, we provide evidence that the gender claim-gap is particularly pronounced at the upper tails of the claim distributions (that is, for large claims), and among high contributors. Figure 5 shows the cumulative density functions (CDFs) of the average claim per individual across the three rounds for each treatment. The gender gap is of moderate size at the means (Cohen's $d = 0.14$ in the Individual Bonus and $d = 0.17$ in the Split Bonus Treatment Cohen (1988)), but becomes larger at the upper tails of the claim distribution.

In the upper panel of Table 3 we estimate the gender gap in each treatment separately for the top quartile and the bottom three quartiles of the claim distribution. In the Split Bonus Treatment, the gender claim-gap is about eight times larger in the top claim-quartile than in the bottom three quartiles ($p<0.01$). In the Individual Bonus Treatment, the point estimate of the gender gap is also larger in the top quartile, but the difference is not statistically significant. We also test for general differences in the shapes of the claim distributions between men and women using Dowd Test Statistics (DTS) (Dowd, 2020). This nonparametric test is sensitive to differences in the tails of distributions and has improved power compared to the Kolmogorov-Smirnov (KS) test or Anderson-Darling (AD) test. The DTS test detects a difference in claim distributions between men and women in the Split Bonus Treatment ($p = 0.01$), but not in the Individual Bonus ($p = 0.08$) or Accuracy treatment ($p = 0.62$).

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

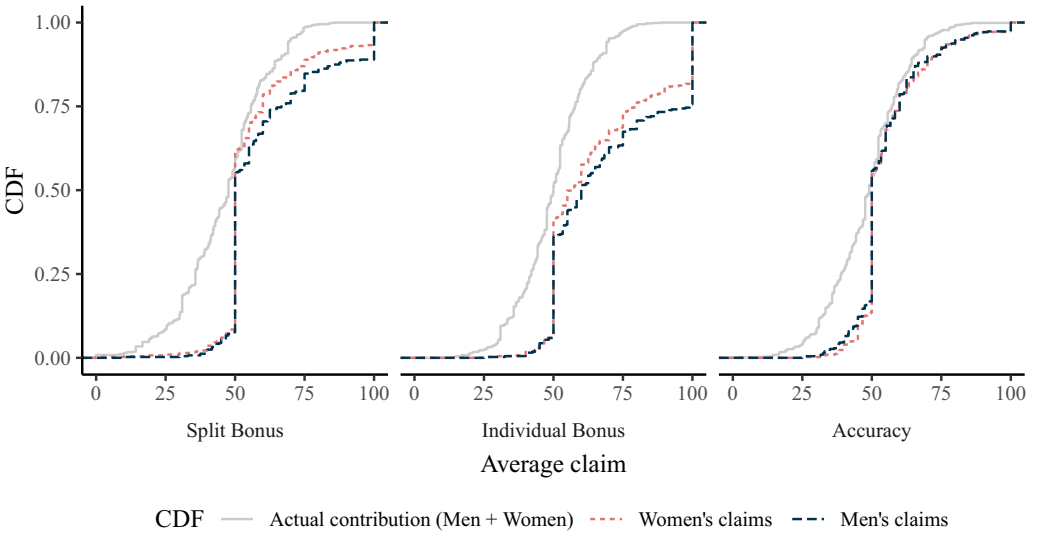


Fig. 5 CDFs of actual contributions and claims
Note: CDF or actual contributions (men and women pooled), and women's and men's average claims. The average claim is the participants' average claim across all (successful) rounds.

The difference in statistical significance between the Split Bonus and Individual Bonus Treatments may be explained by the much larger overall share of 100%-claims in the Individual Bonus Treatment (24.37% in the Individual Bonus Treatment vs. 10.69% in the Split Bonus Treatment), making it more difficult to detect distributional differences in the Individual Bonus Treatment.

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 Accuracy Treatment. We estimate logit regressions on the propensity to claim 100%, controlling for actual contribution and country of residence. Men are statistically significantly more likely to claim 100% in the Individual Bonus Treatment, but not in the Split Bonus or the Accuracy Treatment (see Appendix B.9).

4.2.3. The gender gap among high contributors

In Table 3 we additionally estimate the gender gap in claims within each treatment for high contributors (contribution $>$ 50%) and low contributors (contribution $\leq$ 50%), using the definition of Isaksson (2018). In the Split Bonus as well as the Individual Bonus Treatment, we find a statistically significant gender claim-cap among high contributors, but not among low contributors. When it pays to exaggerate, high-contributing men submit larger claims than high-contributing women (controlling for actual contribution). We cannot reject the null hypothesis of equality of coefficients between high- and low-contributors. In Appendix B.10, we also estimate regressions with an interaction between gender and high contributor as well as top claim-quartile.

4.2.4. The 50–50 norm

Previous research documents a strong norm to share payoffs along a 50-50 split, motivated by the intention to appear fair (Andreoni & Bernheim, 2009). In an identity-based utility framework (Akerlof & Kranton, 2000), individuals who identify as fair derive direct utility from choosing the equal split, making 50–50 a natural focal point. In line with this, Figure 5 reveals that a large share of participants chooses to share credit evenly (i.e., claim to have contributed 50%): The percentage of participants who, on average across all three rounds, claim to have contributed 50% is 39.55% in

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Table 3. The gender gap at the upper tails of the distributions

	Dependent Variable:		
	Claim		
	Split Bonus	Individual Bonus	Accuracy
Top quartile: Women	-8.55***	-4.79***	-0.07
	(1.65)	(0.64)	(1.38)
Bottom three quartiles: Women	-1.06***	-2.87***	0.51
	(0.33)	(0.82)	(0.37)
Z-test for equality of coefficients:	p< 0.01***	p= 0.06*	p= 0.68
High contributors: Women	-3.29***	-4.53***	-0.23
	(1.66)	(1.96)	(1.31)
Low contributors: Women	-1.79	-1.74	1.62
	(1.23)	(1.65)	(0.96)
Z-test for equality of coefficients:	p= 0.47	p= 0.28	p= 0.26

Note: Coefficient estimates for *women* (i.e, estimates of the gender gap in claims) 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. Coefficients in the top and bottom panel are from separate regressions. The top panel shows coefficient estimates from the top and bottom three quartile of claims respectively. The quartiles are constructed separately for each treatment and for men and women. The bottom panel shows coefficient estimates for high contributors (>50%) and low contributors ($\leq$ 50%) respectively. The number of observations in the top quartile are 531 in the Accuracy, 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 Accuracy, 1285 in the Individual Bonus and 1237 in the Split Bonus Treatment. The number of observations of high contributors are 584 in the Accuracy, 583 in the Individual Bonus, and 534 in the Split Bonus Treatment. The number of observations of low contributors are 1253 in the Accuracy, 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.

the Accuracy 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 B.8). When the claim directly affects the team member's payoff, the 50–50 Norm seems to grow in importance. A possible explanation is conditional reciprocity (e.g., Balafoutas et al., 2020): Participants who expect their team member to share credit evenly reciprocate by doing the same. Consistent 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%, believe the other team member claimed 50%. In subsection 4.2.1, we document that in the Split Bonus Treatment, women state wanting to split credit equally as the reason for their claim more often than men. In line with this, the proportion of participants who claim 50% is slightly larger among women than among men in the Split Bonus and Individual Bonus Treatments; yet the difference is not statistically significant (see Appendix B.8). We thus do not find evidence that the 50–50 norm contributes to the gender claim-gap. While women more frequently cite equal splitting as a reason for their claims, we do not observe a corresponding gender difference in the propensity to claim 50%. The gender gap is concentrated among participants who deviate from the 50% focal point, especially among those making large claims and among high contributors (see the previous two subsections).

4.3. Claiming credit after unsuccessful teamwork (exploratory)

So far, the analysis has focused on the sample of successfully solved puzzles. Here, we report exploratory analyses of claims for unsolved puzzles. Credit claiming for failed teamwork may differ from that for successful teamwork (Haynes & Heilman, 2013), and, in particular, men and women may respond to team failure differently. Note the following caveat: The contribution formula is not defined for certain special cases of unsolved puzzles, such as when the team's sum of good and bad

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Table 4. Unsolved puzzles

	Dependent Variable:					
	Claim					
	(1)	(2)	(3)	(4)	(5)	(6)
Unsolved	2.59	-3.78**	-11.49***			
	(1.61)	(1.63)	(1.50)			
Women	-2.98***	-3.12**	1.21	-8.79***	-9.85***	2.98
	(1.11)	(1.49)	(0.86)	(2.22)	(2.64)	(2.06)
UK	-0.98	-0.19	3.16***	-0.94	-4.77*	5.66***
	(1.93)	(1.64)	(0.91)	(4.38)	(2.76)	(2.06)
Unsolved x Women	-4.92**	-6.17***	1.00			
	(2.08)	(2.20)	(1.97)			
Constant	59.41***	65.48***	50.63***	68.81***	66.98***	41.04***
Sample	Split all	Indi. all	Accuracy all	Split unsolved	Indi. unsolved	Accuracy unsolved
Observations	2,351	2,394	2,537	667	613	700

Note: Regressions of the amount claimed on whether the puzzle was solved and gender across all three rounds. For unsolved puzzles the sample is restricted to where the team contribution equals 100. Regressions include participant random effects, round fixed effects and clustered standard errors at the participant level. We control for actual contribution. Columns 1, 2, and 3 are estimated on the samples of the Split, Individual and Accuracy 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. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

moves equals 0 (see Section 2). Here, we exclude observations where the actual contribution is not defined.

In Table 4 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. We find that in the Individual Bonus ($p < 0.05$) and Accuracy Treatments ($p < 0.01$), unsuccessful participants claim less than successful participants. This is in line with earlier research findings that people claim smaller contributions for unsuccessful compared to successful teamwork (Ross and Sicoly, 1979). We find no difference in the Split Bonus Treatment.

Do men and women react differently to team failure? We find that unsuccessful women reduce their claims more than unsuccessful men in both the Split Bonus ($p < 0.05$) and the Individual Bonus Treatment ($p < 0.01$), further widening the gender claim-gap. In the Accuracy Treatment, we find no gender gap and no gender difference in the reaction to team failure.

In columns 4, 5, and 6 in Table 4 we estimate gender differences among unsuccessful teams directly. We observe similar patterns as for solved puzzles (Table 2), but the magnitude of the differences is larger. Women claim less than men in the Split Bonus ($p < 0.01$) and Individual Bonus Treatments ($p < 0.01$), but not in the Accuracy Treatment.

In sum, we find that women react to team failure more strongly than men, further exacerbating the gender claim-gap. Patterns among unsuccessful teams are consistent with our previous findings: The gender gap remains even when claims do not affect the other team member, but is closed when participants are incentivised to claim accurately.

5. Discussion

In this experiment, we find that under misaligned incentives, men claim more credit for their contribution to teamwork than equally contributing women. This result aligns with previous research that finds women evaluate their own performance worse than men (Exley & Kessler, 2022; Haynes

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

& Heilman, 2013). Our experiment complements this research by investigating a teamwork setting with misaligned incentives and imperfectly observable, ambiguous individual contributions.

We test if the gender claim-gap is driven by gender differences in altruism by eliminating concerns towards the other team member by design. The gender gap remains, suggesting that it is not a difference in altruism that drives the gender gap. This is in line with recent results suggesting that, contrary to common beliefs held by both genders, women do not generally show more socially-oriented behaviour than men (Exley et al., 2025).

We further test gender differences in overconfidence as a mechanism for the claim-gap. We find that, on average, both men and women equally overestimate their contribution, providing no support for differences in overconfidence as a main driver of the gap. Thus, using an experiment with high statistical power, we do not find evidence that gender differences in altruism or in overconfidence explain the observed gender claim-gap. There are several other plausible explanations. Exploratory evidence from our post-experiment questionnaire suggests gender differences in stated considerations related to lying aversion, payoff weights and inequality aversion, providing a basis for developing testable hypotheses for future work.

5.1. Why do we not find a gender gap in overconfidence?

Previous research finds that men and women self-evaluate their performance differently (Exley & Kessler, 2022) and hold different beliefs about their contributions to teamwork (Isaksson, 2018). However, in this experiment, we find no evidence for a gender gap in incentivised beliefs.

What explains the contrasting findings? We leverage the fact that the Accuracy Treatment is similar in design to Isaksson (2018) to identify possible reasons for the discrepancy. There are several differences between the experiments that could explain why we do not find a gender gap in overconfidence. For instance, we use a large online sample ($n = 845$ for the Accuracy Treatment) which consists of US and UK residents, while the experiment in Isaksson (2018) was conducted in the laboratory with an all-US-American sample ($n = 197$) and several years earlier. The inclusion of UK participants in our sample does not seem to account for the difference in results, because when we restrict our sample to US participants only, we also find no evidence of a gender claim-gap in the Accuracy Treatment.[11] We are also not aware of any research documenting a decline in gender differences over the past few years.

A more plausible explanation lies in a key difference in the experimental setting: in Isaksson (2018) participants are not anonymous – first names and photos reveal gender – while in our study, they are. Under accuracy incentives, Isaksson (2018) finds gender differences in credit claiming when the other team member's gender is known; we observe no such differences when gender is concealed. This suggests that stereotype activation or social image concerns may play a key role in driving gender disparities, prompting people to project stereotypical beliefs about skills and contributions both onto their team member and onto themselves (Bordalo et al., 2016; Bordalo et al., 2019; Haynes & Heilman, 2013). Our anonymous experiment removes this channel, suggesting that gender gaps in credit claiming may be contingent on social context rather than inherent differences.

Context may also play a role in shaping gender differences in claiming behaviour. Investigating bargaining over a jointly produced outcome under uncertain individual contributions, Cavalan et al. (2022) find that women, but not men, react to differential framing of the instructions asking for their contribution. Similarly, Small et al. (2007) shows that gender differences in self-advocacy depend on whether a situation is framed as negotiation or as "asking", with women more willing to engage in the latter. While our setting does not involve negotiation – claims are unilateral, require no approval, and communication is prohibited – these findings underscore the importance of context when interpreting gender effects.

[11] In the US sample of the Accuracy Treatment, there are 206 women and 110 men. The mean claim is 53.66% for women and 54.07% for men; p= 0.79 for a two-sided two-sample t-test.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Beyond framing, several studies emphasise how social context shapes self-assessments: whether individuals are being watched, judged, or held accountable. Haeckl (2022) and Ludwig et al. (2017) show that women react differently to the observability of their self-assessment's accuracy. Several experiments in the realm of self-evaluations feature a strategic component. These often involve a supervisor (see e.g., Brilon et al., 2024; Exley & Kessler, 2022; Schwardmann & Van der Weele, 2019; Solda et al., 2020) or a peer (Brilon et al., 2024; Charness et al., 2018), who must be convinced of one's claim. In many real-life situations, individuals may "get away" with exaggeration due to limited information or leniency, as is structurally the case in our experiment. We deliberately excluded a third party to isolate altruism towards the other team member. While Exley & Kessler (2022) find that the gender gap in self-promotion persists when there is a chance of "getting caught", an interesting avenue for future research could be to investigate who is believed when (over-)claiming contributions for teamwork. Our finding that men do not report having contributed more than women under accuracy incentives adds to a growing understanding of the contextual effects on gender differences in overconfidence.

A further potential source of overconfidence is motivated reasoning: individuals may derive utility from believing they contributed more, leading to distorted information processing and belief updating (Köszegi, 2006). This mechanism aligns with broader findings on belief distortion for identity maintenance and norm conformity (Akerlof & Kranton, 2000; Babcock et al., 2003). Recent work suggests that motivated reasoning may differ by gender and thereby contribute to gender gaps in overconfidence (Thaler, 2021). In our experiment, however, we find no gender difference in incentivised beliefs in the Accuracy Treatment. Accordingly, we also do not find evidence suggesting a gender difference in motivated reasoning under accuracy incentives. To identify gender differences in overconfidence as a mechanism of the gender claim-gap under misaligned incentives, the model in Section 3 assumes that beliefs are independent of the treatment condition. In fact, identification only requires the weaker assumption that beliefs respond to the treatment in the same way for men and women. However, if motivated reasoning interacts with treatment incentives conditional on gender, our experimental design cannot detect this, which could confound our interpretation of overconfidence in claims. More research is needed on how motivated reasoning responds to different incentives, and whether this interacts with gender.

5.2. Adherence to social norms

We have already discussed gender differences in lying aversion, relative payoff weights and inequality aversion as possible channels for the claim-gap and provided some exploratory evidence in Section 4.2.1. Another possible moderator of claims is adherence to social norms and gender stereotypes. This behaviour is widely documented in the literature. For example, stereotype-based performance expectations have been shown to shape credit attribution, with women engaging in attributional rationalisation and allocating success to others under ambiguity (Haynes & Heilman, 2013). Social identity theory suggests that people derive intrinsic utility from conforming to identity-relevant social norms (Akerlof & Kranton, 2000). Under this framework, adherence to identity norms yields direct utility, while deviations impose negative utility. Women might derive negative utility from claims that violate gender norms around modesty and cooperation, whereas such behaviour is consistent with assertiveness and competition norms for men and may lead them to pursue higher claims. These predictions are consistent with our observation of women more frequently mentioning equality or their contribution, and men emphasising their own payoff as the reason for their claim. Another possibility is that women anticipate extrinsic social penalties – "backlash" – for appearing overly assertive (Babcock et al., 2003). However, in a recent study Haeckl et al. (2025) find no evidence that women are punished more strongly for assertive behaviour than men. While we cannot rule out potential effects of intrinsic direct utility from adherence to social norms, we deliberately minimise

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

the impact of extrinsic social expectations in our study by keeping participants anonymous and their claims private.

5.3. Gender differences at the upper tail of the distribution

The gender gap in credit claiming is particularly pronounced at the upper tails of the claim distribution. A substantial body of experimental literature documents gender differences at the extremes of distributions (see e.g., Andreoni and Vesterlund, 2001; Baye & Monseur, 2016; Liu et al., 2020; Penner, 2008; Penner & Paret, 2008; Thöni & Volk, 2021), emphasising a phenomenon that is also observed empirically (Blau & Kahn, 2017). 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 only find a gender claim-gap among high contributors, that is, individuals who are pivotal to team success. That high-contributing women claim less credit for their contributions than high-contributing men may also exacerbate differences in labour market outcomes at the upper tail of the distribution. Evaluation schemes should thus account for potential exaggeration in self-reports and incentivise accuracy, for example, through the use of objective and transparent performance measures such as task-based metrics. Our study suggests self-evaluations should be separated from bonus payments, promotions, and hiring decisions.

6. Conclusion

In this study, we investigate the drivers of the gender gap in claiming credit for contributions to teamwork, testing both altruism towards the other team member and overconfidence in self-assessments as potential mechanisms. To this end, we conduct a large-scale online experiment with a final sample of 2456 participants. In the experiment, participants solve an interactive task in teams of two and subsequently claim how much they have contributed. We use a between-subject design where we vary the incentives that individuals have for claiming credit for their contribution.

Our experiment shows that when incentives between team members are misaligned, there is a gender gap in claiming contributions to teamwork. The gap persists irrespective of whether exaggeration harms the team member or not, suggesting that altruistic preferences regarding the other team member's payoff are not a driver for the gender claim-gap in this setting. Under accuracy incentives, men and women overestimate their contribution to a similar extent, suggesting that gender differences in overconfidence do not explain the observed gender claim-gap either. Exploratory evidence from our post-experiment questionnaire suggests gender differences in stated considerations related to lying aversion, the relative weight given to payoffs and inequality aversion, which may be important to examine in future work. The gender claim-gap is even more pronounced at the upper tails of the distribution than at the means. Most previous research on behavioural gender differences has focused on differences at the means, possibly overlooking large differences at the extremes. Taken together, our findings suggest that men are more inclined than women to borrow plumes in teamwork settings with misaligned incentives, claiming more credit for their contributions.

Our paper opens avenues for future research, with possible extensions including repeated interaction of team members, varying the degree of anonymity, increasing stake size, or manipulating the salience of social norms, as well as more direct tests of the factors reflected in the stated reasons. As discussed in Section 2.3.5, our design deliberately abstracts from certain contextual factors and thus serves as a mechanism-isolating benchmark for credit claiming when audience and reputational channels are muted, rather than as a universal lower bound for workplace gender gaps. Factors like age, race, ongoing relationships, reputational concerns, social visibility and hierarchies are beyond the

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

scope of this paper. Future research can explore how these complex social interdependencies shape credit claiming, building on our exploratory analyses of stated reasons to develop testable hypotheses.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eec.2026.10063.

Acknowledgement. We thank Mackenzie Alston, Loukas Balafoutas, Roberto Asmat Belleza, Peter Brummund, Nadine Chlaß, Lawrence Choo, Florian Englmaier, Christine Exley, Susann Fiedler, Milos Fisar, Ben Greiner, Simone Haeckl-Schermer, Gergely Hajdu, Lisa Hanzl, Pia Heckl, Stephanie Heger, Alejandro Hirmas, Johannes Hoelzemann, Siri Isaksson, Melanie Koch, Martin G. Kocher, Boon Han Koh, Ada Kovaliukaite, Marco Lambrecht, Olivia Masi, Eva Ranehill, Anna Sandberg, Rupert Sausgruber, Arthur Schram, Anna Schwarz, James Tremewan, Lise Vesterlund, Elisabeth Wurm and the participants at workshops and conferences – ESA European Meeting 2025, EEA Congress 2025, ESA Asia Pacific Meeting 2025, Berlin Workshop on Empirical Public Economics 2024, EALE 2024, SOLE 2024, RES 2024, Essex PhD Workshop 2024, NSE PhD and Post Doc Workshop 2023, Gender Gaps Conference 2023, 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 Vienna University of Economics and Business and Institute for Advanced Studies Vienna – for helpful comments and suggestions. We are particularly grateful to the anonymous referees and the editor for their constructive feedback, which helped improve the paper substantially. Any errors are our own.

Author contributions. All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Klara Kinnl, Jakob Moeller and Anna Walter-Dockx. The first draft of the manuscript was written by all authors, and all of them commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding statement. The project was funded within the Post-DocTrack Programme of the Austrian Academy of Sciences (OeAW).

Competing interest. The authors have no relevant financial or non-financial interests to disclose.

Ethical standards. The experiment reported in this paper has received ethics approval at Vienna University of Economics and Business, approval number WU-HSRP-2021-021.

Replication package. The replication materials for this study are available at https://doi.org/10.17605/OSF.IO/ncq36

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115–1153. https://doi.org/10.3982/ECTA14673.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881.

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607–1636. https://doi.org/10.3982/ECTA7384.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293–312. https://doi.org/10.1162/003355301556419.

Babcock, L., Laschever, S., Gelfand, M., & Small, D. (2003). Nice girls don't ask. *Harvard Business Review*, *81*(10), 14.

Balafoutas, L., Czermak, S., Eulerich, M., & Fornwagner, H. (2020). Incentives for dishonesty: An experimental study with internal auditors. *Economic Inquiry*, *58*(2), 764–779. https://doi.org/10.1111/ecin.12878.

Bartoš, V., Glogowsky, U., & Rincke, J. (2025). The color of knowledge: Impacts of tutor race on learning and performance. Mimeo.

Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-Scale Assessments in Education*, *4*(1), 1–16. https://doi.org/10.1016/j.ssresearch.2007.06.012.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865. https://doi.org/10.1257/jel.20160995.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753–1794. https://doi.org/10.1093/qje/qjw029.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, *109*(3), 739–73. https://doi.org/10.1257/aer.20170007.

Brilon, S., Grassi, S., Grieder, M., & Schulz, J. F. (2024). Strategic competition and self-confidence. *Management Science*, *70*(1), 507–525. https://doi.org/10.1287/mnsc.2023.4688.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Bruhin, A., Petros, F., & Santos-Pinto, L. (2024). The role of self-confidence in teamwork: Experimental evidence. *Experimental Economics*, *27*(3), 687–712. https://doi.org/10.1007/s10683-024-09829-x.

Buser, T., Grimalda, G., Putterman, L., & van der Weele, J. (2020). Overconfidence and gender gaps in redistributive preferences: Cross-country experimental evidence. *Journal of Economic Behavior & Organization*, *178*, 267–286. https://doi.org/10.1016/j.jebo.2020.07.005.

Capraro, V. (2018). Gender differences in lying in sender-receiver games: A meta-analysis. *Judgment and Decision Making*, *13*(4), 345–355. https://doi.org/10.1017/S1930297500009220.

Cavalan, Q., de Gardelle, V., & 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. https://doi.org/10.1016/j.joep.2022.102566.

Charness, G., Rustichini, A., & Van de Ven, J. (2018). Self-confidence and strategic behavior. *Experimental Economics*, *21*(1), 72–98. https://doi.org/10.1007/s10683-017-9526-3.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001.

Coffman, K., Collis, M. R., & Kulkarni, L. (2024). Stereotypes and belief updating. *Journal of the European Economic Association*, *22*(3), 1011–1054. https://doi.org/10.1093/jeea/jvad063.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587.

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, *34*, 1–7. https://doi.org/10.1016/j.joep.2012.10.011.

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

Danz, D., Vesterlund, L., & Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, *112*(9), 2851–2883. https://doi.org/10.1257/aer.20201248.

Dato, S., Feess, E., & Nieken, P. (2024). Lying in competitive environments: Identifying behavioral impacts. *European Economic Review*, *170*, 104844. https://doi.org/10.1016/j.euroecorev.2024.104844.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593–1640. https://doi.org/10.1093/qje/qjx022.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry*, *46*(1), 84–90. https://doi.org/10.1111/j.1465-7295.2007.00082.x.

Dowd, C. (2020). A new ecdf two-sample test statistic. *arXiv preprint arXiv:2007.01360*. https://doi.org/10.48550/arXiv.2007.01360.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, *99*(1), 197–199. https://doi.org/10.1016/j.econlet.2007.06.027.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723–733. https://doi.org/10.1287/mnsc.1110.1449.

Ewers, M., & Zimmermann, F. (2015). Image and misreporting. *Journal of the European Economic Association*, *13*(2), 363–380. https://doi.org/10.1111/jeea.12128.

Exley, C., Hauser, O., Moore, M., & Pezzuto, J. (2025). Believed gender differences in social preferences. *The Quarterly Journal of Economics*, *140*(1), 403–458. https://doi.org/10.1093/qje/qjae030.

Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, *137*(3), 1345–1381. https://doi.org/10.1093/qje/qjac003.

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

Fehr, E., & Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, *112*(478), C1–C33. https://doi.org/10.1111/1468-0297.00027.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547. https://doi.org/10.1111/jeea.12014.

Friesen, L., & Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters*, *117*(3), 624–626. https://doi.org/10.1016/j.econlet.2012.08.005.

Grosch, K., & Fischer, S. (2024). Gender equivalence in overconfidence – a large-scale experimental study in a non-WEIRD country. Working Paper. http://dx.doi.org/10.2139/ssrn.4889383.

Grosch, K., & Rau, H. A. (2017). Gender differences in honesty: The role of social value orientation. *Journal of Economic Psychology*, *62*, 258–267. https://doi.org/10.1016/j.joep.2017.07.008.

Haaland, I., Roth, C., Stantcheva, S., & Wohlfart, J. (2025). Understanding economic behavior using open-ended survey data. *Journal of Economic Literature*, *63*(4), 1244–80. https://doi.org/10.1257/jel.20251780.

Haeckl, S. (2022). Image concerns in ex-ante self-assessments–gender differences and behavioral consequences. *Labour Economics*, *76*, 102166. https://doi.org/10.1016/j.labeco.2022.102166.

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.

Haeckl, S., Moeller, J., & Zednik, A. (2025). Can being competitive but unsuccessful harm you, even more so if you are a woman? *Journal of Economic Behavior & Organization*, *236*, 107108. https://doi.org/10.1016/j.jebo.2025.107108.

Halladay, B., & Landsman, R. (2024). Shame on me: Emotions and gender differences in taking with earned endowments. *Journal of Behavioral and Experimental Economics*, *111*, 102207. https://doi.org/10.1016/j.socec.2024.102207.

Haynes, M. C., & Heilman, M. E. (2013). It had to be you (not me)! women's attributional rationalization of their contribution to successful joint work outcomes. *Personality and Social Psychology Bulletin*, *39*(7), 956–969. https://doi.org/10.1177/0146167213486358.

Hermann, D., & Mußhoff, O. (2019). I might be a liar, but i am not a thief: An experimental distinction between the moral costs of lying and stealing. *Journal of Economic Behavior & Organization*, *163*, 135–139. https://doi.org/10.1016/j.jebo.2019.05.009.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, *80*(3), 984–1001. https://doi.org/10.1093/restud/rdt006.

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

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, *4*(4), 673–707. https://doi.org/10.1162/JEEA.2006.4.4.673.

Liu, R., Alvarado-Urbina, A., & Hannum, E. (2020). Differences at the extremes? Gender, national contexts, and math performance in Latin America. *American Educational Research Journal*, *57*(3), 1290–1322. https://doi.org/10.3102/0002831219876236.

Ludwig, S., Fellner-Röhling, G., & 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. https://doi.org/10.1016/j.euroecorev.2016.11.007.

Meub, L., Proeger, T., Schneider, T., & Bizer, K. (2016). The victim matters–experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters*, *23*(16), 1162–1167. https://doi.org/10.1080/13504851.2016.1139674.

Neyse, L., Bosworth, S., Ring, P., & Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Scientific Reports*, *6*(1), 23294. https://doi.org/10.1038/srep23294.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The Quarterly Journal of Economics*, *122*(3), 1067–1101. https://doi.org/10.1162/qjec.122.3.1067.

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. https://doi.org/10.1086/589252.

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

Ring, P., Neyse, L., David-Barett, T., & Schmidt, U. (2016). Gender differences in performance predictions: Evidence from the cognitive reflection test. *Frontiers in Psychology*, *7*, 1680. https://doi.org/10.3389/fpsyg.2016.01680.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*(3), 322. https://doi.org/10.1037/0022-3514.37.3.322.

Saccardo, S., Pietrasz, A., & Gneezy, U. (2018). On the size of the gender difference in competitiveness. *Management Science*, *64*(4), 1541–1554. https://doi.org/10.1287/mnsc.2016.2673.

Sarsons, H., Gërxhani, K., Reuben, E., & Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, *129*(1), 101–147. https://doi.org/10.1086/711401.

Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, *3*(10), 1055–1061. https://doi.org/10.1038/s41562-019-0666-7.

Small, D. A., Gelfand, M., Babcock, L., & Gettman, H. (2007). Who goes to the bargaining table? the influence of gender and framing on the initiation of negotiation. *Journal of Personality and Social Psychology*, *93*(4), 600. https://psycnet.apa.org/doi/10.1037/0022-3514.93.4.600.

Solda, A., Ke, C., Page, L., & Von Hippel, W. (2020). Strategically delusional. *Experimental Economics*, *23*(3), 604–631. https://doi.org/10.1007/s10683-019-09636-9.

Sutter, M., & Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, *61*(10), 2339–2354. https://doi.org/10.1287/mnsc.2014.1981.

Takahashi, Y. (2025). Gender differences in the cost of corrections in group work. Working Paper.

Thaler, M. (2021). Gender differences in motivated reasoning. *Journal of Economic Behavior & Organization*, *191*, 501–518. https://doi.org/10.1016/j.jebo.2021.09.016.

Thöni, C., & 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. https://doi.org/10.1073/pnas.2026112118.

Cite this article: Kinnl, K., Moeller, J., & Walter-Dockx, A. (2026). Borrowed plumes: the gender gap in claiming credit for teamwork. *Experimental Economics*, 1–25. https://doi.org/10.1017/eec.2026.10063

Downloaded from https://www.cambridge.org/core. 12 Nov 2026 at 12:22:34, subject to the Cambridge Core terms of use.