GRAPE Working Paper #71

Paying for ideal discretion: a framed field experiment on working time arrangements

Magdalena Smyk, Lucas van der Velde, Joanna Tyrowicz

FAME | GRAPE, 2022

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

Paying for ideal discretion: a framed field experiment on working time arrangements

Magdalena Smyk
SGH and
FAME|GRAPE

Joanna Tyrowicz
FAME|GRAPE, University
of Regensburg and IZA

Lucas van der Velde
SGH and FAME|GRAPE

Abstract

The notion of *ideal worker* necessitates being available at the discretion of the employer 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 the discretion to decide about working schedules. We embed our study in the context of gender and compare employee-initiated and employer-initiated request for a change towards more 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 penalty to employee-initiated request for autonomy to decide about working schedules. While our results lend support to the *ideal worker* model, they cast doubt on explanations linking gender wage inequality to labor market flexibility.

Keywords:

non-standard working arrangements, job satisfaction, gender

JEL Classification

Corresponding author

Joanna Tyrowicz, j.tyrowicz@grape.org.pl

Acknowledgements

The authors express gratitude for insightful comments and suggestions by Arthur Alderson, Arne L. Kalleberg, Julia McQuillan, Jekaterina Navicke, Erita Narhetali, Dina Okamoto, and Linas Tarasonis. Kazimierz Czarnocki and Kamil Sekut provided wonderful research assistance. Sam Butterick provided editorial assistance. Financial support by National Science Center grant no. \#2017/27/L/HS4/03219 is appreciated. The remaining errors are ours.

Published by: FAME | GRAPE
ISSN: 2544-2473

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

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

1 Introduction

We study the role of autonomy to decide about the working schedule in determining wages. The *ideal worker* ought to be available at employers' discretion (e.g. Breaugh, 1985; Grote and Raeder, 2009). Building on the theory of *ideal worker*, we theorize that a deviation from the norm of availability leads to wage penalties. The departures from the notion of the *ideal worker* have been extensively studied also in other contexts, involving job satisfaction (e.g. McNall et al., 2009), job turnover (e.g. Batt and Valcour, 2003), and career advancement (e.g. Guillaume and Pochic, 2009).[1] The business case for the *ideal worker* stems from the working process efficiency (Reilly, 1998; Kauffeld et al., 2004): employees who can adapt their working time to workplace conditions provide resources for firms to be in a better position against the competitors in the market. Note that becoming more congruent with the *ideal worker* does not need to imply wage increases.

Against the background of the *ideal worker*, the so-called Goldin conjecture (2014) states that women – as primary care-givers – are prevented from working in jobs that require supply of hours at employers' discretion and thus are paid less (see Goldin and Katz, 2016, for evidence from an occupation where technological innovation relaxed the working hours constraint and simultaneously women's employment shares and wages increased). In fact, using flexible working arrangements reflects constraints arising from household responsibilities rather than preferences as such (Cortes and Pan, 2019). Goldin argues that men receive a premium for being available to the employers, whereas women cannot do that to the same extent. In other words, wage gaps are explained away by a premium offered to *ideal worker*, a norm that women are less likely to comply with.

There is a clear gap between the implications of *ideal worker* on the one hand and Goldin conjecture on the other. There is ample evidence that men request flexibility less frequently than women (e.g. Vandello et al., 2013).[2] However, *ideal worker* postulates wage penalties for not being available at employers discretion and offers no clear predictions for workers who are congruent with this norm. Meanwhile, Goldin conjecture states that workers receive a premium for being available, where women are less frequently able to work at employers' discretion, limiting their ability to reap these premia. Furthermore, neither of the theories is clear on whether the penalty for requesting discretion over own working time differs across genders.

We construct an experiment to bridge the gap between these two theoretical approaches and to test their empirical relevance. Our experiment is designed to address three hypotheses.

Hypothesis 1 *Employee requesting more discretion over own time schedule is penalized 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 the ideal worker would have earned. It is also consistent

[1] This rich literature has been extensively discussed by Davies and Frink (2014), we refer the interested reader to this publication.

[2] Indeed, the *ideal worker* model is in conflict with the *family devotion* schema in the case of women, but not in the case of men (Blair-Loy et al., 2013; Brescoll et al., 2013). Motherhood is viewed as child-centered, emotionally absorbing, and labor-intensive (Hays, 1998; Bear, 2019). The clash between these two mental models is related to the belief that work demands and deserves single-minded focus and allegiance and so does caring (Moen et al., 2005; Feldman et al., 2020).

with the Goldin conjecture that the wage gaps between genders is effectively driven by the fact that women more frequently require (and thus request) discretion over own time schedule more frequently, rather than by gender-specific penalties.

Hypothesis 2 *Employer requesting more discretion over time schedule is expected to raise wages.*

This hypothesis is a clear representation of Goldin conjecture, absent gender context. This hypothesis is consistent also with the theory of *compensating wage differentials* (Smith, 1979), schedules at the discretion of the employer in terms of start and end times raise the disutility of work and thus should be compensated with higher wages (Mas and Pallais, 2017).

Hypothesis 3 *Women face smaller wage decline than men when requesting discretion over own time schedule.*

This hypothesis embeds gender context in the *ideal worker* model. When women request discretion to set own time schedule, this may be viewed as consistent with the *family devotion* schema. In such case, employers may acknowledge the employees' need for more discretion and choose not to penalize the worker (Walby and Olsen, 2002; Cousins and Tang, 2004; Warren et al., 2001). Conversely, when men request discretion over own time schedule, the departure from the *ideal worker* cannot be easily justified, thus penalties are expected to emerge. Recently, Luhr (2020) emphasizes that parenting as a reason for greater ability to set working schedules is relevant for both genders, but women are still stereotypical primary care-givers.

In line with these hypotheses, we provide a framed field experiment to empirically verify the extent to which discretion over working time should be reflected in wages. We operationalize discretion over time schedule as ability to set starting and ending times. In status quo, workers have fixed working time arrangements. Across scenarios, discretion to set starting and ending times falls either to the employer or to the employee. We pitch the employer-initiated against the employee-initiated flexible working time arrangements (WTAs).[3] We purposefully manipulate the gender of employee in order to identify the causal effects of this context on the wage premia and penalties related to WTAs. From an *ideal worker* perspective, the employer is not obliged to reward the availability of employees. Consequently, when the employer wants to increase the scope of discretionary WTAs, no change in wages has to follow. Analogously, if the worker asks for discretion over WTAs, (s)he is moving away from the *ideal worker* model and thus may be penalized as incongruous with the norm (Munsch, 2016; Rudman and Mescher, 2013; Vandello et al., 2013).

Our paper is structured as follows. In section 2 we present the relevant theory. In section 3 we discuss the details of the experimental design and the hypotheses. 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 as well as policy implications of our study.

[3] Our focus in this paper is placed on working time arrangements, we also consider both perspectives on the flexibility (employer's and employee's). We thus rely on the abbreviation WTA. Note that flexible working arrangements may involve a variety of dimensions, including working from home or other forms of tele-work. 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 complexity of flexible work arrangements).

2 Theoretical foundations

For more than a half of a century, the notion of an *ideal worker* has involved commitment, stable career trajectory, permanent and unconstrained availability to the employer (Davies and Frink, 2014). The *ideal worker* is a benchmark for actual workers in terms of relative wages. Workers who become closer to the ideal should expect their compensation to rise whereas the workers who depart from this ideal should observe lower wages. The *ideal worker* may expected to exhibit creativity in how tasks are done, but is not free to determine own schedule (Breaugh, 1985; Gagne and Deci, 2005; Mazmanian et al., 2013; Grote and Raeder, 2009).

This instrumental perspective on workers ignores the individualism and intrinsic motivations. The employees' perspective builds on the work-life balance and on the ability to fulfill various roles in life (see for example Barnett and Hyde, 2001; Byron, 2005; Michel et al., 2011, for the theoretical and empirical treatment of the expansionist theory, derived from social roles theory). Worker autonomy is believed to foster workers' intrinsic motivation and thus their commitment and productivity (Spector, 1986; Heavey et al., 2013; Rubenstein et al., 2018). Flexible working arrangements are rationalized as instruments for improving workers' work-life balance, while also being beneficial to the employers (Fagan et al., 2012). Vast empirical literature suggests that workplaces providing flexible working arrangements outpace the competition (see Dex and Scheibl, 2001; Batt and Valcour, 2003; Beauregard and Henry, 2009; De Menezes and Kelliher, 2011; Azar et al., 2018, to name just a few). There is also a growing body of evidence that flexible start and finish time may substantially improve productivity (Boltz et al., 2020; Angelici and Profeta, 2020).

Although many empirical studies and real world cases demonstrate that WTAs may be neutral or even beneficial to performance (see Azar et al., 2018, for an extensive meta-analysis), workers requesting discretion over WTAs are viewed by their managers and colleagues as less committed (Williams, 2001; Chung, 2020). WTAs are treated as negotiated perks available for top performers rather than universal right (Kossek et al., 1999; Kelly and Kalev, 2006). In a light of *ideal worker* norm, access to flexible WTAs is restricted, and thus unequal, even if providing voluntary and self-controlled flexibility may improve workers well-being and reduce turnover (Kelly et al., 2011; Moen et al., 2011; Kaduk et al., 2019). Moreover, coworkers report reduced job satisfaction when their colleagues benefit from flexible WTAs, arguing that their workload was raised without adequate compensation (Munsch, 2016), also for female coworkers (Teasdale, 2013). This negative perception of WTAs lends support to the *ideal worker* as the mental model used by the workers (especially the managers).

Our paper is perhaps closest to two studies attempting to identify the actual "value" of the WTAs in an experimental context. Mas and Pallais (2017) construct a discrete choice experiment with real consequences, allowing the individuals to choose between inconvenient working hours on a short notice and regular and scheduled working hours, varying the hourly rate accordingly. They find that their subjects have a strong preference against inconvenient working hours. He et al. (2021) deploy a field experiment with job offers and conclude that workers value job flexibility, because they apply more intensively to job offers with flexible working arrangements (holding the salary constant).

Compared to these two studies, we provide several innovations. First, He et al. (2021) study

whether individuals apply for flexible jobs more willingly than for jobs without hourly flexibility, but not what is the actual monetary value of getting those jobs. We address this gap by eliciting the monetary value of flexibility. Mas and Pallais (2017) deploy a discrete choice experiment, which naturally reduces the granularity of measuring the monetary value of flexibility. They also offer a choice between regular and highly inconvenient working hours. We rectify these two issues with a design which allows the individuals to finely adjust wages to changes in working conditions, and with offering minor departures from working time flexibility. Naturally, unlike controlled randomized trials, there are not real consequences to choices made in the vignette experiment. However, we do introduce a real consequence component in the experiment, in order to study if and to what extent individual preferences for working time flexibility affect the choices of the respondents. Another important difference is that Mas and Pallais (2017) ask *job entrants* to choose a bundle of working hours and wages; we ask to value changes for *incumbents*, where status quo bias may exhibit in decisions of experiment participants.

3 Experimental design

We conduct our experiment in Poland, the context is described in Appendix A. Participants were compensated: a flat rate of \$0.5 for participation and up to \$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 recovering respondents views on work-life conflict, gender norms and inequality. The third module is a real-consequence task that serves to recover 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. The experiment was implemented between April 23rd (pre-test)[4] and August 9th 2021 (the actual implementation). To mitigate the risk that some participants of convenience samples are less careful in filling in questionnaires (Cheung et al., 2017; Sharpe Wessling et al., 2017), we included a series of manipulation checks (Porter et al., 2019). Each individual responded to nine manipulation check questions: three for each vignette.

3.1 WTA experiment

We propose a mixed design, combining a 2x2 framed field experiment with a vignette.[5] 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

[4]Before the survey was rolled out on full sample, we have tested the technical features of our survey as well as legibility of all the questions using a pre-test with the help of 40 subject (20 men and 20 women). The pre-test yielded important insights on formulating specific questions and 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.

[5]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.

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 of 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 characterized 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 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 employer or the employee want to change the WTA. Under the new time arrangement, the start and the end times will vary.

Randomization We assign subjects across conditions randomly. With two initiators, two genders and three vignettes, the pool of potential combinations 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 change in WTAs, then the second vignette refers to employee-initiated change in WTAs. Analogously, for the GENDER condition, if in the first vignette the randomized character is a man, then the second vignette portrays a woman. For the third vignette, the algorithm randomized among those GENDER $\times$ 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 sixteen 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 the participants selected one of these three options, 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 50 PLN per month as an interval (approx. \$13).[6] Overall, respondents could choose from 120 categories, of which 60 indicated wage growth and 60 a decline.

In addition to the wage adjustments, we ask participants to disclose their beliefs about the social

[6] All conversion to dollars were made using the exchange rate listed by the National Bank of Poland on the first day of the experiment (April 23rd 2021), \$1 = 3.7855 PLN.

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

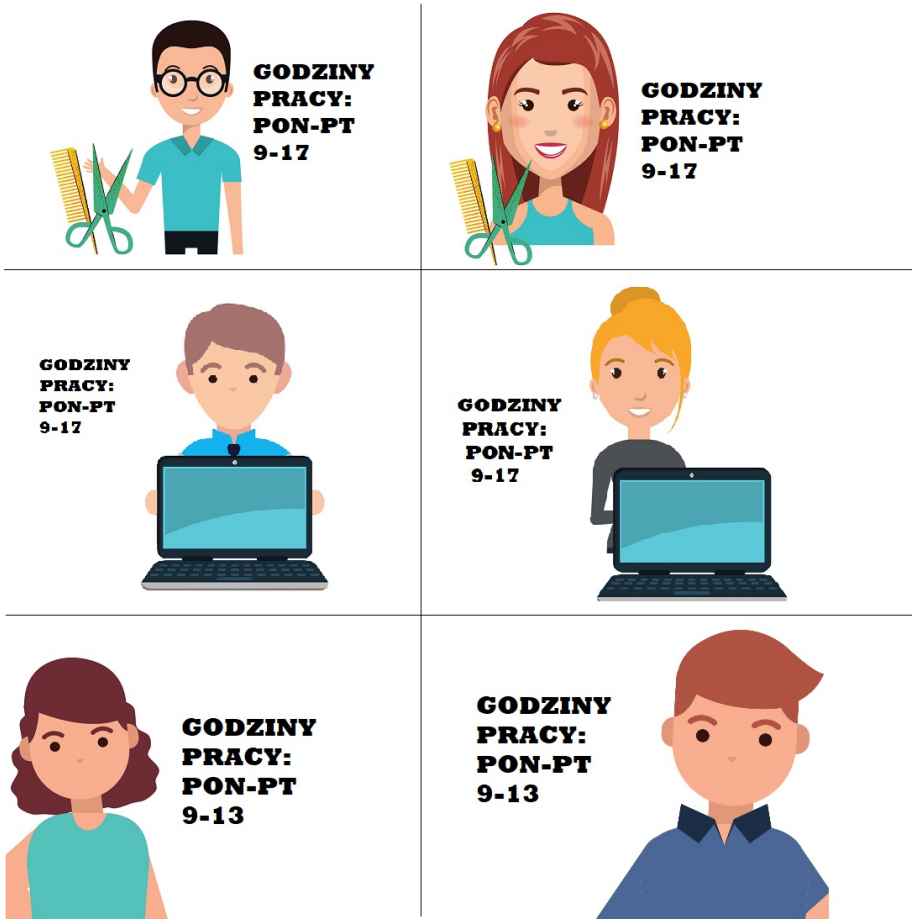
Scenarios The goal of the study is to reveal social norms rather than elicit individual preferences towards discretion in setting working times. To this end, we follow Aguinis and Bradley (2014) and construct third-person vignettes. This allowed us to provide variation in characteristics crucial from 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 et al., 2015, who shows high external validity of such design).

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 Mondary to Friday on a half-time basis (from 9AM to 1PM). Those three occupations were chosen, because all of them are equally likely to employ workers of both genders, but they each require a different skill level and have different social standing. Each of the three occupations 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: 1600 PLN per month (approx. $420) for the shop assistant, 3200 PLN per month (approx. $840) for the hairdresser, and 6400 PLN per month (approx. $1680) 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 in 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 public administration, but on the other hand it may be weakened by the need to meet with the 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, due to the fact that stores are typically open longer than 4 hours on weekdays. It is thus customary to expect rotating the 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.[7] The GENDER treatment is reinforced in text, as characters have distinct male and female names. We incetivized subjects to memorize 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 one dollar. The subjects were not informed about the outcomes of the manipulation checks until 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

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

Figure 1: Scenarios



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

change in discretion over WTAs.

3.2 Social norms

Upon completing the experiment, participants were directed towards a short questionnaire where they answered to what extent they agreed with different statements. These statements were taken from different sources, and aimed at highlighting 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 et al. (1996). Given high values of Cronbach's α, we group the ten items into two variables, where higher values indicate more conflict. The items measuring adherence to traditional gender norms were taken from European Value Survey. As before, Cronbach's α suggested grouping these items into a single variable, where higher values show more adherence. Finally, we recovered views on income inequality using an item

from ISSP. We recode answers such that the resulting variable takes value of one when respondents agree that income differences are required to compensate effort.

3.3 Valuing own time availability

Participants could easily confuse social norm with individual preferences. To be able to adjust estimates of social norms for own preferences of the participants, we include a task where participants reveal the value of own time availability. After completing the vignette questions, the participants were faced with a discrete choice experiment with real consequence. Subjects were informed that in order to complete the survey they will have to wait either 5 minutes or 30 minutes and that this wait time will be randomly chosen. Participants were told that they may choose between various options of the length of window to complete the survey *after* the wait 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 wait times, participants faced the same menu of options: 24 hours to complete the survey after wait time compensated with additional \$0.25 (full discretion), \$0.5 for completing the survey within 25 minutes (less discretion) to \$1.5 for completing the survey within an interval of 5 minutes (the least discretion). Since the subjects made two choices (at 5 and at 30 minutes wait time) we can estimate individual valuation of own time availability. Subjects were informed that if they do not comply with their choice, they will loose money gained in previous rounds and fall back to \$0.5 for participation. Subjects were not informed if the last part of the experiment is long or short. After subject selected their preferred option for each wait time, wait times were randomly assigned.

3.4 Final survey items

After the wait time elapsed, a new screen was visible 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 additional compensation gained in previous modules. All participants complied.

There were two types of questions in the last module. The first type of questions concerned individual and household characteristics: age, gender, education level, managerial experience, and economic situation[8]. The second type of questions sought to identify the subjective importance towards and away from work. 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 Bozewicz et al. (2019).

[8]Given how sensitive questions on income might be, we adopted a question used routinely in household budget surveys; respondents indicate whether their household income is sufficient to make ends meet on a four-level scale (from insufficient to allowing some luxury).

4 The sample

In total, we observe judgment on 963 vignettes from 321 participants. Among them, roughly 60% answered correctly all nine manipulation checks, that is 570 vignettes from 190 respondents. This 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 tertiary education, and a similar percent claims that they never had any managerial experience. Table 1 displays the descriptive statistics for all treatment conditions jointly and across treatment assignments. Given that the four groups do not differ in individual characteristics, we conclude that the randomization across treatments was successful, despite unusual two-step assignment into conditions.

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

	Full sample	Preferred sample				
	Total	Total	Initiator		Gender of employee	
			Employee	Employer	Men	Women
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 for the individuals participating in the experiment. Note that participants reported these characteristics *after* completing the experiment.

Table 1 allows 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 C8 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 enrollment of roughly 55% in the past 20 years in Poland, accompanied by a declining share of high school dropouts. Questions about subjective income, stylized 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 two sets of the estimates. The first one zooms in on the approx. 60% of the collected sample, where individuals replied correctly to all manipulation check questions. This is our preferred sample. The second set of estimates is analogous in terms of methods, but it is based on the entire collected sample. We consider the former our preferred specification, because we have more confidence that the subjects in the study fully understood the trade-offs that they were judging.

5 Results

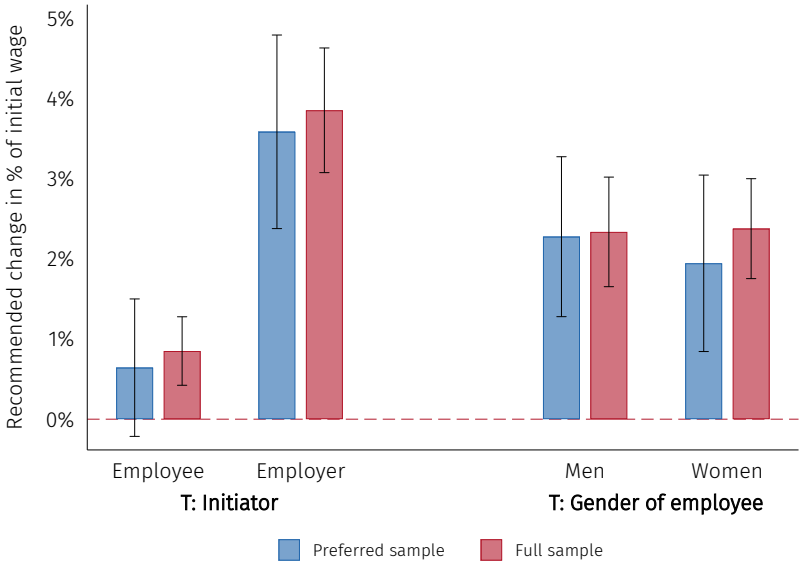
The main interest of this paper is the recommended wage adjustment for a change from fixed to discretionary WTAs across treatments. We discuss the results in four substantive parts. First, we present a comparison of means across treatment conditions, as an unbiased measure of the unconditional average treatment effects. Then, we include confounding variables (deploying a regression model) to provide measures of conditional average treatment effects. We further study heterogeneity of treatment effects. The analysis is complemented by a study of perceived compliance with the social norm: after providing own recommendations, the subjects responded to a number of questions concerning their belief about the overall views in the population. We conclude this section with a discussion of our results vis-a-vis the literature.

5.1 Average treatment effects

Roughly 26% of subjects recommend that wages should increase to accommodate the proposed change in WTAs (from fixed to discretionary), whereas 4% of respondents stated that they should decline. Figure 2 shows the recommended changes in wages, expressed as a fraction of the *status quo* wage (unconditional average treatment effects). Blue bars show the estimates from the preferred sample, while red bars correspond to the full sample. When the employer initiates change to WTAs, the average recommended wage change amounts to 3.5% and this number is statistically higher than roughly 0.8% increase when the employee initiates the change. This 3% recommended wage rise amount to USD 28 per month, which is hardly relevant in wage bargaining. Whenever the employee wants to modify the wage arrangements, the recommended change in wages is much lower, close to \$6 and is not statistically different from zero in the preferred sample. We find no differences across gender treatments. The bars for gender treatments mechanically average across initiator treatments. If recommendations for wage change for male and female characters differed, the bars would not coincide with the average over initiator treatments. The bars show that both male and female characters received on average a 2% wage rise recommendation.

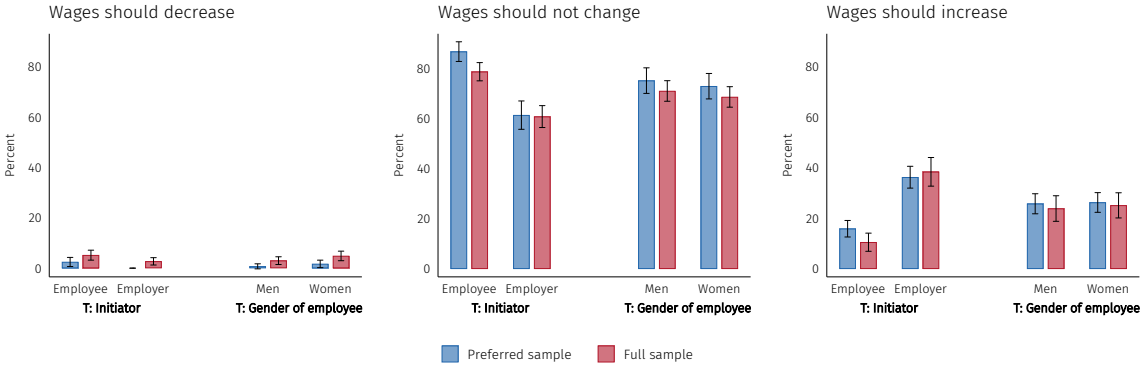
In Figure 3, we delve deeper into the analysis by studying the proportions of recommended wage cuts, wage raises and recommendation to keep wages at *status quo*. No change in wages is the modal recommendation across all the conditions. The proportion recommending no change is considerably higher in the employee treatment, where above 80% of workers opted for this option. When the employer initiates, this proportion drops to 60%.

Figure 2: Treatment effects: what is the 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 B1 in the Appendices.

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 B1 in the Appendices.

We interpret findings presented in Figures 2 and 3 as a confirmation of the *ideal worker* model. Nobody in our preferred sample recommended a wages decrease when the employer requests discretion over WTA, whereas the probability of recommending a decline is of 2.5 pp. when the employee is the one requesting. In the full sample, the frequency of recommending a fall in wages is almost twice as high when the employee requests discretion over WTAs than when the employer wants to renegotiate. In other words, when the employee moves away from the notion of an *ideal worker*, subjects were twice as likely to recommend a reduction in wages (5% vs 3% of the subjects).

However, the *ideal worker* model explains a small fraction of the drop in the proportion of

participants recommending no wage change across initiator treatments. The key contribution comes from subjects who recommend wage increases: only 16% of answers recommended a wage increase when the employee initiates the change in WTA, whereas 36% of answers do so in the employer treatment. Recall that under the *ideal worker* model employees should be generally available for work. This model provides little guidance for understanding wage changes following changes in WTA requested by the employer. Goldin conjecture states that being available at employers' discretion receives a premium, which is reflected in our experimental findings.

In contrast to the initiator, the treatment effects corresponding to gender of the employee are not statistically different from each other. Neither do we find statistically significant differences for the fraction of respondents who recommend wage reduction, increase or change based on the gender of the employee. These results resonate well with the *ideal worker* model which ignores the gender dimension. 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 conjecture that availability is rewarded by premium. On the other hand, however, the premia are not large enough to explain away 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 premia differ across genders.

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

Denote by $y_{i,v}$ the wage change recommended by the 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:

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

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 variables adjusting for individual characteristics. These variables include age, gender, education, household income, managerial experience, and several controls for individual preferences.[9] One should bear in mind that β_I represents the *difference* between employers and employees requested discretion over WTAs, likewise for β_F in the context of gender treatment. 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} informs about the additional effect of female character on vignettes describing change in WTAs initiated by the

[9] Note that in principle we could estimate a model with individual fixed effects:

$$y_{i,v} = \beta + \beta_I T : I_{i,v} + \beta_F T : F_{i,v} + \beta_{inter} T : I_{i,v} \times T : F_{i,v} + \delta_v + \delta_i + e_{i,v}.$$

Our preferred approach is to pursue with a model adjusting for individual characteristics and preferences/norms, because it allows to fully exploit the within-subject and between-subject variation in our data. With individual fixed effects, the estimators of our treatment effects exploit only the within-subject variation.

employer.

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

We estimate equation (1) using a linear regression. Our outcome measure of 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 for the analysis. In practice, less than one percent of the answers corresponded to the largest wage changes (in either direction), so censored regression models should produce similar coefficients, while putting additional strain on the efficiency of the estimators.

Table 2: Recommended change in wage subsequent the change in WTAs: experimental results

	Preferred sample				Full sample			
	in USD per month		in %		in USD per month		in %	
	(1)	(2)	(3)	(4)	(1a)	(2a)	(3a)	(4a)
T: employer = 1	17.20***	18.05***	2.95***	2.87***	22.04***	25.35***	3.13***	3.40***
	(4.85)	(6.54)	(0.56)	(0.69)	(6.80)	(8.16)	(0.93)	(1.00)
T: woman= 1	-3.02	-0.11	0.15	0.44	-4.00	-2.16	0.04	0.35
	(4.87)	(5.30)	(0.37)	(0.54)	(6.59)	(8.60)	(0.82)	(1.05)
T: employer × woman	10.30*	8.96	0.10	-0.02	0.28	-3.02	-0.34	-0.75
	(6.17)	(7.43)	(0.62)	(0.93)	(13.50)	(19.74)	(1.45)	(1.98)
δX_i	Yes	No	Yes	No	Yes	No	Yes	No
δ_v	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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: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 $ per month, whereas in Columns (3) and (4) the dependent variable is 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 B3 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.

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 $17-25 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 characteristics of the participants, and ot the the so-called multiple-hypothesis-testing bias (see Table B2 in the Appendices.) Table B3 in the Appendix reports the full set of coefficients. We do not find evidence for the role of age, income, nor managerial experience. In the preferred subsample, we find that women recommend, on average, lower wage rises than men: roughly $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 approx 3 percent, but note that in this vignette the status quo wage was USD 420 per month, so even a few dollars more already amounts to a relatively high fraction. One could rationalize this outcome as evidence that subjects value time discretion at a fixed price, independent

of income levels.

Tables B4 and B5 (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 the 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 $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.

5.3 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 4. The gray dashed line shows treatment effects from the full sample, while the shadowed area marks the 90% confidence intervals. Figure 4 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.[10]

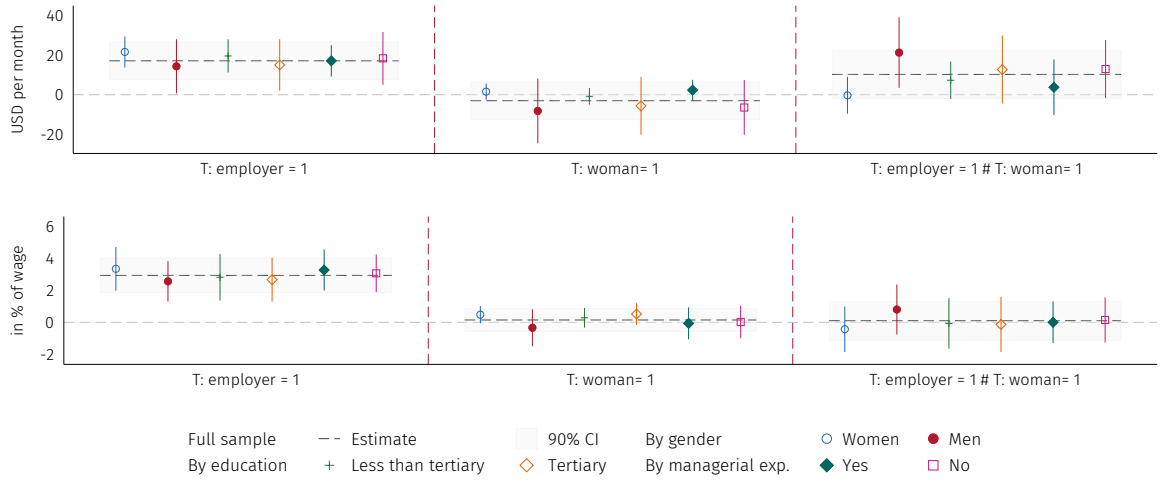
Managerial experience refers to whether a participant in the sample had ability to set wages and negotiate working arrangements. *A priori*, people with managerial experience could identify more with the employer and display greater adherence to the *ideal worker* model. The fact that those who have both experiences evaluate the vignettes no different than those who only had employees perspective suggests that the *ideal worker* perspective is deeply ingrained but also small. Likewise, labor market experience of university graduates and those did not continue education above secondary level is likely to differ, but those differences were not reflected in their responses to vignettes. Finally, due to gender roles and Goldin conjecture, men and women could have been differently exposed to the *ideal worker* norm. Given our results this seems unlikely. The remarkable stability of the estimated treatment effects suggests that the *ideal worker* notion is deeply embedded in the way that our subjects think.

5.4 Social norm on the ideal worker and Goldin conjecture

Our design permits to study the congruence between own recommendations of the respondents and their beliefs about the social norm. Subsequent to each vignette, we asked the subjects to evaluate if

[10] Table B6 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.

Figure 4: Treatment effects: heterogeneity



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.

their recommendation on wage changes is consistent with what the majority of Poles would answer. Responses to this question are indicative of expectations concerning wage changes following changes 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 relatively 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%).

We construct two outcome variables. The first one is a dummy which takes on the value of 1, when the subjects report that the overall norm is to change wage, in any direction, subsequent to a changes in WTAs. We call this variable *Majority in favor of a change*. The second dummy variable takes on the value of one if the subjects believe that the overall norm is to increase the wage and zero otherwise. We call this variable *Majority in favor 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.

Estimates in Table 3 corroborate our inference from Table 2. Indeed, subjects are more likely to expect wages changes when the employer requests discretion over WTA: the proportion of workers who believe that the norm (majority opinion) is in favor of a wage change is between 8 and 10 percentage points higher, when compared to employee initiated changes to working schedules. The results for wage raises are similar. The results hold even after adjusting for their own wage recommendations. The coefficients on own recommendations, presented at the bottom of Table 3 show the extent of agreement between own recommendations and their beliefs of the majority rule. 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 favor of a change		... in favor of a raise	
	(1)	(1a)	(2)	(2a)
T: employer = 1	0.097**	0.078**	0.11***	0.079***
	(0.042)	(0.031)	(0.040)	(0.028)
T: woman= 1	-0.043	0.0092	-0.040	-0.0098
	(0.035)	(0.026)	(0.034)	(0.025)
T: employer = 1 × T: woman= 1	0.074	0.0060	0.082	0.026
	(0.052)	(0.039)	(0.051)	(0.036)
Own recommendation:				
Decrease wages	0.55***	0.56***	0.024	-0.081*
	(0.17)	(0.085)	(0.13)	(0.049)
Increase wages	0.57***	0.68***	0.59***	0.69***
	(0.055)	(0.035)	(0.055)	(0.036)
δX_i	Yes	Yes	Yes	Yes
δ_v	Yes	Yes	Yes	Yes
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 (See Table B7 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 takes on the value of 1 when the subject reports that the rest of the 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.

5.5 Discussion of the results

Our experiment lends support to the *ideal worker* model: when the employer requests discretion over working time arrangements from the workers, most survey respondents recommend no change in wages. When the employee requests discretion, the share of respondents recommending a wage cut doubles, though respondents recommending no change in wages are still a majority. Goldin conjecture finds partial support in our experiment: respondents recommended wage increases when the employers request discretion, but the premia are small. The average change in wages associated with giving employer the right to set start and end times freely is statistically significant, but economically minor: approximately 3 percent of the status quo wages. Moreover, there is no gender dimension to this wage premia. We find that the fraction of individuals recommending wage penalties to employees requesting discretion to set their work schedules is very low.

Our approach is similar to that taken by Gimenez-Nadal et al. (2021) using observational data from time use surveys for the United States. Our estimates are in line with the existing literature: Mas and Pallais (2017) report larger effects, but the responses were censored at 3% of hourly wages. Indeed, among those who recommend wage change, half of the recommended raises falls short of $50 per month. A vast majority of wage cuts falls short of $30 per month. The estimate of the average treatment effects as reported in Tables 2 and B4 signify small overall wage adjustments recommended by participants, rather than a combination of polarized large and small answers. In other words, even those respondents who recommend a wage change following a change in WTA select small increments.

It could be that we find small effects because our respondents did not expect a major disruption

to the lives of the vignette characters and their families. With reference to the gender dimension and Hypothesis 3, the recommended change in wages could be larger if we expanded on the reasons for which the employee could suffer due to increased employer's discretion or benefit from increased own discretion over start and end time, e.g. emphasize caring obligations, leisure, or self-improvement in the vignettes (Vandello et al., 2013). While potentially this approach could boost the gender treatment effect, we would not identify the effects of gender *per se*, but rather a confounding of gender and other factors. Meanwhile, Goldin (2014) hypothesis puts gender and discretion in working time arrangements in stark contradiction: "certain features of occupations that create time demands and reduce the degree of substitution across workers are associated with larger gender earnings gaps" (p. 1117). Our results suggest that the subjects do not automatically invoke caring obligations when the gender of the character in vignettes was a woman. So, it is possible that the unavailability of women to work certain hours puts them at a disadvantage. However, even if this is the case, our experiment shows that this channel would explain only a small part of the prevailing gender wage gap. Our estimates indicated changes of around 3% of mean wages, whereas the adjusted gender wage gap in Poland hovers at around 20% of men's wages (Goraus et al., 2017).

The small size of the effect is also driven by the large number of participants who recommended no change in wages, which could potentially reflect*status quo bias* (Samuelson and 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 a right to sue in car insurance) with 58-60% (Hartman et al., 1990, on electric power consumption preferences) to 75-90% (Jianbiao et al., 2009, on investors' decision making). In our experiment, the cost to participants to recommend no change in wages and a raise or a cut were similar. Moreover, one would not expect respondents to act differently across treatments. Finally, participants who recommended a wage change, typically selected small values. This suggests that change in discretion over WTA was not perceived as a significant amenity (employee initiator) nor inconvenience (employer initiator).

6 Conclusions

The *ideal worker* model provides a coherent framework on penalizing insufficient availability for work. It is not obvious that being flexibly available to the employer should be rewarded, but being unavailable ought to be penalized in wages. The *Goldin conjecture* states implicitly that being available to the employers is rewarded in wages. Further, she argues that women are less frequently able to agree to working time arrangements at the discretion of employers, and that this difference stands behind observed differences in wages between men and women. In other words, Goldin conjecture states employers reward the availability of the employees and *ideal worker* model postulates that employers penalize lack thereof. Goldin further argues that this wage premia (penalties) explain gender wage gaps. Social norms literature further hints that inability to work at employers' discretion may be less penalized for women than form man.

The objective of our experiment was to elicit the pecuniary value and social acceptance for the type of working time arrangements when discretion to set working times is non-standard. We start

from a status quo of regular 9-to-5 arrangement (9-to-1 for part-time) and present the participants with vignettes which portray credible departures from this arrangement. Through the lenses of *ideal worker* model, it seems that employees requesting discretion to set start and end times of their workers should be penalized. Analogously, when discretion is requested by the employer, there ought to be lower social acceptance.

Our results lend support to the *ideal worker* model. For majority of our subjects, employers requesting discretion over WTA necessitate no change in wages. This is a modal answer. We also show that employees requesting discretion over their working time arrangements are twice as likely be recommended a wage cut than when employer initiates the change, though only a small fraction of the participants recommend any wage change. Our results lend partial support to the mechanism of the Goldin conjecture as well. Specifically, when the employer requests more discretion over start and working time, higher fraction of participants recommended wage increase than when the employee initiated it.

However, our results do not support the implications of Goldin conjecture. The recommended wage changes are rare. If participants recommend wage changes at all, they selected small values, akin to 3% of status quo wages or under $ 30 per month. These treatment effects are an order of magnitude smaller than gender wage gaps in the society, where our experiment was implemented. In other words, availability premium (alternatively: unavailability penalty) are too rare and too small to explain away gender wage inequality. Finally, we find no evidence of gender-specific treatment effects.

Our study provides a novel insights on how working time arrangements are socially perceived in the context of the *ideal worker* model. We interpret that according to experiment participants, wage should be related to productivity rather than other aspects of work. The discretion over setting start and end times does not appear to be relevant for wage recommendations. The *ideal worker* model appears to be default in a sense that participants did not think requesting discretion by the employers is universally an inconvenience that ought to be compensated. When the employee requests discretion, participants were even more often recommending no change in wages.

The original concept of *ideal worker* does not account for the remuneration angle, nor workers' productivity. Our experiment hints that more theoretical work is needed, because these dimensions appear to be of relevance for the social perception of flexible working time arrangements. The novel conceptualization of the *ideal worker*, in addition to compensation and its relation to perceived workers' productivity, do not appear to be strongly gender-specific. Non-standard working time arrangements are prevalent, and are likely to become increasingly relevant across various segments of the labor market. Leveraging the *ideal worker* model to theorize deeper about the social role of such arrangements would provide insights into the potential role of such legal initiatives as the *Working Time Directive* in Europe, which specifies the conditions under which the workers may be expected to answer the phone calls or emails. Likewise, more theorizing is needed into the whether or not the perceived productivity is a valid context of flexibility.

As to the gender dimension, our study corroborates that gender wage gaps cannot be explained away by differences in flexible working time arrangements between men and women. We provide novel, causal context for that. It appears that the hopes raised by Goldin conjecture that a final explanation

for gender wage gaps was identified require further theorizing and empirical research.

As to the external validity, this study prioritizes causality through a randomized experiment over representativeness. While we could not work with nationally representative samples in this design, we document that the effect sizes are universally small. We find no heterogeneity of the treatment effects. Hence, our study reveals no reasons to believe that in representative samples the results would be any different. However, this tentative inference needs to be treated with caution and cannot be automatically extrapolated to other countries due to potential differences on their norms regarding work and gender.

References

Aguinis, H. and Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational research methods*, 17(4):351–371.

Angelici, M. and Profeta, P. (2020). Smart-working: Work flexibility without constraints. CESifo Working Papers 8165, CESifo.

Azar, S., Khan, A., and Van Eerde, W. (2018). Modelling linkages between flexible work arrangements' use and organizational outcomes. *Journal of Business Research*, 91:134–143.

Barnett, R. C. and Hyde, J. S. (2001). Women, men, work, and family: An expansionist theory. *American Psychologist*, 56(10):781.

Batt, R. and 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.

Bear, J. B. (2019). The caregiving ambition framework. *Academy of Management Review*, 44(1):99–125.

Beauregard, T. A. and Henry, L. C. (2009). Making the link between work-life balance practices and organizational performance. *Human Resource Management Review*, 19(1):9–22.

Blair-Loy, M., Williams, J. C., and Berdahl, J. L. (2013). Cultural schemas, social class, and the flexibility stigma: Cultural schemas and social class. *Journal of Social Issues*, 69(2):209–234.

Boltz, M., Cockx, B., Diaz, A. M., and Salas, L. (2020). How does working-time flexibility affect workers' productivity in a routine job? Evidence from a field experiment. IZA Discussion Papers 13825, IZA.

Bozewicz, M., Boguszewski, R., Feliksiak, M., and Omyla-Rudzka, M. (2019). Wspolczesna polska rodzina [*Modern Polish family*]. Opinie i Diagnozy 44, CBOS.

Breaugh, J. A. (1985). The measurement of work autonomy. *Human Relations*, 38(6):551–570.

Brescoll, V. L., Glass, J., and Sedlovskaya, A. (2013). Ask and ye shall receive? the dynamics of employer-provided flexible work options and the need for public policy. *Journal of Social Issues*, 69(2):367–388.

Byron, K. (2005). A meta-analytic review of work–family conflict and its antecedents. *Journal of Vocational Behavior*, 67(2):169–198.

Cheung, J. H., Burns, D. K., Sinclair, R. R., and Sliter, M. (2017). Amazon mechanical turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, 32(4):347–361.

Chung, H. (2020). Gender, flexibility stigma and the perceived negative consequences of flexible working in the UK. *Social Indicators Research*, 151(2):521–545.

Cortes, P. and 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.

Cousins, C. R. and 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.

Davies, A. R. and 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.

De Menezes, L. M. and 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.

Dex, S. and Scheibl, F. (2001). Flexible and family-friendly working arrangements in UK-based SMEs: Business cases. *British Journal of Industrial Relations*, 39(3):411–431.

Fagan, C., Lyonette, C., Smith, M., and 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 32, International Labour Organization.

Feldman, E., Reid, E., and Mazmanian, M. (2020). Signs of our time: Time-use as dedication, performance, identity, and power in contemporary workplaces. *Academy of Management Annals*, forthcomming.

Gagne, M. and Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4):331–362.

Gimenez-Nadal, J. I., Molina, J. A., Sevilla, A., et al. (2021). Temporal flexibility, breaks at work, and the motherhood wage gap. IZA Discussion Paper 14578, Institute of Labor Economics (IZA).

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

Goldin, C. and Katz, L. F. (2016). A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, 34(3):705–746.

Goraus, K. and Tyrowicz, J. (2014). Gender wage gap in Poland - can it be explained by differences in observable characteristics? *Ekonomia*, 36:125–148.

Goraus, K., Tyrowicz, J., and van der Velde, L. (2017). Which gender wage gap estimates to trust? a comparative analysis. *Review of Income and Wealth*, 63(1):118–146.

Goraus, K., Tyrowicz, J., and van der Velde, L. A. (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.

Grote, G. and Raeder, S. (2009). Careers and identity in flexible working: Do flexible identities fare better? *Human relations*, 62(2):219–244.

Guillaume, C. and Pochic, S. (2009). What would you sacrifice? access to top management and the work–life balance. *Gender, Work & Organization*, 16(1):14–36.

Hainmueller, J., Hangartner, D., and 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. and List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4):1009–1055.

Hartman, R. S., Donae, M. J., and Woo, C.-K. (1990). Status quo bias in the measurement of value of service. *Resources and Energy*, 12(2):197–214.

Hays, S. (1998). *The cultural contradictions of motherhood*. Yale University Press.

He, H., Neumark, D., and Weng, Q. (2021). Do workers value flexible jobs? a field experiment. *Journal of Labor Economics*, 39(3):709–738.

Heavey, A. L., Holwerda, J. A., and Hausknecht, J. P. (2013). Causes and consequences of collective turnover: A meta-analytic review. *Journal of Applied Psychology*, 98(3):412.

Jianbiao, L., Guangqian, R., Qiuhua, M., and 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., and 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., and Moen, P. (2019). Involuntary vs. voluntary flexible work: Insights for scholars and stakeholders. *Community, Work & Family*, 22(4):412–442.

Kauffeld, S., Jonas, E., and 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.

Kelly, E. L. and 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, P., and 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., Barber, A., and Winters, D. (1999). Managerial use of flexible work schedules: The power of peers. *Human Resource Management Journal*, 38:33–46.

Levitt, S. D. and List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1):1–18.

List, J. A., Shaikh, A. M., and 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):259–283.

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. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–3759.

Mazmanian, M., Orlikowski, W. J., and Yates, J. (2013). The autonomy paradox: The implications of mobile email devices for knowledge professionals. *Organization Science*, 24(5):1337–1357.

McNall, L. A., Masuda, A. D., and Nicklin, J. M. (2009). Flexible work arrangements, job satisfaction, and turnover intentions: The mediating role of work-to-family enrichment. *Journal of Psychology*, 144(1):61–81.

Michel, J. S., Kotrba, L. M., Mitchelson, J. K., Clark, M. A., and Baltes, B. B. (2011). Antecedents of work–family conflict: A meta-analytic review. *Journal of Organizational Behavior*, 32(5):689–725.

Moen, P., Kelly, E. L., and Hill, R. (2011). Does enhancing work-time control and flexibility reduce turnover? A naturally occurring experiment. *Social problems*, 58(1):69–98.

Moen, P., Roehling, P., et al. (2005). *The career mystique: Cracks in the American dream*. Rowman & Littlefield.

Munsch, C. 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.

Netemeyer, R. G., Boles, J. S., and McMurrian, R. (1996). Development and validation of work–family conflict and family–work conflict scales. *Journal of Applied Psychology*, 81(4):400–410.

Porter, C. O., Outlaw, R., Gale, J. P., and Cho, T. S. (2019). The use of online panel data in management research: A review and recommendations. *Journal of Management*, 45(1):319–344.

Reilly, P. A. (1998). Balancing flexibility–meeting the interests of employer and employee. *European Journal of Work and Organizational Psychology*, 7(1):7–22.

Rubenstein, A. L., Eberly, M. B., Lee, T. W., and 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.

Rudman, L. A. and Mescher, K. (2013). Penalizing men who request a family leave: Is flexibility stigma a femininity stigma? *Journal of Social Issues*, 69(2):322–340.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1):7–59.

Sharpe Wessling, K., Huber, J., and Netzer, O. (2017). Mturk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, 44(1):211–230.

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., and 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.

Teasdale, N. (2013). Fragmented sisters? the implications of flexible working policies for professional women's workplace relationships. *Gender, Work & Organization*, 20(4):397–412.

Vandello, J. A., Hettinger, V. E., Bosson, J. K., and Siddiqi, J. (2013). When equal isn't really equal: The masculine dilemma of seeking work flexibility. *Journal of Social Issues*, 69(2):303–321.

Walby, S. and 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., and Whyley, C. (2001). Female finances: Gender wage gaps and gender assets gaps. *Work, Employment and Society*, 15(3):465–488.

Williams, J. (2001). *Unbending gender: Why family and work conflict and what to do about it*. Oxford University Press.

Additional information intended for online dissemination

A Background: the Polish context

We conducted the vignette experiment in Poland. In this country the distribution of hours worked is highly concentrated. According to Labor Force Survey, in 2019 most workers (around 80% of either gender) spent usually 40 hours weekly in their jobs. In the remaining 20%, men and women differ. The proportion of women working (usually) 30 or fewer hours more than doubles the share of men with those arrangements (11 % against 4%). In terms of actual hours, mosst people declared having worked the usual hours. Among the deviants, men are more likely to have worked more hours, either due to overtime (around 2% of men, 1% of women) or due to varying schedules (1% for men, 0.5% for women).

Within this relatively stable work hours, there is room for flexibility. In the same survey, workers report whether they have some autonomy in deciding start and end times. Among men, up to 18 % reported having this possibility, and almost 7% indicated that they can decide their working hours on their own. The proportion of women with the same WTA is almost identical (16% and 6%). When it comes to flexibility requested by the employer, differences are more evident. Around 75% of women report that they are never requested to change their working time to attend an urgent situation. Among men, the percentage is lower (68%), which suggests that men are more often asked to provide labor flexibly. Unlike our experiment, data suggests that both forms of flexibility coexist. Around 27% of the men (and 20% of the women) who have certain autonomy over their schedules report that they have also been asked to modify their schedules to attend an urgency.[11]

Poland ranks relatively low on measures of gender equality. Various estimates indicate that the adjusted gender wage gap is in the vicinity of 20% of men's wages. This value has been stable over the past two decades (Goraus and Tyrowicz, 2014), robust to methods (Goraus et al., 2017), and evenly spread within the country (Majchrowska et al., 2016). In a European context, gender wage gap in Poland is higher than in most other European countries, with the exceptions of Portugal, Spain and Estonia (Goraus et al., 2020). Poland is also a country with quite traditional social norms regarding work, family and gender roles. According to European Values Study, almost 25% of Poles agree that ''Men should have more right to a job than a women when the jobs are scarce", much higher than Germany (7%) or Spain (11%), approximately 5 percentage points higher than the average in Europe. Similarly, 20% of Poles agree with the statement that ''Men make better business executives than women do", again a much higher percentage than in Spain or Germany, with 6.7% and 10% respectively. In the a survey *Modern Polish family* by Bozewicz et al. (2019), 80% of Poles claim that "family happiness" is the most important value in their lives (while ''work" is selected only by 36%). At the same time, majority of household duties is done by women, and men are less likely to admit that they would resign from work to take care of home and family if their household financial situation would allowed for that (32% vs 42%).

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

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

B Additional results

Table B1: Treatment effects: should wages change subsequent new WTAs?

	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 $\bar{y}$ denotes a mean, and $t-stat$ denotes t-static of a test with the null hypothesis that a given mean is equal to zero. Columns entitled *Diff.* report mean differential between treatment conditions. ***, **, * indicate p-values smaller than 10%, 5% and 1% respectively. The sample for all subjects includes 321 subjects and 963 observations. The sample for subjects who passed all the manipulation checks includes 190 subjects and 570 observations. *Reduction* denotes the share of individuals who declare that the wage should decline subsequent the change in WTAs, likewise for *no change* and *increase*. Wage changes reported in USD per month and in % of the status quo wage.

Table B2: Treatment effects: p-values adjusted for multiple hypothesis testing

	Initiator			Gender of character		
	List et al. (2019)	Bonferroni	Holms	List et al. (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 et al. (2019), and the more conservative Holms and Bonferroni adjustments. The p-values correspond to the tests presented in columns titled "Diff." in Table B1.

Table B3: Change in wage subsequent the change in WTA: all coefficients

	Preferred sample				Full sample			
	in USD per month		in %		in USD per month		in %	
	(1)	(2)	(3)	(4)	(1a)	(2a)	(3a)	(4a)
T: employer = 1	17.20***	18.05***	2.95***	2.87***	22.04***	25.35***	3.13***	3.40***
	(4.85)	(6.54)	(0.56)	(0.69)	(6.80)	(8.16)	(0.93)	(1.00)
T: woman= 1	-3.02	-0.11	0.15	0.44	-4.00	-2.16	0.04	0.35
	(4.87)	(5.30)	(0.37)	(0.54)	(6.59)	(8.60)	(0.82)	(1.05)
T: employer × woman	10.30*	8.96	0.10	-0.02	0.28	-3.02	-0.34	-0.75
	(6.17)	(7.43)	(0.62)	(0.93)	(13.50)	(19.74)	(1.45)	(1.98)
V: lawyer = 1	8.84*	8.84	-0.45	-0.45	14.96	14.80	0.20	0.19
	(5.13)	(6.24)	(0.39)	(0.48)	(9.55)	(11.78)	(0.76)	(0.93)
V: salesperson = 1	0.15	0.32	1.86***	1.88***	4.24	4.15	2.28**	2.27**
	(2.80)	(3.43)	(0.52)	(0.64)	(5.75)	(6.99)	(0.92)	(1.12)
Age	0.04		0.01		0.43		0.04	
	(0.23)		(0.03)		(0.30)		(0.03)	
Woman	-10.26**		-0.98*		-3.37		-0.75	
	(4.89)		(0.58)		(9.48)		(1.03)	
Never manager	5.73		0.73		5.12		0.39	
	(5.38)		(0.64)		(9.77)		(1.06)	
Completed tertiary	10.21**		1.08*		4.60		0.46	
	(4.79)		(0.59)		(9.77)		(1.04)	
Income: enough	0.32		0.25		-1.50		-0.29	
	(5.30)		(0.64)		(9.72)		(1.07)	
Income: not enough	9.79		1.35		9.96		2.03	
	(8.45)		(1.12)		(12.12)		(1.57)	
Intercept	-0.44	4.34	-0.95	0.22	-18.79	0.10	-1.98	-0.40
	(9.13)	(4.30)	(1.18)	(0.51)	(18.28)	(7.27)	(1.94)	(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: Table presents results of linear regressions subjects' recommendation about the change in wages portrayed by equation (1). Hairdresser vignette is the base level for the reported coefficients on lawyer and salesperson. Less than tertiary education is the base level for education. Being able to afford luxury is the base level for income. Columns (1) and (2) present regressions where the dependent variable is the absolute wage change in $ per month, whereas in Columns (3) and (4) the dependent variable is 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 those who passed all manipulation checks. Standard errors clustered at subject level presented in parentheses. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1, respectively.

Table B4: Recommended change in wage subsequent to a change in WTAs in the preferred sample: 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***	17.21***	16.90***	17.53***	17.22***	17.27***	2.93***	2.95***	2.90***	2.98***	2.95***	2.96***
	(4.90)	(4.84)	(4.86)	(4.84)	(4.84)	(4.91)	(0.56)	(0.56)	(0.56)	(0.56)	(0.55)	(0.56)
T: woman= 1	-3.04	-3.03	-3.24	-3.20	-3.12	-3.20	0.14	0.14	0.11	0.13	0.13	0.13
	(4.89)	(4.94)	(4.89)	(4.89)	(4.87)	(4.90)	(0.37)	(0.38)	(0.37)	(0.38)	(0.37)	(0.38)
T: employer × woman	10.41*	10.32	10.59*	10.20	10.24*	10.56*	0.14	0.11	0.15	0.09	0.09	0.14
	(6.15)	(6.25)	(6.13)	(6.17)	(6.20)	(6.04)	(0.62)	(0.63)	(0.63)	(0.63)	(0.63)	(0.62)
Valuation of own time	4.02						1.19					
	(19.40)						(2.50)					
Gender norm		0.27						0.11				
		(3.17)						(0.36)				
Work-family conflict			3.46						0.54*			
			(2.46)						(0.30)			
Family-work conflict				7.74***						0.88**		
				(2.80)						(0.36)		
Preference for inequality					7.06						1.10	
					(5.86)						(0.80)	
Importance: away from work						-1.28						-0.19
						(1.38)						(0.17)
Importance: towards work						0.37						0.02
						(1.23)						(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: Table presents results of linear regressions of subjects' recommendation about the change in wages portrayed by equation (1). The estimated model includes the full set of X's, the results are available upon request and are omitted due to being repetitive of Table 2. The estimates reported for the preferred sample. Results for the full sample available on table **??**. Own-time availability based on real-consequence component of our experiment. Higher values indicate greater value attributed to discretion over own time. Gender norm is an index based on items for traditional vs modern norms adopted from World Value Survey. Higher values indicate equitable views. Work-family conflict and family-work conflict are indices based on Netemeyer et al. (1996). Higher values indicate higher sense of conflict. Preference for inequality is based on an item utilized in European Social Survey. Fulfilling life and work & career based on the importance ranking of life components adopted from a standardized opinion poll (Bozewicz et al., 2019). Constant included, not reported. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1, respectively.

Table B5: Recommended change in wage subsequent to a change in WTAs in the full sample: 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***	21.93***	21.59***	21.83***	21.97***	22.07***	3.20***	3.12***	3.05***	3.10***	3.11***	3.14***
	(7.03)	(6.88)	(6.65)	(6.68)	(6.80)	(6.82)	(0.94)	(0.94)	(0.92)	(0.92)	(0.93)	(0.93)
T: woman= 1	-3.72	-4.24	-4.39	-4.46	-4.12	-4.11	0.07	0.02	-0.03	-0.04	-0.00	0.03
	(6.61)	(6.74)	(6.51)	(6.42)	(6.57)	(6.59)	(0.83)	(0.84)	(0.82)	(0.81)	(0.82)	(0.82)
T: employer × woman	-0.67	0.42	0.57	0.45	0.32	0.50	-0.46	-0.32	-0.29	-0.31	-0.32	-0.31
	(13.83)	(13.51)	(13.44)	(13.49)	(13.48)	(13.49)	(1.45)	(1.45)	(1.45)	(1.46)	(1.45)	(1.45)
Valuation of own time	-34.93						-4.48					
	(37.09)						(4.09)					
Gender norm		3.79						0.33				
		(4.93)						(0.56)				
Work-family conflict			5.07						0.90**			
			(3.99)						(0.45)			
Family-work conflict				4.97						0.84		
				(5.32)						(0.56)		
Preference for inequality					3.43						1.21	
					(6.02)						(0.85)	
Importance: away from work						-0.69						-0.11
						(1.28)						(0.16)
Importance: towards work						-1.44						-0.16
						(1.15)						(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: Table presents results of linear regressions of subjects' recommendation about the change in wages portrayed by equation (1). The estimated model includes the full set of X's, the results are available upon request and are omitted due to being repetitive of Table 2. The estimates reported for the full sample. Results for the preferred sample available on table **??**. Own-time availability based on real-consequence component of our experiment. Higher values indicate greater value attributed to discretion over own time. Gender norm is an index based on items for traditional vs modern norms adopted from World Value Survey. Higher values indicate equitable views. Work-family conflict and family-work conflict are indices based on Netemeyer et al. (1996). Higher values indicate higher sense of conflict. Preference for inequality is based on an item utilized in European Social Survey. Fulfilling life and work & career based on the importance ranking of life components adopted from a standardized opinion poll (Bozewicz et al., 2019). Constant included, not reported. ***, ** and * denote significance at p<0.01, p<0.05, and p<0.1, respectively.

Table B6: 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***	17.24***	18.35***	14.41*	2.95***	2.72***	2.69***	2.63***
	(4.85)	(4.67)	(5.05)	(8.34)	(0.56)	(0.80)	(0.86)	(0.76)
× Never manager		-0.10				0.42		
		(8.86)				(1.09)		
× Completed tertiary			-2.11				0.47	
			(8.87)				(1.09)	
× Woman				5.30				0.60
				(9.31)				(1.10)
T: woman	-3.02	1.72	-1.22	-6.84	0.15	0.50	0.25	-0.21
	(4.87)	(3.16)	(2.55)	(9.56)	(0.37)	(0.38)	(0.36)	(0.68)
× Never manager		-9.04				-0.68		
		(8.35)				(0.69)		
× Completed tertiary			-3.27				-0.16	
			(8.50)				(0.68)	
× Woman				7.46				0.68
				(9.59)				(0.74)
T: employer × T: woman	10.30*	4.87	9.23	18.74*	0.10	-0.04	0.13	0.52
	(6.17)	(7.91)	(5.72)	(10.26)	(0.62)	(1.01)	(0.92)	(0.90)
× Never manager		10.16				0.29		
		(10.83)				(1.24)		
× Completed tertiary			1.89				-0.06	
			(10.89)				(1.23)	
× Woman				-16.41				-0.79
				(11.08)				(1.21)
Never manager	5.73	7.84	5.72	5.70	0.73	0.79	0.73	0.73
	(5.38)	(9.18)	(5.38)	(5.38)	(0.64)	(0.76)	(0.64)	(0.64)
Completed tertiary	10.21**	10.22**	12.47	10.10**	1.08*	1.10*	0.94	1.08*
	(4.79)	(4.78)	(8.71)	(4.79)	(0.59)	(0.59)	(0.71)	(0.59)
Woman	-10.26**	-10.30**	-10.26**	-12.59	-0.98*	-0.99*	-0.97*	-1.43*
	(4.89)	(4.90)	(4.88)	(9.56)	(0.58)	(0.58)	(0.58)	(0.77)
Intercept	-0.44	-1.51	-1.56	0.46	-0.95	-0.99	-0.89	-0.71
	(9.13)	(9.42)	(10.49)	(9.42)	(1.18)	(1.18)	(1.28)	(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: Table presents results of linear regressions of subjects' recommendation about the change in wages. These estimates expand on those portrayed by equation (1) by including interactions between treatment and three personal characteristics: managerial experience, whether 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.

Table B7: Beliefs about the social norm: all coefficients

Majority ...	... in favor of a change		in favor of a raise	
	(1)	(1a)	(2)	(2a)
T: employer = 1	0.097**	0.078**	0.11***	0.079***
	(0.042)	(0.031)	(0.040)	(0.028)
T: woman= 1	-0.043	0.0092	-0.040	-0.0098
	(0.035)	(0.026)	(0.034)	(0.025)
T: employer = 1 × T: woman= 1	0.074	0.0060	0.082	0.026
	(0.052)	(0.039)	(0.051)	(0.036)
Vignette				
V: lawyer = 1	-0.081***	-0.090***	-0.076**	-0.089***
	(0.030)	(0.023)	(0.030)	(0.022)
V: salesperson = 1	-0.081**	-0.094***	-0.050	-0.071***
	(0.032)	(0.023)	(0.031)	(0.021)
Proposed change				
Negative change	0.55***	0.56***	0.024	-0.081*
	(0.17)	(0.085)	(0.13)	(0.049)
Positive change	0.57***	0.68***	0.59***	0.69***
	(0.055)	(0.035)	(0.055)	(0.036)
Personal characteristics				
Age	-0.0019	-0.0021	-0.00051	-0.00022
	(0.0024)	(0.0015)	(0.0022)	(0.0014)
Female subject	-0.026	-0.027	-0.019	-0.0064
	(0.047)	(0.034)	(0.043)	(0.031)
Managerial experience	-0.011	-0.023	-0.020	-0.037
	(0.048)	(0.033)	(0.043)	(0.030)
Educ: secondary	0.027	-0.047	0.029	-0.042
	(0.087)	(0.060)	(0.11)	(0.062)
Educ: tertiary	0.044	-0.036	0.035	-0.026
	(0.088)	(0.062)	(0.10)	(0.064)
Income: enough	0.025	0.00093	0.020	-0.00087
	(0.049)	(0.034)	(0.044)	(0.031)
Income: not enough	0.071	0.059	0.036	0.015
	(0.078)	(0.054)	(0.078)	(0.056)
Intercept	0.26*	0.34***	0.17	0.22**
	(0.15)	(0.096)	(0.17)	(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 the 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 the society would raise the wage of the character in the given vignette, and 0 otherwise. Columns with only numbers denotes 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.

C Quality of the sample

Table C8 presents two specifications for studying the features of failures 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 a discretion over WTAs are less frequent than mistakes in identifying working conditions. In fact, 50% of the mistakes referred to specific working hours or weekdays in status quo, 35% related to 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 some extent those with higher earnings) have a greater probability of being successful. Likewise, the vignette where the character is a lawyer have fewer incorrect answers, when compared to the vignette where the character is a hairdresser. These differences might result from the order in which the vignettes were presented to the subjects: the hairdresser vignette was the first, and it followed by the lawyer was the next. Some subjects may have learned what questions are asked in the manipulation checks and pay more attention to these outcomes. Supporting this interpretation, for those who erred once, it was about 1.5 times more likely on the first vignette (hairdresser) than on the other vignettes.

Table C8: Determinants of passing the manipulation checks (marginal effects)

	By individual (1)	By vignette (2)
Age	0.00	0.00
	(0.00)	(0.00)
Woman	0.02	0.02
	(0.06)	(0.03)
Never manager	0.02	0.04
	(0.06)	(0.04)
Completed tertiary	0.17***	0.13***
	(0.06)	(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 presents logit regressions where the dependent variable is the probability of making a mistake. Less than tertiary education is a base level for education. Income sufficient to cover some luxury is the base level for household income. Hairdresser is the base level for the vignettes. 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 missing a manipulation check in a situation. 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 $p<0.1$.