GRAPE
Group for Research in APplied Economics

GRAPE Working Paper #106

# Heuristics and Signals: Experimental Evidence on Information and Wage Discrimination

Katarzyna Bech-Wysocka, Magdalena Smyk

FAME | GRAPE, 2025

Foundation of Admirers and Mavens of Economics
Group for Research in Applied Economics

# Heuristics and Signals: Experimental Evidence on Information and Wage Discrimination

Katarzyna Bech-Wysocka
FAME|GRAPE
Warsaw School of Economics

Magdalena Smyk
FAME|GRAPE
Warsaw School of Economics

Abstract

Statistical discrimination theory explains wage differences between demographic groups by referring to differences in group averages or heuristic-based decision-making. This study investigates whether providing employers with accurate information about individual productivity affects wage-setting practices. We replicate a labor market scenario in which employers determine wages based on perceived productivity differences between male and female workers. Our experimental findings suggest that statistical discrimination influences initial wage decisions, but access to individual performance data reduces reliance on group-based heuristics. The dominant strategy when the actual information about performance is to share the resources according to contribution. We observe that in tasks where women statistically outperform, higher-scoring individuals tend to receive slightly less than their proportional contribution, whereas in tasks where men perform better, they tend to receive slightly more than their contribution. Furthermore, we show that with only statistical information, significant gender-based wage discrimination aligned with performance stereotypes occurs, but there is no gender discrimination under full information about performance. Our results contribute to the broader discussion on labour market inequalities and approaches to reducing statistical discrimination.

Keywords:
statistical discrimination, productivity, information, gender

JEL Classification
J71, J16, C91

Corresponding author
Magdalena Smyk, msmyk@grape.org.pl

Acknowledgements
This project was carried out as part of a research grant (2019/33/B/HS4/01645) funded by the National Science Centre, Poland. The experiment was pre-registered in the OSF Registry. Ethical approval for this study was obtained from the FAME|GRAPE Ethics Committee on December 12, 2024 (decision no. 7/2024). All errors are ours.

Published by: FAME | GRAPE
ISSN: 2544-2473


Foundation of Admirers and Mavens of Economics
Koszykowa 59/7
00-660 Warszawa
Poland

W | grape.org.pl
E | grape@grape.org.pl
TT | GRAPE_ORG
FB | GRAPE.ORG
PH | +48 799 012 202

# 1 Introduction

Despite decades of policy efforts and narrowing gender gaps in education and labor force participation, significant differences in earnings between men and women persist across labor markets. While part of this gap can be attributed to observable characteristics such as occupation, tenure, or hours worked, a substantial portion remains unexplained. One prominent theoretical explanation for this gender wage gap is statistical discrimination, i.e. a mechanism through which employers, operating under incomplete information, use group-level averages as proxies for individual productivity. However, there remains limited understanding of how access to accurate information about individual productivity affects or alters these decision-making strategies. Specifically, it is uncertain whether providing precise performance data reduces reliance on statistical stereotypes and mitigates discriminatory wage-setting practices.

This paper investigates how employers' wage-setting behavior responds to information asymmetries in contexts where they observe either statistical group-level indicators of productivity or precise individual-level performance data. Our objective is to replicate the dynamics of the labor market environment in determining wage levels for men and women. We investigate the role of employers' perceptions of gender-based productivity differences in the wage evaluation process. Neoclassical marginal productivity theory suggests that in competitive labor markets, wages are determined by the marginal productivity of labor, i.e. workers are paid according to their productivity levels (Clark, 1899). In empirical studies, we observe that employers often offer lower wages to women, potentially reflecting an expectation of reduced productivity associated with greater involvement in family responsibilities and a higher likelihood of absences, such as those related to child care. Recent OECD indicators show that the unadjusted (raw) gender wage gap in 2022 was approximately 11.4% (average for OECD countries) and 10.8% [1] (average for the EU countries).

The hypothesis is that employers' perceptions of lower productivity among female workers may be inaccurate and shaped by aggregate labor market observations, societal norms, and generalized public attitudes, and can be categorized as statistical discrimination. This study seeks to investigate whether access to accurate information on actual labor outcomes has the potential to change wage-setting practices. Specifically, the research explores whether employers, when informed of no actual productivity differences between genders, would offer equal wages to male and female employees.

In the context of the labor market, information regarding worker productivity is private prior to hiring. Workers can signal higher productivity through indicators such as educational attainment, but employers cannot directly observe actual productivity outcomes at the outset (Spence, 2002). To address this information asymmetry, employers may invest in acquiring more precise information, such as through screening mechanisms or pre-employment tests, enabling them to offer wages that more accurately reflect individual productivity levels (Autor & Scarborough, 2008; Pallais, 2014). However, obtaining such information is associated with costs, as employers must allocate resources to implement screening processes and tests. Our second hypothesis of interest stays that information acquired through costly efforts is more likely to influence decision-making (Caplin & Leahy, 2001). Specifically, we investigate whether the higher the cost of acquiring information about actual productivity, the greater its impact on final wage offers.

We address these issues through a controlled laboratory experiment designed to simulate wage allocation decisions in a simplified labor market setting. Participants take on the role of managers and are asked to divide earnings between a male and a female worker. In the first stage, managers only receive statistical information about average gender-based performance in a given task. In the second stage, they are provided with actual performance. The experimental tasks include one typically associated with higher female performance (emotion recognition)

[1] Eurostat reports 12.7%.

and one favoring males (math equation solving), allowing us to assess how task-specific gender stereotypes shape decision-making under uncertainty.

Our findings reveal that participants absorb statistical information about group-level demographic differences and incorporate it into their wage-setting decisions. Moreover, we observe that the dominant strategy among participants is to allocate wages in proportion to individual contributions, rather than in pursuit of self-interested profit maximization. What is more, in the solve equation task (where average performance favors male workers) participants in the second stage (when receiving individual performance data) tend to reward the higher-performing worker with slightly more than their proportional contribution. In contrast, in the emotion recognition task managers tend to allocate slightly less than the fair share to the higher-performing worker. Finally, we find gender-based wage discrimination aligned with performance stereotypes: women receive a premium in emotion recognition despite lower average scores, while men are favored in equation solving largely beyond performance differences. However, there are no evidence of gender discrimination under full information scenario.

This study contributes to three strands of the literature. First, it broadens the understanding of statistical discrimination by shifting focus from hiring to wage-setting with distinct strategic incentives and fairness considerations. Second, it investigates how the provision of individual-level productivity data influences managerial wage-setting behavior. Third, it enriches the experimental economics literature by showing how task framing and gender stereotypes interact with information access in shaping economic decisions.

The findings have implications for policy tools aimed at closing gender pay gaps, including pay transparency initiatives, structured performance assessments and pre-employment testing. While greater access to individual performance data can reduce reliance on group stereotypes, our results suggest that information alone may not fully eliminate bias, particularly in domains where gendered expectations are deeply ingrained.

The article is structured as follows. In the next section, we review the relevant literature on the topic. Then, we describe the experimental design and provide details about the participant sample. Subsequently, we present and discuss the experimental results, followed by the conclusion.

# 2 Literature review

Statistical discrimination is a form of decision-making based on group-level characteristics or averages rather than individual attributes (Arrow, 1973; Phelps, 1972). The problem arises from incomplete information, when acquiring detailed, individual-level information is costly or infeasible. However, decisions based on group-averages might be shaped by stereotypes or influenced by prevailing social beliefs and norms. The theory of statistical discrimination serves as a leading social scientific framework for analyzing discrimination in labor markets.

The body of literature on gender discrimination in the labour market is extensive and can be broadly categorized into three strands: hiring discrimination, occupational segregation and wage discrimination. For the *hiring discrimination*, the meta analysis by Galos and Coppock (2023) investigates more than 70 employment audit experiments, carried out from 1983 in more than 26 countries across five continents. They summarize that in higher-paying occupations predominantly occupied by men, being a woman has a negative effect on hiring probability, whereas in lower-paying occupations largely dominated by women, the effect is positive. The impact of statistical discrimination on hiring decisions has been recently analyzed by Tilcsik (2021), who via a survey experiment reveals that exposure to statistical discrimination theory increases managers' belief in the accuracy of stereotypes and their tendency discriminate during a hiring simulation. However, these effects were mitigated when participants read a critical commentary on the theory. In the *occupational segregation* context, the EU Commission report by Verashchagina and Bettio (2009) show that gender-based occupational segregation remains significant and has shown minimal change since the early 1990s. Additionally, the upward trend in gender-based sectoral segregation has become more evident during the current decade.

We add to the broad body of research on *wage discrimination* and its relationship with productivity. The idea that employers use gender as a heuristic, assuming women are less committed or available due to familial responsibilities leading to wage disparities is not new in the literature, and has been already discussed by Bielby and Baron (1986). Becker (1991) theory suggests that women may exhibit lower productivity in the workplace, potentially due to fatigue from domestic responsibilities or the need to conserve energy for anticipated tasks at home. Similarly, mothers may allocate time during work hours to concerns about their children, such as contacting them at home or arranging appointments. They may also take sick leave to care for their children's illnesses. It has been empirically well-established that women disproportionately engage in a greater share of domestic responsibilities (Greenstein, 2000; Sayer, 2016; Tichenor, 2005) and are more likely to provide care for children (Bianchi et al., 2006; Sullivan, 2006). Budig and England (2001) estimate that the motherhood pay gap among young American women is around 7% per child.

Clearly, the anticipation of lower productivity for women may substantially affect their wages. Significant qualitative research indicates that employers often use gender, race and ethnic background as proxies for productivity (Smith, 2002). However, Moss and Tilly (2001) show that their perceptions of applicants are frequently shaped by stereotypes rather than accurate assessments of group-level productivity.

To assure a more equitable wage-setting process, candidates whose productivity is not directly observable by employers may choose to reveal or signal their actual productivity levels (Akerlof, 1970; Spence, 1973). Naturally, this behavior is incentivized primarily among candidates with productivity levels exceeding those anticipated by employers or implied by statistical discrimination, as they stand to negotiate higher wages by distinguishing themselves from the group average. Pinkston (2003) show that workers whose group characteristics are associated with lower average productivity face wage penalties, even if their individual productivity is high. What is more, wage differences persist longer in scenarios where employers have limited opportunities to observe actual worker performance. Our context differs: instead of workers signaling their productivity, employers have the opportunity to acquire information about workers' actual productivity levels, albeit at a cost in terms of time, resources, and effort. Employers must balance the benefits of acquiring precise information with the associated costs to optimize hiring and wage-setting decisions. Altonji and Pierret (2001) explore how employers adjust their wage-setting behavior as they learn more about employees' productivity. The impact of screening, pre-employment skill tests and work trials on various employment outcomes has been recently discussed e.g. in Karan and Mercy (2021), Krekó et al. (2023).

Our experiment can be conceptualized as a continuous extension of the hiring discrimination task, wherein participants make hiring decisions between male and female candidates. We expand this framework by introducing the option to hire both candidates, but with the ability to assign distinct wages to each, thereby adding a layer of complexity and flexibility to the decision-making process.

In an experimental setting where participants were randomly assigned to the roles of employer or employee, Larribeau et al. (2013) demonstrate that employers rely on an employee's gender when assessing their suitability for a job. Regardless of the employer's gender, women were consistently rated significantly lower than men. The main contribution of our study is related to the construction of the experiment design. First, the managers decide about the wage setting rather than simply hiring. Second, the rules of wage setting are sensitive to different strategies of the manager, so we can observe the dominant strategies. Finally, we can observe at the individual (manager) level, how the strategy changes when the statistical information is replaced by the actual one.

Since our goal is to simulate the labor market mechanisms influencing wage setting for both men and women, it is crucial to consider the extensive body of research on gender wage gaps. The gender wage gap has been a central topic in labor economics, with a substantial body of empirical and theoretical research aiming to understand its causes and persistence. Numerous studies document that, on average, women earn less than men across a wide range of countries and labor market settings (Blau & Kahn, 2017). While part of this gap can be explained by observable factors such as education, experience, occupation, and hours worked, a significant portion remains unexplained, often attributed to discrimination or unmeasured productivity differences. Human capital theory has historically suggested that wage differences arise from gender-specific investments in education and work experience (Becker, 1964). However, more recent analyses e.g. by Goldin (2014) show that even as educational attainment among women has surpassed that of men in many countries, the wage gap persists. Occupational segregation (women are overrepresented in lower-paying sectors and roles) also contributes to the gap, with research indicating that jobs predominantly held by women tend to be systematically undervalued (Cortes & Pan, 2017). Moreover, the motherhood penalty and the flexibility of work arrangements have been identified as major contributors to the gender wage gap, particularly in high-skilled professions (Kleven et al., 2019). Experimental studies by Bohnet et al. (2016) have further demonstrated the role of implicit biases and gender norms in wage-setting behavior, revealing that even when productivity is held constant, women may be offered lower compensation. Building on this existing literature, we empirically examine whether, ultimately, female workers receive lower remuneration than their male counterparts. We assess the magnitude of the gender pay gap using both unadjusted comparisons and models that control for relevant productivity-related factors.

# 3 Experimental design and sample

In this section, we describe the experimental design, including a pre-study that served to construct statistical information, main experiment design, and the sample characteristics.

## 3.1 Pre-study

In November 2024, we tested six quizzes on an online sample of respondents aged 19 to 24 via the platform ANSWEO. The age range was chosen to ensure alignment between the pre-study sample and the demographic characteristics of the student population in the main experiment. Each respondent completed six quizzes: solve equation, multiplication, rhymes, emotion recognition, general knowledge, and a mental rotation task (see Appendix A for a detailed description of the tasks). The primary objective of our study is to examine statistical discrimination - specifically, gender-based average differences in outcomes. To this end, we selected tasks that, according to existing literature, are typically associated with better performance by women (emotion recognition and rhymes) and those in which men are reported to perform better (mental rotation, multiplication, solve equation, and general knowledge). Each quiz had a time constraint, with a response time limit of either 5 or 10 seconds per question. The entire session lasted approximately 10 minutes. Participants were compensated with a payment of \$1.50 for completing the survey, which is typical for a study of this length. To further incentivize participation, a \$10 bonus was awarded to the three highest scorers.

We aimed to collect responses from a balanced sample consisting of 50 women and 50 men. Based on the results of the pre-study, we prepared information on the statistical differences between women and men for each quiz. The statistical information showed the median contribution of each gender within two-person teams: comprising one man and one woman, who were randomly paired. That is, we calculate the median of $\frac{x_F}{x_F+x_M}$ and $\frac{x_M}{x_F+x_M}$, where $x_F$ is the woman's score and $x_M$ is the man's score, from randomly assigned mixed gender pairs.

For example, if in the emotion recognition task, a randomly paired woman scored 8/10 and a man scored 6/10, then

$$\frac{x_F}{x_F + x_M} = \frac{8}{6+8} = 0.57. \tag{1}$$

Equivalent fractions are constructed similarly for other pairs and the median of the resulting values is calculated to obtain a composite measure.

To reduce the duration of the main experiment, we included only the two tasks with the largest gender-based differences in contributions: one stereotypically (and empirically in our sample) associated with higher female performance (emotion recognition), and one with higher male performance (math equation solving).

## 3.2 Design

Participants were invited to a computer laboratory to participate in an experiment comprising two parts: (i) a performance assessment and (ii) a decision-making task. Appendix B provides a detailed account of the information displayed to participants during the experiment. At the beginning of the session, participants were presented with a detailed description of the upcoming tasks and were required to disclose their gender.

### 3.2.1 Performance task

In the performance assessment, each participant completed two tasks: *math equation solving* and *emotion recognition*. These tasks were identical to those used in the pre-study (to construct statistical discrimination statements), ensuring consistency across samples. Upon completion of each task, participants received immediate feedback on their individual performance, enabling them to make subjective judgments about the perceived difficulty of the tasks. Clearly, they do not observe the performance of other participants. The individual results of the performance assessment are later referred to as the *actual performance* or *actual outcomes* of the workers.

### 3.2.2 Decision-making task

In the decision-making phase of the experiment, each participant is assigned the role of a manager that oversees a two-worker team, and (at the same time) a worker, in at least one of the two-worker teams. For each manager, two workers of opposite genders are randomly assigned. The worker has a passive role in this stage. However, they receive wages determined by the manager's decisions. The manager serves as the sole decision-maker, determining the wages of their workers in two subsequent scenarios. Wage decisions are made iteratively and are influenced by external information provided to the manager during the task. Initially, the actual performance is unobserved; however, the manager learns this information at subsequent stages. The manager is explicitly informed that their own remuneration is contingent upon the performance of their assigned workers via

$$\alpha x_F + (1 - \alpha) x_M, \tag{2}$$

while their workers will earn, female worker:

$$\alpha (x_F + x_M), \tag{3}$$

and male worker:

$$(1 - \alpha)(x_F + x_M), \tag{4}$$

where $x_F$ is the actual performance of the female worker in a task, and $x_M$ of the male worker (both are initially unobservable to the manager) and $\alpha$ is the choice parameter (chosen by a manager) in $[0, 1]$. $\alpha$ can be interpreted as the proportion of the total output attributed to each worker. The combined output produced by the two workers is $(x_F + x_M)$, with each worker receiving a portion of this total output. The allocation of shares is determined by the manager's decision, which reflects their perspective on the relative contributions of each worker to the overall production. Participants are informed that their final pay will be randomly selected from (2), (3), (4), which means that it can be either from a choice when they were a manager, but also when they were a worker. In addition to any earnings resulting from the manager's decision, each participant received a fixed show-up fee of 10 PLN.

The manager chooses the value of $\alpha$ - share of the total budget allocated to each worker. The manager's remuneration is maximized when $\alpha = 1$ if male worker's actual performance exceeds that of male worker, or $\alpha = 0$ if female worker's actual performance is superior. By the construction of the manager's wage, $\alpha$ can be interpreted as premium for a manager for recognizing the most productive worker. However, the manager does not observe the actual performance of the workers at the beginning of the decision-making process.

Each manager chooses $\alpha$ twice. In the **first stage**, managers are informed about the specific task performed by their workers, but do not have access to the actual performance data of individual workers. Instead, they are presented with statistical statements and are required to select a value for $\alpha$. Managers were randomly assigned to two equally sized groups: in one group, their workers performed the emotion recognition task, while in the other, their workers completed the solve equation task.

Based on the pre-study the following statements were constructed and presented to managers:

- "in one of the previous rounds of the emotion recognition quiz, median contribution of women within randomly assigned mixed-gender pairs of workers, was around 0.60 of the joint performance score, and median contribution of men - around 0.40"

- "in one of the previous rounds of the solve equation quiz, median contribution of women within randomly assigned mixed-gender pairs of workers, was around 0.40 of the joint performance score, and median contribution of men - around 0.60".

Having access only to this statistical statement, managers select the value of $\alpha$.

In the **second stage**, managers decide again about the $\alpha$, but after receiving information about the actual performance of workers. However, managers are informed first that this information will be provided either at no cost (free-information treatment), or with a price (buy-information treatment). Randomly assigned prices are used to prevent managers from self-selection into specific treatment groups. Within the buy-information treatment group, all managers purchase information about actual performance, but the cost is either 2 PLN (cheap) or 6 PLN (expensive). Technically, each manager begins with an initial endowment of 10 PLN, which is then reduced by the price of the information purchased in their remuneration function. We incorporated a cost for acquiring information to investigate whether information obtained at a higher cost has greater influence on managerial decision-making compared to information acquired at a lower cost or freely. In other words, we aimed to assess whether managers place more weight on information that requires greater effort to obtain. Half of the managers were assigned to free information treatment and the prices were evenly spread between the other half. Managers were then presented with the actual performance data of the female and male workers in their team and subsequently made a second selection of the $\alpha$ parameter.

In the **final stage** of the experiment, we ask additional questions about age, attitude toward risk (Holt & Laury, 2014), fairness, and gender norms. The questions about attitude toward fairness were taken from World Values Survey (assess on a scale from 1 to 10 which statement is closer to your view: *Incomes should be made more equal* (1) or There should be greater incentives for individual effort (10)), and European Values Survey (On a scale 1 (fully agree) to 5 (fully disagree), do you agre with a statement *Large differences in people's incomes are acceptable to properly reward differences in talents and efforts.* We included these additional questions in the experiment to examine whether attitudes toward fairness and gender norms correlate with the distribution of $\alpha$, which (in the first stage) reflects managers' perceptions of gender-based performance differences among workers. Our initial hypothesis is that participants with more egalitarian gender norms will be less influenced by the statistical information - being more likely to choose equal shares and less likely to rely on stereotypes indicating that men or women perform better in specific tasks.

## 3.3 Discussion on the manager's strategy

The wage equations for managers and workers were designed to allow for the implementation of various strategic behaviors. Based on participants' decisions, it is possible, in certain cases, to identify the dominant incentive driving each manager's choices.

When the manager selects $\alpha = 0$ or $\alpha = 1$, the scenario resembles a standard hiring decision in which only one worker receives full compensation effectively. However, in this context, the "job" (i.e., the performance assessment) has already been completed. Consequently, such extreme allocations may be less frequently chosen by managers, as this would imply that the second worker receives no wages for their completed contribution to the total output. The managers are aware of that. In the first stage of the experiment, where only statistical information is provided, managers have no clear incentive to choose extreme $\alpha$ values. Lacking precise knowledge of individual worker performance, they tend to diversify the risk of low team output by distributing alpha more evenly, rather than allocating the entire share to one worker. However, in the second stage - once actual performance data are revealed—the dominant strategy, assuming managers seek to maximize their own earnings, is to allocate $\alpha$ entirely to the higher-performing worker. The manager's wage is maximized when 100% of the compensation is assigned to the top performer.

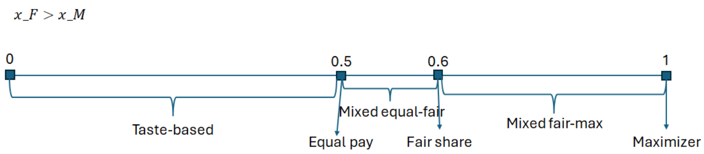
Managers may wish to allocate compensation equally between both workers, assigning $\alpha = 0.5$. This scenario is referred to as "equal pay." Certain participants may have strong preferences towards equality and might be inclined to distribute compensation equally between male and female workers regardless of their respective performance levels. At the end of the experiment, participants are explicitly asked about their preference for equality. This allows us to assess whether people who prioritize equality are more likely to choose equal pay decisions.

The third potential strategy is referred to as the "fair share" allocation. If the manager follows the fair share rule, the parameter $\alpha$ should accurately represent each worker's contribution to the final budget. In this context, "fair share" is defined as a scenario in which each worker is compensated proportionally to their performance: more productive workers receive a larger share, while less productive workers are still guaranteed a non-zero wage. In the initial stage, when the actual performance has not yet been observed, the "fair share" $\alpha$ corresponds to the proportion specified in the statistical discrimination statement. For example, in the emotion recognition task, managers are provided with the information that "within randomly assigned mixed-gender pairs of workers, women, on average, contributed 60% (0.60) of the joint performance score". Hence, the fair-share value of $\alpha = 0.6$. If the manager sets the value of $\alpha$ equal to the corresponding fraction, this can be interpreted as statistical discrimination information being fully absorbed. In this context, the manager believes that in the absence of the actual knowledge about performance it is fair to reward workers based on their group-level productivity, as the actual worker performance is likely to follow same patterns. Subsequently, once managers have observed the performance data, fair share $\alpha$ should adjust to reflect the actual performance differences between male and female workers within the team.

In addition, some managers adopt mixed strategies. The value of $\alpha$ between the "fair share" and 1 (or 0) reflects a combination of two motivations: fairness and own remuneration maximization. Once actual performance is observed, a manager focused solely on maximizing personal remuneration would move from a statistically discriminatory fair share $\alpha$ to an extreme value of 0 or 1. In contrast, managers motivated by fairness would adjust their $\alpha$ to reflect actual performance differences, rather than those inferred from statistical discrimination. Some managers also select $\alpha$ between 0.5 and the fair share allocation, indicating a preference for partially equitable but not fully equal distribution.

The value of $\alpha$ that favors the less productive workers, contrary to the direction suggested by the statistical discrimination statement or actual performace, can be interpreted as indicative of a preference for one gender that cannot be attributed to statistical or observed differences in productivity. We assume such preferences are taste-based and fall outside the scope of this research. Figure 1 provides a comprehensive summary of all possible values of $\alpha$ and discussed manager's strategies.

Figure 1: Possible manager's strategies




*Notes:* The figure presents four main and two mixed manager's strategies for all possible values of $\alpha$ when women outperform men in the emotion recognition task (fair share $\alpha = 0.6$). Similar figures can be drawn for other cases.

We identify four distinct types of managers: own-remuneration maximizers, equal-pay proponents, fair-share adherents, and taste-based decision-makers, plus 2 groups inbetween: mixed equal-fair and mixed fair-max. As a part of the analysis we quantify the proportion of managers within each category and examine how these proportions vary across different experimental scenarios.

## 3.4 Hypotheses of interest

Based on the theoretical and empirical examination of possible managerial decision-making strategies, we derive the following hypotheses:

**H1.** The *equal-share strategy* is more likely to be chosen when only statistical information is available.

**H2.** The *maximizing profit* and *mixed max-fair strategy* is more likely to be chosen when actual information is available.

The hypotheses H1 and H2 are based on the assumption that the statistical information is less precise. Therefore, in this stage, managers may select $\alpha$=0.5 not only when they believe that wages should be equal (regardless of actual contribution) but also as a risk mitigation strategy to avoid lower wages. When actual performance information becomes available, there is no risk of unintentionally reducing their own wages, allowing managers to safely increase their profit by choosing a max-min or maximizing strategy.

**H3.** The *maximizing profit* and *mix max-fair strategy* is more likely to be chosen when the actual information is costly than if it is free.

Finally, if information about actual productivity is costly, managers are more likely to compensate for this cost by choosing an $\alpha$ value that deviates from the actual contribution, depending on which worker's score was higher.

Additionally, regarding gender wage gaps, we hypotheses that:

**H4.** *The adjusted wage gap* is smaller when the actual information on productivity is known.

**H5.** Reduction of the gap is larger in the task in which on average men score higher than women.

If gender wage gaps arise primarily because employers lack direct observability of individual productivity and consequently rely on gender-based stereotypes in wage allocation, then wage differentials should be evident during the statistical information stage. However, upon disclosure of actual performance data, these wage disparities are expected to diminish and be predominantly accounted for by observed productivity differences rather than by prior stereotypical beliefs.

### 3.5 The sample

Participants were invited via email from the pool of students at SGH Warsaw School of Economics and the Faculty of Economics, University of Warsaw. They could voluntarily register using an online tool. The experiment lasted approximately 25 minutes, and participants earned an average of 24 PLN (6 USD). The experiment was conducted between January 20 and January 28, 2025.

There were 16 sessions, with an average of 14 participants per session. In total, data were collected from 223 students. Participants were randomly assigned to the treatments. The first treatment concerned which task formed the basis of payment - *math equation solving* (statistically favoring men) or *emotion recognition* (favoring women). The second treatment addressed the cost of obtaining actual performance information: free, cheap, or expensive. The distribution of participants assignment to treatments is presented in Table 1.

Table 1: Treatment assignment and gender of the participants

| | women | men | Total |
|---|---|---|---|
| *Task:* | | | |
| math equation solving | 49 | 66 | 115 |
| emotion recognition | 48 | 60 | 108 |
| *Cost:* | | | |
| free | 50 | 67 | 117 |
| cheap (2 PLN) | 22 | 37 | 59 |
| expensive (6 PLN) | 25 | 22 | 47 |
| Total | 97 | 126 | 223 |

*Notes:* The table shows the number of participants in the experiment assigned to treatments.

Slightly more participants were assigned to the *solve equation* and *cheap* treatments than originally planned. Additionally, the gender distribution within the sample is unbalanced, with significantly more men than women (a difference exceeding 10%). We control for these differences in the analysis.

# 4 Results

In the results section, we present an analysis of the participant's answers in the experiment. First, we present a raw analysis of $\alpha$ choices in the first (statistical information) and second (actual productivity information) stages. Second, we elaborate on strategy changes between statistical and actual information treatments. Third, we provide heterogeneity analysis.

We begin by presenting the actual performance of participants across the two tasks in Table 2.

Table 2: Actual performance of participants in tasks

| | average | | median | |
|---|---|---|---|---|
| | women | men | women | men |
| math equation solving | 7.98 | 8.63 | 8 | 9 |
| emotion recognition | 4.97 | 5.27 | 5 | 5.5 |

*Notes:* The table presents average and median scores of participants by gender and type of task.

Our observations indicate that the solve equation task was easier, as participants generally achieved higher scores on this task. Consistent with prior findings, men outperformed women in the solve equation task. However, contrary to common stereotypes, men also outperformed women in the emotion recognition task, which is typically associated with superior female performance. Additionally , in Table 3 we see that, in the solve equation task, managers observed worker productivity information consistent with the statistical data in 54% of cases, whereas in the emotion recognition task, this consistency was observed in only 23% of cases. Interestingly, the proportion of teams in which male and female workers performed identically is relatively high (exceeding 20%) and it is similar for both tasks. This suggests a greater possibility of shifts in managerial strategies between stages in the emotion recognition task.

Table 3: Actual performance of female and male workers presented to the manager

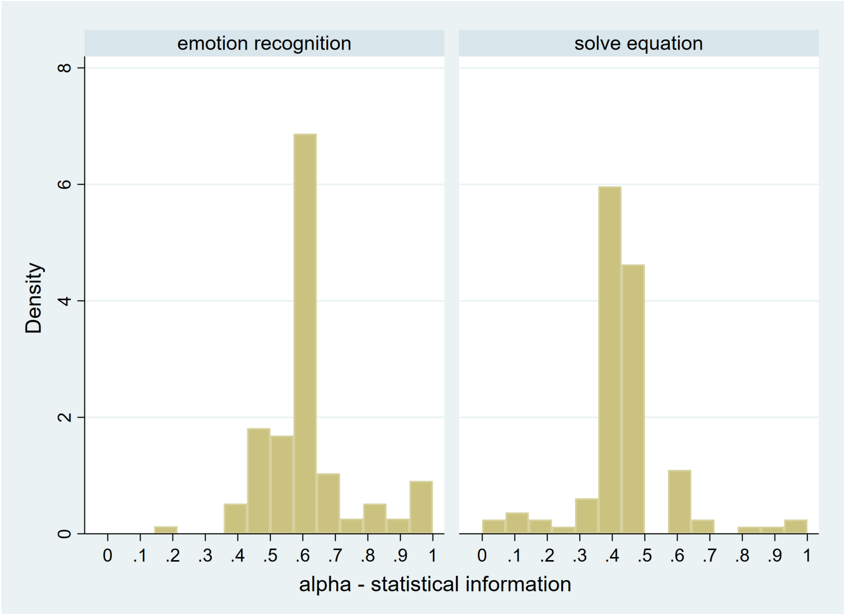
| | women $>$ men | women $=$ men | women $<$ men |
|---|---|---|---|
| solve equation | 25% | 21% | **54%** |
| emotion recognition | **23%** | 24% | 53% |

*Notes:* The table shows how the actual performance of female and male workers in teams corresponds to the statistical information. We present the percentages of cases in which women outperform men, where they achieve the same outcome, or where men outperform women. Cases consistent with the statistical information are highlighted in bold.

## 4.1 Distributions of *alpha's*

The participants' choice of $\alpha's$ in the first stage (only statistical information) is presented in Figure 2. The most frequently chosen answer in both tasks corresponds to the statistic provided by the experimenter (the median contribution of female workers in one of the previous sessions, which was 0.6 in emotion recognition and 0.4 in math equation solving). In the emotion recognition task, 50% of the managers selected $\alpha = 0.6$, and in the math equation solving 41.7% of the managers chose $\alpha = 0.4$. According to our nomenclature, these managers are classified as adopting the fair share strategy i.e. selecting $\alpha$ in a manner that fully incorporates the available statistical information. However, second most frequent answer in the solve equation task was 0.5 (22.6% of managers), while in emotion recognition it was 0.55 or 0.5 (10 and 12%, respectively). This suggests that equal share strategies were more likely among managers in the math equation solving treatment group. In other words, managers were more likely to allocate equal wages to male and female workers in task perceived as male-typed, compared to those perceived as female-typed.

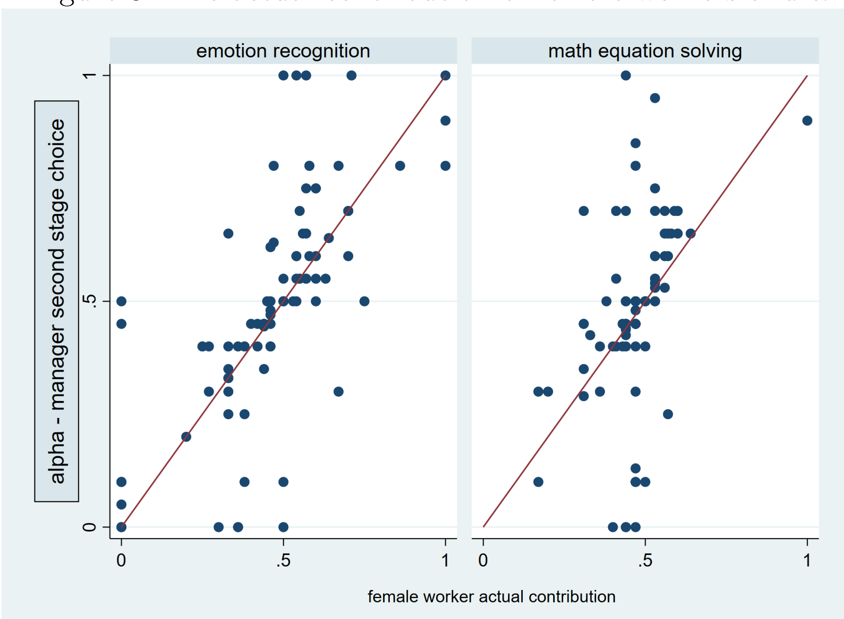
Figure 2: Distribution of $\alpha$'s in statistical information stage




*Notes:* Figures present histograms of $\alpha$'s chosen by managers in the first stage (statistical information) in emotion recognition (on the left) and math solving equation (on the right) treatments.

In the second stage (actual performance information), the analysis becomes less straightforward due to the case-specific nature of individual workers' contributions. In Figure 3, we compare $\alpha$'s with the actual contribution of the female worker in the two worker teams.

Figure 3: The actual contribution of female workers and $\alpha$




*Notes:* Figure presents scatterplots of relation between actual contribution and $\alpha$'s chosen by manager in the second stage (actual contribution information) in emotion recognition (on the left) and math solving equation (on the right) treatments.

On average, the difference between the contribution and $\alpha$ is 0.015 (SD = 0.16), indicating a slight tendency to allocate wages in favor of female workers. However, this difference is statistically insignificant. These results suggest that, on average, participants predominantly adopt a fair share strategy of allocating wages in proportion to the actual contributions of the workers. That is, the wage allocations align with the relative productivity of the workers. The correlation between the actual contribution and $\alpha$ is also quite high (*Pearson's* coeff $= 0.63$), but is significantly stronger in the emotion recognition task (*Pearson's* coeff $= 0.73$) compared to the solve equation task (*Pearson's* coeff $= 0.43$). In the *math equation solving* task, managers are slightly more likely to allocate a lower share of the wage to female workers relative to their actual contribution (31.3% versus 28.7% of such cases), while in the *emotion recognition* - higher (49.1% versus 46.9%). That is, in the male-typed task, female workers tend to be under-paid relative to their actual productivity, while in female-typed over-paid.

Surprisingly, the frequency of extreme $\alpha$ selections (i.e., $\alpha = 0$ or $\alpha = 1$) is relatively low, suggesting that managers place limited emphasis on personal profit maximization when determining wage allocations.

## 4.2 Manager's strategies

A closer examination of the managers' strategies during the statistical information and actual information stages is presented in Table 4. As previously observed in the histograms, the dominant strategy in both the emotion recognition task and the solve equation task when only statistical information was available was the statistical 'fair' strategy. This indicates that the largest group of participants simply followed the statistical differences in contributions. However, the proportion of managers who chose equal shares is significantly larger in the solve equation task than in the emotion recognition task.

Of particular interest is the transition of strategies between experimental stages, specifically examining how managers adjust their $\alpha$ allocations upon disclosure of actual performance information. Our observations indicate that these adjustments differ notably between the solve equation and emotion recognition tasks. In the emotion recognition task, the predominant strategy observed was the "mixed equal-fair" strategy. This indicates that managers generally opted to reward workers who achieved higher performance scores with wage shares that were lower than their exact proportional contributions but nevertheless exceeded an equal division. Such a pattern suggests that managers only partially incorporate performance information while balancing the unevenness in wage shares, potentially reflecting considerations of fairness or equity. Conversely, in the solve equation task, managers disproportionately rewarded the higher-scoring worker by allocating wage shares that exceeded the worker's actual contribution. This overcompensation suggests a departure from the fair share strategy, indicating that performance signals may have been intensified in the allocation process. Such behavior could reflect a stronger influence of task-specific stereotypes in the male-typed task.

Table 4: Distribution of managers strategies

| Strategy | Emotion recognition | | Solve equation | |
|---|---|---|---|---|
| | **Statistical** | **Actual** | **Statistical** | **Actual** |
| taste-based | 5.6% | 10.2% (4.6%) | 13.0% | 9.6% (7%) |
| equal ($\alpha = 0.5$) | 12.0% | 7.4% (19.4%) | **22.6%** | 12.2% (31.3%) |
| mix equal-fair | 12.0% | **30.6%** | 10.4% | 12.2% |
| statistical "fair"/fair ($\alpha$ = contribution) | **49.1%** | 21.3% (9.26%) | **41.7%** | 23.5% (4.35%) |
| mix fair-max | 15.7% | 21.3% | 10.4% | **39.1%** |
| max ($\alpha = 0$ or $\alpha$=1) | 5.6% | 9.3% | 1.7% | 3.5% |

*Notes:* The table presents the distribution of managers strategies in the emotion recognition and solve equation tasks in both stages. The percentages in parentheses present an alternative distribution, where if the workers contributed equally and manager selected $\alpha = 0.5$ such choice is assigned as equal strategy and not fair strategy. Also the number in parenthesis for taste-based strategy include only those who favor men in emotion recognition task, or women in solve equation task, although such choice lowered manager's remuneration.

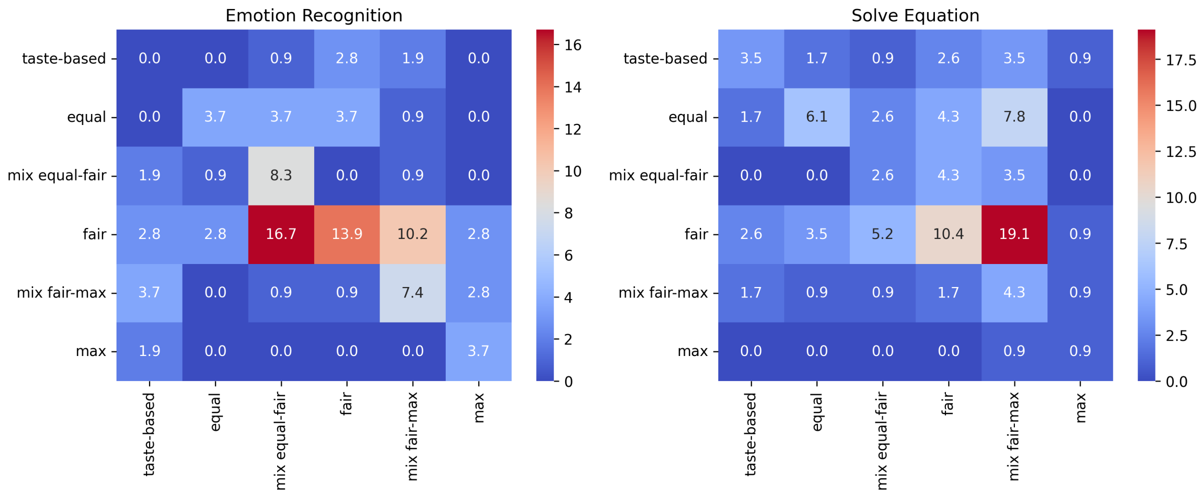
It is important to note that, during the actual productivity information stage, identifying the strategy was sometimes impossible. This difficulty arises from the fact that in 15% of teams in the emotion recognition task, and 20% in the solve equation task, female and male workers contributed equally - each accounting for 50% of the joint score. As a result, distinguishing among the equal, fair, and max strategies becomes infeasible. Consequently, alternative distributions are presented in parentheses to account for this ambiguity.

Table 4 summarizes the distribution of strategies across stages; however, it does not account for the fact that decisions were made repeatedly by the same managers. Therefore, the table reflects the aggregate prevalence of strategies at each stage rather than tracking individual-level transitions between strategies. For this reason, Figure 4 presents the individual-level transitions between strategies across stages, capturing the dynamic adjustments made by each manager. Overall, 37% of managers maintained the same strategy across both stages in the *emotion recognition* task, compared to 27.8% in the *math equation solving* task. Among managers who did not change their strategy, the most prevalent choice was the fair share strategy, that is, selecting $\alpha$ based on statistical information during the first stage and adjusting it in the second stage to reflect the actual contributions of the workers. The most frequent flow observed among managers is from fair share to mix equal - fair strategy in the emotion recognition task (shift toward partially balancing equitable distribution with proportional fairness) and from fair share to mix fair-max strategy in the solve equation task (tendency to combine proportional allocation with more extreme own profit maximizing approach).

Once again, the proportion of managers adopting a self-interested, profit-maximizing strategy remains relatively low. Notably, a subset of managers selected $\alpha$ values that contradict conventional expectations by allocating greater rewards to workers with lower contributions (taste-based strategy). The underlying motivations for this behavior are unclear: it may reflect a personal bias favoring one gender or stem from a misunderstanding of the compensation rules within the experimental framework.

Figure 4 complements the analysis by providing flows between strategies by manager. Interestingly, the most frequent shifts differed between the *emotion recognition* (female-typed) and *math equation solving* (male-typed) tasks. In the female-typed task, 16.7% of managers transitioned from a statistically "fair" strategy to a mixed equal–fair approach in the second stage, 13.9% consistently applied the fair strategy across both stages, and 10.2% shifted toward a mixed fair–maximizing strategy. In contrast, in the male-typed task, 19.1% of managers moved from a statistically fair strategy to a mixed fair–maximizing one, while only 10.4% maintained the fair strategy across both stages. These patterns suggest that, at the individual level, managers were more likely to additionally reward the top-performing worker—regardless of gender—in the male-typed task than in the female-typed one.

Figure 4: Moving between strategies per manager




*Notes:* The figures presents distribution of flows of managers strategies in the emotion recognition (on the left) and solve equation (on the right) tasks between first and second stage. The percentages represent the share of mangers who moved from the strategy described on the y axis (statistical information stage) to the strategy described on the x axis (actual information stage).

In the context of the hypotheses, Table 5 presents the results of tests for H1 through H3. Consistent with our expectations, in both tasks the provision of actual performance information is associated with a decreased prevalence of the equal-share strategy and an increased frequency of the mixed fair-max or max strategies. These differences are more pronounced in the solve equation task.

Table 5: Proportion tests for hypotheses

| | **Emotion recognition** | | **Solve equation** | |
|---|---|---|---|---|
| | z | p | z | p |
| H1. equal-share ↓ | 1.39 | 0.08 | 2.40 | 0.01 |
| H2. mix fair-max and max ↑ | -1.99 | 0.02 | -5.78 | 0.00 |
| H3. costly: mix fair-max and max ↑ | 0.37 | 0.36 | 0.38 | 0.35 |

*Notes:* Table presents z statistic and p-value coefficient from proportions tests.

The cost associated with accessing performance information - whether free, low, or high - appears to have had a negligible effect on participants' wage allocation behavior. While the average difference between actual contribution and $\alpha$ was 0.008 (SD = 0.014) in the free information treatment and 0.023 (SD = 0.017) in the paid information treatment (cheap: 0.0209; expensive: 0.030), these differences are statistically insignificant. Consequently, Hypothesis H3 was rejected for both tasks: the likelihood of selecting a profit-maximizing strategy or choosing an $\alpha$ value between the fair and max strategies did not increase with the cost of accessing actual productivity information. In other words, managers did not assign greater value to performance information obtained at a cost compared to information provided for free, and did not compansate own cost of aquiring such information.

## 4.3 Heterogeneity analyses

Finally, we examine participants' choices of $\alpha$ in the presence of actual worker performance data using a regression-based analytical framework. Table 6 presents the results of OLS regression analyses examining the share of remuneration allocated to the female worker, conducted separately for the *emotion recognition* and *math equation solving* tasks with set of regressors.

First, we observe a substantial difference in the explanatory power of the regression models across the two tasks, as indicated by the R-squared statistics. In the *emotion recognition* task, the simple model explains approximately 52% of the variation in managers' allocation choices ($\alpha$), whereas in the *math equation solving* task, the corresponding figure is only 23%. This suggests that workers' actual contributions account for over half of the variability in wage allocations in the *emotion recognition* task, but only about one-quarter in the *math equation solving* task, indicating a weaker alignment between performance and remuneration in the latter. When additional explanatory variables are added to the model, the R-squared increases to 62% in the *emotion recognition* task and 28% in the *math equation solving* task.

Table 6: OLS regressions of $\alpha$ (share of remuneration for female worker)

| | emotion recognition | | math equation solving | |
|---|---|---|---|---|
| contribution of female worker | 0.828*** | 0.612*** | 0.822*** | 0.574** |
| | (0.0792) | (0.110) | (0.160) | (0.226) |
| first stage $\alpha$ (statistical info) | -0.167 | -0.143 | 0.260** | 0.217** |
| | (0.119) | (0.114) | (0.102) | (0.107) |
| female worker - better | | 0.151*** | | 0.0828* |
| | | (0.049) | | (0.048) |
| male worker - better | | 0.0297 | | -0.00648 |
| | | (0.0487) | | (0.0430) |
| female manager | | 0.0355 | | 0.0299 |
| | | (0.0376) | | (0.0336) |
| gender norms (pro-equal) | | 0.0130*** | | -0.000249 |
| | | (0.00430) | | (0.00405) |
| fairness (pro-equal) | | 0.0131 | | 0.00463 |
| | | (0.009) | | (0.008) |
| risk neutral (vs risk averse) | | -0.0171 | | 0.0371 |
| | | (0.0457) | | (0.0450) |
| risk lover (vs risk averse) | | 0.0170 | | 0.00415 |
| | | (0.0359) | | (0.0336) |
| constant | 0.201** | -0.200 | -0.0245 | 0.0446 |
| | (0.0857) | (0.161) | (0.0885) | (0.166) |
| Observations | 108 | 108 | 115 | 115 |
| R-squared | 0.520 | 0.622 | 0.230 | 0.276 |

*Notes:* Table presents coefficients from OLS regressions of $\alpha$ (share of remuneration for female worker separately for emotion recognition (columns 2 and 3) and math equation solving (columns 4 and 5) treatments. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Second, and as expected, actual performance drives the $\alpha$ allocation in the second stage, indicating that managers predominantly base their wage decisions on observed productivity once this information becomes available. However, the coefficient on this variable is significantly different from 1, suggesting that while actual performance is a key determinant of $\alpha$ allocation, additional factors may also influence managers' decisions.

Furthermore, managers' allocation decisions in the first stage - based solely on statistical information about median differences in worker performance - are significantly correlated with their decisions in the second stage, when actual performance scores are available, but only in the solve equation task. Specifically, the more a manager allocated to the female worker in the first stage (despite the statistical advantage favoring male workers in this task), the more they tended to allocate to the female worker in the second stage, even after controlling for actual individual contributions. This suggests that initial predispositions or fairness considerations may persist and influence allocation behavior even when more precise performance information becomes available. In the emotion recognition task, managers' initial allocation decisions based on statistical information have no significant effect on their subsequent choices when actual performance data are available. This suggests that, in this context, wage allocations are driven almost exclusively by observed productivity, with no detectable influence of prior beliefs or initial allocations.

Higher individual contributions to the team's joint score are positively associated with increased remuneration in both task types. However, this effect is statistical significance only for female workers with higher contributions, and its magnitude is approximately twice as large in the emotion recognition task compared to the solve equation task. This suggests that performance-based rewards for high-contributing female workers are more pronounced in tasks stereotypically associated with female skills.

Among the managers' characteristics examined, only more gender-equal norms are significantly associated with additional remuneration awarded to female workers, after controlling for actual contribution. Other factors, including attitudes toward fairness, risk preferences, and the manager's own gender, do not exhibit a statistically significant correlation with allocation decisions in the second stage.

## 4.4 Gender Pay Gap

To investigate the presence of gender differences in final payments, we employ the parametric Oaxaca-Blinder decomposition (Oaxaca (1973), Blinder (1973)). This method decomposes the observed (raw) gender wage gap in log wages,

$$\ln \bar{W}_M - \ln \bar{W}_F = \underbrace{\beta^*(\bar{X_M} - \bar{X_F})}_{\text{characteristics}} + \underbrace{\overbrace{\bar{X_M}(\beta_M - \beta^*)}^{\text{male advantage}} + \overbrace{\bar{X_F}(\beta^* - \beta_F)}^{\text{female disadvantage}}}_{\text{coefficients}} \tag{5}$$

into a component attributable to differences in observable characteristics (such as performance measures), and a component due to differences in the returns to these characteristics, often interpreted as reflecting discriminatory factors.

A key element in the decomposition involves selecting the reference coefficient vector $\beta^*$, which defines the counterfactual wage structure i.e., the wage distribution that would have prevailed had the returns to characteristics been identical for men and women. This reference is typically constructed as a convex combination of the male and female coefficients:

$$\beta^* = \lambda * \beta^M + (1 - \lambda) * \beta^F. \tag{6}$$

In this analysis, we adopt the approach proposed by Fortin (2008), wherein the counterfactual coefficients are derived from a pooled regression that includes a gender indicator variable.

This approach allows us to decompose the observed wage differentials into a component attributable to differences in observable outcomes (i.e., productivity-related characteristics) and a residual component attributable to differential returns to these characteristics, commonly interpreted as discrimination.

Table 7: Oaxaca-Blinder decomposition of the gender pay gap

| | **Statistical** | | **Actual** | |
|---|---|---|---|---|
| | **Emotion recognition** | **Solve equation** | **Emotion recognition** | **Solve equation** |
| diff (M-F) | -0.2376*** | 0.1484*** | 0.0407 | 0.0591* |
| Explained | 0.0351*** | 0.0329*** | 0.0850*** | 0.0647*** |
| Unexplained | -0.2727*** | 0.1155*** | -0.0443* | -0.0056 |

*Notes:* The table presents the Oaxaca-Blinder decomposition of the gender pay gap. The first row reports the raw (unadjusted) differences in the natural logarithm of wages ($ln(wage)$) between male and female participants, disaggregated by information condition and task type. 'Explained' reflects the mean increase in women's wages if they had the same characteristics as men, i.e. the portion of the gender wage gap attributable to differences in actual task performance outcomes. 'Unexplained' quantifies the change in women's wages when applying the men's coefficients to the women's characteristic, i.e. adjusted wage gap - reflecting differences attributable to discriminatory factors after accounting for performance-related characteristics.

The outcomes presented in Table 7 vary considerably depending on the experimental stage and the type of the task.

In the initial stage of the experiment - when the manager has access only to statistical information - statistically significant gender-based differences in wages are observed across both tasks. In the emotion recognition task, the gender wage gap is negative, indicating that, on average, female workers receive higher compensation than male workers. This occurs despite lower average performance among women (mean score: 4.62) compared to men (mean score: 5.44), unobserved at this stage by managers. This suggests the presence of discrimination against men, wherein women receive a wage premium aligned with the statistical information provided to the manager. A similar pattern emerges in the equation-solving task, but in the opposite direction. A positive gender wage gap is observed, favoring male workers. Although part of this gap can be explained by differences in performance - men outperform women on average - only 22% of the wage differential is attributable to this difference. The remaining 78% reflects discriminatory treatment against female workers. In summary, under the statistical information condition, gender-based wage discrimination is evident in both tasks, and the direction of the discrimination is consistent with the content of the statistical information provided to the manager.

On the other hand, during the second phase - when the manager has access to actual outcomes - there is no statistically significant difference in the final wages paid to male and female workers who completed the emotion recognition task. In contrast, for the equation solving task, a gender-based wage difference is observed, with male participants receiving higher compensation. However, this difference is entirely attributable to differences in task performance: male participants, on average, achieved higher scores (mean score for men: 8.7; for women: 7.9). Thus, under the full information condition, there is no gender wage gap in the emotion recognition task, and although a wage gap appears in the *math equation solving* task, it can be fully explained by performance differences, indicating an absence of gender-based discrimination in this context.

# 5 Conclusions and discussion

This study investigates the mechanisms of statistical discrimination in wage-setting and the role of individual productivity information in reducing reliance on group-based heuristics. Through a carefully designed laboratory experiment simulating a simplified labor market, we provide evidence that when only group-level statistical information is available, decision-makers rely heavily on gender-based performance stereotypes. However, once individual performance data is introduced (even when associated with a monetary cost) this reliance significantly declines, and wage allocations more closely align with actual productivity.

A key finding is that under statistical information, wage-setting reflects performance stereotypes: women receive a wage premium in tasks perceived as female-typed (emotion recognition), while men benefit from higher wages in male-typed tasks (equation solving), even when actual performance contradicts those stereotypes. The Oaxaca-Blinder decompositions confirm that these disparities in wages are largely unexplained by observable productivity differences, suggesting the presence of gender-based discrimination. Importantly, these gaps disappear once individual performance data becomes available, indicating that statistical discrimination is driven not by taste-based preferences but by information constraints.

While managers generally adopt a fair-share strategy (aligning compensation with estimated or actual productivity) our results reveal systematic deviations depending on task type. In the solve equation task, female workers are slightly more frequently under-rewarded relative to their actual contribution. This observation is in line with the tendency among managers to additionally reward superior worker in a mixed-gender team in *math equation solving* task. Conversely, in the emotion recognition task, the better-performing worker (often female) receives slightly less than their proportional share, possibly reflecting implicit fairness preferences or uncertainty about the reliability of performance measures. Surprisingly, introducing a cost to acquire performance information did not significantly affect managerial decision-making. Whether managers received individual productivity data for free or paid for it (at varying price levels), their wage-setting behavior remained consistent.

Our study contributes to the literature by directly demonstrating that wage discrimination driven by statistical expectations can be effectively mitigated when accurate, individual-level productivity information is available. However, residual biases remain in contexts aligned with traditional gender roles, underscoring the stickiness of task-specific stereotypes.

These findings underscore the importance of transparency and performance-based assessment in employment contexts. Employers who rely on heuristics in the absence of individual-level data may perpetuate systemic biases. Policies promoting the collection and disclosure of productivity-related metrics (e.g. pre-employment assessments or structured evaluations) could help to reduce gender-based disparities in wages. However, such measures must be thoughtfully implemented to ensure that they do not introduce new forms of bias or disproportionately disadvantage groups with historically limited access to such signaling mechanisms.

Further investigation should explore how repeated exposure to individual performance data affects long-term wage-setting behavior, particularly whether initial reliance on group-level stereotypes diminishes over time with continued access to individual-level information. Such research could reveal whether behavioral adjustments observed in one-off experimental settings persist or deepen in repeated or real-world decision-making contexts. Moreover, longitudinal designs could help identify whether managers internalize performance signals in a way that eventually alters their prior beliefs about demographic groups, thereby reducing not only statistical discrimination but also implicit bias. In addition, exploring the interplay between statistical and taste-based discrimination across different types of labor market environments could provide valuable insights into the persistence and sources of inequality. While statistical discrimination arises from informational asymmetries and may be mitigated through better data on individual performance, taste-based discrimination (rooted in personal preferences or biases against certain groups) may persist even when full information is available. Future research could examine how these two forms of discrimination interact under varying institutional conditions, such as competitive versus monopsonistic labor markets, anonymous versus identifiable hiring processes, or short-term versus long-term employment relationships. Finally, extending this experimental framework to real-world hiring or promotion settings would also help to validate the external applicability of our results and assess their relevance beyond the laboratory context. While controlled experiments allow for the isolation of causal mechanisms, real-world labor markets introduce additional complexities (institutional constraints, strategic interactions, legal regulations, and organizational cultures) that may influence decision-making in ways not captured in a lab environment. By embedding similar experimental structures into field experiments or audit studies (e.g. by using randomized controlled trials in firms, recruitment platforms or HR processes) we could test whether access to individual productivity information has comparable effects on wage-setting and selection decisions under natural conditions. Such extensions would also show the practical challenges and ethical considerations involved in collecting, sharing, and using performance data in organizational settings.

## References


Ackerman, P., Bowen, K., Beier, M., & Kanfer, R. (2001, 12). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*, 797-825. DOI: 10.1037/0022-0663.93.4.797

Akerlof, G. A. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. DOI: 10.2307/1879431

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, *116*(1), 313–350. DOI: 10.1162/003355301556329

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in labor markets* (pp. 3–33). Princeton, NJ: Princeton University Press.

Autor, D. H., & Scarborough, D. (2008). Does job testing harm minority workers? evidence from retail establishments. *The Quarterly Journal of Economics*, *123*(1), 219–277.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press.

Becker, G. S. (1991). *A treatise on the family*. Cambridge, MA: Harvard University Press.

Bianchi, S. M., Robinson, J. P., & Milkie, M. A. (2006). *Changing rhythms of american family life*. New York: Russell Sage Foundation.

Bielby, W. T., & Baron, J. N. (1986). Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology*, *91*(4), 759–799. DOI: 10.1086/228350

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

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, *8*(4), 436–455.

Bohnet, I., van Geen, A., & Bazerman, M. H. (2016). When performance trumps gender bias: Joint versus separate evaluation. *Management Science*, *62*(5), 1225–1234. DOI: 10.1287/mnsc.2015.2186

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, *66*(2), 204–225. DOI: 10.2307/2657415

Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, *116*(1), 55–79.

Clark, J. B. (1899). *The distribution of wealth: A theory of wages, interest and profits*. New York: Macmillan.

Cortes, P., & Pan, J. (2017, March). *Occupation and Gender* (IZA Discussion Papers No. 10672). Institute of Labor Economics (IZA). Retrieved from https://ideas.repec.org/p/iza/izadps/dp10672.html

Fortin, N. M. (2008). The gender wage gap among young adults in the united states: The importance of money versus people. *Journal of Human Resources*, *43*(4), 884–918.

Galos, D. R., & Coppock, A. (2023). Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiments. *Science Advances*, *9*(18), eade7979. Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.ade7979 DOI: 10.1126/sciadv.ade7979

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119. DOI: 10.1257/aer.104.4.1091

Greenstein, T. N. (2000). Economic dependence, gender, and the division of labor in the home: A replication and extension. *Journal of Marriage and Family*, *62*(2), 322–335. DOI: 10.1111/j.1741-3737.2000.00322.x

Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 97–118). Cambridge, UK: Cambridge University Press.

Herts, J., & Levine, S. C. (2020, 03). *Gender and math development.* Oxford University Press. Retrieved from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-1186 DOI: 10.1093/acrefore/9780190264093.013.1186

Holt, C. A., & Laury, S. K. (2014). Assessment and estimation of risk preferences. *Handbook of the economics of risk and uncertainty*, *1*, 135–201.

Karan, C., & Mercy, F. (2021). A critical review of literature on employment selection tests. *Journal of Human Resource and Sustainability Studies*, *9*, 451–469. DOI: 10.4236/jhrss.2021.93029

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209. DOI: 10.1257/app.20180010

Krekó, J., Munkácsy, B., Csillag, M., & Scharle, (2023). A job trial subsidy for youth: Cheap labour or a screening device? *Social Policy & Administration*, *57*(7), 1089-1116. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/spol.12924 DOI: https://doi.org/10.1111/spol.12924

Larribeau, S., Masclet, D., & Peterle, E. (2013). Une mesure expérimentale de la discrimination homme-femme à l'embauche. *Revue d'économie politique*, *123*(3), 333–351.

Moss, P., & Tilly, C. (2001). *Stories employers tell: Race, skill, and hiring in america.* New York: Russell Sage Foundation.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709.

Pallais, A. (2014). Inefficient hiring in entry-level labor markets. *American Economic Review*, *104*(11), 3565–3599.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, *62*(4), 659–661.

Pinkston, J. C. (2003). Screening discrimination and the determinants of wages. *Labour Economics*, *10*(6), 643–658. DOI: 10.1016/S0927-5371(03)00043-6

Sayer, L. C. (2016). Trends in women's and men's time use, 1965–2012: Back to the future? In S. McHale, V. King, J. V. Hook, & A. Booth (Eds.), *Gender and couple relationships* (pp. 43–77). Berlin, Germany: Springer.

Smith, R. A. (2002). Race, gender, and authority in the workplace: Theory and research. *Review of Sociology*, *28*, 509-542. Retrieved from https://api.semanticscholar.org/CorpusID:28737141

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374. DOI: 10.2307/1882010

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American economic review*, *92*(3), 434–459.

Sullivan, O. (2006). *Changing gender relations, changing families: Tracing the pace of change over time.* Lanham, MD: Rowman & Littlefield.

Tichenor, V. J. (2005). *Earning more and getting less: Why successful wives can't buy equality.* New Brunswick, NJ: Rutgers University Press.

Tilcsik, A. (2021). Statistical discrimination and the rationalization of stereotypes. *American Sociological Review*, *86*(1), 93-122. Retrieved from https://doi.org/10.1177/0003122420969399 DOI: 10.1177/0003122420969399

Verashchagina, A., & Bettio, F. (2009). *Gender segregation in the labour market – root causes, implications and policy responses in the eu.* Publications Office. DOI: doi/10.2767/1063

Wang, H., Li, L., & Zhang, P. (2023, August). Gender differences in mental rotational training based on computer adaptive tests. *Behavioral Sciences (Basel)*, *13*(9), 719. DOI: 10.3390/bs13090719

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., & Zhou, X. (2012, March). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychological Science*, *23*(3), 320–330. DOI: 10.1177/0956797611427168

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, *8*(2), 187–203. Retrieved from http://iej.com.au

# Appendix A - Description of tasks

In the pre-study, we collected responses from participants across six standardized quizzes. For the main experiment, we selected two quizzes from the initial six based on the pronounced differences in performance between female and male participants: emotion recognition (favoring females) and equation solving (favoring males).

**Emotion recognition** tasks assess participants' ability to accurately identify emotions from facial expressions. Participants are presented with photographs showing individuals expressing various emotions. They are required to identify the emotions displayed from a predefined list (e.g., happiness, sadness, anger, surprise, fear, or disgust). We used a generative AI tool, specifically ChatGPT-4, to produce images for this task. Various meta-analyses and reviews on gender differences in emotion recognition have shown a small to moderate female advantage (Hall et al., 2000).

The **rhyming** task is used to investigate cognitive abilities and language skills and can serve as a proxy for verbal fluency, mental flexibility or cultural knowledge. Participants are presented with a pair of words and are required to judge whether the words rhyme within a limited time frame. We used ChatGPT-4 to produce questions for this task. Wei et al. (2012) show that girls outperform boys in word-rhyming tasks.

The **general knowledge** quiz is a tool used to assess participants' knowledge across a range of topics to study their cognitive ability. Participants are presented with a set of multiple-choice questions (randomize question selection) covering various topics, such as history, geography, science, literature, or current events. We used ChatGPT-4 to produce questions for this task. We explicitly designed the inquiry to focus on general knowledge questions for Generation Z participants. Gender differences on knowledge tests favoring men are among the most stable gender differences found in cognitive ability measures (Ackerman et al., 2001).

The **mental rotation** task is a cognitive experiment widely used to measure spatial visualization ability. Participants are presented with similar shapes and the task requires to determine whether the two objects are identical (one is a rotated version of the other), mirror images or completely different. Wang et al. (2023) show that males perform better than females on mental rotation tasks.

The **multiplication** and **algebra** (solve equation) tasks are commonly used cognitive exercises in experimental economics. These tasks are designed to assess computational skills. In the first task, participants solve basic multiplication problems, while in the second task, they solve simple algebraic equations involving a single variable. Performance is evaluated based on accuracy and speed. Zhu (2007) claims that a substantial body of literature indicates the existence of gender differences in mathematical problem-solving, with males often demonstrating an advantage. Additionally, there are differences in math attitudes and stereotypes such that teachers and parents believe that males are better at math than females (Herts & Levine, 2020).

# Appendix B - Experiment design

**Welcome to the experiment!**

Thank you for participating in our study. The experiment consists of three parts. The total duration will not exceed 30 minutes.

In the first part, you will be asked to complete two quizzes—one focused on mathematics and the other on emotion recognition. The results of these quizzes will be treated as your performance outcomes for the second part.

In the second part, you will take part in an experimental labor market. You will take on the role of a manager who must decide how to allocate a budget for employee wages.

In the final part, we will ask you to complete a questionnaire about yourself, your characteristics, and your beliefs (demographic survey).

For participating in the experiment, you will receive compensation: 10 PLN for entering the study, and up to an additional 40 PLN based on your performance in the second part. Your total earnings will depend on your results and the decisions made during the experiment.

**Before we begin**

What is your gender: Female or Male?

**Part I. Solve the Quizzes**

Quiz 1. In a moment, 10 questions will appear on the screen. Each will contain an equation with one unknown. Your task is to solve the equation for the given variable xx. Enter your answer in the field below each question. The correct answer will always be an integer. You will have 10 seconds to answer each question.

Your result: xxx. You answered correctly xxx out of 10 questions. This result will be part of your compensation in the experiment.

Quiz 2. In a moment, an image showing a person's face will appear on the screen. Your task is to determine which emotion best describes the person's expression in the image. Choose your answer from the four proposed emotion options. The images were generated by ChatGPT-4o. You will have 5 seconds to answer each question.

Your result: xxx. You answered correctly xxx out of 10 questions. This result will be part of your compensation in the experiment.

**Part II Experimental Labour Market**

This part of the study simulates an experimental labor market. You will soon be randomly assigned to at least three teams. In one of them, you will take on the role of a manager; in the others, you will be a worker.

As a worker, you don't have to do anything - your "work" is represented by the result of one of the quizzes you completed in Part I. In the team where you act as the manager, your compensation will depend on your decision regarding how to allocate wages between your two employees - a woman and a man - based on their quiz results from Part I.

Payments in the experiment will be based on randomly selected teams and a single manager's decision. This means your final earnings may come either from the team where you acted as the manager (and made a decision) or from a team where you were a worker.

Employee compensation is calculated using the following formula:

Wage of the female employee: $\alpha(x_F + x_M)$

Wage of the male employee: $(1-\alpha)(x_F + x_M)$

where: $x_F$ is the female employee's test score, multiplied by 2 (each point is worth 2 PLN), $x_M$ is the male employee's test score, calculated analogously, and $\alpha$ is a value between 0 and 1, determined by the manager.

Manager's wage: $\alpha x_F + (1-\alpha)x_M$.

Note: Choosing extreme values $\alpha$ means that one of your employees will receive no pay!

**Control Question**

Given that:

Wage of the female employee: $\alpha(x_F + x_M)$

Wage of the male employee: $(1 - \alpha)(x_F + x_M)$

where: $x_F$ is the female employee's test score, multiplied by 2 (each point is worth 2 PLN), $x_M$ is the male employee's test score, calculated and $\alpha$ is a value between 0 and 1, determined by the manager.

Manager's wage: $\alpha x_F + (1 - \alpha)x_M$;

and knowing that $x_F = 10$, $x_M = 10$ and $\alpha = 0.5$ calculate:

Wage of the female employee:

Wage of the male employee:

Manager's wage:

**Choose** $\alpha$

You are a manager. Your team's salaries will be based on the results of a quiz: emotion recognition.

Knowing that in one of the previous rounds of the emotion recognition quiz, the median score among female employees accounted for approximately 60% of the combined male-female team score, and the median score among male employees accounted for 40%, what alpha will you choose for yourself and your employees?

OR

You are a manager. Your team's salaries will be based on the results of a quiz: solve equation.

Knowing that in one of the previous rounds of the solve equation quiz, the median score among female employees accounted for approximately 40% of the combined male-female team score, and the median score among male employees accounted for 60%, what alpha will you choose for yourself and your employees?

Enter a number between 0 and 1, using a dot as the decimal separator.

Remember that:

Wage of the female employee: $\alpha(x_F + x_M)$;

Wage of the male employee: $(1 - \alpha)(x_F + x_M)$;

Manager's wage: $\alpha x_F + (1 - \alpha)x_M$.

If you had the opportunity to decide on the price of purchasing information about the exact results of your employees, what is the maximum amount you would be willing to pay for such information?

Enter an integer. Remember that the maximum salary in this experiment is 50 PLN.

There is an option to find out the exact quiz results of your employees. This information may be free or come at a cost (2 or 6 PLN). Half of the managers in this round will receive this information for free, while the other half will pay either the lower or higher price (in equal proportions).

You received the information for free / for 2 PLN / for 6 PLN.

The cost of the information will be deducted from your final salary.

In your team: The female employee scored XXX points, and the male employee scored XXX points.

What alpha do you choose now?

Remember that:

Wage of the female employee: $\alpha(x_F + x_M)$;

Wage of the male employee: $(1 - \alpha)(x_F + x_M)$;

Manager's wage: $\alpha x_F + (1 - \alpha)x_M$.

**Part III**

How old are you?

Below are 10 pairs of lotteries, in which participation is free. For each pair, which option – A or B – would you be more inclined to choose, taking into account the changing probabilities of winning a given amount?

| Lottery A | Lottery B |
|---|---|
| 20 PLN with 10% or 16 PLN with 90% | 38.5 PLN with 10% or 1 PLN with 90% |
| 20 PLN with 20% or 16 PLN with 80% | 38.5 PLN with 20% or 1 PLN with 80% |
| 20 PLN with 30% or 16 PLN with 70% | 38.5 PLN with 30% or 1 PLN with 70% |
| 20 PLN with 40% or 16 PLN with 60% | 38.5 PLN with 40% or 1 PLN with 60% |
| 20 PLN with 50% or 16 PLN with 50% | 38.5 PLN with 50% or 1 PLN with 50% |
| 20 PLN with 60% or 16 PLN with 40% | 38.5 PLN with 60% or 1 PLN with 40% |
| 20 PLN with 70% or 16 PLN with 30% | 38.5 PLN with 70% or 1 PLN with 30% |
| 20 PLN with 80% or 16 PLN with 20% | 38.5 PLN with 80% or 1 PLN with 20% |
| 20 PLN with 90% or 16 PLN with 10% | 38.5 PLN with 90% or 1 PLN with 10% |
| 20 PLN with 100% | 38.5 PLN with 100% |

On a scale from 1 to 10, respond to the following statements:

Income should be more equal – Income should more strongly reward individual effort

On a scale from 1 to 5, where 1 means ”strongly agree” and 5 means ”strongly disagree,” respond to the statement: Large income differences are acceptable in order for people to be properly rewarded for their abilities and effort.

On a scale from 1 to 4, where 1 means ”strongly agree” and 4 means ”strongly disagree,” respond to the following statements:

When a mother works, the children suffer.

Women can work, but what most of them really want is a home and children.

Generally speaking, when a woman works full-time, family life suffers.

Men's role is to earn money; women's role is to take care of the home and family.

Generally speaking, men are better political leaders than women.

Generally speaking, men manage business better than women.

When jobs are scarce, men should have greater right to work than women.

**Thank you**

Today you earned xxx PLN.

Thank you for participating in the experiment!