GOSPODARKA NARODOWA

2(322)2025, 1–28
DOI: 10.33119/GN/202289
gnpje.sgh.waw.pl
SGH

The Polish Journal of Economics
Istnieje od / Published since 1931

Magdalena Smyk
Warsaw School of Economics, Poland
FAME|GRAPE, Poland

Lucas van der Velde
Warsaw School of Economics, Poland
FAME|GRAPE, Poland

Joanna Tyrowicz
FAME|GRAPE, Poland
IZA Institute of Labor Economics, Germany
University of Regensburg, Germany

Paying for Ideal Discretion: A Framed Field Experiment on Working Time Arrangements*

Premia za pełną dyspozycyjność. Eksperyment winietkowy dotyczący organizacji czasu pracy

Abstract

The notion of the ideal worker entails being available at the employer's discretion in terms of time. By contrast, the ability to set one's own schedule is widely considered a cornerstone of work-life balance and job satisfaction. We provide causal evidence on the pecuniary and social valuation of discretion over work schedules. We embed our study in the context of gender and compare employee- and employer-initiated requests for changes towards greater discretion over working hours. We show that employer-initiated availability should be reflected in higher wages, but the premium is small. There appears to be no wage penalty to employee-initiated requests for work schedule autonomy. While our results lend support to the ideal worker model, they cast doubt on explanations linking gender wage inequality to labour market flexibility.

Keywords:
gender inequality, flexibility penalty, ideal worker, vignette experiment

JEL classification codes:
J32, J71, J16

Article history:
submitted: July 9, 2024
revised: January 24, 2025
accepted: February 13, 2025

Streszczenie

Pojęcie idealnego pracownika wiąże się z byciem dostępnym na życzenie pracodawcy bez względu na organizację czasu pracy. Z kolei możliwość samodzielnego ustalania harmonogramu pracy jest powszechnie uznawana za fundament równowagi między życiem zawodowym a prywatnym oraz satysfakcji z pracy. Zbadaliśmy, jaką wartość (materialną i społeczną) przypisuje się możliwości decydowania o własnym czasie wykonywania pracy. Osadziliśmy nasze badanie w kontekście równości płci. Porównaliśmy inicjowane przez pracowników i pracodawców prośby o większą autonomię w zakresie ustalania godzin pracy. Uczestnicy badania wskazali, że

Słowa kluczowe:
nierówności płci, kara za elastyczność, idealny pracownik, eksperyment winietkowy

Kody klasyfikacji JEL:
J32, J71, J16

Historia artykułu:
nadesłany: 9 lipca 2024 r.
poprawiony: 24 stycznia 2025 r.
zaakceptowany: 13 lutego 2025 r.

* The authors are grateful to Katrin Auspurg, Adrian Chadi, Jose Garcia-Louzao, Thomas Hinz, Julia McQuillan, Jekaterina Navicke, David Ritter and Linas Tarasonis for their valuable insights. Participants of the ICEA 2021, AIEL 2022 and LabFam 2022 conferences gave excellent comments. Erita Narhetali provided wonderful research assistance. This project was supported by the joint NCN-LMT DAINA initiative (grant #2017/27/L/HS4/03219). Lucas van der Velde acknowledges the support in National Science Center grant no. \#2019/33/B/HS4/01645 for work in revision. The experiment was pre-registered in the AEA RCT registry (the unique # AEARCTR-0007642). Ethical approval for this study was obtained from the FAME|GRAPE Ethics Committee on April 27, 2021 (decision #1/2021).
All files required to replicate the results of this research are available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YHCI85.

dyspozycyjność inicjowana przez pracodawcę powinna być rekompensowana wyższym wynagrodzeniem, jednak premia ta jest niewielka. Nie zaobserwowaliśmy jednocześnie istnienia kary płacowej za prośby pracowników dotyczące autonomii w ustalaniu czasu pracy. Nasze wyniki wspierają teoretyczny model idealnego pracownika, ale podważają wyjaśnienia, które łączą nierówność płac ze względu na płeć z elastycznością na rynku pracy pod względem dostępności.

Introduction

We study the role of autonomy to decide about the working schedule in determining wages. Deciding about wages and working conditions is a key element of employer-employee relations. We zoom in on decisions concerning working time arrangements, as departures from the standard 9–5 arrangement can be credibly requested by both employers and employees. By manipulating who requests power over working time arrangements, we can identify the value of discretion over work schedule arrangements.

The importance of working time arrangements have been brought to the forefront of the literature by, among others, **Goldin** [**2014**]. Her research concludes that the remaining inequalities between men and women in the labour market are rooted in differences in the ability to work flexible hours. Goldin argues that workers receive a premium for being available to their employers. Primary care givers, usually women, cannot reap these benefits to the same extent given their time constraints [**Cortes, Pan, 2019**].

However, the premise that being available to the employer results in a wage premium is questionable. Literature suggests that wage decisions are based on comparison to the *ideal worker*, who ought to be available at employers' discretion. (e.g. **Breaugh** [**1985**]; **Grote, Raeder** [**2009**]). The business case for the *ideal worker* stems from the efficiency of the work process [**Reilly, 1998**; **Kauffeld, Jonas, Frey, 2004**]: employees who can adapt their working hours to the demands of the workplace provide firms with a competitive advantage in the market. Importantly, the *ideal worker* model promotes an asymmetric approach to flexibility. Flexibility requests by employees are penalised, whereas being asked to demonstrate flexibility is not rewarded with a wage increase since workers are expected to adapt to the changing needs of the business.

Both the *ideal worker* model and the Goldin conjecture agree that employees who are available to work flexibly should be rewarded more than those who require flexibility for themselves. However, the two theories differ on whether this difference reflects a penalty for lacking flexibility or rather a premium for being available. The *ideal worker* model posits wage penalties for employees who are not available at the employer's discretion, but offers no explicit predictions for those who conform to this norm. Meanwhile, the Goldin conjecture states that workers earn a premium for being available, without addressing whether those requesting flexibility face penalties.

Theories also diverge on the role of gender as a moderating factor. In **Goldin** [**2014**], the wage effects of greater flexibility are identical for men and women. In contrast, the *ideal worker* model competes with the *family devotion* schema. For women, both work and caring demand (and deserve) single-minded focus and allegiance. It is theoretically unclear whether the availability of alternative schema hampers or reinforces the penalties for departing from the *ideal worker* model.

We construct an experiment to bridge the gap between these two theoretical approaches and to test their empirical relevance. Our experiment aims to empirically verify the extent to which discretion over working time should be reflected in wages. We operationalise discretion over work schedules as the ability to set starting and ending times. In the status quo scenario, workers have fixed working time arrangements. Across scenarios, discretion to set starting and ending times is assigned either to the employer or to the employee. We contrast employer-initiated and employee-initiated flexible working time arrangements (TAs).[1] Our experi-

[1] This paper focuses on working time arrangements. We consider both perspectives on flexibility: the employer's and employee's. We thus rely on the abbreviation TA. Note that flexible time arrangements may involve a variety of dimensions, including working from

ment also manipulates a second dimension: the employee's gender. This allows us to identify whether gender moderates the wage premiums and penalties associated with TAs.

Our paper is close to two studies attempting to identify the actual "value" of the TAs in an experimental context: **Mas and Pallais [2017]** and **He, Neumark, and Weng [2021]**, both of which provide field experiments on job offers and wage negotiations. Compared to these two studies, we introduce several innovations. First, our study focuses on the actual monetary value associated with working flexible hours, rather than on whether people apply for such jobs. Second, our participants decide on wage changes for *incumbents*, which may elicit stronger preferences for the status quo compared to *job entrants* analysed in **Mas and Pallais [2017]** and **He, Neumark, and Weng [2021]**. Finally, while **Mas and Pallais [2017]** reveal individual preferences for flexible work, our experiment captures social perceptions. Vignette experiments lack real-world consequences, which limits their use to learn actual behaviors. However, these experiments help to understand what participants consider appropriate in hypothetical situations. To highlight the social dimension, we included a question asking participants about their perceptions of the prevailing social norm.

Our experiment provides moderate support for Goldin's conjecture [2014]: being available to meet employer demands is associated with a wage premium. However, there are two caveats. First, the magnitude of the premium for flexible work is small – less than 3 percent of wages— an order of magnitude lower than estimates of the gender wage gap in the country. Second, while the average premium for providing flexible work was positive, most respondents suggested that wages should remain unchanged. This result provides some support for the *ideal worker* model. When employees request flexible work, the average change in wages is much smaller, as most respondents indicated no change in wages. Only two percent of respondents recommended a wage penalty for employees requesting flexible TAs.

Our paper is structured as follows. In section 2, we present the relevant theory and our working hypotheses. In section 3, we discuss the details of the experimental design. In section 4, we describe the sample. We present the results in section 5, along with an extensive discussion. The paper concludes with a discussion of external validity and the policy implications of our study.

Theoretical foundations

For more than half a century, the notion of an *ideal worker* has involved commitment, a stable career trajectory, and permanent and unconstrained availability to the employer [**Davies, Frink, 2014**]. This ideal serves as a benchmark for wage determination: workers who more closely align with this model typically receive higher compensation, while those who deviate face wage penalties. While the *ideal worker* is expected to demonstrate creativity in task execution, they lack autonomy in determining their work schedule [**Breaugh, 1985**; **Gagne, Deci, 2005**; **Mazmanian, Orlikowski, Yates, 2013**; **Grote, Raeder, 2009**]. Idiosyncratic working time arrangements (TAs) are treated as negotiated perks available for top performers rather than a universal right [**Kossek, Barber, Winters, 1999**; **Kelly, Kalev, 2006**]. Under the *ideal worker* model, access to flexible TAs is restricted, and thus unequal, even if providing voluntary and self-controlled flexibility may improve workers' well-being and reduce turnover [**Kelly, Moen, Tranby, 2011**; **Moen, Kelly, Hill, 2011**; **Kaduk et al., 2019**].

The instrumental perspective on workers embedded in the *ideal worker* model neglects individualism and intrinsic motivations as drivers of productivity. The employees' perspective builds on the work-life balance and on the ability to fulfil various roles in life that are enabled by greater work flexibility (see, for example, **Barnett, Hyde [2001]**; **Byron [2005]**; **Michel et al. [2011]**, for the theoretical and empirical treatment of the expansionist theory, derived from the social roles theory). Worker autonomy is believed to foster workers' intrinsic motivation and thus their commitment and productivity [**Spector, 1986**; **Heavey, Holwerda, Hauskknecht, 2013**; **Rubenstein et al., 2018**]. Flexible working TAs are often justified as tools to enhance workers'

home or other forms of telework. In addition, arrangements flexible for the worker may be inflexible for the employer and vice versa (see **Soga et al. [2022]**, for a systematic review on the complexity of flexible work arrangements).

work-life balance, while also being beneficial to the employer [**Fagan et al., 2012**]. Empirical evidence suggests that workplaces providing flexible working arrangements outpace the competition (see **Dex, Scheibl** [**2001**]; **Batt, Valcour** [**2003**]; **Beauregard, Henry** [**2009**]; **De Menezes, Kelliher** [**2011**], to name just a few), though the causal link has not been established [**Azar, Khan, Van Eerde, 2018**]. Controlled experiments demonstrate – albeit in highly specific contexts – that flexible start and finish times improve productivity [**Boltz et al., 2020**; **Angelici, Profeta, 2020**]. Two reasons might explain the higher productivity of flexible workers: work intensification [**Kelliher, Anderson, 2009**; **Cañibano, 2018**] and the unlocking of productive potential [**Chung, Horst, 2017**]. Work intensification reflects a gift exchange process: workers reciprocate the acquired flexibility by exerting more effort during working hours. Unlocking the productive potential concerns workers who could not exert effort in standard working arrangements (see **Chung, Horst** [**2017**], for an example of work intensification among working mothers).

Although many empirical studies demonstrate that flexible TAs may be neutral or even beneficial to performance (see **Azar, Khan, Van Eerde** [**2018**], for an extensive meta-analysis), the "ideal worker" model remains a powerful benchmark. Managers and colleagues perceive workers who request discretion over TAs as less committed [**Williams, 2001**; **Chung, 2020**]. Co-workers report reduced job satisfaction when their colleagues benefit from flexible TAs, arguing that their workload was raised without adequate compensation [**Munsch, 2016**]. This effect is not restricted to male co-workers; female co-workers display similar patterns [**Teasdale, 2013**].

This tension creates a flexibility paradox: while flexible TAs appear to entail little economic cost, they carry significant reputational costs for employees. These costs arise from the conflict between the *ideal worker* model and an individualised approach to work. In this research, we ask how flexible TAs requested by the employer would be rewarded and which perspective dominates when the employee requests flexibility. Based on the literature, we propose three research questions:

Hypothesis 1. *An employee requesting more discretion over their work schedule is penalised with a reduction in wages.*

This hypothesis stems from the *ideal worker* model: the further the worker departs from this model, then the lower the wage relative to what an ideal worker would have earned.

Hypothesis 2. *An employer requesting more discretion over the work schedule is expected to raise wages.*

This hypothesis is a clear representation of Goldin's conjecture. Leaving schedules at the discretion of the employer in terms of start and end times raises the disutility of work and thus should be compensated with higher wages (see also **Smith** [**1979**]).

Hypothesis 3. *Women face a smaller wage decline than men when requesting discretion over their work schedules.*

This hypothesis embeds a gender dimension in the *ideal worker* model. When women request discretion to set their schedule, this may be viewed as consistent with the *family devotion* schema. In such a case, employers may acknowledge the employees' need for more discretion and choose not to penalise the worker [**Walby, Olsen, 2002**; **Cousins, Tang, 2004**; **Warren, Rowlingson, Whyley, 2001**]. Conversely, when men request discretion over their schedule, the departure from the *ideal worker* cannot be easily justified; thus penalties are expected to emerge. Recently, **Luhr** [**2020**] emphasized that while both genders may seek greater control over their work schedules, only women are still stereotypically viewed as the primary caregivers

Experimental design

We conducted our experiment in Poland; the context is described in Appendix 8. Participants were compensated: a flat rate of USD 0.50 for participation and up to USD 3 depending on choices during the experiment (described in detail below).

Our tool consisted of four modules. The first module contains our vignette experiment. The second module is a questionnaire aimed at learning respondents' views on work-life conflict, gender norms and inequality. The third module is a real-consequence task that serves to estimate respondent's value of time. The four and last module is a questionnaire on personal characteristics.

The experiment was administered online in Poland using ANSWEO convenience sample (a panel where participants pre-register for experiments and surveys). Participation was voluntary. A pre-test was carried out in late April[2] while the final experiment was implemented on August 9. To mitigate the risk that some participants of the convenience samples were less careful in filling out the questionnaires [**Cheung et al., 2017**; **Sharpe Wessling, Huber, Netzer, 2017**], we included a series of manipulation checks [**Porter et al., 2019**]. Each individual responded to nine manipulation check questions: three for each vignette.

The median time to complete the survey was 8 minutes. Given that the experiment was at the start of the survey, most respondents answered all experimental questions within the first few minutes.

WTA experiment

We propose a mixed design, combining a 2×2 framed field experiment with a vignette.[3] The framed field experiment randomly assigns subjects to treatment conditions, where they evaluate three vignettes. By varying treatment across vignettes, the design provides both within- and between-subject variation to exploit.

At the beginning of each vignette, subjects learn the story of a worker who initially works in a regular, fixed time schedule, five days a week. Subjects are informed that there will be a change in this arrangement. We state that the *average* number of weekly hours remains unchanged and the start and end times for each day will be communicated with some notice. The new, changed WTAs to be evaluated by the participants in the experiment are characterised by "flexible start and end hours with cumulative average weekly working time unchanged (40 hours)." By keeping the average number of hours explicitly constant, we secure that subjects will not confound discretion over WTAs with part-time or overtime. This form of WTA (flexible start and end times) could credibly be requested by both the employee and the employer. Such WTA encompasses two possible arrangements: (i) a constant number of hours per day, but with varying start and end times; (ii) a constant number of hours per week, but a varying number of hours on each weekday.

Treatments

The two treatments include the initiator of the change in WTAs (worker or employer) and the gender of the worker (a man or a woman). In the GENDER treatment, the subjects are faced with a man or a woman as a worker in the vignette. In the INITIATOR treatment, either the employer or the employee want to change the WTA. Under the new time arrangement, the start and end times will vary.

Randomization

We assign subjects across conditions randomly. With two initiators, two genders and three vignettes, the pool of potential combinations, including variation in ordering the vignettes, is 64 ($2^3 \times 2^3$), which would not be feasible. We constrain potential combinations to exploit within-subject variation in each treatment. If, in the first vignette, the respondent was assigned to employer-initiated changes in WTAs, the second vignette refers to employee-initiated changes in WTAs. Similarly, for the GENDER condition, if, in the first vignette, the employee is a man, then the second vignette portrays a woman. For the third vignette, the algorithm

[2] Before rolling out the survey on the full sample, we tested the technical features of our survey as well as the legibility of all the questions using a pre-test in a sample of 40 subjects (20 men and 20 women). The pre-test yielded important insights on formulating specific questions and the overall features of our survey. Given that the changes were substantial in some cases, the 40 subjects from the pre-test are not included in the analyses.

[3] Following the terminology of **Harrison and List [2004]** as well as **Levitt and List [2009]**, this is a between-subject framing design in the field with a within-subject vignette survey design.

randomised among those GENDER × INITIATOR conditions that did not appear in the first two vignettes (it picks one of the two remaining treatment combinations). This procedure limits the number of possible combinations to 16 and ensures that for each participant of our experiment we observe both GENDERs, both INITIATORs, and no repeated combinations.

Outcome measures

After reading the vignette and the information on the proposed changes in WTAs, subjects were asked whether they believe that the wage should change as well. We offered three categories: "increase," "remain the same," and "decrease." Once participants selected an option, a question about the magnitude of wage adjustment appeared for increases and decreases. The recommended wage adjustment was measured in a quasi-continuous way: the participants chose their preferred amount from the list, with PLN 50 per month as an interval (approx. USD 13).[4] Overall, respondents could choose from 120 categories, of which 60 indicated wage growth and 60 a decline.[5]

In addition to the wage adjustments, we asked participants to disclose their beliefs about the social norm. Specifically, we asked if they believed that their evaluation is shared by the majority of Poles. For positive answers, this was the end of this module in the experiment. For individuals who reported that their evaluation was not shared by the majority of Poles, we additionally asked what they thought the majority of Poles would prefer (increase, decrease or no change).

Scenarios

The goal of the study is to reveal social norms rather than elicit individual preferences towards discretion in setting working times. To this end, we follow **Aguinis and Bradley [2014]** and construct third-person vignettes. This approach allowed us to capture variation in characteristics that are crucial from the perspective of social norms: gender and working time arrangements. This design is particularly suitable, when asking participants hypothetical questions about themselves (first-person vignettes, see **Hainmueller, Hangartner, Yamamoto [2015]**, who shows high external validity of such designs).

The three vignettes faced by each subject differ by context in terms of occupation: a hairdresser, a lawyer and a retail salesperson. In the first story, the character was a hairdresser working regularly from Monday to Friday in a 9–5 schedule. In the second story, the character was a lawyer in a large law firm with the same working hours. In the third story, the character was a retail salesperson working Monday to Friday on a part-time basis (from 9 a.m. to 1 p.m.). These occupations were chosen because all of them are likely to employ workers of both genders, but require a different skill level and have a different social standing. Each occupation has its own base wage in the status quo (identical for both genders). The base wages were set in line with the market averages at the time of the experiment: PLN 1,600 per month (approx. USD 420) for the shop assistant, PLN 3,200 per month (approx. USD 840) for the hairdresser, and PLN 6,400 per month (approx. USD 1,680) for the lawyer. For the hairdresser, it is conventional to assume that this service should be provided to the customers outside their working hours (after or before their work). Workers in this occupation, despite frequently being women, are expected to work early and late hours, with much less traffic within the 9-to-5 schedule. For the lawyer, the regularity of WTA is, on the one hand, strengthened by the 9-to-5 schedule of courts and the public administration, but, on the other, it may be weakened by the need to meet with clients at their convenience or to work long hours in order to prepare the case in a short period of time. For the shop assistants, the strict 9-to-1 schedule is a complement of WTAs of another worker, because stores

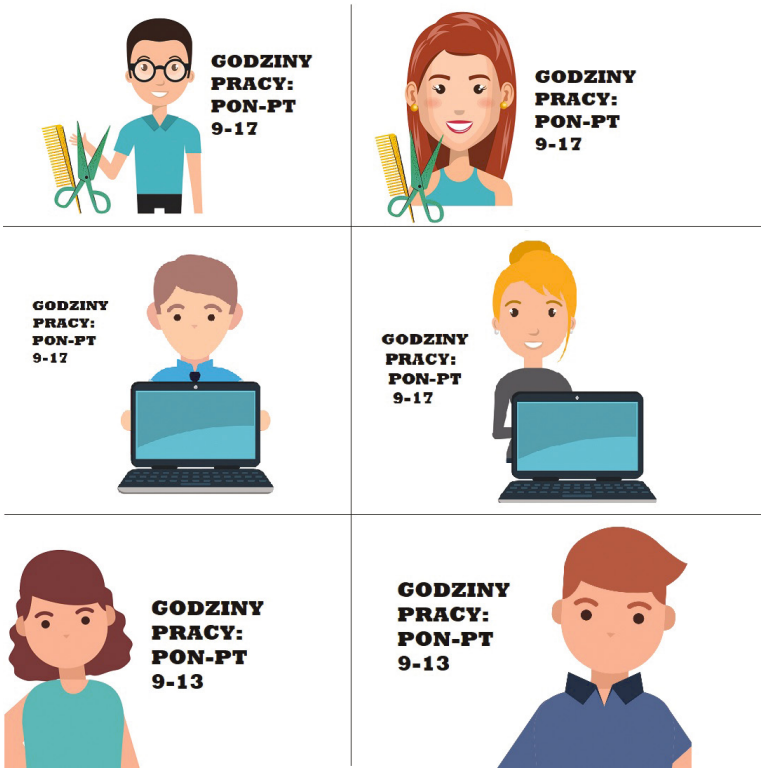
[4] All conversions to US dollars were made using the exchange rate listed by the National Bank of Poland on the first day of the experiment (April 23, 2021), USD 1 = PLN 3.7855.

[5] From an *ideal worker* perspective, allowing respondents to provide positive answers following an employee-initiated renegotiation would not make much sense. However, leveraging the flexibility paradox, one could expect no changes or even positive changes, depending on the nature and extent of work intensification.

are typically open longer than four hours on weekdays. It is thus customary to expect rotating workers across shifts to fully schedule the operating hours.

The relevant information is communicated graphically and in text. The vignettes use cartoons to demonstrate the gender and the occupation of the worker; see Figure 1.[6] The GENDER treatment is reinforced in text, as characters have distinct male and female names. We incentivised subjects to memorise the content of each vignette by offering them additional compensation for answering nine manipulation checks, three after each vignette. Participants who answered correctly all nine questions received an additional dollar. The subjects were informed about the outcomes of the manipulation checks after the end of the experiment. The manipulation questions were related to the relevant features of work arrangements: the working hours and days in the status quo and the party who initiated the change in discretion over WTAs.

Figure 1. Scenarios



Notes: The vignette stories were visualised with the pictures presented in this figure. In each picture, the inscription says: "Working hours: MONDAY-FRIDAY...". Each picture complemented a story that introduced these characters Adam and Anna (hairdressers), Marek and Maria (lawyers), and Karol and Karolina (salespeople).

Source: Authors' own elaboration.

Social norms

Upon completing the experiment, participants answered a short questionnaire where they indicated to what extent they agreed with different statements. These statements were taken from different sources and emphasised three dimensions that could potentially confound the experiment results: "work-family" and "family-work" conflict, adherence to traditional gender norms, and views on income inequality.

The work-family and family-work conflict items were taken from **Netemeyer, Boles, and McMurrian [1996]**. Given the high values of Cronbach's α, we group the 10 items into two variables, where higher values indicate more conflict. The items measuring adherence to traditional gender norms were taken from the European Value Survey. As before, Cronbach's α suggested grouping these items into a single variable, where

[6] The text next to each cartoon indicates the working days and hours in the status quo.

higher values show more adherence. Finally, we captured views on income inequality using an item from the International Social Survey Programme, recoding responses so that the resulting variable equals 1 if respondents agree that income differences are necessary to reward effort.

Valuing personal time autonomy

Participants could easily confuse social norm with individual preferences. To adjust our estimates of social norms for individual preferences, we included a task in which participants revealed the value they placed on personal time autonomy. After completing the vignette questions, participants engaged in a discrete choice experiment with real consequences. Subjects were informed that in order to complete the survey they will have to wait either five or 30 minutes and that this waiting time will be randomly chosen. Participants were told that they may choose between various options of the length of the window to complete the survey *after* the waiting time – each associated with a different rate. In other words, the participants were free to leave the computer and come back to complete the survey at any convenient time, according to their reported preference. The longer the interval (more discretion to choose convenient time), the lower the pay.

For both waiting times, participants were presented with the same set of compensation options: they could choose to complete the survey within 24 hours after the wait time for an additional USD 0.25 (full discretion), within 25 minutes for USD 0.50 (less discretion), or within a five-minute window for USD 1.50 (least discretion). Since each participant made two choices – following a 5-minute and a 30-minute wait – we were able to estimate individual valuations of personal time autonomy. Subjects were informed that failure to comply with their chosen option would result in forfeiting the bonus earned in previous rounds, reverting their compensation to a base amount of USD 0.50. They were not told whether the final portion of the experiment would be long or short. After they selected their preferred option for each waiting time, waiting times were randomly assigned.

Final survey items

After the waiting time elapsed, a new screen was shown to participants. The screen contained six additional questions that subjects had to answer to complete the survey. As described above, the participants were informed that if they fail to complete the last module, they will not receive the additional compensation gained in previous modules. All participants complied.

The last module included two sets of questions. The first set referred to individual and household characteristics: age, gender, education level, managerial experience, and household income.[7] The second set sought to identify the subjective importance of work against other domains. The participants were asked to order a list of 16 values according to their importance. The set of values was adopted from the cyclical study "Modern Polish Family" by **Bożewicz et al. [2019]**. The questions and the available alternatives are listed in Appendix 7.

The sample

In total, we observed judgment on 963 vignettes from 321 participants. Roughly 60% of participants answered correctly all nine manipulation checks totalling 570 vignettes from 190 respondents. This subsample is our preferred sample. The descriptive statistics for the preferred and the full sample are shown in Table 1. In the preferred sample, the average age is 38.8 years. Around 55% of the sample were subjects with a tertiary education, and a similar percentage claimed they had no managerial experience. Table 1 provides descriptive statistics for all treatment conditions jointly and across treatment assignments. Given that individual characteristics appear similar across the four treatments, we conclude that the randomisation was successful, despite the unusual two-step assignment into conditions.

[7] Given the sensitivity of questions about income, we adopted a measure commonly used in household budget surveys; we asked respondents to rate whether their household income is sufficient to make ends meet on a four-point scale (ranging from "insufficient" to "allows for some luxuries").

Table 1. Descriptive statistics of respondents: full sample and preferred sample

	Full sample	Preferred sample				
	Total	Total	Initiator		Gender of employee	
			Employee	Employer	Man	Woman
Age	38.51	38.76	38.87	38.64	38.88	38.65
% of women participants	0.50	0.52	0.52	0.51	0.53	0.51
Managerial experience	0.54	0.54	0.53	0.56	0.54	0.54
Education						
% with primary education	0.10	0.06	0.05	0.07	0.06	0.07
% with secondary education	0.41	0.38	0.38	0.38	0.39	0.37
% with tertiary education	0.49	0.55	0.56	0.54	0.55	0.56
Income level						
% can afford some luxury	0.42	0.42	0.43	0.40	0.41	0.42
% can make ends meet	0.46	0.46	0.44	0.48	0.46	0.45
% cannot afford living	0.12	0.13	0.13	0.12	0.12	0.13
Passed all manipulation checks	0.59	1.00	1.00	1.00	1.00	1.00

Notes: Table reports the socio-economic characteristics of individuals participating in the experiment. Note that participants reported these characteristics *after* completing the experiment.

Source: Authors' own elaboration.

Table 1 allows for comparing the full sample of all subjects to the preferred sample of subjects who completed all manipulation checks. This comparison reveals few differences between the preferred subsample and full sample. The most noticeable difference corresponds to education levels: in the full sample, the share with tertiary studies is lower than in the preferred sample. To verify the extent to which failure at manipulation checks may affect our results, we estimate a series of logit models, where the dependent variables indicate the probability of making a mistake. In the interest of brevity, the marginal effects are reported in Table 11 in Appendix. We found no evidence that failure at manipulation checks is systematic.

The descriptive statistics are in line with a representative sample. Around half of respondents are women. The education distribution is consistent with a tertiary enrolment of roughly 55% in the past 20 years in Poland, accompanied by a declining share of high school dropouts. Questions about subjective income, stylised after standard items in household budget surveys around the world, report shares similar to the Polish population. While our sample does not come from a random sampling of the entire population, mimicking these basic structural characteristics is a desirable feature.

In the remainder of this study, we report estimates obtained from the subsample of individuals who replied correctly to all manipulation check questions. As shown in Table 1, the subsample includes around 60% of all respondents. In this sample, we are more confident that participants understood the features of each vignette, and the trade-off they were judging. Estimates obtained from the full sample are left to the appendix.

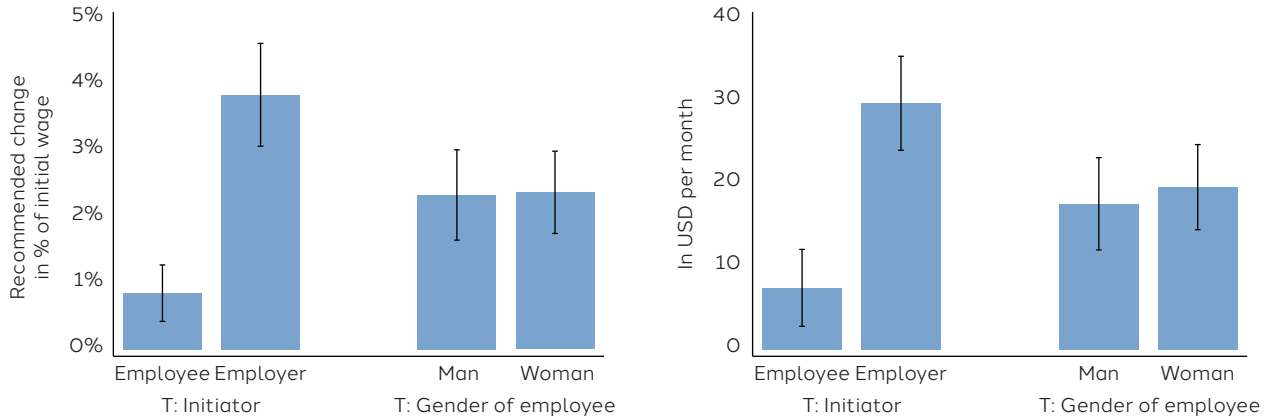
Results

We discuss the results in four substantive parts. First, we present a comparison of mean recommended wage changes across treatment conditions. Given randomisation, these differences are an unbiased measure of the average treatment effects. Second, we deploy a regression model that addresses differences in characteristics across groups (as small as they are) to provide more precise measures. Third, we study heterogeneity of treatment effects by estimating the regressions in different subsamples. Fourth, we study the perceived compliance with the social norm, where participants indicate what answers they believe is the most common among Polish respondents, i.e. their answers indicate what is socially acceptable, We conclude this section with a discussion of our results vis-à-vis the literature.

Average treatment effects

Roughly 26% of subjects recommend increasing wages to accommodate the proposed shift from fixed to discretionary WTAs, while 4% suggested a wage reduction. Figure 2 shows the recommended wage changes, expressed as a fraction of the *status quo* wage. When the employer initiates the change in WTAs, the average recommended wage increase is 3.5% – significantly higher than the roughly 0.8% increase recommended when the change is initiated by the employee. The 3.5% rise corresponds to roughly USD 28 per month, a relatively minor amount in the context of wage bargaining. In contrast, when employees propose the change, the average recommended increase is just USD 6 and is not statistically different from zero. No significant differences were found across gender treatments. On average, both male and female characters received a 2% recommended wage increase, based on an average across initiator conditions.

Figure 2. Treatment effects: what is the recommended wage change?



Notes: Bars indicate unconditional average treatment effects. Whiskers show 95% confidence intervals. Proportions obtained from the preferred sample, which includes 570 observations from the 190 subjects who answered correctly all manipulation check questions. We report detailed test results in Table 4 in the Appendices.

Source: Authors' own elaboration.

Figure 3 provides a breakdown of the proportions recommending wage cuts, wage increases, or no change. Keeping wages unchanged is the most common recommendation across all scenarios. This preference is especially strong when the employee initiates the change – over 80% of respondents chose to keep wages unchanged in that case, compared to 60% when the employer initiates.

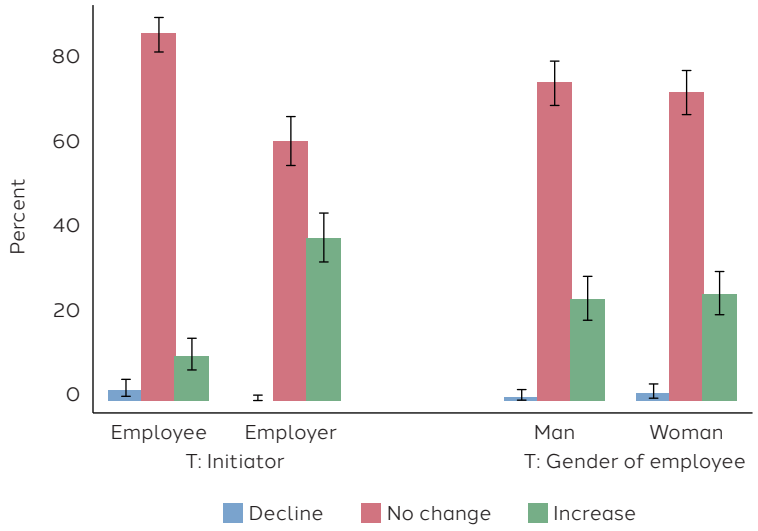
The findings presented in Figures 2 and 3 lend support to both the *ideal worker* model and Goldin's conjecture. Wage declines are only perceived as possible outcomes when the employee requests additional control over the WTA. The frequency of this answer is low but statistically significant. Nobody in our preferred sample recommended a wage decrease when the employer requests discretion over WTA.

Recall that under the *ideal worker* model, employees are expected to be generally available for work. For the subset of participants who recommended no wage change, this model appears to have limited predictive power. The key finding lies in the shift in wage recommendations: only 16% of responses supported a wage increase when the employee initiated the request for greater control over working time, compared to 36% in the employer treatment.

The *ideal worker* model has predictions for employee-initiated changes, but for employer-initiated changes, the effects on wages are not theorised. Here, the Goldin conjecture states that being available at employers' discretion receives a premium, which is reflected in our experimental findings. Goldin goes on to argue that women are more likely to request discretion over working time, thus ending up being penalised in wages. In contrast to the initiator, the treatment effects corresponding to the gender of the employee are not statistically different from each other. Neither do we find statistically significant differences for the proportion of

respondents who recommend a wage reduction, increase or change based on the gender of the employee. It is the availability of the worker that matters, and not the alternative uses of time that one could presuppose for men and women. On the one hand, differences in recommended wage changes across initiators could substantiate Goldin's conjecture that availability is rewarded by a premium. On the other hand, the premiums are not large enough to explain away the gender wage gaps of roughly 20% as observed in Poland, where we run the experiment. In the next section, we employ a linear regression to study if availability premiums differ across genders.

Figure 3. Treatment effects: direction of recommended wage change



Notes: Bars indicate unconditional average treatment effects. Whiskers show 95% confidence intervals. The preferred sample includes 570 observations from the 190 subjects who answered correctly all manipulation check questions. The full sample includes 963 from 321 subjects. We report detailed test results in Table 4 in the Appendices.

Source: Authors' own elaboration.

Regression model: is there a role for the confounding variables?

Denote by $y_{i,v}$ the wage change recommended by respondent i in vignette v. This outcome variable is measured in absolute terms (the wage change in USD) and in relative terms (the wage change as % of the base wage described in each vignette). We estimate the following models:

$$y_{i,v} = \beta + \beta_I T:I_{i,v} + \beta_F T:F_{i,v} + \beta_{inter} T:I_{i,v} T:G_{i,v} + \delta_v + \delta X_i + e_{i,v} \quad (1)$$

where $T:I_{v,i}$ and $T:F_{v,i}$ refer to the treatment conditions, i.e. who requests discretion over WTA (employee is the reference category) and the gender of the character (man is the reference category). We adjust for vignette characteristics by including fixed effects denoted δ_v. Finally, δX_i indicates individual characteristics. These variables include age, gender, education, household income, managerial experience, and several controls for individual preferences. In the second specification, we replace X_i with individual fixed effects. While in the first model estimates are based on variation within and between subjects, in the second the parameters are identified from answers provided by the same individuals to different vignettes. This second approach is more robust to idiosyncratic behaviour on the side of respondents that cannot be easily captured by individual controls. In these specifications, β_I represents the *difference* in recommended wage changes between the two initiators, and not the levels. Given interactions, the interpretation for β_F is the role of the character's gender in the employee-initiated vignettes, and β_I captures the role of initiator in the context of vignettes about men. Finally, β_{inter} provides information about the additional effect of the female character on vignettes describing the change in WTAs initiated by the employer.

Following our hypotheses, we expect β_I to be positive, since the coefficient captures differences across initiator treatments (Hypotheses Hypothesis 1 and Hypothesis 2). In the *ideal worker* model, there is no specific assignment for genders. Goldin's conjecture does not require β_F to differ from zero either; however, the family devotion schema and gender norms hint towards positive β_F and β_{inter} (Hypothesis Hypothesis 3).

We estimate equation (1) using a linear regression. Our outcome measure, recommended wage change, is quasi-continuous, which validates the use of a linear approximation. In principle, one could be worried that top-codes would restrict possible answers, in which case a truncated model would be more adequate. In practice, less than 1 percent of the answers corresponded to the largest wage changes (in either direction), so censored regression models should produce similar coefficients, while putting an additional strain on the efficiency of the estimators.

Table 2. Recommended wage changes subsequent to changes in WTAs: experimental results

	in USD per month		in percent	
	(1)	(2)	(3)	(4)
T: employer = 1	17.20*** (4.85)	18.05*** (6.54)	2.95*** (0.56)	2.87*** (0.69)
T: woman= 1	−3.02 (4.87)	−0.11 (5.30)	0.15 (0.37)	0.44 (0.54)
T: employer × woman	10.30* (6.17)	8.96 (7.43)	0.10 (0.62)	−0.02 (0.93)
δX_i	Yes	No	Yes	No
δ_v	Yes	Yes	Yes	Yes
Observations	570	570	570	570
R^2	0.10	0.54	0.13	0.58

Notes: Table presents results of linear regressions of subjects' recommendation about the change in wages portrayed by Equation (1). Columns (1) and (2) present regressions where the dependent variable is the absolute wage change in USD per month, whereas in Columns (3) and (4) the dependent variable is the percentage change relative to status quo wages in each vignette. Columns (2) and (4) include subject fixed effects. The full specification is available in Table C3 in the Appendices. Standard errors clustered at subject level presented in parentheses. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1, respectively.

Source: Authors' own elaboration.

Consistent with the earlier results, Table 2 shows significant treatment effects for the initiator condition and insignificant effects for the gender condition. Overall, the wage is recommended to increase by roughly USD 17–18 per month (or 3% of initial wages) if the employer requests additional discretion over WTAs. These results are remarkably resilient to the inclusion of individual fixed effects, and to the so-called multiple-hypothesis-testing bias (see Table 5 in the Appendices.)

Table C3 in the Appendix reports the full set of coefficients, as well as the results in the full sample. We do not find evidence for the role of age, income or managerial experience. We find that women recommend, on average, lower wage rises than men: roughly USD 10 or 1% of the status quo wages. We further observe that the estimates of δ_v are significant only for the salesperson, and only in the case of relative wage changes. The estimates of approximately 2% appear large relative to the treatment effect of approximately 3 percent, but in this vignette the status quo wage was USD 420 per month, so even a few dollars more already amounts to a relatively higher fraction. One could rationalise this outcome as evidence that subjects value time discretion at a fixed price, independent of income levels.

Tables C4 and C5 (for preferred and full sample, respectively) in the Appendix expand further the results by adjusting for individual preferences. We described these variables in Sections 3.2 and 3.3. The set of variables includes their valuation of their own time availability, their agreement with traditional gender norms, measures of work-family and family-work conflict, the preference for inequality *vis a vis* redistribution, and their stated importance of care and work related values. Given that these variables are likely correlated, we estimate the model with each one of them separately, to avoid the risk that imperfect multicolinearity inflates

our t-statistics. Of these additional variables, only the indices of work-family and family-work conflict are related to the recommended changes in wages. Subjects who reported higher levels conflict tend to recommend higher changes in wages (by around USD 8 per month, or 0.6–0.9% relative to status quo). This result is in line with the general tendency that individuals who have experienced difficulty in managing the boundaries are more sensitive to others being exposed to the same risks.

Table C6 presents an specification that allows for interactions between treatments and vignettes. The finding that employer initiated requests require a wage increase seems to be consistent across specifications, and in the same range as those presented in Table 2, though precision seems to be lower. When we include individual fixed effects, none of the estimated interactions is statistically significant. In models with personal characteristics, we observe some positive interactions between the third vignette (Salesperson) and both treatments in levels (initiator and gender). However, we are cautious in interpreting these findings as pure attributes of the occupation because the order of vignettes was not randomized. Hence, one cannot disentangle pure vignette effects from respondents' fatigue.

Regression model: heterogeneity of treatment effects

While we find stable treatment effects across specifications and samples split according to manipulation checks, heterogeneity is still possible across relevant confounders. To explore heterogeneity we run Equation (1) for six sub-samples: people with (without) managerial experience, people with (without) university diploma, as well as men and women. We plot the estimated coefficients for treatment effects in Figure C1. The grey dashed line shows treatment effects from the full sample, while the shadowed area marks the 90% confidence intervals. Figure C1 shows that while point estimates for these sub-samples might deviate from the average effects, the confidence intervals overlap suggesting that treatment effects do not differ across sub-samples.[8]

The consistency of results is remarkable. It suggests that people with very different backgrounds, and potentially different labour market experience expect only minor adjustments in wages following the introduction of new TA. These results point towards the ideal worker model being deeply embedded in the way subjects think.

Social norm on the ideal worker and Goldin conjecture

Our design makes it possible to study the congruence between respondents' recommendations and their beliefs about the social norm. Subsequent to each vignette, we asked the subjects whether their recommendation on wage changes is consistent with what the majority of Poles would answer. Responses to this question are indicative of expectations about wage changes following shifts in WTAs. In most cases (83% of the responses), respondents claimed that their own recommendation is consistent with what most people in Poland would recommend. Differences were more frequent among those who recommended a decrease in wages (28% of responses) and the least frequent among those who recommended a wage increase (12%). As a result, respondents overestimate the frequency with which the average Pole would recommend a wage increase.

We construct two outcome variables. The first outcome is a binary variable that equals one when the subjects report that the overall norm is to change wages, in any direction, subsequent to changes in TAs. We call this variable *Majority in favour of a change*. The second binary variable equals one when the subjects believe that the overall norm is to increase the wage, and zero otherwise. We call this variable *Majority in favour of a raise*. We estimate a linear regression model analogous to equation (1) for these two outcome variables. The results are presented in Table 3.

The estimates in Table 3 corroborate our inference from Table 2. Indeed, subjects are more likely to expect wage changes when the employer requests discretion over TA: the proportion of workers who believe that the norm (majority opinion) is in favour of a wage change is between 8 and 10 percentage points higher, when

[8] Table C7 in the appendix provides a formal test of this inference: we interact treatment and the three personal characteristics: managerial experience, education and gender. None of the interactions is significant.

compared to employee initiated changes to work schedules. The results for wage increases are similar. The results hold even after adjusting for their own wage recommendations. The coefficients on recommendations, presented at the bottom of Table 3, show the extent of agreement between respondents' recommendations and their beliefs about the socially accepted rules. Those who recommend a wage increase are around 60 percentage points more likely to state that the majority believes that wages should increase.

Table 3. Beliefs about the social norm

	majority in favour of...	
	... change	... increase
T: employer = 1	0.097** (0.042)	0.11*** (0.040)
T: woman= 1	−0.043 (0.035)	−0.040 (0.034)
T: employer = 1 × T: woman= 1	0.074 (0.052)	0.082 (0.051)
Proposed change		
Negative change	0.55*** (0.17)	0.024 (0.13)
Positive change	0.57*** (0.055)	0.59*** (0.055)
Observations	570	570
P(y = 1)	0.360	0.326

Notes: All specifications were estimated using linear probability models and include the full set of *X*'s (See Table C8 in the Appendix). Columns (1) and (1a) report the estimates where the dependent variable takes on the value of 1 when the subject reports that the majority would change the wage of the character in the given vignette, and 0 otherwise. Columns (2) and (2a) the dependent variable equals one when the subject reports that the rest of society would raise the wage of the character in the given vignette, and 0 otherwise. The letter *a* denotes specifications for the full sample; the remaining columns show coefficients for the preferred sample. Standard errors clustered at the individual-level presented in parentheses. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

Discussion of the results

Our experiment lends partial support to the *ideal worker* model. When the employer requests discretion over working time arrangements, most survey respondents recommend no change in wages. Among those who recommend an increase in wages, the premium for granting such discretion is small. Conversely, when the employee requests discretion, the proportion of respondents recommending a wage cut increases, although the majority still suggest no change.

The share of respondents advocating wage penalties for employees requesting discretion to set their work schedules is very low, which challenges the predictions of the *ideal worker* model. One possible explanation is that participants expect work intensification: employees increase their effort during working hours to offset potential costs to the employer associated with increased flexibility.

However, our results do not fully support the implications of Goldin's conjecture. The recommended wage changes are rare and modest. If participants recommended wage changes at all, they selected small values – typically around 3% of status quo wages or under USD 30 per month. These treatment effects are an order of magnitude smaller than gender wage gaps in the society where our experiment was conducted. In other words, the availability premium – or conversely, the unavailability penalty – is neither frequent nor substantial enough to explain away gender wage inequality. Finally, we find no evidence of treatment heterogeneity across genders: men and women recommend similar wage changes when the employer initiates the negotiation.

Even though the estimated parameters are similar across specifications, we also observe that adding participant fixed effects substantially improves the fit of the model as measured by R^2. This suggests that other demographic characteristics have a strong bearing on proposed changes. One such characteristic is previous

experience working flexibly. Respondents who worked flexible hours can better assess the costs and benefits of flexible TA than respondents for whom this is a hypothetical exercise. However, given the low prevalence of flexible TAs in Poland before the pandemic (around 3% of workers, see Appendix B), the sample size required for any meaningful statistical analysis would have had to be an order of magnitude larger than what was collected.

Our estimates lie at the lower bound of those presented in the existing literature: **Mas and Pallais [2017]** report larger effects, although in their paper responses were capped at 3% of hourly wages. In our study half of the respondents who recommend wage changes suggested increases were under USD 50 per month, while the vast majority of recommended wage cuts were below USD 30 per month. The estimated average treatment effect reported in Tables 2 and C4 indicate modest overall wage adjustments, rather than a combination of answers at both extremes of the distribution. In other words, even those who supported wage increases or cuts following employer-initiated changes in TA generally proposed small modifications.

While we interpret these limited wage increases – and the modal response of no change – as evidence consistent with the ideal worker model, alternative explanations are also possible. The low premium could also indicate that respondents did not expect a major disruption to the lives of the vignette characters and their families. Another possibility is that participants viewed flexible working time arrangements as context-dependent: benefiting employers in some circumstances and employees in others As shown in Appendix B, empirically, flexibility seldom benefits only one side.

With reference to the gender dimension and Hypothesis 3, the recommended change in wages could be larger if we were to elaborate on how employer's discretion affects employees work-life balance; or how having additional discretion over own schedule could benefit the employee. We could emphasize how caring obligations, leisure, or self-improvement in the vignettes would be affected by changes in how schedules are set **[Vandello et al., 2013]**. However, doing so would risk conflating gender with other variables, thereby undermining the role of gender per se.

The small size of the effect is also driven by the large number of participants who recommended no change in wages, which may reflect *status quo bias* **[Samuelson, Zeckhauser, 1988]**. Our estimates fall within the range of standard estimates of *status quo bias* in the literature: around 50% (**Johnson et al. [1993]**, on the right to sue in car insurance), 58–60% (**Hartman, Donae, Woo [1990]**, on electric power consumption preferences) and 75–90% (**Jianbiao et al. [2009]**, on investors' decision-making). An alternative explanation would be that respondents found it easier to answer "no change." However, in our experiment, the effort required to recommend no change in wages was comparable to that required to recommend either an increase or a cut. More importantly, participants who recommended a wage change typically selected small values. This suggests that changes in discretion over TA were not perceived as a significant amenity (employee initiator) or a major inconvenience (employer initiator).

While the overall effect sizes tend to be small, the positive and marginally significant interaction between treatments observed in the first column of Table 2 raises some interesting questions. Respondents appear to expect a higher wage increase when the employer requests more control over TA from female employees. The coefficient is large, though imprecisely estimated. These coefficients suggest a different link between working time flexibility and gender. If employers anticipate higher rewards to women, they might be more reluctant to request such arrangements – potentially contributing to the penalty identified by **Goldin [2014]**. However, given the size of the standard errors, our discussion of interaction effects should be treated as speculative and an agenda for future research.

As with other experiments, external validity requires discussion. While we could not work with nationally representative samples in this design, we document effect sizes that are universally small. We find no heterogeneity in treatment effects. Hence, our study reveals no reasons to believe that, in representative samples, the results would be any different. However, extrapolating results to other countries with potentially different norms regarding work and gender should be done with caution.

On the positive side, our experimental approach allows us to isolate the effects of interest. In observational studies, it is typically impossible to identify which party initiated negotiations over working time arrangements. Moreover, as discussed by **Gascoigne and Kelliher** [**2017**] in their analysis of idiosyncratic deals, such negotiations often involve multiple dimensions: time arrangements, wages, tasks and (implicitly) effort. Our experimental designs offers a controlled environment where negotiation can be conveniently reduced to a single outcome: wages. Our vignette experiment also avoids self-selection problems inherent to observational studies.

Conclusions

The *ideal worker* model provides a coherent framework for understanding wage penalties associated with insufficient availability for work. While it is not obvious that being flexibly available to the employer should be rewarded, the model suggests that being unavailable ought to be penalised in wages. The *Goldin conjecture* implicitly states that being available to the employer is rewarded in wages. Goldin also argues that women are less frequently able to agree to working time arrangements at the discretion of employers, and that this difference stands behind the observed differences in wages between men and women.

The objective of our experiment was to elicit the pecuniary value and social acceptability of the type of working time arrangements where discretion over work schedules deviates from the standard. Starting from the status quo of a regular 9-to-5 arrangement (or 9-to-1 for part-time), participants were presented with vignettes depicting credible departures from this arrangement. From the perspective of the *ideal worker* model, it seems that employees requesting discretion over their start and end times should be penalised, while employer discretion should face lower social acceptance

Our study provides novel insights into how working time arrangements are perceived socially in the context of the *ideal worker* model. We interpret our results as evidence that wages should be related to the productivity of individual workers rather than some standardised parameters of work arrangements. Discretion over start and end times does not appear to be a relevant factor in wage recommendations. Although the *ideal worker* model seems to be the default reference (with participants not thinking that granting employers more discretion to set schedules should result in higher wages for the employee), departures from this model did not significantly influence wage recommendations.

The original *ideal worker* concept does not account for remuneration or worker productivity. Our experiment shows that more theoretical and empirical work is needed because these dimensions appear to be of relevance for the social perception of flexible working time arrangements. A novel conceptualisation of the *ideal worker*, that incorporates compensation and its relation to perceived worker productivity, does not appear to need an important gender dimension. Non-standard working time arrangements are prevalent, and are likely to become increasingly relevant across various segments of the labour market. Leveraging the *ideal worker* model to theorise more deeply about the social role of such arrangements could provide insights into the potential role of legal initiatives such as the EU's *Working Time Directive*, which specifies the conditions under which workers can be expected to answer phone calls or emails outside of work hours. More theorising is also needed to determine whether perceived productivity is a valid benchmark for evaluating flexibility.

From a policy perspective, our results suggest that employees attach little monetary value to flexible time arrangements, so even moderate productivity gains could justify their use. The trade-off between employee protection and productivity gains does not set the bar too high for the latter. However, our results should be interpreted with caution. Respondents did not decide about their own working time arrangements in the experiment, and the valuations refer to fictional characters. Our findings are consistent with European evidence provided in **Bech-Wysocka et al.** [**2024**], which emphasises that removing flexible work arrangements would be detrimental for workers.

Our study provides several avenues for future research. First, there is a rationale for repeating the experiment today. When the experiment was conducted, the prevalence of work flexibility was low. One would

expect that the pandemic changed the landscape. A greater share of respondents has first-hand experience working from home, where employer control over start and end times is typically looser. Second, our experiment sought to understand wage changes associated with departures from standard TAs, and not those related to the adoption of standard TAs. It is unclear whether transitioning to a standard TA would result in symmetric wage effects of the opposite sign.

Regarding gender, our study corroborates the conclusion that gender wage gaps cannot be explained solely by differences in access to flexible working time arrangements between men and women. We provide novel causal evidence. It appears that the raised by Goldin conjecture, that a final explanation for gender wage gaps was identified, require further theorizing and empirical research.

References

Aguinis H., Bradley K.J. [2014], Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies, *Organizational Research Methods*, 17(4): 351–371.

Angelici M., Profeta P. [2020], *Smart-Working: Work Flexibility Without Constraints*, CESifo Working Papers No. 8165.

Azar S., Khan A., Van Eerde W. [2018], Modelling Linkages Between Flexible Work Arrangements' Use and Organizational Outcomes, *Journal of Business Research*, 91: 134–43, https://doi.org/10.1016/j.jbusres.2018.06.004.

Barnett R.Ch., Hyde J.Sh. [2001], Women, Men, Work, and Family: An Expansionist Theory, *American Psychologist*, 56(10): 781, https://doi.org/10.1007/978-3-319-21635-5_4.

Batt R., Valcour P.M. [2003], Human Resources Practices as Predictors of Work-Family Outcomes and Employee Turnover, *Industrial Relations: A Journal of Economy and Society*, 42(2): 189–220.

Beauregard T., Henry L. [2009], Making the link between work-life balance practices and organizational performance, *Human Resource Management Review*, 19(1): 9–22, https://doi.org/10.1016/j.hrmr.2008.09.001.

Bech-Wysocka K., Smyk M., Tyrowicz J., Velde L. van der [2024], Matching It up: Non-Standard Work and Job Satisfaction, *Journal of Policy Analysis and Management*, January, https://doi.org/10.1002/pam.22565.

Boltz M., Cockx B., Diaz A.M., Salas L. [2020], *How Does Working-Time Flexibility Affect Workers' Productivity in a Routine Job? Evidence from a Field Experiment*, IZA Discussion Papers No. 13825.

Bożewicz M., Boguszewski R., Feliksiak M., Omyla-Rudzka M. [2019], *Współczesna polska rodzina*, Opinie i Diagnozy No. 44, CBOS, https://www.cbos.pl/PL/publikacje/diagnozy.php, (accessed on 26.05.2025).

Breaugh J.A. [1985], The Measurement of Work Autonomy, *Human Relations*, 38(6): 551–570, https://doi.org/10.1177/001872678503800604.

Byron K. [2005], A Meta-Analytic Review of Work – Family Conflict and Its Antecedents, *Journal of Vocational Behavior*, 67(2): 169–198, https://doi.org/10.1016/j.jvb.2004.08.009.

Cañibano A. [2018], Workplace Flexibility as a Paradoxical Phenomenon: Exploring Employee Experiences, *Human Relations*, 72(2): 444–470, https://doi.org/10.1177/0018726718769716.

Cheung J.H., Burns D.K., Sinclair R.R., Sliter M. [2017], Amazon Mechanical Turk in Organizational Psychology: An Evaluation and Practical Recommendations, *Journal of Business and Psychology*, 32(4): 347–361, https://doi.org/10.1007/s10869-016-9458-5.

Chung H. [2020], Gender, Flexibility Stigma and the Perceived Negative Consequences of Flexible Working in the UK, *Social Indicators Research*, 151(2): 521–545, https://doi.org/10.1007/s11205-018-2036-7.

Chung H., Horst M. van der [2017], Women's Employment Patterns After Childbirth and the Perceived Access to and Use of Flexitime and Teleworking, *Human Relations*, 71(1): 47–72, https://doi.org/10.1177/0018726717713828.

Cortes P., Pan J. [2019], When Time Binds: Substitutes for Household Production, Returns to Working Long Hours, and the Skilled Gender Wage Gap, *Journal of Labor Economics*, 37(2): 351–398, https://doi.org/10.1086/700185.

Cousins Ch.R, Tang N. [2004], Working Time and Work and Family Conflict in the Netherlands, Sweden and the UK, *Work, Employment and Society*, 18(3): 531–549, https://doi.org/10.1177/0950017004045549.

Davies A.R., Frink B.D. [2014], The Origins of the Ideal Worker: The Separation of Work and Home in the United States from the Market Revolution to 1950, *Work and Occupations*, 41(1): 18–39, https://doi.org/10.1177/0730888413515893.

De Menezes L.M., Kelliher C. [2011], Flexible Working and Performance: A Systematic Review of the Evidence for a Business Case, *International Journal of Management Reviews*, 13(4): 452–474, https://doi.org/10.1111/j.1468-2370.2011.00301.x.

Dex S., Scheibl F. [2001], Flexible and Family-Friendly Working Arrangements in UK-Based SMEs: Business Cases, *British Journal of Industrial Relations*, 39(3): 411–431, https://doi.org/10.1111/1467-8543.00207.

Fagan C., Lyonette C., Smith M., Saldana-Tejeda A. [2012], *The Influence of Working Time Arrangements on Work-Life Integration or 'Balance': A Review of the International Evidence*, Conditions of Work and Employment No. 32, International Labour Organization.

Gagne M., Deci E.L. [2005], Self-Determination Theory and Work Motivation, *Journal of Organizational Behavior* 26(4): 331–362, https://doi.org/10.1002/job.322.

Gascoigne Ch., Kelliher C. [2017], The Transition to Part-Time: How Professionals Negotiate 'Reduced Time and Workload' i-Deals and Craft Their Jobs, *Human Relations*, 71(1): 103–125, https://doi.org/10.1177/0018726717722394.

Goldin C. [2014], A Grand Gender Convergence: Its Last Chapter, *American Economic Review*, 104(4): 1091–1119, https://doi.org/10.1257/aer.104.4.1091.

Goraus K.., Tyrowicz J. [2014], Gender Wage Gap in Poland – Can It Be Explained by Differences in Observable Characteristics?, *Ekonomia*, 36: 125–148, http://ekonomia.wne.uw.edu.pl/ekonomia/getFile/733.

Goraus K., Tyrowicz J., Velde L.A. van der [2017], Which Gender Wage Gap Estimates to Trust? A Comparative Analysis, *Review of Income and Wealth*, 63(1): 118–146, https://doi.org/10.1111/roiw.12209.

Goraus K., Tyrowicz J., Velde L.A. van der [2020], How Rankings Disguise Gender Inequality: A Comparative Analysis of Cross-Country Gender Equality Rankings Based on Adjusted Wage Gaps, *PLOS One*, 15(11): e0241107, https://doi.org/10.1371/journal.pone.0241107.

Grote G., Raeder S. [2009], Careers and Identity in Flexible Working: Do Flexible Identities Fare Better?, *Human Relations*, 62(2): 219–244, https://doi.org/10.1177/0018726708100358.

Hainmueller J., Hangartner D., Yamamoto T. [2015], Validating Vignette and Conjoint Survey Experiments Against Real-World Behavior, *Proceedings of the National Academy of Sciences*, 112(8): 2395–2400.

Harrison G.W., List, J.A. [2004], Field Experiments, *Journal of Economic Literature* 42(4): 1009–1055, https://doi.org/10.2139/ssrn.698961.

Hartman R.S., Donae M.J., Woo Ch.-K. [1990], Status Quo Bias in the Measurement of Value of Service, *Resources and Energy*, 12(2): 197–214.

He H., Neumark D., Weng Q. [2021], Do Workers Value Flexible Jobs? A Field Experiment, *Journal of Labor Economics*, 39(3): 709–738, https://doi.org/10.1086/711226.

Heavey A.L., Holwerda J.A, Hausknecht J.P. [2013], Causes and Consequences of Collective Turnover: A Meta-Analytic Review, *Journal of Applied Psychology*, 98(3): 412, https://doi.org/10.1037/a0032380.

Jianbiao L., Guangqian R., Qiuhua M., Lüke L. [2009], An Experimental Study on Investors' Status Quo Bias and Its Determinants, *Frontiers of Business Research in China*, 3(4): 543–565.

Johnson E.J., Hershey J., Meszaros J., Kunreuther H. [1993], Framing, Probability Distortions, and Insurance Decisions, *Journal of Risk and* Uncertainty, 7(1): 35–51.

Kaduk A., Genadek K., Kelly E.L., Moen Ph. [2019], Involuntary vs. Voluntary Flexible Work: Insights for Scholars and Stakeholders, *Community, Work & Family*, 22(4): 412–442.

Kauffeld S., Jonas E., Frey D. [2004], Effects of a Flexible Work-Time Design on Employee-and Company-Related Aims, *European Journal of Work and Organizational Psychology*, 13(1): 79–100, https://doi.org/10.1080/13594320444000001.

Kelliher C., Anderson D. [2009], Doing More with Less? Flexible Working Practices and the Intensification of Work, *Human Relations*, 63(1): 83–106, https://doi.org/10.1177/0018726709349199.

Kelly E.L., Kalev A. [2006], Managing Flexible Work Arrangements in US Organizations: Formalized Discretion or 'a Right to Ask', *Socio-Economic Review*, 4(3): 379–416.

Kelly E. L, Moen Ph., Tranby E. [2011], Changing Workplaces to Reduce Work-Family Conflict: Schedule Control in a White-Collar Organization, *American Sociological Review*, 76(2): 265–290.

Kossek E.E., Barber A., Winters D. [1999], Managerial Use of Flexible Work Schedules: The Power of Peers, *Human Resource Management Journal*, 38: 33–46, https://doi.org/10.1002/(SICI)1099-050X(199921)38:1<33::AID-HRM4>3.0.CO;2-H.

Lesley H. [2009], Making the Link Between Work-Life Balance Practices and Organizational Performance, *Human Resource Management Review*, 19(1): 9–22, https://doi.org/10.1016/j.hrmr.2008.09.001.

Levitt S.D., List J.A. [2009], Field Experiments in Economics: The Past, the Present, and the Future, *European Economic Review*, 53(1): 1–18, https://doi.org/10.1016/j.euroecorev.2008.12.001.

List J.A., Shaikh A.M., Xu Y. [2019], Multiple Hypothesis Testing in Experimental Economics, *Experimental Economics*, 22(4): 773–793.

Luhr S. [2020], Signaling Parenthood: Managing the Motherhood Penalty and Fatherhood Premium in the US Service Sector, *Gender & Society*, 34(2): 259–83.

Majchrowska A., Strawiński P. et al. [2016], Regional Differences in Gender Wage Gaps in Poland: New Estimates Based on Harmonized Data for Wages, *Central European Journal of Economic Modelling and Econometrics*, 8(2): 115–141.

Mas A., Pallais A. [2017], Valuing Alternative Work Arrangements, *American Economic Review*, 107 (12): 3722–3759, https://doi.org/10.1257/aer.20161500.

Mazmanian M., Orlikowski W. J., Yates J. A. [2013]. The Autonomy Paradox: The Implications of Mobile Email Devices for Knowledge Professionals, *Organization Science*, 24 (5): 1337–1357, https://doi.org/10.1287/orsc.1120.0806.

Michel J. S., Kotrba L. M., Mitchelson J. K., Clark M. A., Baltes B. B. [2011], Antecedents of Work – Family Conflict: A Meta-Analytic Review, *Journal of Organizational Behavior*, 32 (5): 689–725.

Moen Ph., Kelly E. L., Hill R. [2011], Does Enhancing Work-Time Control and Flexibility Reduce Turnover? A Naturally Occurring Experiment, *Social Problems*, 58 (1): 69–98.

Munsch Ch. L. [2016], Flexible Work, Flexible Penalties: The Effect of Gender, Childcare, and Type of Request on the Flexibility Bias, *Social Forces*, 94 (4): 1567–1591, https://doi.org/10.1093/sf/sov122.

Netemeyer R. G., Boles J. S., McMurrian R. [1996], Development and Validation of Workfamily Conflict and Familywork Conflict Scales, *Journal of Applied Psychology*, 81 (4): 400–410, https://doi.org/10.1037/0021-9010.81.4.400.

Porter Ch. O. L.H., Outlaw R., Gale J. P., Cho Th. S. [2019], The Use of Online Panel Data in Management Research: A Review and Recommendations, *Journal of Management*, 45 (1): 319–344, https://doi.org/10.1177/0149206318811569.

Reilly P. A. [1998], Balancing Flexibility – Meeting the Interests of Employer and Employee, *European Journal of Work and Organizational Psychology*, 7 (1): 7–22, https://doi.org/10.1080/135943298398934.

Rubenstein A. L., Eberly M. B., Lee Th. W., Mitchell T. R. [2018], Surveying the Forest: A Meta-Analysis, Moderator Investigation, and Future-Oriented Discussion of the Antecedents of Voluntary Employee Turnover, *Personnel Psychology*, 71 (1): 23–65, https://doi.org/10.1111/peps.12226.

Samuelson W., Zeckhauser R. [1988], Status Quo Bias in Decision Making, *Journal of Risk and Uncertainty*, 1 (1): 7–59.

Sharpe Wessling K., Huber J., Netzer O. [2017], MTurk Character Misrepresentation: Assessment and Solutions, *Journal of Consumer Research*, 44 (1): 211–230, https://doi.org/10.1093/jcr/ucx053.

Smith R. S. [1979], Compensating Wage Differentials and Public Policy: a Review, *ILR Review*, 32 (3): 339–352.

Soga L. R., Bolade-Ogunfodun Y., Mariani M., Nasr R., Laker B. [2022], Unmasking the Other Face of Flexible Working Practices: A Systematic Literature Review, *Journal of Business Research*, 142: 648–662.

Spector P. E. [1986], Perceived Control by Employees: A Meta-Analysis of Studies Concerning Autonomy and Participation at Work, *Human Relations*, 39 (11): 1005–1016, https://doi.org/10.1177/001872678603901104.

Teasdale N. [2013], Fragmented Sisters? The Implications of Flexible Working Policies for Professional Women's Workplace Relationships, *Gender, Work & Organization*, 20 (4): 397–412, https://doi.org/10.1111/j.1468-0432.2012.00590.x.

Vandello J. A., Hettinger V. E., Bosson J. K., Siddiqi J. [2013], When Equal Isn't Really Equal: The Masculine Dilemma of Seeking Work Flexibility, *Journal of Social Issues*, 69 (2): 303–321, https://doi.org/10.1111/josi.12016.

Walby S., Olsen W. [2002], *The Impact of Women's Position in the Labour Market on Pay and Implications for UK Productivity*, Report, Department of Trade and Industry, Women and Equality Unit.

Warren T., Rowlingson K., Whyley C. [2001], Female Finances: Gender Wage Gaps and Gender Assets Gaps, *Work, Employment and Society*, 15 (3): 465–488, https://doi.org/10.1177/09500170122119110.

Williams J. [2001], *Unbending Gender: Why Family and Work Conflict and What to Do about It*, Oxford University Press.

Appendix

A. Survey materials

Besides the vignette experiment, the survey included two short modules to highlight social attitudes and work-family conflict. Below we reproduce the questions used in those sections.

Social attitudes

The question on social attitudes was taken from the cyclical study "Modern Polish family" by **Bożewicz et al. [2019]**. The original questions were in Polish. We present them translated into English.

1. People guide their lives based on different values. Please, sort the values presented below according to their importance in your life. Please start with the most important.

- Material well-being, wealth
- Friendship
- Contact with culture
- Participation in a democratic society
- Prosperity of the homeland
- Professional success
- Peace
- Fame
- Respect from others
- Family happiness
- Honest life
- Religious faith
- Freedom to express personal views
- Education
- Living in good health
- Living an adventurous life

Based on the sorting, we indicate that a person is socially oriented when *friendship*, *family happiness*, or *living an adventurous life* are listed first. A person is career-oriented when *material well-being*, *professional success* or *fame* come first.

To find out about the relation between work and family life, we asked participants to what extent they agreed with several statements (listed below). The answers were presented using a Likert scale with five levels: strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree. The questions below are reproduced from **Netemeyer, Boles, and McMurrian [1996]**.

1. Work-family conflict
   - The demands of my work interfere with my home and family life.
   - The amount of time my job takes up makes it difficult to fulfill family responsibilities.
   - Things I want to do at home do not get done because of the demands my job puts on me.
   - My job produces strain that makes it difficult to fulfill family duties.
   - Due to work-related duties, I have to make changes to my plans for family activities.
2. Family-work conflict
   - The demands of my family interfere with work-related activities.
   - I have to put off doing things at work because of demands on my time at home
   - Things I want to do at work don't get done because of the demands of my family.
   - My home life interferes with my responsibilities at work such as getting to work on time, accomplishing daily tasks, and working overtime
   - Family-related strain interferes with my ability to perform job-related duties.

To inquire about gender values and social norms, we adapted five questions used by the International Social Survey Program. Participants were asked to indicate to what extent they agreed with different statements. There were four possible answers: strongly agree, agree, disagree, and strongly disagree. The set of questions was the following:

- A man's job is to earn money; a woman's job is to look after the home and family.
- All in all, family life suffers when the woman has a full-time job.
- When jobs are scarce, men should have priority over women.
- A preschool child is likely to suffer if his or her mother works.
- On the whole, men make better business managers than women.

B. Background: the Polish context

We conducted the vignette experiment in Poland, a country where the distribution of hours worked is highly concentrated. According to the Labour Force Survey, in 2019, most employees (around 80% of both genders) worked 40 hours per week. Among the remaining 20%, gender differences emerge. The proportion of women working 30 hours or fewer more than doubles that of men (11% vs. 4%). In terms of actual hours worked, most people declared working their usual hours. Among those who deviated, men were more likely to have worked more hours, either due to overtime (around 2% of men vs. 1% of women) or varying schedules (1% for men vs. 0.5% for women). Within the EU, Poland ranks in the lower half in terms of flexible working hours. In leading countries – mostly those in Northern Europe – the proportion of individuals reporting flexible work schedules is four to seven times higher. In these countries, between 16% and 21% of workers report having flexible time arrangements.

Even within this relatively stable pattern of working hours, there is some room for flexibility. In the same survey, workers were asked whether they had autonomy in deciding their start and end times. Among men, up to 18% reported having such autonomy, and almost 7% said they could fully decide their working hours. The proportions for women were similar (16% and 6% respectively). However, when it comes to flexibility requested by the employer, gender differences become more evident. Around 75% of women reported that they were never asked to change their working time to attend urgent matters. For men, this proportion was lower (68%), suggesting that men are more frequently asked to work flexibly. Unlike our experiment, data suggests that both forms of flexibility – employee- and employer-initiated – coexist in actual employment arrangements. Around 27% of men (and 20% of the women) who said they had a certain measure of autonomy over their schedules reported being asked to modify them for urgent tasks.[9]

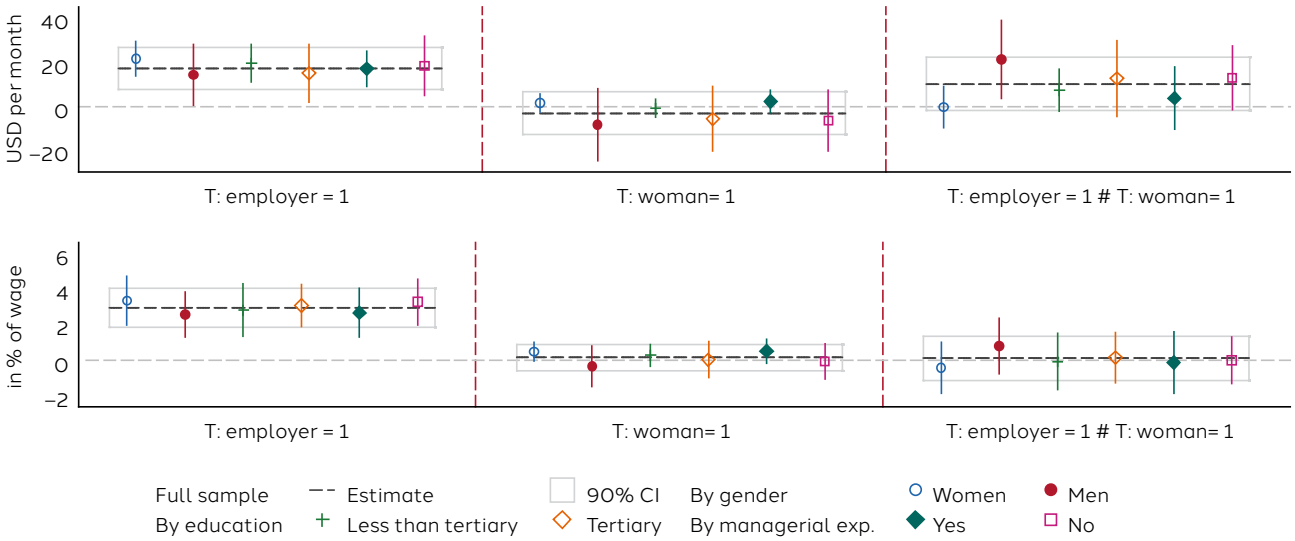
Poland ranks relatively low on measures of gender equality. Various estimates place the adjusted gender wage gap at around 20% of men's wages. This gap has been stable over the past two decades [**Goraus, Tyrowicz, 2014**], robust across methods [**Goraus, Tyrowicz, Velde, 2017**], and evenly spread within the country [**Majchrowska, Strawiński et al., 2016**]. In the European context, Poland's gender wage gap is higher than in most other European countries, with the exceptions of Portugal, Spain and Estonia [**Goraus, Tyrowicz, Velde, 2020**]. Social norms in Poland also reflect more traditional views on work, family and gender roles. According to the European Values Study, almost 25% of Poles agree that "men should have more right to a job than women when jobs are scarce" – a figure much higher than in Germany (7%) or Spain (11%), and approximately 5 percentage points above the European average. Similarly, 20% of Poles agree with the statement that "men make better business executives than women," compared to 6.7% in Spain and 10% in Germany. In the survey *Modern Polish Family* by **Bożewicz et al. [2019]**, 80% of Poles said that "family happiness" is the most important value in their lives, while only 36% selected "work." At the same time, most household duties are performed by women, and men are less likely to say they would be willing to leave their job to care for the home and family if their household finances made that a viable option (32% vs. 42%).

When it comes to the *ideal worker* notion, two-thirds of Poles claim that work is very important in their lives, compared to 46.6% in Germany and 35.4% in the Netherlands. Nearly 40% of Poles agree that "work should always come first, even if it means less spare time," compared to 30% in Germany and 23% in the Netherlands, according to the European Values Study. Additionally, the European Social Survey shows that Poles lead Europe in agreeing with the statement that income inequality is acceptable if it reflects rewards for talent and effort, with 60% of respondents in agreement.

[9] Unfortunately, the sample size is too small and the occupation codes are too coarse to provide similar estimates for the occupations used in the experiment. These estimates are simple averages that do not account for differences in other job and personal characteristics.

C. Additional results

Figure C1. Treatment effects: heterogeneity is virtually absent



Notes: Treatment effects on wage changes (in USD) and 95% confidence intervals. The preferred sample (190 participants) is divided along gender, managerial experience, and education. The regressions include demographic characteristics and cluster standard errors at the individual level.

Source: Authors' own elaboration.

Table C1. Treatment effects: Should wages change subsequent to new TAs?

	Initiator					Gender of the character				
	Employer		Employee		Diff.	Man		Woman		Diff.
	(1)	(2)	(3)	(4)	(1)–(3)	(1)	(2)	(3)	(4)	(1)–(3)
	$\bar{y}$	*t-stat*	$\bar{y}$	*t-stat*		$\bar{y}$	*t-stat*	$\bar{y}$	*t-stat*	
Preferred sample: Subjects who passed all manipulation checks										
Decrease	0.000	.	0.025	2.67***	−0.02***	0.007	1.42	0.017	2.25**	−0.01
No change	0.615	21.35***	0.870	43.47***	−0.25***	0.754	29.00***	0.731	28.24***	0.02
Increase	0.385	13.35***	0.106	5.78***	0.28***	0.239	9.30***	0.252	9.93***	−0.01
in USD / month	29.788	10.31***	7.441	3.15***	22.35***	17.611	6.21***	19.633	7.50***	−2.02
in %	3.852	9.76***	0.850	3.92***	3.00***	2.335	6.73***	2.376	7.50***	−0.04
Full sample: All subjects participating in the study										
Decrease	0.027	3.65***	0.052	5.13***	−0.02*	0.030	3.79***	0.049	5.02***	−0.02
No change	0.610	27.32***	0.789	42.53***	−0.18***	0.712	34.03***	0.688	32.99***	0.02
Increase	0.363	16.51***	0.159	9.56***	0.20***	0.258	12.76***	0.263	13.27***	−0.01
in USD / month	28.595	4.75***	5.895	1.71*	22.70***	19.855	5.27***	14.652	2.55**	5.20
in %	3.583	5.84***	0.642	1.47	2.94***	2.275	4.48***	1.944	3.47***	0.33

Notes: In the table, *t-stat* denotes the mean, and $\bar{y}$ refers to the t – statistic from a test of the null hypothesis that the given mean is equal to zero. Columns entitled *Diff.* report the mean differential between treatment conditions. ***, **, * indicate p-values less than 10%, 5% and 1% respectively. The full sample includes 321 subjects and 963 observations, while the sample for subjects who passed all manipulation checks includes 190 subjects and 570 observations. *Reduction* denotes the share of individuals recommending a wage decrease following a change in TAs; *no change* and *increase* are defined analogously. Wage changes are reported in both USD per month and as a percentage of the status quo wage.

Source: Authors' own elaboration.

Table C2. Treatment effects: p-values adjusted for multiple hypothesis testing

	Initiator			Gender of character		
	List, Shaikh, and Xu (2019)	Bonferroni	Holms	List, Shaikh, and Xu (2019)	Bonferroni	Holms
Subjects who passed all manipulation checks						
Decrease	0.03	0.15	0.03	0.66	1.00	1.00
No change	0.00	0.00	0.00	0.85	1.00	1.00
Increase	0.00	0.00	0.00	0.90	1.00	1.00
in USD / month	0.00	0.00	0.00	0.87	1.00	1.00
in %	0.00	0.00	0.00	0.93	1.00	0.93
All subjects participating in the study						
Decrease	0.04	0.21	0.04	0.41	0.73	0.73
No change	0.00	0.00	0.00	0.75	1.00	1.00
Increase	0.00	0.00	0.00	0.86	1.00	0.86
in USD / month	0.00	0.00	0.00	0.74	1.00	1.00
in %	0.00	0.00	0.00	0.87	1.00	1.00

Notes: The table presents p-values for the null hypothesis that treatment effects are equal to zero. Three corrections for multiple hypothesis testing are proposed: **List, Shaikh, and Xu [2019]**, and the more conservative Holms and Bonferroni adjustments. The p-values correspond to the tests presented in columns titled "Diff." in Table 4.

Source: Authors' own elaboration.

Table C3. Wage changes subsequent to changes in TAs: all coefficients

	(1)	(2)	(3)	(4)	(1a)	(2a)	(3a)	(4a)
T: employer = 1	17.20*** (4.85)	18.05*** (6.54)	2.95*** (0.56)	2.87*** (0.69)	22.04*** (6.80)	25.35*** (8.16)	3.13*** (0.93)	3.40*** (1.00)
T: woman= 1	−3.02 (4.87)	−0.11 (5.30)	0.15 (0.37)	0.44 (0.54)	−4.00 (6.59)	−2.16 (8.60)	0.04 (0.82)	0.35 (1.05)
T: employer × woman	10.30* (6.17)	8.96 (7.43)	0.10 (0.62)	−0.02 (0.93)	0.28 (13.50)	−3.02 (19.74)	−0.34 (1.45)	−0.75 (1.98)
V: lawyer = 1	8.84* (5.13)	8.84 (6.24)	−0.45 (0.39)	−0.45 (0.48)	14.96 (9.55)	14.80 (11.78)	0.20 (0.76)	0.19 (0.93)
V: salesperson = 1	0.15 (2.80)	0.32 (3.43)	1.86*** (0.52)	1.88*** (0.64)	4.24 (5.75)	4.15 (6.99)	2.28** (0.92)	2.27** (1.12)
Age	0.04 (0.23)		0.01 (0.03)		0.43 (0.30)		0.04 (0.03)	
Woman	−10.26** (4.89)		−0.98* (0.58)		−3.37 (9.48)		−0.75 (1.03)	
Never manager	5.73 (5.38)		0.73 (0.64)		5.12 (9.77)		0.39 (1.06)	
Completed tertiary	10.21** (4.79)		1.08* (0.59)		4.60 (9.77)		0.46 (1.04)	
Income: enough	0.32 (5.30)		0.25 (0.64)		−1.50 (9.72)		−0.29 (1.07)	
Income: not enough	9.79 (8.45)		1.35 (1.12)		9.96 (12.12)		2.03 (1.57)	
Intercept	−0.44 (9.13)	4.34 (4.30)	−0.95 (1.18)	0.22 (0.51)	−18.79 (18.28)	0.10 (7.27)	−1.98 (1.94)	−0.40 (0.89)
Observations	570	570	570	570	963	963	963	963
R^2	0.10	0.54	0.13	0.58	0.02	0.46	0.03	0.50

Notes: The table presents the results of linear regressions analysing subjects' wage change recommendations, as specified in equation (1). The base category for occupation is "hairdresser;" thus, coefficients for "lawyer" and "salesperson" are interpreted relative to this group. For education, "less than tertiary education" serves as the reference category, and for income, "being able to afford luxury" is the baseline. Columns (1) and (2) display regressions where the dependent variable is the absolute wage change in USD per month, whereas in Columns (3) and (4) the dependent variable is the percentage change relative to status quo wages in each vignette. Columns (2) and (4) include subject fixed effects. The letter *a* denotes specifications for the full sample; the remaining columns show coefficients for subjects who passed all manipulation checks. Standard errors, clustered at the subject level, are reported in parentheses. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

Table C4. Recommended wage changes subsequent to changes in TAs: additional controls (preferred sample)

	Dependent variable: Wage change in USD per month						Dependent variable: Wage change in % of status quo					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
T: employer = 1	17.16*** (4.90)	17.21*** (4.84)	16.90*** (4.86)	17.53*** (4.84)	17.22*** (4.84)	17.27*** (4.91)	2.93*** (0.56)	2.95*** (0.56)	2.90*** (0.56)	2.98*** (0.56)	2.95*** (0.55)	2.96*** (0.56)
T: woman= 1	−3.04 (4.89)	−3.03 (4.94)	−3.24 (4.89)	−3.20 (4.89)	−3.12 (4.87)	−3.20 (4.90)	0.14 (0.37)	0.14 (0.38)	0.11 (0.37)	0.13 (0.38)	0.13 (0.37)	0.13 (0.38)
T: employer × woman	10.41* (6.15)	10.32 (6.25)	10.59* (6.13)	10.20 (6.17)	10.24* (6.20)	10.56* (6.04)	0.14 (0.62)	0.11 (0.63)	0.15 (0.63)	0.09 (0.63)	0.09 (0.63)	0.14 (0.62)
Valuation of own time	4.02 (19.40)						1.19 (2.50)					
Gender norm		0.27 (3.17)						0.11 (0.36)				
Work-family conflict			3.46 (2.46)						0.54* (0.30)			
Family-work conflict				7.74*** (2.80)						0.88** (0.36)		
Preference for inequality					7.06 (5.86)						1.10 (0.80)	
Importance: away from work						−1.28 (1.38)						−0.19 (0.17)
Importance: towards work						0.37 (1.23)						0.02 (0.14)
δX_i	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
δ_v	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	570	570	570	570	570	570	570	570	570	570	570	570
R^2	0.10	0.10	0.10	0.11	0.10	0.10	0.13	0.13	0.13	0.14	0.13	0.13

Notes: The table presents the results of linear regressions of subjects' recommendations about wage changes, as specified in equation (1). The estimated model includes the full set of *X*'s; the results are available upon request and omitted here for brevity, as they closely mirror those in Table 2. The estimates are reported for the preferred sample. The results for the full sample are available in Table 7. "Own-time availability" is based on real-consequence component of our experiment. Higher values indicate a greater value attributed to discretion over personal time. "Gender norm" is an index based on items for traditional vs. modern norms adopted from the World Value Survey. Higher values indicate equitable views. "Work-family conflict" and "family-work conflict" are indices based on **Netemeyer, Boles, and McMurrian [1996]**. Higher values indicate a higher sense of conflict. "Preference for inequality" is based on an item utilised in the European Social Survey. "Fulfilling life" and "work & career" are based on a ranking of life components adopted from a standardised opinion poll [Bożewicz et al., 2019]. A constant was included in all regressions but is not reported. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

Table C5. Recommended wage changes following a change in TAs: additional controls (full sample)

	Dependent variable: Wage change in USD per month						Dependent variable: Wage change in % of status quo					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
T: employer = 1	22.55*** (7.03)	21.93*** (6.88)	21.59*** (6.65)	21.83*** (6.68)	21.97*** (6.80)	22.07*** (6.82)	3.20*** (0.94)	3.12*** (0.94)	3.05*** (0.92)	3.10*** (0.92)	3.11*** (0.93)	3.14*** (0.93)
T: woman= 1	−3.72 (6.61)	−4.24 (6.74)	−4.39 (6.51)	−4.46 (6.42)	−4.12 (6.57)	−4.11 (6.59)	0.07 (0.83)	0.02 (0.84)	−0.03 (0.82)	−0.04 (0.81)	−0.00 (0.82)	0.03 (0.82)
T: employer × woman	−0.67 (13.83)	0.42 (13.51)	0.57 (13.44)	0.45 (13.49)	0.32 (13.48)	0.50 (13.49)	−0.46 (1.45)	−0.32 (1.45)	−0.29 (1.45)	−0.31 (1.46)	−0.32 (1.45)	−0.31 (1.45)
Valuation of own time	−34.93 (37.09)						−4.48 (4.09)					
Gender norm		3.79 (4.93)						0.33 (0.56)				
Work-family conflict			5.07 (3.99)						0.90** (0.45)			
Family-work conflict				4.97 (5.32)						0.84 (0.56)		
Preference for inequality					3.43 (6.02)						1.21 (0.85)	

cont. Table C5

	Dependent variable: Wage change in USD per month						Dependent variable: Wage change in % of status quo					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Importance: away from work						−0.69 (1.28)						−0.11 (0.16)
Importance: towards work						−1.44 (1.15)						−0.16 (0.13)
δX_i	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
δ_v	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	963	963	963	963	963	963	963	963	963	963	963	963
R^2	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03

Notes: The table presents the results of linear regressions of subjects' wage change recommendations as portrayed by equation (1). The estimated model includes the full set of *X*s; the results are available upon request and are omitted here to avoid redundancy with Table 2. The estimates are reported for the full sample. The results for the preferred sample are available in Table 6. "Own-time availability" based on the real-consequence component of our experiment. Higher values indicate a greater value attributed to discretion over one's own time. "Gender norm" is an index based on items for traditional vs. modern norms adopted from the World Value Survey. Higher values indicate equitable views. "Work-family conflict" and "family-work conflict" are indices based on **Netemeyer, Boles, and McMurrian [1996]**. Higher values indicate a higher sense of conflict. "Preference for inequality" is based on an item utilised in the European Social Survey. "Fulfilling life" and "work & career" are based on a ranking of life components adopted from a standardised opinion poll [Bożewicz et al., 2019]. A constant was included, but is not reported. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

Table C6. Heterogeneity of treatment effects across vignettes

	in USD per month		in percent	
	(1)	(2)	(3)	(4)
T: employer = 1	17.98^**** (7.57)	24.06^**** (9.89)	2.11^**** (0.89)	2.22 (1.45)
T: woman= 1	−8.63 (5.55)	8.62 (11.39)	−1.06 (0.65)	0.71 (1.32)
T: employer × woman	9.24 (10.33)	−8.91 (17.31)	1.19 (1.21)	−0.71 (2.02)
V: lawyer = 1	6.44 (14.89)	11.77 (17.92)	−0.11 (1.00)	0.05 (1.67)
... × T: employer = 1	−8.04 (17.24)	−11.89 (21.11)	−1.53 (1.30)	−1.20 (2.46)
... × T: woman= 1	2.99 (16.41)	−13.91 (21.13)	0.63 (1.13)	−1.35 (1.90)
... ×T: employer = 1 × T: woman= 1	20.67 (23.33)	41.50 (29.81)	0.63 (1.76)	3.46 (2.43)
V: salesperson = 1	−6.23 (5.26)	5.50 (9.91)	−0.46 (0.69)	0.18 (1.26)
... ×T: employer = 1	6.53 (9.87)	−6.79 (15.20)	3.77^**** (1.71)	3.44 (2.32)
... ×T: woman= 1	14.34^*** (8.55)	−11.67 (18.97)	2.57^** (1.38)	0.15 (2.18)
... ×T: employer = 1 × T: woman= 1	−17.78 (16.11)	14.10 (26.38)	−3.46 (2.76)	−0.55 (3.78)
Intercept	3.33 (11.63)	1.82 (6.78)	−0.20 (1.35)	0.60 (0.93)
δX_i	Yes	No	Yes	No
Observations	570	570	570	570
H0: vignettes x initiator	0.50	0.39	0.02	0.13
H0: vignettes x gender	0.23	0.68	0.16	0.70
H0: vignettes x interactions	0.22	0.44	0.30	0.34

Notes: The table presents the results of linear regressions of subjects' recommendations about changes in wages. These estimates expand on those portrayed by equation (1) by including interactions between treatment and vignettes. The estimates are obtained from the preferred sample. Standard errors clustered at subject level presented in parentheses. The bottom includes tests for whether interactions between vignettes and (a) initiator, (b) gender of the employee, and (c) initiator × gender of the employee are statistically different from zero. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

Table C7. Heterogeneity of treatment effects for gender, educational attainment and managerial experience

Dependent variable:	Wage change in USD per month				Wage change in % of status quo			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
T: employer	17.20*** (4.85)	17.24*** (4.67)	18.35*** (5.05)	14.41* (8.34)	2.95*** (0.56)	2.72*** (0.80)	2.69*** (0.86)	2.63*** (0.76)
× Never manager		−0.10 (8.86)				0.42 (1.09)		
× Completed tertiary			−2.11 (8.87)				0.47 (1.09)	
× Woman				5.30 (9.31)				0.60 (1.10)
T: woman	−3.02 (4.87)	1.72 (3.16)	−1.22 (2.55)	−6.84 (9.56)	0.15 (0.37)	0.50 (0.38)	0.25 (0.36)	−0.21 (0.68)
× Never manager		−9.04 (8.35)				−0.68 (0.69)		
× Completed tertiary			−3.27 (8.50)				−0.16 (0.68)	
× Woman				7.46 (9.59)				0.68 (0.74)
T: employer × T: woman	10.30* (6.17)	4.87 (7.91)	9.23 (5.72)	18.74* (10.26)	0.10 (0.62)	−0.04 (1.01)	0.13 (0.92)	0.52 (0.90)
× Never manager		10.16 (10.83)				0.29 (1.24)		
× Completed tertiary			1.89 (10.89)				−0.06 (1.23)	
× Woman				−16.41 (11.08)				−0.79 (1.21)
Never manager	5.73 (5.38)	7.84 (9.18)	5.72 (5.38)	5.70 (5.38)	0.73 (0.64)	0.79 (0.76)	0.73 (0.64)	0.73 (0.64)
Completed tertiary	10.21** (4.79)	10.22** (4.78)	12.47 (8.71)	10.10** (4.79)	1.08* (0.59)	1.10* (0.59)	0.94 (0.71)	1.08* (0.59)
Woman	−10.26** (4.89)	−10.30** (4.90)	−10.26** (4.88)	−12.59 (9.56)	−0.98* (0.58)	−0.99* (0.58)	−0.97* (0.58)	−1.43* (0.77)
Intercept	−0.44 (9.13)	−1.51 (9.42)	−1.56 (10.49)	0.46 (9.42)	−0.95 (1.18)	−0.99 (1.18)	−0.89 (1.28)	−0.71 (1.20)
δX_i	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
δ_v	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	570	570	570	570	570	570	570	570
R^2	0.10	0.10	0.10	0.10	0.13	0.13	0.13	0.13

Notes: The table presents the results of linear regressions of subjects' recommendations about changes in wages. These estimates expand on those portrayed by equation (1) by including interactions between treatment and three personal characteristics: managerial experience, whether a respondent completed university studies, and gender. The estimates are obtained from the preferred sample. Other characteristics and vignette fixed effects are included, but not reported. The results from the full sample are available upon request. Standard errors clustered at subject level presented in parentheses. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

Table C8. Beliefs about the social norm: all coefficients

Majority...	... in favour of a change		in favour of a raise	
	(1)	(1a)	(2)	(2a)
T: employer = 1	0.097** (0.042)	0.078** (0.031)	0.11*** (0.040)	0.079*** (0.028)
T: woman= 1	−0.043 (0.035)	0.0092 (0.026)	−0.040 (0.034)	−0.0098 (0.025)
T: employer = 1 × T: woman= 1	0.074 (0.052)	0.0060 (0.039)	0.082 (0.051)	0.026 (0.036)

Majority...	... in favour of a change		in favour of a raise	
	(1)	(1a)	(2)	(2a)
Vignette				
V: lawyer = 1	−0.081*** (0.030)	−0.090*** (0.023)	−0.076** (0.030)	−0.089*** (0.022)
V: salesperson = 1	−0.081** (0.032)	−0.094*** (0.023)	−0.050 (0.031)	−0.071*** (0.021)
Proposed change				
Negative change	0.55*** (0.17)	0.56*** (0.085)	0.024 (0.13)	−0.081* (0.049)
Positive change	0.57*** (0.055)	0.68*** (0.035)	0.59*** (0.055)	0.69*** (0.036)
Personal characteristics				
Age	−0.0019 (0.0024)	−0.0021 (0.0015)	−0.00051 (0.0022)	−0.00022 (0.0014)
Female subject	−0.026 (0.047)	−0.027 (0.034)	−0.019 (0.043)	−0.0064 (0.031)
Managerial experience	−0.011 (0.048)	−0.023 (0.033)	−0.020 (0.043)	−0.037 (0.030)
Educ: secondary	0.027 (0.087)	−0.047 (0.060)	0.029 (0.11)	−0.042 (0.062)
Educ: tertiary	0.044 (0.088)	−0.036 (0.062)	0.035 (0.10)	−0.026 (0.064)
Income: enough	0.025 (0.049)	0.00093 (0.034)	0.020 (0.044)	−0.00087 (0.031)
Income: not enough	0.071 (0.078)	0.059 (0.054)	0.036 (0.078)	0.015 (0.056)
Intercept	0.26* (0.15)	0.34*** (0.096)	0.17 (0.17)	0.22** (0.095)
Observations	570	963	570	963
P(y=1)	0.360	0.382	0.326	0.328

Notes: All specifications were estimated using linear probability models and include the full set of *X*'s. Columns (1) and (1a) report the estimates where the dependent variable takes on the value of 1 when the subject reports that the rest of society would change the wage of the character in the given vignette, and 0 otherwise. Finally, in columns (2) and (2a), the dependent variable takes on the value of 1 when the subject reports that the rest of society would raise the wage of the character in the given vignette, and 0 otherwise. Columns with only numbers denote specifications for the preferred sample. The letter *a* denotes specifications for the full sample. Standard errors clustered at the individual-level presented in parentheses. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1 respectively.

Source: Authors' own elaboration.

D. Quality of the sample

Table 11 presents two specifications for studying the determinants of failure in manipulation checks. In the first specification, we look at the probability of making at least one mistake in any manipulation check. This variable varies only between individuals. In the second column, we split mistakes by vignettes. Hence, the dependent variable is the probability of making a mistake in a manipulation check in vignette *v*. Overall, 60% of those who fail a manipulation check fail only once. Mistakes in identifying who requests discretion over TAs are less frequent than mistakes in identifying working conditions. In fact, 50% of the mistakes referred to specific working hours or weekdays in the status quo; 35% related to the initiator of the change, and in the remaining 15% of the cases subjects erred on both dimensions. The logit models reveal that passing all manipulation checks is not fully random. Better educated respondents – and, to an extent, those with higher earnings— were more likely to answer correctly. Additionally, participants made fewer mistakes in the lawyer vignette compared to the hairdresser vignetteThese differences might result from the order in which the vignettes were presented to the subjects: the hairdresser vignette was presented first, followed by the lawyer vignette. Some subjects may have learned what aspects are more relevant after the first manipulation

checks and begun to pay more attention to these outcomes. Supporting this interpretation, among participants who made a single error, it was about 1.5 times more likely to occur in the first vignette (hairdresser) than in subsequent ones

Table D1. Determinants of passing manipulation checks (marginal effects)

	By individual	By vignette
	(1)	(2)
Age	0.00 (0.00)	0.00 (0.00)
Woman	0.02 (0.06)	0.02 (0.03)
Never manager	0.02 (0.06)	0.04 (0.04)
Completed tertiary	0.17*** (0.06)	0.13*** (0.03)
Income: enough	0.02 (0.06)	0.06* (0.04)
Income: not enough	0.10	0.09
	(0.09)	(0.05)
T: employer = 1		0.03
		(0.03)
T: woman= 1		−0.01
		(0.03)
T: employer = 1 × T: woman=1		−0.01
		(0.05)
V: lawyer = 1		0.04
		(0.03)
V: salesperson = 1		0.03
		(0.03)
Observations	321	963
Pr (Y=1)	0.59	0.79

Notes: The table reports results from logit regressions where the dependent variable is the probability of making a mistake. "Less than tertiary education" is the base level for education, and "income sufficient to cover some luxury" is the base level for household income. The hairdresser vignette serves as the base level for vignette comparisons. Column (1) presents the restriction used in the preferred specifications, i.e. whether a subject erred in at least one manipulation check. In column (2), the dependent variable is making a mistake in a specific vignette. Standard errors in parentheses. In (1), we used robust standard errors, in (2) standard errors are clustered at the individual level. *** denotes $p<0.01$, ** denotes $p<0.05$, and * denotes $p<0.1$.

Source: Authors' own elaboration.