Strategic Incompetence and Gender Stereotypes*

Sorravich Kingsuwankul† Christina Rott‡ Janneke Oostrom§

Job Market Paper

Version: February 18, 2026

Click here for the latest draft

Abstract

We examine whether individuals claim to be incompetent to avoid costly public goods tasks such as non-promotable tasks or household chores – a behavior widely described as strategic incompetence. Using a laboratory experiment, we exogenously vary the gender stereotype associated with a task (neutral, female, or male) to study its impact on communicated ability and task assignment. We find that individuals consistently claim incompetence and successfully evade task assignment. When tasks are stereotyped, gender gaps in communicated ability emerge over time in stereotype-consistent directions. These gaps are not driven by gender differences in strategic incompetence, but by self-stereotyping: when stereotypes become salient, individuals lower their private self-assessments if their gender is viewed as less competent. Our results underscore the importance of stereotype salience and highlight how communication of incompetence in strategic situations can perpetuate gender disparities by reinforcing unequal task assignment.

Keywords: Communication, Gender, Stereotypes, Beliefs, Non-Promotable Task, Lab experiment

JEL: C91, D83, J16

*Acknowledgements: The authors are grateful to Katharina Bruett, Klarita Gerxhani, Astrid Hopfensitz, Bernd Irlenbusch, Nicolas Jacquemet, Dorothea Kuebler, Warn Lekfuangfu, Diego Marino Fages, Eugenio Verrina, Lise Vesterlund, Marie Claire Villeval, Georg von Weizsaecker, and Andreas Ziegler for their very helpful comments. The authors thank participants at presentations at the ADRES conference in Paris, the 2nd ASFEE graduate meeting in Aussois, the Experimental Seminar at SBE Vrije Universiteit Amsterdam, Economics seminar at the University of Portsmouth, the BBE Seminar in Berlin, IMEBESS in Valencia, ESA World meeting in Beijing, ASFEE in Nancy, TIBER Symposium in Tilburg, ESA European Meeting in Brno, and Matterhorn Symposium in Brig, for useful comments. We are thankful to GATE-Lab, and to Quentin Thevenet and Chaya Senichaut for their assistance. The study received an ethics approval from the SBE Vrije Universiteit Amsterdam (# SBE/7/ski322). All mistakes remain our own.

†School of Business and Economics, Vrije Universiteit Amsterdam, HV Amsterdam, The Netherlands; Tinbergen Institute, MS Amsterdam, Netherlands. Email: s.kingsuwankul@vu.nl

‡School of Business and Economics, Vrije Universiteit Amsterdam, HV Amsterdam, The Netherlands; Tinbergen Institute, MS Amsterdam, Netherlands. Email: c.e.rott@vu.nl

§Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands. Email: j.k.oostrom@tilburguniversity.edu

1 Introduction

Workplace inequalities can stem from employee characteristics, beliefs, and preferences or biased hiring and promotion decisions; but they can also emerge from intra-organizational dynamics such as systemically biased assignments of non-promotable tasks (Babcock et al., 2017a). These tasks are often tedious, time-consuming, and low-visibility, contributing minimally to career advancement. Yet, they are essential for the functioning of an organization, benefiting many employees when completed. When disproportionately distributed and routinely performed by certain groups, the impact of such non-promotable tasks can be costly by diverting time and effort away from highly promotable tasks. For example, women have been shown to perform these tasks more frequently than men (e.g., De Pater et al., 2010; Chan and Anteby, 2016; Babcock et al., 2017b).[1] For instance, a survey by McKinsey & Company in 2021 finds that female senior managers are more likely to consistently dedicate time to employees' well-being and emotional support. Although most companies deem these tasks essential for collective success, few formally recognize them. Similar logic applies to household chores in the private domain.

Understanding who performs non-promotable tasks and *why* is crucial in promoting workplace equity. Although previous studies have paid much attention to individuals who inflate competence to attract promotable opportunities (e.g., Schwardmann and van der Weele, 2019; Soldà et al., 2020; Exley and Kessler, 2022), little is known about individuals who claim incompetence to avoid non-promotable tasks. We study this key mechanism in the context of a costly public goods task that must be assigned and performed by a group member. We define *strategic incompetence* as the understatement of one's own ability (i.e., claiming to be incompetent) to evade assignment and shift the task onto others.[2] We argue that the communication of inability can systematically distort who is assigned costly public goods tasks.

We also examine whether this behavior is shaped by stereotypes (Bordalo et al., 2016).[3] Specifically, we investigate how gender stereotypes about ability impact men's and women's propensity to communicate inability. When a task is stereotyped (i.e., men or women are believed to perform better or worse), claims of incompetence may be seen as more or less

[1] Miller and Roksa (2020), Cubel et al. (2024), Chu et al. (2022) and Nelson et al. (2023) provided more recent evidence on gender inequality in non-promotable tasks among PhD students, economists and medical attendings.

[2] Similar constructs are discussed in psychology and organizational behavior e.g., *feigned incompetence* or *weaponized incompetence* (e.g., McLuhan, 2020a,b) and *skilled incompetence* (Argyris, 2003). The topic recently received considerable media attention (see, e.g., Huffpost and Forbes, accessed on 7th November 2025), but lack empirical support (see a recent review by Stadnicka, 2024). Alternative definitions of strategic incompetence are possible. For example, exerting low effort in a task to generate 'poor' performance and avoid future assignment. In this paper, we focus on communication of inability.

[3] Bordalo et al. (2016) model stereotypes, based on the representativeness heuristic of Kahneman and Tversky (1972), as distorted probability distributions that exaggerated perceived group differences even when actual differences may be small or insignificant.. In other words, judgments are based on mental representations of what is 'typical' of a group.

credible depending on the worker's gender, shaping task assignment.[4] Men may 'play dumb' in female-typed tasks (e.g., arts tasks), and women in male-typed tasks (e.g., technical tasks). These gender gaps in communication shift non-promotable tasks onto those who claim and are perceived to be more competent.

To study these mechanisms, we design a laboratory experiment that elicits participants' beliefs about their ability, and measures their communicated ability when a costly public goods task has to be assigned within the group. All members benefit from task completion, but only the assigned individual bears the effort cost. We implement a modified public goods game that captures key features of such settings: the performer's ability affects all group members' payoffs, and ability is private information.

The experimental session has two blocks, each with four parts. In part 1, participants complete a timed trivia quiz spanning multiple topics. Quiz performance determines ability type: in groups of four, the top two are High-types, and the bottom two are Low-types; participants are not informed about their actual performance or type. In part 2, we elicit participants' beliefs about their type. In part 3, participants play the modified public goods game in groups. Success probability depends on the performer's type, with High-types being more likely to succeed. While the group benefits most in expectation if a *High-type* performs the task, doing so is personally costly for the assigned individual.[5] Consequently, performing the task is 'non-promotable' for all payoff-maximizing members regardless of types. In part 4, we elicit participants' first-order beliefs about others' types and second-order beliefs about their own type.

Our modified public good game mirrors how such tasks are assigned in groups outside of the lab — for instance, in meetings where team members communicate perceived ability before deciding who will perform the task. Group members first simultaneously send a *cheap-talk* message about the likelihood of being a High-type. After observing all messages, each member privately and simultaneously casts a vote for who should perform the task; the individual with the most votes is assigned. In this strategic setting, individuals can claim incompetence to reduce their selection probability, thereby shifting the task burden onto others.

We further manipulate *perceived* gender differences in ability– gender stereotypes– by varying the trivia question categories. In the first block of the session, participants are randomly assigned to one of three stereotype treatments:

- **Neutral**: Questions on neutral topics where men and women are expected to perform equally well (e.g., cities and countries).
- **Female**: Questions on female-typed topics where women are believed to outperform men (e.g., art and craft).

[4] While ability is some times assumed irrelevant for non-promotable tasks, we argue that it vary and affect assignment. However, once assigned, ability might be less relevant than effort for the final performance. In this paper, we focus only on ability and rule out effort provision.

[5] We induce the cost of "effort" (see Smith, 1976). This cost is fixed across types and can be interpreted as a fixed time commitment, with type affecting probability of success.

- **Male**: Questions on male-typed topics where men are believed to outperform women (e.g., fixes and repairs)

The trivia quizzes are calibrated so that no significant gender differences in performance are expected, but participants generally believe gender gaps exist in female- or male-typed topics. This design isolates the effects of *inaccurate* stereotype beliefs from *actual* differences in ability.[6] In the second block, participants perform a trivia quiz from a different treatment than in the first block (e.g., Neutral in Block 1 → Female or Male in Block 2), resulting in six treatment sequences that are balanced.

We find that individuals claim to be incompetent to evade assignment of a costly public goods task. Signaling 'incompetence' strongly reduces selection probability: around 95% of the groups select the individual signaling highest ability. Stereotypes influence communication: men and women are more likely to claim incompetence when the task is gender-incongruent (i.e., when their gender is believed to be less competent). Surprisingly, these gender gaps in communicated inability only emerge over time.

Contrary to our expectations, these gender gaps in communication are not driven by gender differences in strategic incompetence. Rather, they are largely explained by gender gaps in private self-assessments. When tasks are gender-stereotyped and incongruent, participants distort their beliefs downward, internally aligning with the stereotypes. Consistent with 'self-stereotyping' (Coffman, 2014), our results suggest that individuals internalize stereotypes, affecting private self-assessment, and ultimately shaping communication about ability. This internal mechanism occurs after the task's stereotypical nature becomes salient.

Analysis of the first- and second-order beliefs further reveals how stereotypes shape credibility. Individuals who self-stereotype perceive their claims to be credible and the others view them as credible. Once salient, stereotypes are incorporated into individuals' belief system, shaping not only personal evaluation and behavior but also expectations about how others view them. Reducing gender disparities in the gendered non-promotable task assignment therefore requires challenging stereotypes that reinforce apparent strategic incompetence.

Our findings contribute to the literature on workplace inequality, strategic communication and stereotypes. To the best of our knowledge, our study provides the first experimental evidence on strategic incompetence in costly public good tasks. Further, we demonstrate that gender gaps in communicated inability in stereotyped tasks arise from stereotype-driven distortions of private self-assessment, rather than deliberate misrepresentation.

The remainder of the paper is organized as follows. Section 2 provides an overview of the related literature. Section 3 outlines the experimental design and procedure. Section 4 provides the theoretical framework and formulates the hypotheses. Section 5 reports the results. Section 6 discusses the findings and concludes.

[6] We confirm (ex-post) no systematic gender gaps in the trivia performances. More details of the experimental design and calibration of trivia questions are described in section 3.

2 Related Literature

Our study relates to the literature on misrepresentation of ability, which shows that people overtly exaggerate their competence for strategic reasons. People adjust communicated competence by reacting to monetary incentive (Charness et al., 2018) and to social perception about underlying type (Thoma, 2016; Kurschilgen and Marcin, 2019). Related research on confidence shows that self-confidence can be motivated in strategic settings, providing instrumental benefits in persuading others (see Soldà et al., 2020; Schwardmann and van der Weele, 2019; Schwardmann et al., 2022). A study by Exley and Kessler (2022) shows that individuals self-promote by over-claiming their competence in a context of desirable tasks, with men exaggerating more than women in a male-typed task. More recent works show that the gender gap in self-promotion is driven by women's reluctance to promote oneself, a lower assertiveness when doing so (Chang et al., 2025), and fear of social judgment or norm violation (e.g., Ludwig et al., 2017; Tradenta et al., 2025). Further, Chaudhuri (2025) shows that this gap is largely modulated by underlying ability differences, with individuals self-promoting more in domains where their gender has a comparative advantage.

While these prior works have focused on displaying self-confidence and communicating competence to increase the likelihood of receiving promotable tasks, little is known whether individuals claim to be incompetent in non-promotable tasks. Our study contributes by focusing on how individuals claim incompetence as an excuse to evade assignment of such tasks. An example is a low-promotablity task (see e.g., Babcock et al., 2017a,b; Villas-Boas et al., 2019; Banerjee and Mustafi, 2025), which is deemed essential for organizations but does not increase an individual's likelihood of receiving a promotion. Therefore, we complement the existing literature by showing that an inequality in task assignment can be influenced not only by displaying bolstered competence when tasks are promotable but also by "playing dumb" when they are non-promotable.[7]

Our work also relates to the growing literature of gender economics and in particular stereotypes (see Bordalo et al., 2016, 2019) by examining how gender stereotypes about ability differences shape the gender gap in communicated ability.[8] This literature shows that gender stereotypes cause disparity in the labor market outcomes. On the demand side, stereotypical beliefs – accurate or not – create a gender bias in hiring and promotion against individuals whose gender is believed to be less competent. In a male-typed task (typically STEM-based such as an arithmetic task), managers tend to believe women to be less competent than men and thus prefer to hire men (e.g., Reuben et al., 2014; Kübler et al., 2018; Coffman et al., 2021; Bohren et al., 2023; Barron et al., 2024). In team settings where individual

[7] Relatedly, Chen et al. (2024) shows that people vary facial expressions depending on whether the task is desirable or not, suggesting a form of nonverbal strategic communication independent of ability.

[8] Stereotypical beliefs about gender could also pertain to traits and attributes, as documented in psychology (see e.g., Fiske et al., 2018; Diekman and Eagly, 2000). In economics, Exley et al. (2025) show that despite no gender difference in other-regarding preferences, the vast majority of people believe that women are more prosocial than men. Such misperceptions could reinforce social expectations about what one ought to be or do based on gender identity.

contributions are obscure, evaluators attribute more credits to men than women in a male-typed task, but attribute equally when the task is female-typed (Sarsons et al., 2021). On the supply side, stereotypes affect labor segregation. Buser et al. (2014) and Coffman et al. (2024) document that women are less likely to self-select into a male-dominated sector or those perceived to be highly competitive (see also Niederle and Vesterlund, 2007). Stereotypes also affect the willingness to participate in gender-incongruent domains due to self-stereotyping (Coffman, 2014) and anticipated discrimination (Lepage et al., 2025). This creates inefficiency as unrealized talent or resource misallocation arise when individuals do not receive training based on their innate abilities but on stereotypes (Hopfensitz et al., 2025).

In relation to this stream of literature, our work demonstrates how individuals claim incompetence to evade costly public goods task assignment by *hiding* behind stereotypes. Our experimental design exogenously varies the gender stereotypes about ability differences in the task to examine their impact on the gender gaps in communicated ability. Despite no meaningful gender difference in actual ability, individuals may hold biased beliefs that exaggerate true underlying gender differences (Bordalo et al., 2016). These *perceived* ability differences may make it easier for individuals to claim incompetence in gender-incongruent domains as that very claim is seen by others as more credible.[9] Therefore, a man can hide behind a generalized belief that women are better in female-typed tasks (e.g., employees' well-being support), while a woman can do the same in male-typed tasks (e.g., tasks that demand technical skills). If stereotypes create a gender gap in how men and women claim incompetence, the implication follows that to minimize gender inequality through the assignment of non-promotable stereotyped tasks, it is essential to challenge misperception about gendered abilities to prevent them from being used as justification.

Lastly, our work relates to a larger literature from psychology and organizational behavior on impression management (see Bolino et al., 2016) which looks at behavioral tactics used by employees to manage social perceptions and relations in organizational contexts. Most prior works in this field typically rely non-incentivized survey measures and qualitative data from focus groups and interviews. To the best of our knowledge, we are the first to quantitatively examine playing dumb and to isolate the impact of stereotypical beliefs about ability on the propensity in playing dumb using incentivized experimental methods in a controlled environment.

3 Experimental Design and Procedures

Overview

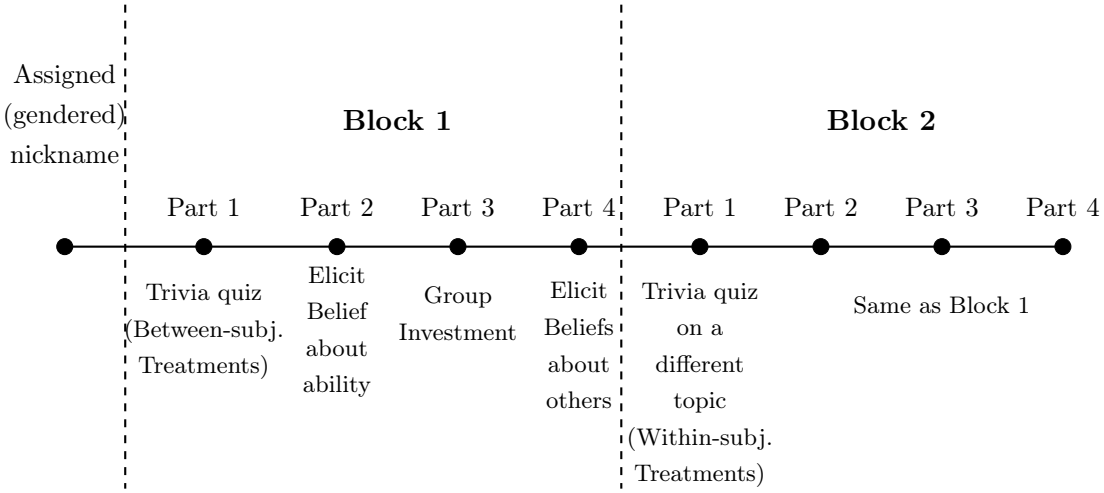
Each session consists of two blocks with four parts each. At the beginning of the session, participants are informed about the number of blocks and parts, but receive relevant instructions

[9] This could be exacerbated further by self-stereotyping: individuals who internalize stereotypes may hold beliefs that are distorted vis-à-vis their *actual* beliefs. Consequently people who claim incompetence in gender-incongruent domains may not only believe their claims to be credible but hold stereotyped beliefs that support their behavior.

before each part. They are informed that one of the four parts per block will be randomly selected for their additional payment at the end of the session. Group matching is fixed, and participants are informed in the instruction as the experiment proceeds. All feedback is given at the end of the session.

We reveal participants' gender information using gendered nicknames. Precisely, prior to any instructions, each participant is assigned an anonymous ID, comprising of a gendered nickname and a string of random three-digit number. The assignment of gendered nicknames is based on self-reported gender identity in the survey. The assigned IDs are then displayed during the experiment. We elaborate our design choices and discuss trade-offs in Section 3.2. Figure 1 depicts the timeline of the experimental session.

Figure 1: Timeline of the session.



3.1 Design

Part 1: Trivia

In part 1, inspired by Coffman (2014), participants complete a timed quiz, consisting of 20 multiple choice trivia questions. Each question has five answer options, one of which is correct. Participants have five minutes to provide answers to as many questions as possible. No performance feedback is given. If this part is selected for additional payment, participants earn 12 points per correct answer.

The trivia topic depends on the treatment, which varies in the gender stereotype associated with the topic (Neutral, Female, or Male).[10] We designed these question sets so that men and women perform equally well in each treatment on average, but participants *believe* that

[10]The instructions do not refer to the topic as *Neutral*, *Female* or *Male*. Participants are informed of the question categories that could be asked in the assigned treatment. For instance, neutral topics included 'cities and countries', female-typed topics included 'cooking and home', and male-typed topics included 'fixes and repairs'.

the performance differs across gender when the trivia is stereotyped. That is, our treatment effects aim to isolate the effect of *inaccurate* stereotypical beliefs from true ability differences.

To this end, we conducted a separate experiment (N=65) with an independent sample drawn from the same subject pool, allowing us to calibrate our trivia used in the experiment. We confirm (ex-post) that we find no systematic performance differences across gender in each treatment, though participants believe that men (women) perform better in male-typed (female-typed) trivia.[11]

Treatment sequences

The two experimental blocks are identical except for the stereotype of the trivia quiz. In Block 1, participants are either assigned to the Neutral, Female or Male treatment (between-subjects). In Block 2, participants are assigned to a treatment different than in Block 1 (within-subjects). This results in a mixed experimental design with the following six treatment sequences namely NeutralFemale, NeutralMale, FemaleMale, FemaleNeutral, MaleFemale, MaleNeutral. We elaborate this design choice in section 3.2. Besides the variations in the trivia topic, the remaining three parts remain identical across the two blocks.[12]

Part 2: Elicitation of beliefs about ability

In part 2, we elicit participants' belief about their ability in the trivia. They are informed that they would be matched with three other participants in the session (anonymously identified by their gendered nicknames) and their trivia performances will be ranked from the lowest to the highest. They have to guess the likelihood of being in the top half, by selecting an interger on a scale from 0% (certain to be in the bottom half) to 100% (certain to be in the top half). This is incentivized; participants are informed that if this part is selected for payment, the chance of receiving the payoff in this part (240 points) is maximized by being as accurate as possible.[13] Additionally, participants are informed that their response will not be shown to the other participants. We use this as a measure of participants' belief about ability.

Part 3: Group Investment

[11] All materials used in the calibration experiment (e.g., instructions, question sets and tests to select questions for the final trivia) are reported in the Online Appendix E.

[12] Additionally, in Block 1 participants answer to some understanding questions before entering their responses in part 2 and 3. These understanding questions are not present in Block 2.

[13] We incentivized our belief measures as per Karni (2009) and closely followed the instruction used in Danz et al. (2022). Precisely, if part 2 is selected, the computer program randomly selects a number between 0 and 100. If the guess is lower or equal to the random number, the participant receives 240 points if they are in the top half, and otherwise nothing. If the guess is higher than the random number, the program conducts a lottery where the chance to receive 240 points is equal to that random number. As in Danz et al. (2022), we provide minimum information about the incentivization (i.e., that being as accurate as possible maximizes the chance to receive the points without mathematical formulas). Further, participants are informed that the exact payment rule for this part can be obtained at the end of the session in case they wish to know. Of all participants, only three asked for this instruction. For best practices in eliciting beliefs using incentive compatible methods, see Healy and Leo (n.d.).

In part 3, participants play a modified one-shot public good game (framed as a group investment decision). In our configuration, only one person must incur a personal cost in an investment that benefits the entire group. This cost is fixed and captures the provision of effort in producing a public good. In our design, the person who is assigned to be the investor 'takes one for the team', reflecting the opportunity cost of performing a non-promotable task rather than a promotable one.

Participants are informed that they are re-matched with the same participants as in part 2. Each group consists of two High types and two Low types, determined by the trivia performance rankings. Group members do not know their own type, nor the type of others, and this is common knowledge. The two types differ in the probability that the investment will succeed. If the investor is a Low type, the investment succeeds with 25% and fails with 75%. If the investor is a High type, the investment succeeds with 75% and fails with 25%.

Investment outcomes affect the payoffs of all group members. When the investment succeeds, each group member earns 240 points; when it fails, each earns 160 points. Regardless of the outcome, the investor incurs a cost of 100 points. Thus, when the investment succeeds, the investor earns 140 points (240-100) and the other group members earn 240 points each. If the investment fails, the investor earns 60 points (160-100) and the other group members earn 160 points each. Given these parameters, the group's expected payoff is higher when a High type invests, but neither types have an incentive to be the investor.

The key feature of our game is a pre-play communication stage in which group members send a signal about their ability type before jointly deciding who should be the investor. This mirrors a stylized situation outside the laboratory where a public goods task is assigned to an individual based on self-reported ability.

To determine the investor, each group member simultaneously sends a *cheap-talk* message about the likelihood of being a High type by selecting an integer between 0 and 100 to complete a pre-formulated message: *"I think that I am a High type with ...%"*, which is our measure of *communicated ability*. After all messages are revealed to the group, participants privately and simultaneously vote for one member (including themselves) to be the investor. We implement a majority voting rule: the member with the most votes becomes the investor, and in the case of a tie, the computer randomly selects one of the tied members to invest.

After the investor has been selected, Nature determines the investment outcome according to the success probability associated with the investor's type. Participants receive feedback at the end of the session about the voting outcome, the selected investor, the investment outcome and their payoffs.[14]

Part 4: Elicitation of beliefs about others

In part 4, we elicit participants' beliefs about others in the Group Investment. We are interested in exploring the mechanism of credibility in communication and how that changes with stereotypes of the task. Precisely, we elicit two types of beliefs: what participants think

[14] Given fixed group matching, this rules out any potential reciprocity that may arise across the two blocks.

about the type of the other three group members (first-order belief), and what they think others think about their type (second-order belief).

Participants submit in total four guesses. We first elicit participants' first-order beliefs. They see again the (cheap-talk) messages sent by the three group members during the Group Investment, each accompanied by the sender's anonymous ID. They are asked to guess the likelihood that each member is a High type. After submitting these three guesses, we elicit their second-order belief. Each participant is reminded of the message they sent to their group and is asked to guess the average guesses submitted by the three group members about them (i.e., the mean first-order beliefs of others).[15]

3.2 Design choices

We elaborate in more detail our design choices, in particular revealing participant's gender information, and the use of mixed design.

Revelation of gender information: To study the mechanism of credibility in communication, participants need to be aware of the co-participant's gender identity. To preserve anonymity, we reveal participants' gender information using gendered nicknames assigned at the beginning of the session based on the self-reported gender identity.

Precisely, after an informed consent and prior to receiving any instructions, participants are asked to fill out a short questionnaire indicating their age, gender identity (male, female or other) and university. After completing this short survey, they are informed about the assignment of the anonymous ID, which comprises of a (gendered) nickname chosen from a set of common names and a random three-digit number. Unbeknownst to the participants, we assign the nickname based on self-reported gender identity.[16]

The gender revealing anonymous ID is displayed whenever participants are matched into groups (i.e., part 2, 3 and 4). In addition, to increase statistical power, participants are matched to form gender balanced groups, with two men and two women. We do not explicitly indicate the nickname assignment procedure nor group matching protocol to avoid any experimenter demand effect.

[15] The incentivization and the task instruction are similar to the one used in the belief elicitation in part 2. Precisely, participants are informed that if part 4 is selected, one of the four guesses would be randomly selected for additional payment. For the first-order beliefs, the incentivization is the same to that used for the belief elicitation in part 2. For the second-order belief, the following protocol is implemented; the computer randomly selects a number between 0 and 100. If the guess is lower or equal to the random number, the computer conducts a lottery where the chance to receive 240 points is equal to the mean guesses submitted by the group members, and nothing otherwise. If the guess is higher than the random number, the computer conducts a lottery where the chance of receiving 240 points is equal to the random number.

[16] We refrain from allowing participants to freely select nicknames to ensure that the nickname matches their actual gender identity. We observe that the fraction of self-reported gender in the experiment and the gender information in the recruitment system do not differ at the session level. This implies that our nickname procedure perfectly matches participants' gender identity.

Mixed design: We use a mixed design with alternating trivia topics across the two blocks to enable both between-subjects and across-block comparisons. We can test the impact of gender stereotypes by comparing Block 1 behavior, where participants have no prior experience of the game. The addition of the second block allows us to further test whether prior experience in Block 1 affects behavior in Block 2, conditional on the trivia topic.

In addition, the within-subjects dimension, realized through alternating the trivia topics, allows us to observe how stereotypes operate. Each trivia quiz contains questions belonging to a single stereotype (i.e., Neutral, Female or Male). This design makes it possible for participants' awareness of these stereotypes to change over time, as they are exposed to different trivia topics in the second block. A purely between-subjects design would not allow us to capture the impact of experience and the stereotype awareness on behavior and belief.

Procedures

The experiment was run at GATE-Lab in Lyon, France. We ran a total of 25 sessions with 492 participants (50.81% female). All were recruited through Hroot (Bock et al., 2014) from local universities and various educational background, including business, economics and engineering. The experiment was programmed with oTree (Chen et al., 2016). Recruitment was gender balanced; however, due to no shows, the gender composition varies from session to session. In all of our analyses, we focus on gender balanced groups to ensure that beliefs about ability are formed consistently, leaving us with N=440.[17]

Upon arrival, participants randomly drew a ticket from an opaque bag which assigned them to computer terminals. After an informed consent and the assignment of the anonymous ID, instructions were distributed at the beginning of relevant part and read aloud. In Block 2, instructions were displayed on the computer screen and not read aloud to avoid repetition. The full set of translated instructions are in Online Appendix A.

At the end of Block 2, after receiving feedback on part 3 of both blocks (i.e., vote outcome, selected investor, investment outcome, and realized payoffs), participants responded to the final questionnaire, including standard demographic questions, self-reported risk attitudes, measures of lying type from Schudy et al. (2024), a 16-item Personal Attribute Questionnaire taken from Spence et al. (1975), and beliefs about ability differences in the two trivia performed (similar to Coffman, 2014) – all of which were non-incentivized. Finally, participants received feedback on the payoffs of all parts and blocks, and were informed which parts had been randomly selected for their additional payment by the program . The experimental session lasted for approximately 75 minutes and the average earnings were 20 Euros (including a show-up fee of 5 Euros).

[17] This excludes 52 individual participants. The fraction of female participants ranges between 37.5% and 65%. Including these observations do not change the main conclusions of the paper. For summary statistics, see Table B1 to B4 in the Online Appendix.

4 Hypotheses

We formalize the Group Investment described in Section 3.1 and formulate our hypotheses.

There are four group members indexed by $i \in \{1, 2, 3, 4\}$. Each group member GM_i is assigned a type $t \in \{High, Low\}$ based on the trivia performance ranking. The distribution of types (two Highs and two Lows) is common knowledge, but GMs do not observe their own type nor the types of others. Types differ in investment success probabilities $p(t_i)$, with $p(High) > p(Low)$.

Each GM takes two actions. First, they simultaneously send a cheap-talk message $m_i \in \{0, 1, 2, ..., 100\}$ about the likelihood of being a High type (i.e., *communicated ability*). Messages are publicly displayed. Then, each GM privately casts a vote $v_i \in \{1, 2, 3, 4\}$ for the member who should become the investor. The member with the most votes becomes the investor. Ties are broken by Nature at random. Let I denote the selected investor.

The payoff function of GM_i is:

$$\Pi_i(I, t_I) = \begin{cases} p(t_i)\pi_S + (1 - p(t_i))\pi_F - c & \text{if } I = i \\ p(t_j)\pi_S + (1 - p(t_j))\pi_F & \text{if } I = j \text{ and } j \neq i \end{cases}$$

where, π_S, π_F, and c are the success payoff, failure payoff, and the cost of investment, respectively. Given our parameters choices, a GM's expected payoff is strictly higher when someone else invests, regardless of types.[18] Therefore, regardless of types, rational GMs strictly prefer to not be selected as the investor.

In our game, messages are cheap-talk and do not directly impact payoffs. Because investing is costly, both types have incentives to claim incompetence. Rational GMs anticipate this, and therefore do not update beliefs upon observing any type of messages. As a result, messages are treated as uninformative by rational agents and any message is a best response. Consequently, rational GMs are expected to randomize their message and randomly cast their vote on one of the three other members independently of their messages. This gives rise to a continuum of *babbling equilibria* in which any message is sent and votes are cast uniformly among the other group members.

However, messages may serve as a coordination device if we assume GM believes that others will condition their votes on them. Since each GM prefers to minimize own chance of being selected as the investor, they have an incentive to downplay their ability. In particular, sending lower messages is the best response to beliefs that votes are increasing in communicated ability. Given that voting constitutes a coordination game, GM prefers to vote for the member they expect others will vote for. In this setting, relatively higher messages (i.e., stronger claims to be a High type) naturally become focal points for coordination.

While the "standard" model predicts that messages are treated as uninformative, GMs respond to the incentive by claiming incompetence if they believe that votes respond pos-

[18] When another group member invests, a GM earns an expected payoff of 180 points if the investor is a Low type and 220 points if the investor is a High type. In contrast, when a GM invests themselves, their expected payoff is 80 points if they are a Low type and 120 points if they are a High type.

itively to the level of communicated ability. This may not hold if individuals experience non-pecuniary costs from misrepresentation (Gneezy et al., 2018; Dufwenberg and Dufwenberg, 2018; Abeler et al., 2019). These costs may stem from self and social image concerns which generate disutility from claiming to be incompetent (e.g., Bénabou and Tirole, 2005, 2011). If these costs are positive, individuals face a trade-off between monetary incentives and misrepresentation costs, leading some to refrain from exaggerating their incompetence to the fullest. This introduces heterogeneity in the messages depending on behavioral types. Given such heterogeneity, GMs who send relatively higher messages become salient focal points and are more likely to be selected as investors, even if messages do not carry information in a strict game theoretic sense.

Taken together, the structure of our game provides strong monetary incentives for individuals to strategically claim incompetence to shift the investment cost onto others. As long as these incentives dominate image concerns for most individuals,[19] we should observe systematic engagement in strategic incompetence, leading to our first hypothesis.

Hypothesis 1. (Strategic incompetence): Individuals claim to be incompetent to avoid being assigned the costly public goods task, shifting it onto others in the group.

We further expect experience in the strategic environment to strengthen this behavior, which can be tested by comparing behavior across blocks. If experience reinforces strategic behavior, communicated ability is expected to be lower in Block 2 compared to Block 1, conditional on the treatment, which leads us to the second hypothesis.

Hypothesis 2. (The impact of experience): Individuals claim more incompetence in Block 2 compared to Block 1.

Finally, we expect gender stereotypes associated with the task to generate gender gaps in communicated inability. If one gender is stereotypically viewed as *less* competent, claims of incompetence are perceived more credible for that group. Thus, they can effectively avoid selection by playing more dumb, as their claim aligns with the stereotyped belief. This leads us to the last two hypotheses.

Hypothesis 3. (The impact of female stereotype): Relative to gender neutrality, men claim more incompetence than women when the task is female-typed.

Hypothesis 4. (The impact of male stereotype): Relative to gender neutrality, women claim more incompetence than men when the task is male-typed.

[19] We believe that image concerns from playing dumb in our context are, if any, minimal and do not overpower the monetary incentives. The findings of Kurschilgen and Marcin (2019) suggest little image concerns for the type of non-academic trivia questions.

Our design and hypotheses were pre-registered prior to the data collection at AsPredicted.org (#223442).

5 Results

We start by first assessing the gender gap in the trivia performance within each block and treatment, then we evaluate the prevalence of strategic incompetence and the role of experience. We then examine the impact of stereotypes on the gender gap in communication and investigate their mechanisms. Finally, we provide additional evidence by analyzing our belief data to explore credibility in communication.

Throughout this section, and unless otherwise specified, we use Mann-Whitney U tests (for between-subject comparisons) and Wilcoxon signed-rank tests (for within-subjects comparisons) as our main statistical tools, treating each participant as one independent observation. All p-values are reported from two-sided tests. We further support our claims via econometric analyses using OLS regression models with clustering at the matching group level.

5.1 Gender gap in performance

Before testing our hypotheses, we first check for any gender gap in the trivia performance. Table 1 displays the mean number of correct answers in the trivia by men and women in each block and treatment.[20] Overall, we see that men and women perform equally well across all blocks and treatments. All comparisons do not reveal significant gender gap at the conventional levels. An exception is in the Female treatment, where women slightly outperform men (Block 1, $p = 0.063$; Block 2, $p = 0.110$), although these differences are marginally significant. This implies that our treatment manipulation principally varies *inaccurate* gender stereotypes about ability differences between men and women without meaningful actual differences.[21]

5.2 Strategic incompetence

Prevalence

We begin by documenting the prevalence of strategic incompetence. Table 2 places side-by-side beliefs about ability and communicated ability by block and treatment.

Given our design with no gender gap in performance and a gender balanced group composition, and assuming that participants form rational expectation about their ability and communicate it honestly, communicated ability is expected to be 50% on average. The observed pattern (see Table 2) shows that this is not the case, as average communicated ability are well below the 50% benchmark across all blocks and treatments (the largest p-value is

[20] We refer to male participants as *men* and female participants as *women* to distinguish from the treatment terminology (Neutral, Male, Female). No participants self-reported to be other.

[21] Looking at the fraction of men and women classified as High types by block and treatment leads to a similar conclusion: pairwise comparisons reveal no significant differences (the smallest p-value is 0.066). Furthermore, across the two blocks, we find no significant differences in trivia performance or the fraction of High types given treatment for either gender (see Tables B5 and B6 in the Online Appendix).

Table 1: Trivia performance by gender, block and treatment.

	Block 1		Block 2		Men vs. Women ($p-$value)	
	Men	Women	Men	Women	Block 1	Block 2
Neutral	13.09	13.11	12.97	13.30	$p = 0.613$	$p = 0.783$
	(3.75)	(3.11)	(3.58)	(2.93)		
Female	11.42	12.59	12.28	13.28	$p = 0.063$	$p = 0.110$
	(3.65)	(3.09)	(3.79)	(2.70)		
Male	11.43	11.49	11.80	11.18	$p = 0.836$	$p = 0.286$
	(3.48)	(3.08)	(3.17)	(3.07)		

Note: This table reports the mean number of correct answers in the trivia quizzes, with standard deviations shown in parentheses. $p-$values reported are from ranksum tests.

0.005). Given we elicited participants' belief about their ability, we observe that communicated ability is consistently lower than belief in a given treatment and block (all $p < 0.001$). This provides strong evidence of strategic incompetence to avoid being assigned a costly public goods task.

Result 1: Individuals claim incompetence to avoid being assigned an undesirable task.

Table 2: Belief & communicated ability by block and treatment.

	Block 1		Block 2		Belief vs. Comm. ($p-$value)	
	Belief	Comm.	Belief	Comm.	Block 1	Block 2
Neutral	61.26	39.58	57.27	41.89	$p < 0.001$	$p < 0.001$
	(23.63)	(25.41)	(23.54)	(26.67)		
Female	61.39	43.12	51.82	33.00	$p < 0.001$	$p < 0.001$
	(24.27)	(23.42)	(27.48)	(25.25)		
Male	54.68	35.00	44.25	33.32	$p < 0.001$	$p < 0.001$
	(26.31)	(25.09)	(23.88)	(22.07)		

Note: This table reports the mean belief and communicated ability about being a High type, with standard deviation in parentheses. Belief refers to the belief about one's likelihood of being a High type. Communicated ability (Comm.) refers to the cheap-talk message about one's likelihood of being a High type, and is the inverse measure of strategic incompetence. $p-$values reported are from signrank tests.

The role of experience

If people are predisposed to play dumb to evade a costly public goods task assignment, do they engage more with experience? We find no systematic across-block differences in communicated ability conditional on the treatment. In particular, the average communicated ability in the Neutral and Male treatments remain constant across the two blocks ($p = 0.413$

and $p = 0.649$, respectively), while the average communicated ability in the Female treatment do become lower (moving from 44% to 33%, $p < 0.001$). However, this is inconsistent with our hypothesized effect of experience as the experience effect should play out independently of the treatment.

Further, we observe an unexpected pattern in beliefs about ability across blocks. Our priors were that communicated ability systematically lower with experience of the game (which is not the case), while private self-assessment remains constant. This would be indicative of experience-induced strategic incompetence. This is however not the case (see Table 2). The beliefs about ability in Block 2 are statistically lower than in Block 1 in the Female and Male treatments; both by about 10 percentage points ($p = 0.002$ and $p < 0.001$, respectively). The difference is not significant for the Neutral treatment ($p = 0.226$). This result is striking: as show previously the trivia performance across blocks are indistinguishable. We will revisit this finding when we discuss about the mechanism of strategic incompetence when the task is stereotyped.[22]

Given all evidence, we fail to reject the null hypothesis that experience reinforces strategic incompetence.

Result 2: Experience has no impact on individuals' propensity to engage in strategic incompetence.

The impact of gender stereotypes

To evaluate how gender stereotypes shape strategic incompetence, we first look more closely at the gender gap in communicated ability across treatments. Figure 2 displays the average communicated ability by men and women by block and treatment. We find a very small to no differences in the way men and women communicate about their ability for all treatments in Block 1 (Panel a). However, gender gaps in communication emerge in Block 2 when the task is stereotyped (Panel b). While men and women do not communicate their ability differently in the Neutral treatment (46% vs. 38% to be a High type, $p = 0.092$), we find that men communicate less competence than women in the Female treatment (26% vs. 40%, $p < 0.001$). Similarly in the Male treatment, women communicate less competence than men (26% vs. 41%, $p < 0.001$).[23]

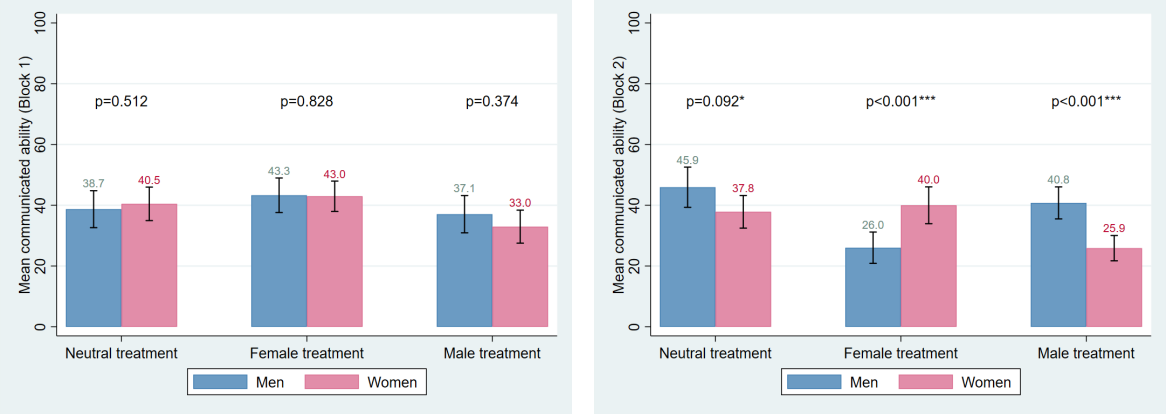
Further, we observe how gender stereotypes affect men's and women's communication relative to the Neutral treatment. We find that stereotypes impact communication for those whom stereotypes prescribe as 'less competent' (see Panel b of Figure 2). Conditional on the same gender group, men (women) communicate less competence in the Female (Male)

[22] Given beliefs differ across block, we compare the difference between belief and communicated ability across blocks in a given treatment and find no clear pattern that supports our hypothesized effect of experience on communicated ability (see Table B7 in the Online Appendix; Wald tests from OLS regressions return $p = 0.054$, $p = 0.876$ and $p = 0.001$ for Neutral, Female and Male treatment, respectively).

[23] As a robustness check, we perform the same analysis by treatment sequence, and find that the observed gender gaps in Block 2 in the Female and Male treatment occur independently of the assigned treatment in Block 1. See this complementary analysis in the Online Appendix D.

Figure 2: Communicated ability by gender and treatment.

(a) Block 1 (b) Block 2



Note: The figure displays communicated ability by men and women across treatments in Block 1 (Panel A) and Block 2 (Panel B). Communicated ability refers to the cheap-talk message about one's likelihood of being a High type, and is the inverse measure of strategic incompetence. $p-$values reported are from ranksum tests.

treatment relative to the Neutral treatment (both $p < 0.001$). On the other hand, communicated ability remain largely unaffected when the treatment is gender-congruent. Women communicate on average 40% to be a High type in the Female treatment (vs. 38% in the Neutral treatment, $p = 0.721$), while men communicate 41% to be a High type in the Male treatment (vs. 46% in the Neutral treatment, $p = 0.229$).

These results imply that when the task is stereotyped and gender-incongruent (i.e., when own gender is believed to be less able), individuals claim incompetence – though we find this only in Block 2 of the experiment.

Our (ex-post) interpretation is that the stereotype embedded in the trivia might have been subtle in Block 1. and might not be immediately salient. As the design alternates the trivia topics at the within-subject level in Block 2, it allows us to observe *when* stereotypes begin to impact behavior and belief. Our results suggest that for stereotypes to have any impact, they may need to become salient first.[24]

Result 3: A gender gap in communicated inability emerges when the task is stereotyped, as individuals claim to be incompetent in gender-incongruent tasks. These gaps arise only in Block 2 of the experiment, suggesting the importance of stereotype salience for it to have any impact.

[24] To support this interpretation, we find that participants' belief about gender differences in ability (measured in a non-incentivized question in the final questionnaire) in the female- or male-typed trivia is *more* stereotypical when it's assigned in Block 2 compared to when the same trivia is assigned in Block 1. This is driven by a shift from less extreme beliefs (e.g., 'men and women are equally good' or 'slightly better') towards extreme ones (e.g., 'men/women are better'). See Table B11 in the Online Appendix.

5.3 Origins of strategic incompetence in stereotyped tasks

To understand why the gender gaps in communicated (in)ability in stereotyped tasks only emerge in the latter half of the experiment, we examine the role of beliefs about ability. Recall that we elicit participants' beliefs about their ability after the trivia quiz without any performance feedback and before the group investment decision.

We find the observed gender gaps in communication are not driven by gender differences in strategic incompetence. Instead, they are primarily explained individuals' private self-assessment of their *own* ability. Further, we show that this internal mechanism arises from belief distortion by those individuals, whom stereotype prescribes as 'less competent' in that domain.

Gender gaps in belief largely explain gender gaps in communication

To examine the origin of the observed gender gaps in communicated ability, Table 3 presents the coefficients from OLS regressions, where the dependent variable is the communicated ability in Block 2.[25] In model 1, the independent variables include treatment dummies (with Neutral treatment as the reference category), a sex dummy (1 for women, 0 otherwise) and their interaction terms. Consistent with the previous section, post-estimation tests (see at the bottom of Table 3) confirm the presence of gender gaps in claimed competence between men and women in the Female and Male treatment (both $p < 0.001$). In addition, for a given gender, stereotypes cause men (women) to claim less competence in the Female (Male) treatment relative to the Neutral treatment (both $p < 0.001$).

To examine the role of private self-assessment, model 2 controls for individual's belief about ability. The results show that individuals' beliefs *largely* account for the observed gender differences in the way men and women communicate ability. After accounting for beliefs, the gender gap in communicated ability is fully closed in the Female treatment ($p = 0.280$), but remains somewhat in the Male treatment ($p = 0.030$). Similarly, accounting for beliefs largely explains the impact of stereotype on women's and men's communicated ability in the Male and Female treatment relative to the Neutral treatment (see Wald tests indicated at the bottom of Table 3).

This result suggests that beliefs act as an important mediating channel through which the treatment affects communicated ability. In the Female treatment, including beliefs reduces the gender gap in communicated ability from 13.96 to 4.52, implying that about 68% of the gap is explained by beliefs. Similarly, in the Male treatment, the reduction is from -14.89 to -6.21, or about 58%. This pattern indicates that beliefs serve as a key mediating mechanism through which gender stereotypes shape communication.

Lastly, our results are robust to a set of control variables, namely trivia performance, age, risk attitude, personality traits, and lying aversion measures (see model 3). We find that risk loving attitudes and age to be positively associated with claimed competence (both

[25] Given our results, we focus on Block 2 where the gender gaps are observed. For completeness, Table B8 in the Online Appendix reports the coefficients from OLS regressions where the dependent variable is the communicated ability in Block 1, using the same regression specifications.

Table 3: Determinants of communicated ability in Block 2

Dependent variable: Communicated ability	(1)	(2)	(3)
Neutral T.	*ref.*	*ref.*	*ref.*
Female T.	-19.905***	-10.277**	-9.791**
	(4.724)	(4.148)	(4.004)
Male T.	-5.162	-0.734	-1.454
	(4.838)	(3.903)	(3.904)
Women	-8.095**	-3.846	-3.317
	(3.475)	(2.756)	(3.128)
Women X Female T	22.053***	8.362*	7.730
	(5.066)	(4.984)	(4.931)
Women X Male T	-6.797	-2.362	-2.241
	(4.715)	(4.009)	(3.968)
Belief about ability		0.510***	0.524***
		(0.053)	(0.055)
Constant	45.932***	14.578***	1.163
	(3.761)	(3.593)	(13.905)
Control Variables	No	No	Yes
N	440	440	440
Clusters	110	110	110
R-Square	0.091	0.326	0.364
Post-estimation Wald tests			
(Women - Men) in Female T.	13.96***	4.52	4.41
	(3.69)	(4.16)	(4.19)
(Women - Men) in Male T.	-14.89***	-6.21**	-5.56*
	(3.19)	(2.82)	(2.88)
Women in Male T. - in Neutral T.	-11.96***	-3.10	-3.70
	(3.80)	(3.21)	(3.04)
Men in Neutral T. - in Female T.	-19.91***	-10.28**	-9.79**
	(4.72)	(4.15)	(4.00)

Notes: This table reports presents the coefficients from OLS regressions in which the dependent variable is communicated ability in Block 2. Independent variables in model (1) include treatment dummies (with Neutral treatment as the reference category), sex dummy (1 for women, 0 otherwise) and their interaction terms. Model (2) includes beliefs about being a High type. Model (3) further include control variables namely trivia performance, age, self-reported risk attitudes, personality traits (agentic and communal), and lying type dummies (high intrinsic cost and high social cost). Standard errors are clustered at the group level. * $p<0.10$; ** $p<0.05$; *** $p<0.01$.

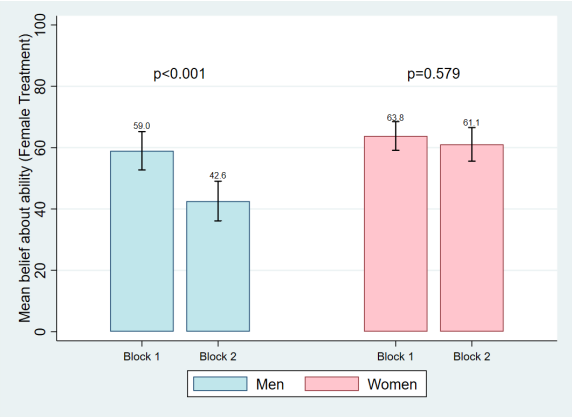
significant at 1%). For personality traits,[26] the measure of agency (i.e., 'masculine' traits such as assertiveness and competition) is negatively correlated with claimed competence (significant at 5% level).

These results imply that the observed gender gaps in communication in stereotyped tasks are largely explained by beliefs about ability. When the task is stereotyped and gender incongruent (i.e., their gender is believed to be 'less able'), individuals tend to hold *low* beliefs about their ability, consistent with *self-stereotyping* (Coffman, 2014) Taken together, this implies that the gender gaps in communicated ability are internally driven by beliefs. It's not that men and women differ in their engagement in strategic incompetence along stereotypical domains; rather differences in private self-assessments drive the observed gender gaps in communicated (in)ability.

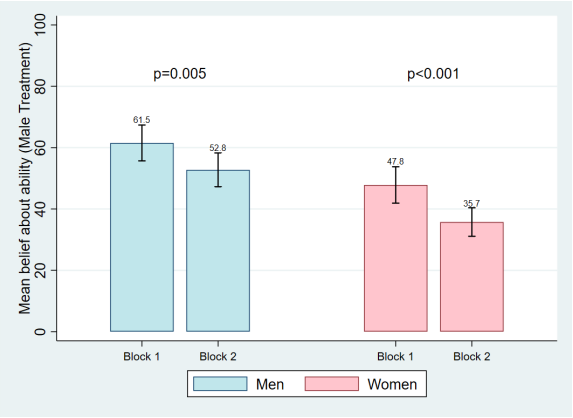
Stereotypes distort beliefs about ability

Figure 3: Beliefs in stereotyped trivia by gender across blocks.

(a) Female Treatment



(b) Male Treatment



Note: The figure displays beliefs about being a High type by men and women in the Female treatment (Panel A) and Male treatment (Panel B) across blocks. p-values reported are from ranksum tests.

To better understand how self-stereotyping (i.e., low private self-assessment in gender incongruent domains) occurs, Figure 3 displays the mean beliefs about ability by gender in the Female and Male treatment across the two blocks. Because these beliefs are elicited before communication decisions and without performance feedback, changes across blocks reflect belief formation rather than ex-post rationalization. The beliefs pattern of men and women across blocks suggest that these 'low' private self-assessments in gender-incongruent tasks have been *downwardly* distorted to align with the stereotypes.

[26] Table B12 in the Online Appendix reports the factor analysis of the 16-item Personal Attributes Questionnaire. The results show a clear two-factor structure with acceptable internal consistency (Cronbach's alpha) for both personality dimensions (i.e., agency and communion). We therefore averaged item responses within each trait and used in the regression analyses.

When facing a female-typed trivia in Block 1 (Panel a of Figure 3), men believe to be a High type with 59% chance on average, yet when the same quiz is presented in Block 2, men's beliefs drop to 43% ($p < 0.001$). A similar pattern is observed among women in the Male treatment (Panel b of Figure 3): when facing a male-typed trivia in Block 1, women on average believe to be a High type with 48% chance, but this falls to 36% when the same quiz is presented in Block 2 ($p < 0.001$).[27] This is particularly striking given neither gender gaps in performance nor within-gender performance differences across blocks. Taken together, these findings suggest that self-stereotyping may not occur by default but requires stereotype to be salient, leading to distortion of private self-assessment. Thus, one ends up believing that one is incompetent in gender-incongruent tasks, even though true ability remains unaffected.

Finally, we note that belief distortion prevails among those whom stereotype prescribes to be 'less competent' and only occurs when the trivia is stereotyped. The beliefs of women in the Female treatment remain highly constant across blocks (64% in Block 1 vs. 61% in Block 2, $p = 0.579$). Interestingly, the private beliefs of men in the Male treatment *do* differ across blocks (62% in Block 1 vs. 53% in Block 2, $p = 0.005$).[28] In the Neutral treatment, where ability is believed to be gender neutral, we do not observe any belief distortion across blocks (men $p = 0.350$; women $p = 0.428$; see Figure C1 in Online Appendix).

In summary, the mechanism behind the gender gaps in communication observed in Block 2 is two-fold. (i) Gaps in communicated (in)ability are largely explained by the gaps in beliefs, consistent with self-stereotyping. (ii) These 'self-stereotyping' beliefs are not innate but have been distorted *downward* to confirm the stereotype in which *their* gender is believed to be less competent. Our findings suggest that claims of incompetence in stereotyped tasks are not driven by gender differences in strategic incompetence, but by distorted beliefs that one is "incompetent" in gender-incongruent tasks.

5.4 Credibility and perceived credibility

We examine how stereotypes shape the perception of communication. This analysis speaks to the credibility of claims and the informational environment in which communication takes place, rather than to the primary mechanism driving behavior. Because beliefs themselves are distorted by stereotypes, we cannot separately identify whether second-order beliefs causally drive communication choices.

Recall that participants are asked to estimate the likelihood that each group member is a High type after having observed their claims (i.e., first-order beliefs about others). Then, we elicit their second-order belief: they need to estimate the average first-order beliefs submitted by their group members about them. These statistics are reported in Table B9 and B10 in the Online Appendix.

[27] These downward shifts in beliefs are independent of ability types. Conditioned on gender and type, the beliefs in Block 2 are significantly lower than in Block 1 when the gender does not match the task's stereotype (the largest p-value is 0.026).

[28] The difference is mainly driven by High type men (Block 1 vs. 2, High type $p = 0.025$; Low type $p = 0.196$). This pattern, where men distort beliefs in their own domain, may relate to gender differences in the formation of motivated belief (see Thaler, 2021).

We first look at the first-order beliefs. If the average beliefs about an individual aligns closely to the cheap-talk message, it suggests that the message is credible. Given a group of four, we calculate the mean first-order beliefs held by the other three members about each participant. We find that the mean first-order beliefs are consistently higher than what has been communicated in the cheap-talk message. This is true for all treatments and blocks, suggesting general skepticism (the largest p-value from pairwise comparisons is 0.034).

However, even after discounting, stereotypes still shape how men and women are perceived in Block 2. When the task is stereotyped and gender-incongruent, on average group members believe men and women to be a High type with only 38% and 36%, respectively. This is far below a benchmark of 50% in the case where believed ability is gender neutral (both $p < 0.001$). This suggests that, although individuals are aware that others might have played dumb, perceptions about ability remain strongly shaped by gender stereotypes. By contrast, in the Neutral treatment, the mean first-order beliefs about men and women do not differ significantly from the 50% benchmark (men, $p = 0.132$; women, $p = 0.062$).[29]

If judgment of abilities are gender-biased, an important question is how messengers expect their claims to be interpreted. This is reflected in the second-order beliefs held by men and women. When the task is stereotyped and gender-incongruent, individuals expect others to judge them stereotypically: men (women) believe they are more likely to be perceived as a Low type when the trivia topic is female-typed (male-typed) ($p = 0.001$ and $p < 0.001$, respectively). These expectations are well-aligned with the views of others: comparing their second-order beliefs with the actual first-order beliefs (i.e., average views held by group members) reveals no significant differences ($p = 0.533$ for men in Female treatment, and $p = 0.351$ for women in Male treatment). In other words, individuals who have claimed to be incompetent in gender-incongruent domains perceive their claim to be believed by others, which also aligns with the actual views of others. This points to a shared understanding of stereotype-based inference among both senders and receivers, which are consistent with and may sustain self-stereotyping.

6 Conclusion

In this paper, we provide novel evidence from a laboratory experiment showing that individuals claim incompetence to shift a costly public goods task onto others in the group. While prior studies focus on the promotion of ability in the positive domain (where competence pays off), we address individuals' ability communication in the negative domain. Our results highlight the importance of stereotypical beliefs about gendered ability for men's and women's willingness to claim incompetence in strategic situations. When there is stereotypical belief ascribing either gender to be 'less able', a gender gap in playing dumb emerges. Our findings

[29] To further support our interpretation, we point out that in Block 1 where we find no impact of stereotypes on communication, the mean first-order beliefs about men and women are close to 50%. Pairwise comparisons reveal no significant differences (Neutral, $p = 0.480$; Female $p = 0.746$; Male $p = 0.112$), suggesting that they are perceived to be equally able in all treatments in Block 1.

highlight the key role of stereotype salience in triggering self-stereotyping to arise, which in turn explains the observed gaps in communication.

Our experimental findings enrich the literature on stereotypes and strategic communication. While prior work documents self-stereotyping as a mechanism behind a gender gap in group contribution (e.g., Coffman, 2014), our findings suggest that such low self-assessments in a gender-incongruent domain may not occur by default. Instead, self-stereotyping emerges once stereotypes are sufficiently salient to decision-makers. Unlike Coffman (2014) who employed the trivia combining both female and male-typed questions, our design separates different stereotypes at a within-subject block level, enabling us to identify the conditions under which stereotypes begin to shape beliefs and behavior. Taken together, our findings show that individuals internalize stereotypes despite being largely inaccurate: once salient, stereotypes distort individuals' private self-assessments, communicated ability and judgment of others' abilities.

While we remain cautious about the generalizability of our findings, they offer insights for policies aimed at reducing gender inequality at task allocation. Given evidence on how undesirable, non-promotable tasks are disproportionately borne by women, our work suggests that this may partly stem from the fact that the majority of these public goods tasks are perceived as female-typed. To mitigate such imbalances, organizations can work to reduce stereotypes: by correcting inaccurate beliefs about ability, individuals prone to believe they are 'incompetent' will find it harder to justify their incompetence.

Our work opens up numerous future research avenues. While we demonstrate how gender stereotypes about ability differences shape gender gaps in communicated (in)ability in task with non-promotable features, future research could examine it in other contexts such as stereotypes about age, race or their intersectionality. Studying strategic incompetence through the channel of effort to evade future task assignment instead of communication is another promising direction. Finally, looking at the managerial responses to incompetence claims is worth exploring. If managers fail to account the gender gap in communication, this can further reinforce gender disparities in task assignment in a top-down fashion.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Argyris, Chris, "A Life Full of Learning," *Organization Studies*, September 2003, *24* (7), 1178–1192.

Babcock, Linda, Maria P. Recalde, and Lise Vesterlund, "Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash," *American Economic Review*, 2017, *107* (5), 131–135.

_ , _ , _ , and Laurie Weingart, "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability," *American Economic Review*, 2017, *107* (3), 714–747.

Banerjee, Ritwik and Priyoma Mustafi, "Using social recognition to address the gender difference in volunteering for low-promotability tasks," *Management Science*, 2025.

Barron, Kai, Ruth Ditlmann, Stefan Gehrig, and Sebastian Schweighofer-Kodritsch, "Explicit and Implicit Belief-Based Gender Discrimination: A Hiring Experiment," *Management Science*, May 2024.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch, "hroot: Hamburg Registration and Organization Online Tool," *European Economic Review*, October 2014, *71*, 117–120.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope, "Inaccurate Statistical Discrimination: An Identification Problem," *The Review of Economics and Statistics*, September 2023, pp. 1–45.

Bolino, Mark, David Long, and William Turnley, "Impression management in organizations: Critical questions, answers, and areas for future research," *Annual Review of Organizational Psychology and Organizational Behavior*, 2016, *3* (1), 377–406.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *Quarterly Journal of Economics*, November 2016, *131* (4), 1753–1794.

_ , _ , _ , and _ , "Beliefs about Gender," *American Economic Review*, March 2019, *109* (3), 739–773.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness, and Career Choices*," *Quarterly Journal of Economics*, August 2014, *129* (3), 1409–1447.

Bénabou, Roland and Jean Tirole, "Self-Confidence and Personal Motivation," in Bina Agarwal and Alessandro Vercelli, eds., *Psychology, Rationality and Economic Behaviour: Challenging Standard Assumptions*, International Economic Association Series, London: Palgrave Macmillan UK, 2005, pp. 19–57.

_ and _ , "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

Chan, Curtis K. and Michel Anteby, "Task Segregation as a Mechanism for Within-job Inequality: Women and Men of the Transportation Security Administration," *Administrative Science Quarterly*, June 2016, *61* (2), 184–216.

Chang, Jenny, Silvia Saccardo, and Jana Gallus, "Leaning In Softly: On Gender and Self-Promotion," August 2025.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven, "Self-confidence and strategic behavior," *Experimental Economics*, March 2018, *21* (1), 72–98.

Chaudhuri, Kalyani, "Gender Similarities in Self-Promotion: The Role of Stereotypes," 2025.

Chen, Daniel, Astrid Hopfensitz, Boris van Leeuwen, and Jeroen van de Ven, "The Strategic Display of Facial Expressions," 2024.

Chen, Daniel L., Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chu, Bryan, Ben Handel, Jon Kolstad, and Ulrike Malmendier, "Gender Differences in Non-Promotable Tasks: The Case of Clinical Note-Taking," 2022.

Coffman, Katherine B., Christine L. Exley, and Muriel Niederle, "The Role of Beliefs in Driving Gender Discrimination," *Management Science*, June 2021, *67* (6), 3551–3569.

_ , Manuela R. Collis, and Leena Kulkarni, "Whether to Apply," *Management Science*, July 2024, *70* (7), 4649–4669.

Coffman, Katherine Baldiga, "Evidence on Self-Stereotyping and the Contribution of Ideas*," *Quarterly Journal of Economics*, November 2014, *129* (4), 1625–1660.

Cubel, Maria, Laura Hospido, and Judit Vall, "Survey on the Work Environment in Economics," 2024.

Danz, David, Lise Vesterlund, and Alistair J. Wilson, "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, September 2022, *112* (9), 2851–2883.

Diekman, Amanda B. and Alice H. Eagly, "Stereotypes as Dynamic Constructs: Women and Men of the Past, Present, and Future," *Personality and Social Psychology Bulletin*, November 2000, *26* (10), 1171–1188.

Dufwenberg, Martin and Martin A. Dufwenberg, "Lies in disguise – A theoretical analysis of cheating," *Journal of Economic Theory*, May 2018, *175*, 248–264.

Exley, Christine L and Judd B Kessler, "The Gender Gap in Self-Promotion*," *The Quarterly Journal of Economics*, August 2022, *137* (3), 1345–1381.

_ , Oliver P Hauser, Molly Moore, and John-Henry Pezzuto, "Believed Gender Differences in Social Preferences*," *The Quarterly Journal of Economics*, February 2025, *140* (1), 403–458.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu, "A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition," in "Social Cognition" 2018. Num Pages: 53.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel, "Lying Aversion and the Size of the Lie," *American Economic Review*, February 2018, *108* (2), 419–453.

Healy, Paul J and Greg Leo, "BELIEF ELICITATION: A USER'S GUIDE."

Hopfensitz, Astrid, Christina Rott, and Ernesto Reuben, "Arbitrary stereotypes cause gender segregation in labor markets," 2025. mimeo.

Kahneman, Daniel and Amos Tversky, "Subjective probability: A judgment of representativeness," *Cognitive Psychology*, July 1972, *3* (3), 430–454.

Karni, Edi, "A Mechanism for Eliciting Probabilities," *Econometrica*, 2009, *77* (2), 603–606. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA7833.

Kurschilgen, Michael and Isabel Marcin, "Communication is more than information sharing: The role of status-relevant knowledge," *Games and Economic Behavior*, January 2019, *113*, 651–672.

Kübler, Dorothea, Julia Schmid, and Robert Stüber, "Gender discrimination in hiring across occupations: a nationally-representative vignette study," *Labour Economics*, December 2018, *55*, 215–229.

Lepage, Louise-Pierre, Xiaomeng Li, and Basit Zafar, "Anticipated Discrimination and Major Choice," 2025. mimeo.

Ludwig, Sandra, Gerlinde Fellner-Röhling, and Carmen Thoma, "Do women have more shame than men? An experiment on self-assessment and the shame of overestimating oneself," *European Economic Review*, February 2017, *92*, 31–46.

McKinsey & Company, "Women in the Workplace 2021," Technical Report, McKinsey Company and LeanIn.Org 2021.

McLuhan, Arthur, "Adopting a Cloak of Incompetence: Impression Management Techniques for Feigning Lesser Selves," *Sociological Theory*, 2020, *38* (2), 122–141.

_ , "Feigning Incompetence in the Field," *Qualitative Sociology Review*, 2020, *16* (2), 62–74.

Miller, Candace and Josipa Roksa, "Balancing Research and Service in Academia: Gender, Race, and Laboratory Tasks," *Gender & Society*, February 2020, *34* (1), 131–152.

Nelson, Laura K., Alexandra Brewer, Anna S. Mueller, Daniel M. O'Connor, Arjun Dayal, and Vineet M. Arora, "Taking the Time: The Implications of Workplace Assessment for Organizational Gender Inequality," *American Sociological Review*, August 2023, *88* (4), 627–655.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away From Competition? Do Men Compete Too Much?*," *The Quarterly Journal of Economics*, August 2007, *122* (3), 1067–1101.

Pater, Irene E. De, Annelies EM Van Vianen, and Myriam N. Bechtoldt, "Gender differences in job challenge: A matter of task allocation," *Gender, Work & Organization*, 2010, *17* (4), 433–453.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales, "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences*, March 2014, *111* (12), 4403–4408.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram, "Gender Differences in Recognition for Group Work," *Journal of Political Economy*, January 2021, *129* (1), 101–147.

Schudy, Simeon, Susanna Grundmann, and Lisa Spantig, "Individual Preferences for Truth-Telling," November 2024.

Schwardmann, Peter and Joël van der Weele, "Deception and self-deception," *Nature Human Behaviour*, October 2019, *3* (10), 1055–1061. Number: 10.

_ , Egon Tripodi, and Joël J. Van Der Weele, "Self-Persuasion: Evidence from Field Experiments at International Debating Competitions," *American Economic Review*, April 2022, *112* (4), 1118–1146.

Smith, Vernon L., "Experimental Economics: Induced Value Theory," *The American Economic Review*, 1976, *66* (2), 274–279.

Soldà, Alice, Changxia Ke, Lionel Page, and William von Hippel, "Strategically delusional," *Experimental Economics*, September 2020, *23* (3), 604–631.

Spence, J. T., R. Helmreich, and J. Stapp, "Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity," *Journal of Personality and Social Psychology*, July 1975, *32* (1), 29–39.

Stadnicka, Agnieszka, "Feigned Digital Incompetence as a New Managerial Challenge: A Systematic Literature Review and Future Research Agenda," *Annales Universitatis Mariae Curie-Skłodowska, Sectio H Oeconomia*, 2024, *58* (4), 223–248.

Thaler, Michael, "Gender differences in motivated reasoning," *Journal of Economic Behavior & Organization*, November 2021, *191*, 501–518.

Thoma, Carmen, "Under- versus overconfidence: an experiment on how others perceive a biased self-assessment," *Experimental Economics*, March 2016, *19* (1), 218–239.

Tradenta, Julio Mancuso, Ananta Neelim, and Joseph Vecci, "Gender differences in the self-promotion of prosocial behaviour: exploring the female modesty constraint," *Experimental Economics*, January 2025, *28* (1), 1–21.

Villas-Boas, Sofia B., Rebecca L. C. Taylor, and Elizabeth Deakin, "Effects of peer comparisons on low-promotability tasks: Evidence from a university field experiment," *Journal of Economic Behavior & Organization*, February 2019, *158*, 351–366.

ONLINE APPENDIX

Online Appendix A Instructions (Translated from French)

At the beginning of the session, participants received the general instructions, and the part instructions before each relevant part. They were read aloud using an audio recording. Apart from the trivia topics, which depended on the assigned treatment, all instructions remained identical.

General Instructions *(printed)*

Hello and welcome to today's experiment. We will go through the instructions together and read them aloud. Please turn off your phone and put it away. You are not allowed to use your phone or talk with others during the experiment. You are not allowed to open tabs or computer programs other than the one opened for you. If you violate any of these rules, you will be excluded from the experiment immediately and you will not receive any payment.

During this experiment, you can earn money. The amount of money you earn depends on your decisions as well as those made by others in the session. Therefore, please read the instructions carefully.

The experiment is divided into two blocks, followed by a questionnaire at the end. Each block has four parts. Thus, there will be a total of eight parts. You will receive the part instructions at the beginning of the respective part. The instructions are identical for all participants. You will have the opportunity to ask questions privately to the experimenter in case you do not understand the instructions. To do so, please press the red button on the side of your table or raise your hand. The experimenter will come to answer in private. Please do not ask questions aloud.

How will your decisions affect your earnings?
For today's experiment, **you receive 5 euros for participating in the study plus additional earnings**. The exact amount of additional earnings is determined at the end of today's session. The computer program randomly selects one part from each block. Your collected points in the two selected parts determine the additional earnings. **100 points are worth 5 euros**. Any decimal points will be rounded. Since all parts have an equal chance to be selected, it is in your best interest to make decisions as if that part is selected for reward and determines the additional earnings.

Anonymity of your decisions
All decisions and responses will be anonymous. No one will be able to determine which decisions you made. You will not know the identity of the other participants before, during, or after the experiment. The other participants will not know your identity either.

You will now receive the instructions for Block 1 - Part 1.

Block 1 - Part 1 *(printed)*

You will now respond to a trivia quiz with **20 multiple-choice questions**. These questions are fun facts about various topics such as *[Neutral treatment: cities and countries, Earth and nature, food, animals, board games, popular movies, and popular books; Female treatment: cooking, home, art and crafts, fashion and clothes, mind-body sports, musical films, and romance books; Male treatment: fixes and repairs, information and technology, video games, cars, competitive sports, action movies, and sci-fi books]*. For each question, you will see a label, indicating the respective topic.

Each question has five answer options and one of them will be correct. You will have **5 minutes** in total to submit your answers to the questions. The remaining time will be displayed at the top-right corner of the screen.

If this part is selected at the end of the experiment, you will receive an additional payoff based on your score. **For each correct answer, you receive 12 points**. There is no point deduction for an incorrect or no answer.

If you have questions, please press the red button. The experimenter will answer them in private.

Block 1 - Part 2 *(printed)*

You will now answer a question about your performance in the trivia quiz you just completed.

If this part is selected at the end of the experiment, your guess determines whether you receive a prize of 240 points or not. **The payment rule is designed so that you can secure the largest chance of winning the prize by reporting your most-accurate guess.** The precise payment rule details are available upon request at the end of the experiment.

You will be matched with three other participants in the session, whose anonymous IDs will be shown on the screen. Your trivia performances will be ranked from the lowest score (rank 4) to the highest score (rank 1). Any ties will be broken randomly. This ranking means that ranks 4 and 3 are in the bottom half, and ranks 2 and 1 are in the top half.

Your guess: What do you think is the likelihood (in percent) that you are in the top half?

You can enter any value between 0 and 100(%).

- 0(%) means that you are completely certain that you are in the bottom half.
- 50(%) means that you believe it's equally likely that you are in the bottom half or the top half.
- 100(%) means that you are completely certain that you are in the top half.

This guess will **not** be shown to the other participants.

Before you submit your guess, you will answer a few understanding questions about this part on the computer screen. If you have any questions, please press the red button and we will answer in private.

–Understanding questions for part 2 (on screen)–

1. Your trivia performance will be compared and ranked with three other participants in this session. (True/False)
2. Which statement is correct?
   (a) The trivia performance ranking is randomly determined.
   (b) The trivia performance is ranked from the lowest (rank 4) to the highest score (rank 1).
   (c) The trivia performance is ranked from the highest (rank 4) to the lowest score (rank 1).
3. If your guess to be in the top half is a number between 1 and 49, what does it mean?
   (a) You are completely certain that you are in the bottom half.
   (b) You are completely certain that you are in the top half.
   (c) You are not completely certain but believe it's more likely that you are in the top half.
   (d) You are not completely certain but believe it's more likely that you are in the bottom half.
4. If your guess to be in the top half is a number between 51 and 99, what does it mean?
   (a) You are completely certain that you are in the bottom half.
   (b) You are completely certain that you are in the top half.
   (c) You are not completely certain but believe it's more likely that you are in the top half.

(d) You are not completely certain but believe it's more likely that you are in the bottom half.

Block 1 - Part 3 *(printed)*

You are matched again with the same three other participants to form a group of four. If this part is selected at the end of the experiment, the decisions made by you and your group members will affect your additional payoff. Therefore, please read the following instructions carefully.

Group Investment
You will interact with your group through the computer towards a group investment decision. The outcome of this investment will affect the points earned by each group member. But only one group member will invest in the group account and bear the investment cost of **100 points**.

Return from Investment and Participant Types
The outcome of investment depends on the type of the investor. Each participant is assigned a type, either **'Low'** or **'High'** based on the trivia performance in part 1. The trivia performances of all group members will be ranked from the lowest to the highest score. Any ties will be broken randomly.

- The **two lower performing group members** are **Low** types.
- The **two higher performing group members** are **High** types.
  Neither you nor your group members will be informed about your own or each other's type.

The investment can either **succeed or fail**. The chance of success is higher if the investor is a **High** type.

- **If the investor is a Low type**:
  - 25% chance the investment succeeds → each member earns 240 points.
  - 75% chance the investment fails → each member earns 160 points.
- **If the investor is a High type**:
  - 75% chance the investment succeeds → each member earns 240 points.
  - 25% chance the investment fails → each member earns 160 points.

Regardless of the outcome and the investor's type, the investor will pay the investment cost of 100 points.

Your Decisions
Before the group investment outcome is determined, there will be **two stages**: first **Communication of type** and then **Selection of investor**.

1. Communication of type
Each participant has been assigned an anonymous ID and will have the opportunity to send a message to their group. This message allows each member to indicate how likely they think they are a High type.

A screenshot of the communication stage is displayed below. The anonymous IDs were chosen for illustrative purposes and likely vary for your group.

(screenshot of communication stage displayed here)

Each participant will select a number between 0 and 100 to complete the sentence: "I think that I am a High type with ...%."

- 0(%) means that you are completely certain that you are a Low type

- 50(%) means that you believe it's equally likely that you are a Low type or a High type.
- 100(%) means that you are completely certain that you are a High type.

Once all group members submit their messages, they will be displayed to the entire group and each member will select whom they would like to make the investment.

2. Selection of investor
The investor is selected through majority voting.

- Each participant will select one of the four group members (including themselves).
- The participant with most votes will be selected to invest in the group account.
- In case of a tie, the computer program will select one of the tied participants randomly.

A screenshot of the selection stage with *hypothetical messages* is displayed below. The percentages for the messages were chosen randomly between 0 and 100. The anonymous IDs are displayed in a random order. This example does not indicate what you should do or what you can expect others to do.

(screenshot of selection stage displayed here)

Investment Outcomes and Final Payoffs
After the investor is chosen by the group, the computer program determines whether the group investment succeeds or fails based on the investor's type.

- If the investment **succeeds**:
  - the investor receives 140 points (that is, 240 points - investment cost of 100 points).
  - the other group members receive 240 points each.
- If the investment **fails**:
  - the investor receives 60 points (that is, 160 points - investment cost of 100 points).
  - the other group members receive 160 points each.

At the end of the experiment
You will be informed about the number of votes each group member received and the chosen investor. Furthermore, you will be informed whether the investment succeeded or failed, and the final payoffs of all group members. You will **not** be informed whether the investor is a Low type or a High type.

SUMMARY
1. One of the four group members must invest in the group account.
2. Each group member is either a Low type or a High type, based on the ranking of the trivia performance.
3. Low types have a 25% chance of investment success, while High types have a 75% chance of investment success.
4. Each group member sends a message about how likely they think to be a High type.
5. All group members observe all messages and privately select a group member to invest.
6. The group member with the most votes invests and pays the cost of 100 points.
7. If the investment succeeds, the investor receives 140 points and the others receive 240 points each.
8. If the investment fails, the investor receives 60 points and the others receive 160 points each.
9. At the end of the experiment, you will learn which group member invested, whether the investment succeeded and the final payoffs for this part.

Please take a few minutes to re-read the instructions. Do not hesitate to press the red button if you need further explanation, or have a question. The experimenter will answer them in private. After this, you will answer a few understanding questions about the group investment on the computer screen.

– Understanding questions for part 3 (on screen)–

1. If Florence-969 received 2 votes, Jules-013 received 1 vote, Béatrice-477 received 1 vote and Matheiu-268 received 0 vote, who will be selected as the investor?
   (a) Florence-969
   (b) Jules-013
   (c) Béatrice-477
   (d) Matheiu-268
2. How many points will be deducted from the investor's payoffs regardless of the investment outcome? (100)
3. If the group selects a participant who is a Low type, what is the probability that the investment ...
   - Succeeds? (25%)
   - Fails? (75%)
4. If the group selects a participant who is a High type, what is the probability that the investment ...
   - Succeeds? (75%)
   - Fails? (25%)
5. At the end, if you learn that the investment succeeded, what can be inferred about the investor's type?
   (a) The investor must be a Low type.
   (b) The investor must be a High type.
   (c) The investor could be either a Low type or a High type; both types can result in the investment to succeed.

Block 1 - Part 4 *(printed)*

You will now answer some questions related to the group investment in the previous part.

In this part, you will make a total of four guesses. If this part is selected at the end of the experiment, one of the four guesses will be randomly selected by the program and used to determine whether you receive a prize of 240 points or not. **The payment rule is designed so that you can secure the largest chance of winning the prize by reporting your most-accurate guess**. The precise payment rule details are available by request at the end of the experiment.

In the previous part, each group member sent a message indicating their likelihood to be a High type. You will see a summary of the messages sent by your group member on the computer screen.

Your guess: What do you think is the likelihood (in percentage) that each member is a High type?

You can enter any value between 0 and 100(%).

- 0(%) means that you are completely certain that the group member is a Low type.
- 50(%) means that you believe it's equally likely that the group member is a Low type or a High type.
- 100(%) means that you are completely certain that the group member is a High type.

You will submit three guesses, one about each group member. Then, the instruction of the fourth guess will appear.

These guesses will **not** be shown to the other participants.

Please press the red button if you need further explanation or have a question.

–Instructions for the fourth guess (i.e., second-order belief) (on screen)–

We have asked your group members namely (anonymous IDs of the group members displayed here) about the likelihood they believe you are a High type.

Recall that they have seen the following message in the previous part: *"I think I am a High type with (communicated belief displayed here))%."*

Your guess: What do you think is the average value of the three guesses submitted by your group members?

You can enter any value between 0 and 100(%).

- 0(%) means that you think they are completely certain that you are a Low type.
- 50(%) means that you think they believe it's equally likely that you are a Low type or a High type.
- 100(%) means that you think they are completely certain that you are a High type.

Please press the red button if you have a question. Otherwise, you can enter your guess and click on the Next button to continue.

—

Block 2 instructions

At the end of Block 1, participants were informed that the instructions of Block 2 would be displayed on the computer screen. They received part instructions before each relevant part. There were no understanding questions for the belief (part 2) and the group investment (part 3) to avoid repetition.

Final Questionnaire *(on screen)*

- *(SOEP general risk attitude)* In general, how willing are you to take risks on a scale of 0 to 10? 0 means you are "completely unwilling to take risks" and 10 means you are "completely willing to take risks."
- *(Non-incentivized lying aversion)* Imagine that the host of a live radio shows calls you to participate in a raffle.
  Your role is as follows: you must toss a coin 4 times and indicate the number of tails you got. You will receive 10 euros for each tail you got. You know that the host cannot under any circumstances cehck whether you have indicated the exact number of tails.
  - Situation 1: Imagine the unlikely case where you get 4 tails. How many talks will you indicate to the host?
  - Situation 2: Imagine the unlikely case where you get 0 tails. How many talks will you indicate to the host?
- *(16-item Personal Attributes Questionnaire)* The questions below ask you about the type of person you think you are. Each question consist of two characteristics. Between the two, a scale of 1 to 5 is displayed

  You must select a number to indicate the characteristics that best describe you, 1 mean you think the characteristic on the left best describe you , while 5 means that the characteristic on the right best describe you.

- Not at all independent/Very independent
- Not at all emotional/Very emotional
- Very passive/Very active
- Not at all able to devote self completely to others/Able to devote self completely to others
- Very rough/Very gentle
- Not at all helpful to others/Very helpful to others
- Not at all competitive/Very competitive
- Not at all kind/Very kind
- Not at all aware of others' feelings/Very aware of others' feelings
- Can make decisions easily/Has difficulty making decisions
- Give up very easily/Never give up easily
- Not at all self-confident/Very self-confident
- Feel very inferior/Feel very superior
- Not at all understanding of others/Very understanding of others
- Very cold in relations with others/Very warm in relations with others
- Go to pieces under pressure/Stand up well under pressure.

- *(Non-incentivized belief about ability)* In Part 1 of each block, you took the quiz on various topics. **For each quiz, on average, do you think men are better, men and women are equally good or women are better?**

  - **Block 1**: the topics were [a list of topics displayed here, depending on treatment].
    -Men are better than women.
    -Men are slightly better than women.
    -Men and women are equally good.
    -Women are slightly better than men.
    -Women are better than men.
  - **Block 2**: the topics were [a list of topics displayed here, depending on treatment].
    -Men are better than women.
    -Men are slightly better than women.
    -Men and women are equally good.
    -Women are slightly better than men.
    -Women are better than men.

- *(Ego-relevance of the quiz)* Not counting the money you can receive for your trivia score, how important is it to you to do well on these quizzes? Please rate from 1 "Not at all important" to 7 "Very important"

Explanation of incentivization of beliefs provided upon request at the end of the session

(printed)

Payment rule for Part 2

At the end of the experiment, if Part 2 in either Block 1 or 2 is selected by the program, the following rule applies.

To determine whether you get the additional payment of 240 points, we will compare your guess to be in the top half to a random draw between 0 and 100. All numbers are equally likely to be selected. We refer to your guess as G and the random draw as R.

- If $R \leq G$, you will receive 240 points if you are in top half (i.e., Ranks 1 or 2) and nothing if you are in the bottom half (i.e., Ranks 3 or 4).
- If $R > G$, you will receive 240 points with R% and nothing with (1-R)%.

For example:

- Suppose your guess to be in the top half is 75.
- Now suppose that the randomly drawn number is 46.
- Since 46 is lower than 75, you will receive 240 points if you are in the top half and nothing if you are in the bottom half.

Another example:

- Suppose your guess to be in the top half is 75.
- Now suppose that the randomly drawn number is 80.
- Since 80 is higher than 75, you will receive 240 points with 80% chance and nothing with 20% chance.
- In this case, the program conducts a lottery where the chance of receiving 240 points is equal to 80%.

It is always better for you to guess what you really believe:

Imagine you are rather confident to be in the top half (say, with 75%), but instead of submitting a 75, you submit a guess of 50.

- If the random draw is 55, which is higher than your guess (50), you earn 240 points with 55% and nothing with 45%.
- However, had you submitted 75 (your true belief), you will receive 240 points if you are in the top half, and nothing if you are in the bottom half. Since your true belief is 75% to be in the top half, this gives you a higher chance to receive the payment (75% vs. 55%).
- This means that answering your true belief maximizes the chance to receive the payment.

Payment rule for Part 4

At the end of the experiment, if Part 4 in either Block 1 or 2 is selected by the program, the following rule applies.

The program first randomly selects one of the four guesses and used to determine your additional payment. Recall that the first three guesses were about the likelihood each group member is a High type, while for the fourth guess, you were asked to estimate the average guesses made by the three group members about you being a High type.

If the selected guess is one of the first three guess:

The payment rule is similar to the one in Part 2 **EXCEPT** that when the random draw is lower than or equal to your guess, then you receive 240 points if the group member is a High type, and nothing if that group member is a Low type.

If the selected guess is the fourth guess:

The program first calculates the average (mean) of the guesses submitted by the three group members about you being a High type. We refer to this mean as X.

To determine whether you get 240 points, we will compare your guess to a random draw between 0 and 100. All numbers are equally likely to be selected. We refer to your guess as G and the random draw as R.

- If $R \leq G$, you will receive 240 points with X% and nothing with (1-X)%.
- If $R > G$, you will receive 240 points with R% and nothing with (1-R)%.

Online Appendix B Appendix Tables

Table B1: Details of treatment sequences and sessions conducted.

Treatment Sequence	Men	Women	*N*	Sessions
NeutralFemale	37	43	80	4
NeutralMale	42	42	84	4
FemaleNeutral	42	42	84	5
FemaleMale	40	40	80	4
MaleNeutral	40	40	80	4
MaleFemale	41	43	84	4
Total	242	250	492	25

Notes: This table summarizes the treatments sequences and sessions conducted with all observations. We have excluded N=52 from our analysis due to gender imbalance in the group, leaving us with N=440. Including these observations do not change the conclusion of the paper.

Table B2: Number of observations by block and treatment (gender balanced only)

	Block 1			Block 2		
Treatment	Men	Women	*N*	Men	Women	*N*
Neutral	74	74	148	74	74	148
Female	74	74	148	72	72	144
Male	72	72	144	74	74	148
Total	220	220	440	220	220	440

Notes: This table summarizes the number of observation by block and treatment, keeping only gender balanced groups.

Table B3: Summary statistics: Block 1 by treatment (gender balanced only)

	(1) Neutral		(2) Female		(3) Male		(1-2)	(1-3)	(2-3)
								$p-$value	
	Mean	SD	Mean	SD	Mean	SD			
Age (in years)	21.48	2.37	21.59	3.52	21.47	2.55	$p = 0.274$	$p = 0.936$	$p = 0.351$
Business & Economics (dummy)	0.55	0.50	0.47	0.50	0.52	0.50	$p = 0.163$	$p = 0.569$	$p = 0.413$
SOEP risk attitude (0-10)	6.53	1.79	6.61	2.14	6.48	1.92	$p = 0.613$	$p = 0.989$	$p = 0.586$
High intrinsic lying cost (dummy)	0.34	0.47	0.33	0.47	0.27	0.45	$p = 0.902$	$p = 0.214$	$p = 0.262$
High social lying cost (dummy)	0.14	0.34	0.16	0.37	0.10	0.31	$p = 0.513$	$p = 0.415$	$p = 0.145$
Agentic trait (1-5)	3.48	0.61	3.42	0.64	3.47	0.62	$p = 0.487$	$p = 0.846$	$p = 0.417$
Communal trait (1-5)	3.89	0.54	3.92	0.58	3.92	0.54	$p = 0.535$	$p = 0.626$	$p = 0.896$
Observations	148		148		144				

Notes: This table displays summary statistics of participants by treatment in Block 1. The p-values reported are from chi-square tests for dummy variables and ranksum tests for continuous variables.

Table B4: Summary statistics: Block 2 by treatment (gender balanced only)

	(1) Neutral		(2) Female		(3) Male		(1-2)	(1-3)	(2-3)
								$p-$value	
	Mean	SD	Mean	SD	Mean	SD			
Age (in years)	21.34	2.15	21.35	2.60	21.83	3.61	$p = 0.567$	$p = 0.997$	$p = 0.621$
Business & Economics (dummy)	0.52	0.50	0.53	0.50	0.49	0.50	$p = 0.805$	$p = 0.642$	$p = 0.478$
SOEP risk attitude (0-10)	6.45	1.95	6.58	1.97	6.59	1.94	$p = 0.377$	$p = 0.374$	$p = 0.960$
High intrinsic lying cost (dummy)	0.30	0.46	0.27	0.45	0.36	0.48	$p = 0.531$	$p = 0.268$	$p = 0.085$*
High social lying cost (dummy)	0.11	0.32	0.12	0.33	0.16	0.37	$p = 0.790$	$p = 0.239$	$p = 0.366$
Agentic trait (1-5)	3.42	0.61	3.48	0.66	3.47	0.60	$p = 0.351$	$p = 0.688$	$p = 0.588$
Communal trait (1-5)	3.94	0.58	3.91	0.54	3.88	0.54	$p = 0.703$	$p = 0.342$	$p = 0.564$
Observations	148		144		148				

Notes: * $p < 0.10$. This table displays summary statistics of participants by treatment in Block 2. The p-values reported are from chi-square tests for dummy variables and ranksum tests for continuous variables.

Table B5: Trivia performance difference across blocks given gender.

	Block 1		Block 2		Block 1 vs. 2 ($p-$value)	
	Men	Women	Men	Women	Men	Women
Neutral	13.09	13.11	12.97	13.30	p = 0.680	p = 0.787
	(3.75)	(3.11)	(3.58)	(2.93)		
Female	11.42	12.59	12.28	13.28	p = 0.113	p = 0.147
	(3.65)	(3.09)	(3.79)	(2.70)		
Male	11.43	11.49	11.80	11.18	p = 0.851	p = 0.562
	(3.48)	(3.08)	(3.17)	(3.07)		

Note: This table reports the mean number of correct answers provided in the trivia quizzes (standard deviation in parentheses). $p-$values reported are from ranksum tests.

Table B6: Fraction of men and women classified as High types by block and treatment.

	Block 1		Block 2		Men vs. Women ($p-$value)	
	Men	Women	Men	Women	Block 1	Block 2
Neutral	50.00%	50.00%	54.05%	45.95%	p = 1.000	p = 0.441
Female	43.24%	56.76%	47.22%	52.78%	p = 0.139	p = 0.617
Male	58.33%	41.67%	56.76%	43.24%	p = 0.066	p = 0.139

Note: This table reports the fraction of men and women who are classified as High types (i.e., top two performers in the trivia quizzes) by block and treatment. Across-block comparisons reveal no significant differences (Neutral p = 0.742, Female p = 0.740, Male p = 0.868). $p-$values reported are from exact tests.

Table B7: Diff-in-Diff comparisons of belief and communicated ability across blocks given treatment.

Dependent variable	(1)
Diff.(Belief & Comm.)	Coeff.
	(St.Err.)
Neutral T.	*ref.*
Female T.	-3.405
	(3.009)
Male T.	-2.002
	(2.944)
Block 2	-6.291*
	(3.231)
Block 2 X Female T.	6.833
	(5.389)
Block 2 X Male T.	-2.457
	(4.768)
Constant	21.676***
	(2.129)
N	880
Clusters	110
R-Square	0.019
$p-$values from Wald tests	
Block 1 vs. 2 in Neutral T.	p = 0.054
Block 1 vs. 2 in Female T.	p = 0.876
Block 1 vs. 2 in Male T.	p = 0.001

Notes: This table reports presents the coefficients from OLS regressions in which the dependent variable is the difference between belief and communicated ability. Independent variables include treatment dummies (with Neutral treatment as the reference category), a block dummy (1 for Block 2, 0 otherwise) and their interaction terms. Standard errors are clustered at the group level. * p<0.10; ** p<0.05; *** p<0.01.

Table B8: Determinants of communicated ability in Block 1

Dependent variable	(1)	(2)	(3)
Communicated ability	Coeff.	Coeff.	Coeff.
	(St.Err.)	(St.Err.)	(St.Err.)
Neutral T.	*ref.*	*ref.*	*ref.*
Female T.	4.581	7.644*	8.145*
	(4.768)	(4.444)	(4.276)
Male T.	-1.647	0.331	3.118
	(4.664)	(4.598)	(4.298)
Women	1.757	5.944	6.937*
	(3.875)	(4.023)	(4.075)
Women X Female T	-2.081	-8.322	-8.578*
	(5.151)	(5.120)	(4.969)
Women X Male T	-5.854	-4.228	-6.477
	(5.442)	(5.644)	(5.342)
Belief about ability		0.424***	0.371***
		(0.049)	(0.052)
Constant	38.703***	10.607**	-24.577
	(3.503)	(4.166)	(15.529)
Control Variables	No	No	Yes
N	440	440	440
Clusters	110	110	110
R-Square	0.021	0.191	0.241
Post-estimation Wald tests			
(Women - Men) in Female T.	-0.32	-2.38	-1.64
	(3.39)	(3.18)	(3.39)
(Women - Men) in Male T.	-4.10	1.72	0.46
	(3.82)	(3.97)	(3.80)
Women in Male T. - in Neutral T.	-7.50*	-3.90	-3.36
	(3.89)	(3.18)	(3.06)
Men in Neutral T. - in Female T.	4.58	7.64*	8.15*
	(4.77)	(4.44)	(4.28)

Notes: This table reports presents the coefficients from OLS regressions in which the dependent variable is communicated ability in Block 1. Independent variables in model (1) include treatment dummies (with Neutral treatment as the reference category), sex dummy (1 for women, 0 otherwise) and their interaction terms. Model (2) includes private beliefs about being a High type. Model (3) further include control variables namely trivia performance, age, self-reported risk attitudes, personality traits (agentic and communal), and lying type dummies (high intrinsic cost and high social cost). Standard errors are clustered at the group level. * p<0.10; ** p<0.05; *** p<0.01.

Table B9: First-order beliefs about men and women by block and treatment.

	Block 1		Block 2		Men vs. Women ($p-$value)	
	Men	Women	Men	Women	Block 1	Block 2
Neutral	49.41	51.40	53.74	44.84	$p = 0.480$	$p = 0.023$
	(21.11)	(18.82)	(23.33)	(22.72)		
Female	47.78	48.78	37.56	58.68	$p = 0.746$	$p < 0.001$
	(22.55)	(22.55)	(19.22)	(23.46)		
Male	49.77	42.93	53.14	35.88	$p = 0.112$	$p < 0.001$
	(22.26)	(23.88)	(20.44)	(18.83)		

Note: This table reports the mean first-order beliefs about men and women (standard deviation in parentheses). First-order beliefs refer to the likelihood that each participant is perceived as a High type by the other three group members in the group. $p-$values reported are from ranksum tests.

Table B10: Second-order beliefs by men and women by block and treatment.

	Block 1		Block 2		Men vs. Women ($p-$value)	
	Men	Women	Men	Women	Block 1	Block 2
Neutral	51.30	49.46	52.45	46.18	$p = 0.616$	$p = 0.104$
	(24.10)	(23.80)	(29.42)	(23.17)		
Female	46.01	50.57	37.83	51.89	$p = 0.322$	$p = 0.001$
	(28.10)	(23.05)	(28.39)	(24.63)		
Male	49.57	40.80	48.53	33.82	$p = 0.066$	$p < 0.001$
	(30.46)	(24.51)	(24.40)	(20.85)		

Note: This table reports the second-order beliefs held by men and women (standard deviation in parentheses). Second-order beliefs refer to each participant's belief about the average first-order beliefs that the other three group members hold about them. $p-$values reported are from ranksum tests.

Table B11: Believed gender differences in ability by block and treatment.

	Block 1					Block 2					p-value
	$M >> W$	$M > W$	$M = W$	$M < W$	$M << W$	$M >> W$	$M > W$	$M = W$	$M < W$	$M << W$	Block 1 vs. 2
Neutral	1.35%	7.43%	64.86%	22.97%	3.38%	0%	8.78%	64.19%	22.97%	4.05%	$p = 0.816$
Female	0.68%	0.68%	35.14%	43.24%	20.27%	0%	2.08%	18.75%	48.61%	30.56%	$p = 0.006$
Male	12.5%	47.22%	34.03%	6.25%	0%	21.62%	43.92%	31.76%	1.35%	1.35%	$p = 0.029$

Notes: This table reports the distribution of believed gender differences in ability (non-incentivized) elicited at the end of the experiment (N=440). The response ranges from 'men are better than women', 'slightly better', 'equally good' to 'women are slightly better than men', 'women are better than men'. Modal responses are highlighted in gray. Pairwise comparisons across treatments in a given block reveal significant differences (all $p < 0.001$). p-values from exact tests.

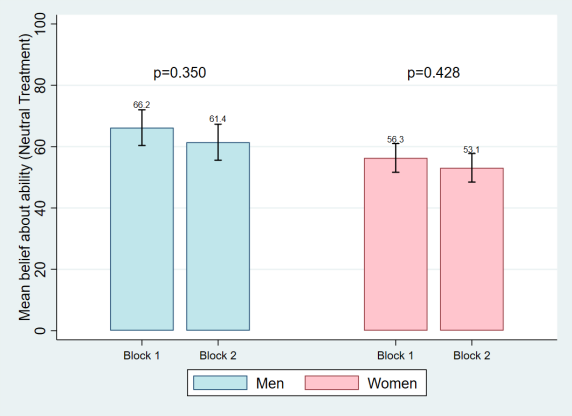
Table B12: Factor analysis of 16-item Personal Attributes Questionnaire

Item	Factor loading	Cronbach's alpha
Agentic traits		0.7369
Independent	0.2741	
Active	0.3412	
Competitive	0.4733	
Difficulty making decision (Reversed)	0.4160	
Never give up easily	0.3066	
Self-confident	0.5866	
Feel superior	0.4965	
Standup well under pressure	0.5078	
Communal traits		0.7183
Emotional	0.1581	
Devoted to others	0.3625	
Gentle	0.2964	
Helpful	0.4223	
Kind	0.4855	
Aware of others' feelings	0.4222	
Understanding of others	0.5026	
Warm in relations	0.5066	

Notes: This table presents the factor loadings and reliability estimates (Cronbach's alpha) for the 16-item Personal Attributes Questionnaire (N = 440). The two-factor structure shows a clear pattern of loadings corresponding to Agentic and Communal traits, each demonstrating acceptable internal consistency. Based on these results, we computed the mean scores of the items within each construct and used them as control variables in subsequent regression analyses.

Online Appendix C Appendix Figures

Figure C1: Beliefs about ability in Neutral treatment by gender across blocks.



Note: The figure displays beliefs about being a High type by men and women in the Neutral treatment across blocks. p-values reported are from ranksum tests.

Online Appendix D Robustness check: Ordering effect

In the main analysis, we report treatment effects pooling together participants who faced the same quiz in Block 2 but had faced a different quiz in Block 1 (or vice versa). However, it's plausible that the impact of stereotype on communicated ability observed in Block 2 may depend on the treatment assigned in Block 1 (e.g., individuals might react more strongly in Block 2 if they were exposed to a stereotyped treatment in Block 1, compared to when the treatment was neutral in Block 1).

Table D1 reports the mean communicated ability by men and women for each treatment sequence. Panel (a) corresponds to the Female treatment in Block 2, Panel (b) to the Male treatment in Block 2, and Panel (c) to the Neutral treatment in Block 2. These figures are consistent with in the main text. The gender gaps in communication in stereotyped tasks emerge in Block 2 (see Panel a and b). Comparing the difference in communicated ability in Block 1 and 2 by men (or by women) across NeutralFemale and MaleFemale, and NeutralMale and FemaleMale do not yield any significant differences (the smallest p-value is 0.289). This suggests that the stereotype effects found in Block 2 are independent of the treatment assigned in Block 1.

Table D1: Communicated ability by men and women by treatment sequence.

	Block 1		Block 2		Men vs. Women ($p-$value)	
	Men	Women	Men	Women	Block 1	Block 2
Panel (a)						
NeutralFemale	38.64	41.14	30.03	41.44	$p = 0.577$	$p = 0.055$
	(28.91)	(23.59)	(24.04)	(23.53)		
MaleFemale	29.69	32.39	22.03	38.53	$p = 0.652$	$p = 0.010$
	(22.02)	(23.33)	(19.94)	(28.95)		
Panel (b)						
NeutralMale	38.76	39.82	32.95	23.11	$p = 0.758$	$p = 0.039$
	(24.81)	(25.07)	(21.79)	(18.75)		
FemaleMale	47.17	43.69	49.03	28.81	$p = 0.488$	$p < 0.001$
	(22.91)	(20.66)	(21.82)	(17.56)		
Panel (c)						
FemaleNeutral	39.61	42.26	42.97	30.53	$p = 0.688$	$p = 0.060$
	(26.54)	(23.33)	(28.58)	(21.00)		
MaleNeutral	44.42	33.53	49.06	45.56	$p = 0.099$	$p = 0.600$
	(28.91)	(24.03)	(29.65)	(23.90)		

Note: This table reports the mean communicated ability by men and women across treatment sequences. Panel (a) corresponds to the Female treatment in Block 2, Panel (b) to the Male treatment in Block 2, and Panel (c) to the Neutral treatment in Block 2. Standard deviations are shown in parentheses.
Communicated ability refers to the cheap-talk message about one's likelihood of being a High type, and is the inverse measure of strategic incompetence. $p-$values reported are from ranksum tests.

We do observe, however, an interesting pattern when the Neutral treatment is assigned in Block 2 (see Panel c). Among men, changes in communicated ability in Block 1 and 2 do not differ significantly between FemaleNeutral and MaleNeutral ($p = 0.840$, ranksum test). In contrast, the difference is highly significant

among women ($p < 0.001$, ranksum test). This suggests that women seem to respond differently in the Neutral treatment in Block 2 depending on whether they were assigned to the Female or Male treatment in Block 1.

This pattern of communicated ability suggests that the neutral quiz was perceived as *female-typed* when it followed a Male treatment, compared to when it followed a Female treatment. This interpretation is consistent with the (non-incentivized) beliefs about gender differences in ability elicited at the end of the experiment. The distribution of stereotypical belief in Block 2 shifts from 'neutral' in FemaleNeutral to slightly female-typed in MaleNeutral ($p = 0.025$, exact test). Overall, these results indicate that the treatment ordering matters only when the Neutral treatment is presented in Block 2.

Online Appendix E Calibration of Trivia Quizzes

Separate Experiment: Three sessions (N=65) were conducted with an independent sample drawn from the same subject pool prior to the main experiment with the goal of calibrate the trivia quizzes. The experiment consisted of two parts. In part 1, participants answered six sets of 20 trivia questions (two sets per treatment). The sets were randomly assigned at an individual level. In part 2, participants performed six estimation tasks, which corresponded to each trivia quiz in part 1. In each estimation task, they saw each trivia question and were asked to guess the number of men and women in the session who have answered correctly to the question. Participants were incentivized for the accuracy of their response. At the end of the experiment, two tasks (one from each part) were randomly selected for an additional payment. See full instructions on page 29 of this appendix.

Question Categories: Questions belong either to neutral, female- of male-typed topics. The question categories for the neutral topics were cities and countries, earth and nature, food, animals, boardgames, popular films and books. Those for the female-typed topics were cooking, home, art and craft, fashion and clothes, mind-body sports, musical films and romance novels. Those for the male-typed topics were fixes and repairs, information and technology, video games, cars, competitive sports, action films and sci-fi books. All questions tested are listed on page 26 of this appendix.

Selection Criteria: Table E2 to E7 display, for each question, the fractions of men and women who answered correctly (i.e., actual performance difference), and the *believed* fractions of men and women who answered correctly (i.e., believed performance difference). We selected the trivia questions for the main experiment following these criteria:.

1. In terms of performance, we considered questions without significant differences between the fraction of men and women who answered the question correctly. We also checked that the mean trivia scores by men and women in the selected questions did not significantly differ.

2. In terms of belief about performance:

   - For the Neutral treatment, we considered questions in which participants believed there was no significant differences between the fraction of men and women who answered the question correctly.
   - For the Female and Male treatment, we considered questions in which participants believed the gender gap existed (i.e., believed that the fraction of women (men) who answered correctly to be higher than that of men (women) for the Female (Male) treatment).

Table E1 shows the mean trivia scores by men and women in the selected questions, and the believed fraction of men and women correctly.

Table E1: Trivia performance and believed gender gap in selected questions by treatment.

	Neutral		Female		Male		Men vs Women (p-value)		
	Men	Women	Men	Women	Men	Women	Neutral	Female	Male
Trivia performance	13.53	13.30	12.69	12.73	13.15	12.39	$p = 0.588$	$p = 0.6708$	$p = 0.2168$
	(2.96)	(2.91)	(2.63)	(2.11)	(2.46)	(2.82)			
Believed performance	0.74	0.74	0.65	0.75	0.73	0.67	$p = 0.684$	$p < 0.001$	$p < 0.001$
	(0.11)	(0.10)	(0.11)	(0.09)	(0.09)	(0.09)			

Notes: This table reports mean trivia scores (actual performance) and mean *believed* fractions of correct responses by men and women for each treatment. All statistics are based on the subset of trivia questions selected after the calibration experiment. Standard deviations are shown in parentheses. $p-$values are from ranksum tests for trivia performance and signrank tests for believed performance.

Table E2: Actual and believed gender gaps in trivia performance: Neutral Set 1

Question	Name	Actual Gap		p-val	Believed Gap		p-val
		Men	Women	(exact)	Men	Women	(signrank)
1	Eurovision	0.41	0.15	0.028**	0.51	0.54	0.020**
2	Kiwi	0.44	0.39	0.804	0.55	0.53	0.010**
3	Switzerland	0.19	0.06	0.149	0.51	0.51	0.993
4	Strawberry	0.88	0.70	0.130	0.82	0.85	0.479
5	Deadsea	0.94	0.73	0.044**	0.76	0.76	0.998
6	Desert rock	0.44	0.33	0.450	0.56	0.56	0.641
7	Pasta	0.56	0.67	0.450	0.63	0.68	0.003***
8	Durian	0.59	0.42	0.218	0.56	0.58	0.426
9	Unpopular pizza	0.19	0.09	0.303	0.63	0.63	0.788
10	Flamingo	0.84	0.94	0.258	0.60	0.62	0.487
11	Wombat	0.53	0.36	0.216	0.50	0.50	0.751
12	Octopus	0.53	0.39	0.324	0.55	0.55	0.398
13	Charades	0.53	0.67	0.317	0.74	0.74	0.849
14	Jenga	0.66	0.61	0.798	0.75	0.73	0.147
15	Monopoly	0.91	0.85	0.708	0.88	0.88	0.509
16	Lion King	0.91	0.88	1.000	0.89	0.89	0.4732
17	Titanic	0.94	0.88	0.672	0.77	0.82	<0.001
18	Pixar	0.56	0.36	0.138	0.60	0.59	0.566
19	Harry Potter	0.91	0.88	1.000	0.79	0.80	0.556
20	Willy Wonka	1.00	0.82	0.024**	0.87	0.89	0.003***
N		32	33		65	65	

Notes:This table reports, for each trivia question, (i) the actual fractions of men and women who answered correctly, and (ii) participants' *beliefs* about these fractions. The column "Actual Gap" refers to observed performance differences between men and women, with corresponding p-values from exact tests. The column "Believed Gap" refers to the perceived performance differences elicited in the belief tasks, with p-values from Wilcoxon signed-rank tests.

Table E3: Actual and believed gender gaps in trivia performance: Female Set 1

Question	Name	Actual Gap		p-val	Believed Gap		p-val
		Men	Women	(exact)	Men	Women	(signrank)
1	Bechamel	0.44	0.67	0.083*	0.64	0.77	<0.001
2	Carbonara	0.56	0.48	0.622	0.72	0.78	<0.001
3	Braising	0.19	0.06	0.149	0.51	0.62	<0.001
4	Scrub daddy	0.44	0.36	0.617	0.54	0.64	<0.001
5	Dryer	0.75	0.94	0.044**	0.69	0.83	<0.001
6	Disinfectant	0.53	0.64	0.452	0.59	0.71	<0.001
7	Andy Warhol	0.66	0.64	1.000	0.64	0.69	0.003
8	Origami	0.97	0.94	1.000	0.89	0.90	0.340
9	Glitter	0.28	0.39	0.434	0.67	0.62	0.006
10	Yoga meaning	0.59	0.52	0.620	0.58	0.64	<0.001
11	Yoga block	0.31	0.42	0.310	0.54	0.66	<0.001
12	Pregnancy sport	0.59	0.85	0.028**	0.68	0.82	<0.001
13	Coco Chanel	0.44	0.70	0.025**	0.55	0.73	<0.001
14	Converse	0.50	0.52	1.000	0.66	0.73	<0.001
15	Levi's	0.59	0.48	0.459	0.70	0.74	0.011
16	La La Land	0.75	0.64	0.422	0.58	0.69	<0.001
17	Frozen	1.00	0.94	0.492	0.96	0.97	<0.001
18	The Wicked	0.72	0.79	0.574	0.57	0.66	<0.001
19	Bridgerton	0.50	0.58	0.622	0.41	0.67	<0.001
20	Twilight	0.94	0.94	1.000	0.70	0.86	<0.001
N		32	33		65	65	

Notes: This table reports, for each trivia question, (i) the actual fractions of men and women who answered correctly, and (ii) participants' *beliefs* about these fractions. The column "Actual Gap" refers to observed performance differences between men and women, with corresponding p-values from exact tests. The column "Believed Gap" refers to the perceived performance differences elicited in the belief tasks, with p-values from Wilcoxon signed-rank tests.

Table E4: Actual and believed gender gaps in trivia performance: Male Set 1

Question	Name	Actual Gap		p-val	Believed Gap		p-val
		Men	Women	(exact)	Men	Women	(signrank)
1	IKEA	0.53	0.33	0.136	0.77	0.66	<0.001
2	Car tire	0.81	0.73	0.558	0.81	0.72	<0.001
3	Leaky pipe	0.53	0.58	0.805	0.72	0.65	<0.001
4	Youtube video	0.47	0.12	0.003***	0.62	0.55	<0.001
5	Keyboard shortcut	0.78	0.45	0.010***	0.80	0.77	0.005
6	Siri	0.19	0.33	0.260	0.53	0.50	<0.001
7	Super Mario	1.00	0.88	0.114	0.95	0.91	0.001
8	Minecraft	0.94	0.91	1.000	0.92	0.83	<0.001
9	Esport	0.84	0.42	0.001***	0.84	0.61	<0.001
10	Mustang	0.66	0.39	0.048**	0.81	0.64	<0.001
11	Lamborghini	0.56	0.30	0.046**	0.66	0.51	<0.001
12	Ferrari	0.88	0.73	0.215	0.87	0.76	<0.001
13	Golf	0.22	0.27	0.775	0.54	0.46	<0.001
14	Ronaldo	0.72	0.39	0.013**	0.76	0.55	<0.001
15	Baseball	0.19	0.27	0.558	0.53	0.43	<0.001
16	007 films	0.69	0.42	0.046**	0.79	0.74	0.002
17	Ironman	0.59	0.30	0.025**	0.78	0.63	<0.001
18	The Terminator	0.47	0.317	0.202	0.62	0.53	<0.001
19	Ender's Game	0.22	0.12	0.339	0.55	0.48	<0.001
20	The Hunger Games	0.84	0.79	0.751	0.76	0.73	0.003
N		32	33		65	65	

Notes: This table reports, for each trivia question, (i) the actual fractions of men and women who answered correctly, and (ii) participants' *beliefs* about these fractions. The column "Actual Gap" refers to observed performance differences between men and women, with corresponding p-values from exact tests. The column "Believed Gap" refers to the perceived performance differences elicited in the belief tasks, with p-values from Wilcoxon signed-rank tests.

Table E5: Actual and believed gender gaps in trivia performance: Neutral Set 2

Question	Name	Actual Gap		p-val	Believed Gap		p-val
		Men	Women	(exact)	Men	Women	(signrank)
1	Australia	0.47	0.27	0.127	0.85	0.87	0.138
2	Amsterdam	0.66	0.79	0.277	0.76	0.77	0.657
3	Italy	0.97	0.94	1.000	0.93	0.93	0.569
4	Ostrich	0.84	0.79	0.751	0.73	0.72	0.222
5	Sea otter	0.56	0.73	0.200	0.56	0.59	0.051*
6	Baby hippo	0.22	0.36	0.277	0.55	0.53	0.130
7	Sunflower	1.00	1.00	-	0.83	0.87	0.004***
8	Mont St. Michel	0.72	0.70	1.000	0.67	0.68	0.560
9	Antarctica	0.81	0.73	0.558	0.71	0.70	0.619
10	Saffron	0.88	0.82	0.733	0.70	0.75	<0.001***
11	Tofu	0.84	0.79	0.751	0.67	0.74	<0.001***
12	Sushi	0.97	0.94	1.000	0.90	0.92	0.068*
13	Pictionary	0.72	0.76	0.783	0.78	0.80	0.321
14	Escape Room	0.97	0.91	0.613	0.89	0.88	0.782
15	Carcassone	0.41	0.39	1.000	0.65	0.55	<0.001***
16	Shrek	0.91	0.91	1.000	0.86	0.86	0.356
17	Star Wars I	0.94	0.79	0.149	0.90	0.85	0.0002***
18	Finding Nemo	0.69	0.79	0.408	0.88	0.88	0.012**
19	Petit Prince	0.53	0.48	0.806	0.69	0.73	0.005***
20	Harry Potter II	0.78	0.73	0.775	0.67	0.67	0.926
N		32	33		65	65	

Notes: This table reports, for each trivia question, (i) the actual fractions of men and women who answered correctly, and (ii) participants' *beliefs* about these fractions. The column "Actual Gap" refers to observed performance differences between men and women, with corresponding p-values from exact tests. The column "Believed Gap" refers to the perceived performance differences elicited in the belief tasks, with p-values from Wilcoxon signed-rank tests.

Table E6: Actual and believed gender gaps in trivia performance: Female Set 2

Question	Name	Actual Gap		p-val	Believed Gap		p-val
		Men	Women	(exact)	Men	Women	(signrank)
1	Ratatouille	0.78	0.88	0.339	0.78	0.85	<0.001
2	Chocolate mousse	0.84	0.97	0.105*	0.70	0.83	<0.001
3	Meringue	0.84	0.85	1.000	0.74	0.84	<0.001
4	Hygge	0.31	0.27	0.789	0.48	0.53	0.003
5	Indoor plant	0.09	0.06	0.672	0.48	0.60	<0.001
6	Home DIY	0.72	0.76	0.783	0.76	0.70	<0.001
7	Glass work	0.97	0.97	1.000	0.88	0.85	0.046
8	Paper marché 0.16	0.06	0.258	0.68	0.72	0.017	
9	Orange color	0.91	0.97	0.355	0.89	0.90	0.448
10	Yoga pose	0.25	0.36	0.422	0.51	0.61	<0.001
11	Meditation	0.94	0.88	0.672	0.71	0.76	0.008
12	Headspace	0.72	0.76	0.783	0.65	0.71	<0.001
13	Baggy jeans	0.25	0.21	0.775	0.54	0.71	<0.001
14	Shein	0.94	0.97	0.613	0.87	0.91	<0.001
15	Crocs	0.97	0.91	0.613	0.80	0.83	<0.001
16	Taylor Swift	0.50	0.21	0.020	0.50	0.62	<0.001
17	Encanto	0.53	0.55	1.000	0.48	0.57	<0.001
18	Lady Gaga	0.66	0.82	0.166	0.63	0.78	<0.001
19	Christian Grey	0.97	0.91	0.613	0.75	0.85	<0.001
20	Call me by your name	0.50	0.58	0.622	0.50	0.64	<0.001
N		32	33		65	65	

Notes: This table reports, for each trivia question, (i) the actual fractions of men and women who answered correctly, and (ii) participants' *beliefs* about these fractions. The column "Actual Gap" refers to observed performance differences between men and women, with corresponding p-values from exact tests. The column "Believed Gap" refers to the perceived performance differences elicited in the belief tasks, with p-values from Wilcoxon signed-rank tests.

Table E7: Actual and believed gender gaps in trivia performance: Male Set 2

Question	Name	Actual Gap		p-val	Believed Gap		p-val
		Men	Women	(exact)	Men	Women	(signrank)
1	Patch work	0.69	0.76	0.587	0.75	0.67	<0.001
2	Clogged drain	1.00	0.97	1.000	0.93	0.91	0.042
3	Door hinge	1.00	0.94	0.492	0.86	0.81	<0.001
4	Facebook	0.75	0.30	<0.001	0.76	0.74	0.010
5	Alexa	0.88	0.88	1.000	0.83	0.80	0.009
6	Augmented reality	0.69	0.61	0.606	0.69	0.62	<0.001
7	Sony	0.94	0.52	<0.001	0.91	0.73	<0.001
8	Health points	0.94	0.39	<0.001	0.87	0.72	<0.001
9	Zombie game	0.66	0.48	0.213	0.79	0.58	<0.001
10	Car's engine	0.66	0.55	0.450	0.82	0.71	<0.001
11	BMW	0.91	0.73	0.108	0.83	0.68	<0.001
12	Toyota	0.81	0.70	0.389	0.71	0.58	<0.001
13	Michael Phelps	0.81	0.61	0.102	0.78	0.64	<0.001
14	Cricket	0.84	0.70	0.240	0.74	0.65	<0.001
15	Soccer	0.84	0.55	0.015	0.90	0.78	<0.001
16	Louvre	0.72	0.64	0.598	0.63	0.60	0.011
17	Star Wars II	0.97	0.85	0.197	0.94	0.90	<0.001
18	Tom Cruise	0.97	0.82	0.105	0.79	0.71	<0.001
19	Dune	0.56	0.45	0.460	0.70	0.66	0.001
20	Valerian	0.44	0.30	0.310	0.58	0.52	<0.001
N		32	33		65	65	

Notes: This table reports, for each trivia question, (i) the actual fractions of men and women who answered correctly, and (ii) participants' *beliefs* about these fractions. The column "Actual Gap" refers to observed performance differences between men and women, with corresponding p-values from exact tests. The column "Believed Gap" refers to the perceived performance differences elicited in the belief tasks, with p-values from Wilcoxon signed-rank tests.

All trivia questions

Below reports all questions used in the calibration exercise, and which are selected for the treatment manipulation. Each question and set correspond to those reported in Table E2 to E7.

Neutral categories Set 1:

1. Which city will be hosting the Eurovision song contest in 2025? (Cities & Countries)
2. Which country has a law that makes it illegal to own just one guinea pig, because they must have a companion? (Cities & Countries)
3. Which country has a law that makes it illegal to own just one guinea pig, because they must have a companion? (Cities & Countries)
4. What is the only fruit that has its seeds on the outside? (Earth & Nature)
5. What makes the Dead Sea unique? (Earth & Nature)
6. What weird thing can happen to rocks in the desert? (Earth & Nature) **[Selected]**
7. Which type of pasta is shaped like little 'ears'? (Food) **[Selected]**
8. Which fruit is known as the 'king of fruits' due to its strong smell? (Food) **[Selected]**
9. What unusual pizza topping is popular in Sweden? (Food)
10. What strange thing happens to flamingos if they do not eat enough shrimp? (Animals)
11. Which animal's poop is cube-shaped? (Animals)
12. How many hearts does an octopus have? (Animals)
13. Which of the following best describes the Charades game setup? (Board games) **[Selected]**
14. Which board game requires players to remove pieces without making the structure collapse? (Board games) **[Selected]**
15. What is the main objective of Monopoly? (Board games)
16. What is the name of Simba's father in The Lion King? (Popular films) **[Selected]**
17. What was the issue with the lifeboats in the movie Titanic? (Popular films)
18. What was the first Pixar movie ever released? (Popular films)
19. Which of the following book series has sold over 500 million copies worldwide? (Popular Books) **[Selected]**
20. In "Charlie and the Chocolate Factory," what is the name of the eccentric factory owner?

Female categories Set 1:

1. Which of these ingredients is essential for making a classic French bechamel sauce? (Cooking)
2. What is the key ingredient in a traditional carbonara pasta? (Cooking) **[Selected]**
3. What is the cooking technique used to slowly cook food in liquid at a low temperature for a long period of time? (Cooking)
4. Which of the following is the famous cleaning product brand known for its smiley face sponge? (Home) **[Selected]**
5. Which of these items should never be put in the dryer, as it may easily get damaged? (Home)
6. Which of the following household items can act as a natural disinfectant? (Home) **[Selected]**
7. Which artist is known for colorful paintings of soup cans and Marilyn Monroe? (Art & Craft) **[Selected]**
8. What is the Japanese word for the traditional art of folding paper to create shapes? (Art & Craft)
9. Which of the following materials is commonly added to craft projects to create a sparkling effect? (Art & Craft)
10. What is the traditional meaning of the word "yoga"? (Mind-Body Sports) **[Selected]**
11. Which of these is a common accessory used in yoga practice for support? (Mind-Body Sports)
12. Which of the following statements is NOT true about exercises for pregnant women? (Mind-Body Sports) **[Selected]**
13. Which fashion designer is known for creating the little black dress? (Fashion & Clothes)

14. Which Converse sneaker model is considered the most iconic and remains a popular choice for casual footwear worldwide? (Fashion & Clothes)
15. Which brand is known for its high-quality denim and casual wear, featuring a logo with two horses? (Fashion & Clothes)
16. Which musical movie features a romantic dance scene in front of a sunset-lit Los Angeles skyline? (Musical Movies) **[Selected]**
17. Which Disney musical movie features a young queen who learns to embrace her true powers and the song Let It Go? (Musical Movies)
18. Which Broadway musical adaptation is based on The Wizard of Oz but tells the story from the witches' perspective? (Musical Movies)
19. In 'The Duke and I', the first novel to the Bridgerton series by Julia Quin, who is the main protagonist? (Romance Books)
20. Which novel series features the character Bella Swan, a young woman caught in a love triangle between a vampire and a werewolf? (Romance books) **[Selected]**

Male categories Set 1:

1. What tool is commonly used when assembling IKEA furniture? (Fixes & Repairs)
2. What should you do first when you notice a flat tire on your car? (Fixes & Repairs)
3. What household item can be used temporarily to stop a leaking pipe before calling a plumber? (Fixes & Repairs)
4. What was the first video ever uploaded on YouTube? (Info. & Tech)
5. Which of the following keyboard shortcuts is incorrect? (Info. & Tech)
6. In 2016, what bizarre thing happened when people asked Siri 'What's 0 divided by 0'? (Info. & Tech) **[Selected]**
7. In the Super Mario universe, what is the name of Mario's brother? (Video Games) **[Selected]**
8. Which type of game is Minecraft known as? (Video Games) **[Selected]**
9. Which of these competitive games is considered to have the most iconic and established esports scene, known for its massive global tournaments? (Video Games)
10. Which car company is known for producing the iconic Mustang, a symbol of American muscle cars?(Cars)
11. Before Lamborghini became known for high-performance sports cars, what did the company originally produce? (Cars)
12. Which Italian sports car brand is famous for its prancing horse logo? (Cars)
13. Which sport was banned in early Scotland because it distracted from military training? (Competitive Sports) **[Selected]**
14. Which soccer player has a statue in their hometown that looks nothing like them? (Competitive Sports)
15. Which sport once had a rule banning players from having beards? (Competitive Sports) **[Selected]**
16. What is the name of the spy played by Daniel Craig in the recent 007 films? (Action Movies)
17. Which of the following superhero characters is NOT from the DC Universe? (Action Movies)
18. In 'The Terminator' (1984), why does a cyborg assassin travel back in time to target Sarah Connor? (Action Movies) **[Selected]**
19. Which sci-fi novel features a military academy where children are trained to fight in an intergalactic war? (Sci-fi Books) **[Selected]**
20. Which dystopian novel features a competitive survival game in a post-apocalyptic world, where participants fight for their lives in a controlled environment? (Sci-fi Books) **[Selected]**

Neutral categories Set 2:

1. What is the only country that is also a continent? (Cities & Countries) **[Selected]**
2. Which capital city is known for having more bicycles than people? (Cities & Countries) **[Selected]**
3. Which country is shaped like a boot on the map? (Cities & Countries) **[Selected]**
4. Which bird is unable to fly but is the fastest on land? (Animals) **[Selected]**
5. Which animal holds hands while sleeping to avoid drifting apart? (Animals) **[Selected]**

6. What specie is Moo Deng, the famous baby animal that went viral online? (Animal) **[Selected]**
7. Which flower is known for following the sun during the day? (Earth & Nature)
8. Which of these real islands can be cut off during high tide? (Earth & Nature) **[Selected]**
9. Which of the following is a desert that receives snowfall? (Earth & Nature) **[Selected]**
10. Which spice is often called "red gold" due to its price? (Food)
11. What is tofu primarily made from? (Food)
12. In which country is sushi a traditional dish? (Food) **[Selected]**
13. Which classic board game involves drawing and guessing words or phrases based on illustrations? (Board games) **[Selected]**
14. In which game do players work together in teams to solve puzzles within a limited time to escape from a locked room or scenario? (Board games)
15. What popular game in France involves laying tiles to build a medieval landscape with towns, roads, and fields? (Board games) **[Selected but for Male treatment]**
16. What kind of creature is Shrek? (Popular films) **[Selected]**
17. Which film series is famous for the phrase "May the Force be with you"? (Popular films)
18. In the movie "Finding Nemo," what type of fish is Nemo? (Popular films) **[Selected]**
19. In Le Petit Prince, what does the prince ask the narrator to draw? (Popular books)
20. Which Harry Potter book features a giant snake living in the school's basement? (Popular books) **[Selected]**

Female categories Set 2:

1. Which of these ingredients is typically used to create a classic French ratatouille? (Cooking) **[Selected]**
2. What is the main ingredient that gives chocolate mousse its airy texture? (Cooking)
3. Which dessert is made from whipped egg whites and sugar, baked until crisp? (Cooking) **[Selected]**
4. What is the Danish concept that refers to creating a cozy, comfortable, and content atmosphere at home? (Home)
5. Which popular indoor plant is so low-maintenance it could probably survive a month without water and still look cute on your shelf? (Home) **[Selected]**
6. Which of these materials is commonly used in home DIY projects for creating custom furniture? (Home) **[Selected but for Male treatment]**
7. What material is typically used to create stained glass windows, often seen in churches or cathedrals? (Art & craft)
8. What is papier-mâché made of? (Art & Craft) **[Selected]**
9. Which two colors make orange when mixed? (Art & Craft)
10. Which of these is NOT a yoga pose? (Mind-Body Sports)
11. What is the primary benefit of meditation in mind-body sports? (Mind-Body sports) **[Selected]**
12. Which of these apps is specifically used for meditation and mindfulness practices? (Mind-Body sports)
13. Which type of pants were made iconic in the 1990s by celebrities like Britney Spears and the Spice Girls? (Fashion & Clothes) **[Selected]**
14. Which of these fashion brands has been criticized for its unsustainable business practices? (Fashion & Clothes) **[Selected]**
15. Which brand made a viral "ugly shoe" trend with its foam-like clogs? (Fashion & Clothes) **[Selected]**
16. Which famous pop star acted in the 2019 movie adaptation of Cats? (Musical Movies)
17. In Disney's Encanto, what does the song "We Don't Talk About Bruno" reveal about his visions? (Musical Movies) **[Selected]**
18. Who plays the lead role in the musical film A Star Is Born (2018)? (Musical movies) **[Selected]**
19. What is the name of the rich and mysterious man that falls for Anastasia Steele in Fifty Shades of Grey? (Romance books) **[Selected]**
20. In the film adaptation of Call Me by Your Name, based on the novel by André Aciman, where does the story take place? (Romance books) **[Selected]**

Male categories Set 2:

1. Which of these items can be used to patch a hole in a wall? (Fixes & repairs) **[Selected]**
2. When should you use a plunger? (Fixes & repairs) **[Selected]**
3. Which of these household items can be used to fix a squeaky door hinge as a lubricant? (Fixes & repairs) **[Selected]**
4. Which of these social media platforms was originally launched as a networking site for university students? (Info. & Tech)
5. What is the name of the popular robot assistant created by Amazon for controlling smart devices? (Info. & Tech) **[Selected]**
6. Which technology enables your phone's camera to identify and interact with real-world objects through digital overlays? (Info. & Tech) **[Selected]**
7. Which company is known for developing the PlayStation consoles? (Video games)
8. Which of these colors is typically associated with health restoration in video games? (Video Games)
9. Which game is set in a post-apocalyptic world and revolves around surviving against zombies? (Video games)
10. What is the first thing you should check if your car's engine won't start? (Cars) **[Selected]**
11. Which of these car brands is famous for its "German engineering" and known for producing luxury vehicles? (Cars)
12. Which Japanese automaker, originally a division of a textile company, became famous for its innovation in mass production and hybrid technology? (Cars) **[Selected]**
13. Which sport did Michael Phelps dominate, winning a record number of Olympic gold medals? (Competitive sports)
14. Which of these sports is played with a bat and a ball, and is most popular in countries like India and Australia? (Competitive sports)
15. In which sport can you get a "hat trick" by scoring three goals in a single game? (Competitive sports)
16. In the movie 'The Da Vinci Code', what famous Parisian landmark is featured in one of the key scenes? (Action films) **[Selected]**
17. Which movie features the famous quote, "I am your father"? (Action films) **[Selected]**
18. Which movie features Tom Cruise performing his own stunts while hanging from the side of a flying airplane? (Action films)
19. In Dune, what precious resource is only found on the desert planet Arrakis? (Sci-fi books) **[Selected]**
20. Which French comic series is considered a sci-fi classic and influenced the film The Fifth Element? (Sci-fi books)

Full instructions of this separate experiment (translated from French):

General Instructions

Hello and welcome to today's experiment. We will go through the instructions together and read them aloud. Please turn off your phone and put it away. You are not allowed to use your phone or talk with others during the experiment. You are not allowed to open tabs or computer programs other than the one opened for you. If you violate any of these rules, you will be excluded from the experiment immediately and you will not receive any payment.

During this experiment, you can earn money. The amount of money you earn depends on your decisions as well as those made by others in the session. Therefore, please read the instructions carefully.

The experiment is divided into two blocks, followed by a questionnaire at the end. Each block has six sets of tasks. You will receive the instructions for each block at the beginning of the block. The instructions are the same for all participants. You will have the opportunity to ask questions privately to the experimenter in case you do not understand the instructions. To do so, please press the red button on the side of your table or raise your hand. The experimenter will come to answer in private. Please do not ask questions aloud.

How will your decisions affect your earnings?
For this experiment, you receive 5 euros as your participation fee, plus additional earnings. The exact amount of the additional earnings is determined at the end of the session. The computer program randomly selects one task from each block. The points you have accumulated in these two selected tasks constitute your additional earnings. 100 points equal 4 euros. Decimals will be rounded. Since all tasks have an equal chance of being selected, it is in your best interest to make your decisions as if each task were selected and determined your additional earnings.

—

Block 1
Your task:
In Block 1, you will answer six sets of quizzes. Each set will contain 20 multiple-choice questions. These questions cover trivia related to various topics. You will see the topics in each set, and each question will be labeled with a label indicating the corresponding theme. Each question has five answer options and one of them will be correct. You will have **5 minutes** in total to submit your answers to the questions. The remaining time will be displayed at the top-right corner of the screen.

Your earnings in this block:
At the end of the experiment, the computer program will randomly select a quiz set. You will win based on your score in the selected set. Each correct answer earns 12 points. No penalty is applied for an incorrect answer or no answer. If you have any questions, press the red button. The experimenter will answer them privately. Otherwise, please click "Next" to begin this block.

—

Block 2
Your task:
In Block 2, there are six sets of estimation tasks, each corresponding to the quiz in the previous block. In each estimation task, you will see the same 20 questions. In addition, you will be informed of the distribution of participants by gender (male or female) in the current session. For each question, you must estimate how many male and female participants answered correctly. Therefore, you must provide two estimates for each question: one for male participants and one for female participants. You must therefore make 40 estimates in each set.

Your earnings in this block:
At the end of the experiment, the program randomly selects one of the six estimation tasks. Ten out of twenty questions will then be drawn to determine your additional gain in this block.

- If your estimate exactly matches the actual number of male/female participants who answered correctly, you receive 12 points.
- If your estimate deviates by one point from the actual value, you receive 6 points.
- Otherwise, you receive nothing.

It is therefore in your best interest to make all estimates as accurate as possible. If you have any questions, press the red button. The experimenter will answer them privately. Otherwise, please click "Next" to begin this block.

—