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

Magdalena Smyk
FAME|GRAPE
Warsaw School of Economics

Joanna Tyrowicz
FAME|GRAPE, IZA
University of Warsaw

Lucas van der Velde
FAME|GRAPE
Warsaw School of Economics

October 2021

Abstract

The notion of *ideal worker* necessitates being available at the discretion of the employer in terms of time. We embed our study in the context of gender. Discretion over working time, meanwhile, is widely considered one of the cornerstones of work-life balance and job satisfaction. We study the pecuniary and social valuation of the autonomy to decide about working schedules. We compare employee-initiated and employer-initiated request for a change towards more flexible working time arrangement. We provide plausibly causal evidence that an *ideal worker* indeed ought to be available, but requesting this availability should be reflected in wage rise on average. There appears to be no penalty to employee-initiated request for flexibility and this result is common for men and women.

JEL codes: J32, J71, J16

Key words: labor market inequality, flexibility penalty, vignette experiment, gender norms

*The authors are grateful to Katrin Auspurg, Adrian Chadi, Jose Garcia-Louzao, Thomas Hinz, Jekaterina Navicke, David Ritter and Linas Tarasonis. Erita Narhetali provided wonderful research assistance. This project was supported by joint NCN-LMT DAINA initiative (grant #2017/27/L/HS4/03219). The experiment was pre-registered in the AEA RCT Registry (the unique # AEARCTR-0007642). Ethical approval for this study was obtained from FAME|GRAPE Ethics Committee on 27.04.2021 (decision #1/2021).

1 Introduction

We study the role of autonomy to decide about the working schedule in determining wages. Deciding about wages and working conditions is a key element of employer-employee relations. We zoom in on these decisions, manipulating the working time arrangements. Building on the existing literature, we theorize that the workers who depart further away from the notion of an *ideal worker* experience wage penalties, whereas becoming more congruent with that model implies wage increases. We provide a framed field experiment to empirically verify the extent to which working time flexibility should be reflected in wages. We operationalize working time flexibility as discretion over starting and ending times, which can be exerted by the employer or by the employee. We also provide a gender context.

For more than a half of a century, the notion of an *ideal worker* involves commitment, stable career trajectory, permanent and wide availability to the employer (Davies and Frink, 2014). The *ideal worker* is a benchmark for the actual workers in terms of relative wages. The 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.[1] The scholarship in the field recognizes also a notion of *worker autonomy*, which involves the worker's ability to determine method, schedule and criteria of evaluation (Breaugh, 1985). Worker autonomy is believed to foster intrinsic motivation of the workers and thus their commitment and productivity (Spector, 1986; Heavey et al., 2013; Rubenstein et al., 2018). However, the *ideal worker* may be autonomous in terms of method, but not in determining own schedule (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 an ability to fulfill the 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, descendant from social roles theory). Flexible working arrangements (FWAs) are rationalized as instruments of improving the work-life balance of the workers to the benefit of the employers (Fagan et al., 2012). While they are typically advocated as instruments costly to employers and beneficial to employees, 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).

Indeed, there are actually multiple reasons for why both employers and employees may prefer FWAs. In an online experiment with real effort, Boltz et al. (2020) demonstrate that flexible start and finish time may substantially improve productivity. Two main channels drive this finding: first, with the ability to choose start and finish time, the workers take fewer breaks; second, more productive workers sort into working flexible hours. Bloom et al. (2015) show in a controlled field experiment that workers' productivity is not reduced when they work from home. Likewise, evidence on the employee side favors flexible working hours rather than fixed schedules in standard hours. Angelici and Profeta (2020) provide a controlled field experiment, where workers were randomly assigned to fixed hours condition or flexible hours condition one day a week. The workers in the flexible hours treatment report higher levels of well-being and work-life balance, especially women. This is consistent with the notion of workers deriving value from own autonomy (Hayman, 2009; Peters et al., 2009; Shagvaliyeva and Yazdanifard, 2014).[2]

Although many empirical studies and real world cases demonstrate that FWAs may be neutral or even beneficial to performance (see Azar et al., 2018, for an extensive meta-analysis), the workers requesting FWAs

are viewed as less committed by their managers and colleagues (Williams, 2001; Chung, 2020). Moreover, managers approve FWAs for top performers and are less likely to do so if they perceive that FWAs may proliferate to other employees (Kossek et al., 1999). Coworkers report reduced job satisfaction when their colleagues benefit from FWAs, arguing that their workload was raised without adequate compensation (Munsch, 2016), including female coworkers (Teasdale, 2013). This negative perception of FWAs lends support to the *ideal worker* as the mental model used by the workers (especially the managers).

The literature in this field emphasizes the gender angle to FWAs. There is ample evidence that men request flexibility less frequently then women (Vandello et al., 2013). The so-called Goldin's conjecture (2014) states that women – as primary care givers – are prevented from working in jobs that require flexible supply of hours at employers discretion.[3] 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). 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). Indeed, men are less likely to have their flexibility request accepted if it is for family reasons (Brescoll et al., 2013). As a consequence, women are effectively prevented from converging to the *ideal worker* model.

Building on this theoretical literature and empirical insights we study how the notion of *ideal worker* is related to compensating wages. Is *ideal worker* proportionally rewarded for being available to the employer? Are women rewarded the same way as men? We pitch the employer-initiated against the employee-initiated nonstandard working time arrangements (WTAs).[4] We also manipulate the gender dimension in order to identify the causal effects of gender of the worker on the wage premia and penalties related to WTAs.

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 summarize 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, ours is an online vignette experiment rather than controlled randomized trial. However, we do introduce a real consequence component to the experiment 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.

Summarizing, our study provides a novel approach to study flexible working time arrangements in the

context of who initiates a change from standard to non-standard arrangements; and in the context of worker's gender. To the best of our knowledge ours is the first study to analyze these two contexts jointly. Our results suggest that when purposefully manipulated and isolated from other potential effects, then employee-requested discretion over WTAs is not associated with wage penalty, regardless of the workers' gender. By contrast, when the employer requests more discretion to set working schedules of the workers, we find a robust wage premium. This premium is similar for both genders and does not vary with context.

The paper is structured as follows. Next, in section 2 we present some background about flexible working time arrangements, as well as characterize the case of Poland, where we conducted the experiment. In section 3 we discuss the details of the experimental design and the hypotheses. Section 4 describes the sample. We present the results in section 5, along with an extensive discussion. The paper concludes with discussion of external validity as well as policy implications of our study.

2 Background

According to the European Working Conditions Survey (EWCS), approx. 69 per cent of workers declare that their working time arrangement is set exclusively by the employer.[5] By comparison, roughly 5 per cent of the salaried workers report that they have full autonomy over their working time arrangements. However, setting working hours – by the employee or by the employer – bundles together several dimensions of WTAs: varying start and end times, different number of hours a day, shift work, and long hours. Inspecting the incidence of specific WTAs reveals that shift work remains a separate phenomenon. Shift work tends to be closely associated with specific industries and occupations and thus is not fit for experimental manipulation *without a job change*.

By contrast to long hours and shift work, varying hours within week, varying start and end time often occur jointly. We define a dummy variable for varying hours within week, taking on the value of 1 when this WTA is reported and 0 otherwise. Analogously, we construct a dummy variable for varying start and end times. The correlation between these two dummy variables exceeds 0.5 (with $p < 0.00$). The correlation with the prevalence of long hours is lower, but nonetheless statistically significant (the correlation coefficient of approx. 0.19 with $p < 0.00$). Given the high coincidence rate of the two WTAs, they are a suitable case for experimental manipulation.

Among the four WTAs, long hours tends to be the least prevalent in Europe. By contrast, varying start and end times as well as varying number of hours across weekdays are relatively prevalent, with 15 per cent to as much as 80 per cent of salaried workers in Europe reporting this form of WTAs in their job. Figure 1 portrays the distribution of the four WTAs across the countries in Europe. This scatter plot reveals that long hours are systematically more prevalent among men and generally infrequent in Europe. By contrast the other WTAs tend to be reported by both men and women and appear to be much more frequent.

[Figure 1 around here]

In line with the coincidence of two most prevalent WTAs, we construct our experiment to reflect realistic employment conditions faced by workers and employers.[6] In our vignettes, the status quo involves standard employment: fixed start and end hours with an even number of hours per day. The alternative to be evaluated by the participants involves a combination of the three WTAs: holding the total weekly hours

constant, the new working time arrangement will involve changing the start and end time on a short notice. Implicitly this working time arrangement permits long hours on a given day, but not permanent overtime.

The additional advantage of working with these WTAs is that we are credible in putting for consideration that they are requested by the employer as well as the employee. It would be unrealistic for the subjects in the experiment to consider cases where the worker requests shift work or explicitly long hours. By comparison, discretion over start and end times can be requested realistically by both parties, because it is conceivable that the employee may prefer such arrangements for a wide range of reasons.

2.1 The context of Poland

We conduct a vignette experiment in Poland. Compared to other European countries, the share of workers reporting WTAs is low. Poland stands out also in terms of WTAs composition: shift work continues to be the most prevalent.[7] It is also evenly reported by men and women.

When compared to other European countries, Polish men report the highest prevalence of long hours: 32% report working for longer than 48 hours per week at least once a month. According to EWCS, long hours is the least prevalent WTA among women, but at par with all the other WTAs among men. The distribution of hours worked in Poland is highly concentrated. According to Labor Force Survey, between 2015 and 2019 most workers spent 40 hours weekly in their jobs. Among working men, 70% declared working exactly 40 hours, the fraction of men reporting less than 40 hours per week is negligible. Among women, 74% report even 40 hours per week, a substantial fraction reports part-time arrangements and a small fraction reports long hours.[8]

High prevalence of both shift work and long hours suggests that the model of *ideal worker* is strong in Poland, especially among men. However, this model is combined with relatively low share of workers who report that the working hours are set exclusively by the employer: 70% against the 65% on average in Europe. Roughly 26% of the workers report some discretion over setting the working hours, either by choosing from a menu of options or by the ability to adapt hours on a daily basis. By contrast, working with varying start and end times is reported by merely 17% of the workers (against the average of 33% for other European countries in EWCS).

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 average in Europe. Similarly, 20% of Poles agree also with a statement that ''Men make better business executives than women do", again much higher than 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%).

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.

3 Experimental design

We elicit the value of working time arrangements in terms of changes to wages, relative to the status quo. Each participant assess three vignettes, which gives us the within-subject variation. The participants are randomly assigned across the treatment conditions. However, each participant answers at least one female and one male character and also at least one change was initiated by worker and one by employee. With this design, we can ignore systematic differences between the participants of the experiment (technically, we can estimate the models with individual fixed effects; naturally, we also cluster the standard errors at individual-level).

3.1 Treatments

We propose a mixed design, combining a 2x2 framed field experiment with a vignette.[9] The framed field experiment randomly assigns subjects to treatment conditions, where they evaluate three vignettes. The two treatments include the initiator of the change in WTAs (worker or employer) and a gender of the worker (a man or a woman). In the INITIATOR treatment, the subjects learn a story of a worker, who currently works in a regular, fixed time schedule, five days a week. Either employer or the worker want to change that WTA: the start and the end times will vary. We state that the *average* number of weekly hours stays the same and the start and end times for each day will be communicated with some notice. In the GENDER treatment, the subjects are faced with a man or a woman as a worker in the vignette.

The new, changed WTAs to be evaluated by the participants in the experiment is 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. As discussed earlier, this form of WTAs could credibly be initiated by both the employee and the employer. Obviously, such WTAs combine 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.

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 desirable wages adjustment was measured in a quasi-continuous way, the participants were choosing their preferred amount from the list, with 50 PLN per month as an interval (approx. 13 USD).

In addition to wage adjustment, we ask the participants to disclose their beliefs about the social 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 that majority of Poles would prefer (increase, decrease or no change).

3.2 Hypotheses

With the *ideal worker* model, the employer is not obliged to reward the availability of the 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). In line with the *ideal worker* model, we hypothesize that:

Hypothesis 1 *Employer-requested increase in discretion over WTAs is associated with no change in wages.*

Hypothesis 2 *Employee-requested increase in discretion over WTAs is associated with a reduction in wages.*

Hypothesis 3 *Women experience no wage decline when requesting increase discretion over WTAs, but men do.*

Naturally, there are alternative mechanisms. First, working at the discretion of the employer in terms of start and end times raises the disutility of work and thus should be compensated (Mas and Pallais, 2017). Second, when the employer requests more discretion in WTAs, presumably this is driven by better business opportunities and thus encourages workers to renegotiate the wages. Furthermore, when women request WTAs, this may be viewed as consistent with the *family devotion* schema. In such case, the employers may acknowledge the need for more discretion and choose not to penalize the worker (Walby and Olsen, 2002; Cousins and Tang, 2004; Warren et al., 2001). The opposite should hold when a man requests WTAs.

3.3 Scenarios

The three vignettes faced by each subject differ by context in terms of occupation: a hairdresser, a lawyer and a shop assistant. 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 half-time (from 9AM to 1PM). The characterization of the worker is communicated graphically, with the use of a cartoon signifying the gender and the occupation of the worker in the vignette, see Figure 2. We inform the subjects to memorize the specific content of each vignette. After each vignette, subjects were asked three manipulation check questions. For answering correctly all nine questions participants received additional one dollar. The subjects were not informed about the outcomes of the manipulation checks until after the end of the experiment.

[Figure 2 around here]

The design with two genders, two initiators and three vignettes implies 64 different combinations ($2^3 \times 2^3$). We randomly assign subjects across conditions, but the randomization follows a pattern that ensures that we will be able to exploit within subject variation in each treatment. To this end, we restrict the pool of situations such that if in the first vignette the respondent was assigned to employer-initiated change in

WTAs, then we restrict that the second vignette refers to employee-initiated change in WTAs. Analogously, for the GENDER condition, if in the first vignette randomization implied that the character is that of a man, then the second vignette is forced to portray a woman. For the third vignette, the algorithm randomized among INITIATOR $\times$ GENDER conditions that did not appear in the first two vignettes (the two remaining treatments). This procedure restricts the number of possible combinations to sixteen and ensures that for each participant of our experiment we observe at both GENDERs and both INITIATORs, though admittedly not each of their combinations.

These three vignettes were constructed to invoke three specific contexts. 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 during 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.

Each occupation has its own base wage in the status quo (identical for both genders, note that for each occupation, participants see only one of the two genders). The base wages were set in line with the market averages: 1600 PLN per month (approx. USD 420) for the shop assistant, 3200 PLN per month (approx. USD 840) for the hairdresser, and 6400 PLN per month (approx. USD 1680) for the lawyer.

3.4 The conflict between work and family lives and attitude towards inequality

After the experimental treatment, all participants take part in a short survey about their experience of work-family conflict, inequality and gender norms. The questions on work-family conflict (and family-work conflict) are adopted from Netemeyer et al. (1996). The authors developed a ten-item questionnaire that explores to what extent work interferes with family life and to what extent family duties collide with work responsibilities. Each item offered five possible levels of answer, ranging from "strongly agree" to "strongly disagree".

In our sample, consistency was high for both work-family conflict (α equals 0.926), and the family-work conflict ($\alpha = 0.899$).[10] Given these high values of α, we aggregate these questions into two standardized indices. For both indices, higher values indicate higher sense of conflict.

Besides the work-family conflict and family-work conflict items, the second part of the questionnaire asked for adherence to traditional gender norms. We asked respondents to indicate to what extent they agreed with the following statements: "A man's job is to earn money; a woman's job is to look after the home and family;' "All in all, family life suffers when the woman has a full-time job;" "When jobs are scarce, men have more right to a job than women;" "When a mother works for pay, the children suffer;" "On the whole, men make better business executives than women do." The possible answers for each statement was scored on a five-level scale ranging from "strongly agree" to "strongly disagree," without a neutral condition. These items work well together ($\alpha = 0.855$). The questions are thus aggregated into a single index, which is then standardized. Higher values indicate more progressive attitudes towards gender equality.

The second part of the questionnaire also included an item on views towards income inequality. Concretely, we asked for agreement with the statement that "Differences in income are necessary to compensate people's productivity and efforts." This question was used already in other surveys concerning inequality, notably in the International Social Survey Program. Possible answers ranged from "strongly agree" to "strongly disagree," with a neutral condition. For the analysis, the variable was recoded into a binary variable where the value of one shows agreement (or strong agreement) with the statement, and zero corresponds to the remaning alternative.

3.5 Valuing own time availability

After the participants share their valuation of changes in WTAs of hypothetical characters in the vignettes, we elicit the value of own availability by the means of a real consequence discrete choice experiment. After completing the vignettes, the subjects were informed that we will randomly choose their wait times between 5 minutes and 30 minutes. Either of these wait times was unavoidable to be able to complete the final part of the experiment. The subjects were not informed if the last part of the experiment is long or short, they were only advised that they will not be able to complete it sooner than after 5 or 30 minutes.

For each of those wait times, the subjects were asked to choose which trade-off between availability and compensation suits them the most. For each of the wait times the menu of options was the same: full flexibility (24 hours) to complete the survey compensated with additional 0.25 dollars, inflexible time span to complete the survey after waiting compensated with additional 0.5 dollar (for completing the survey in 25 minutes) to 1.5 dollar (for completing the survey in 5 minutes). Subjects were informed that if they do not comply with their selection, they fall back to 0.5 dollars for the entire experiment (no additional compensation). After the subject selected their preferred option for each of the wait times, the surveying system randomly assigned the wait times.

Since the subjects make two choices (at 5 and at 30 minutes wait time), we are able to estimate individual valuation of own time availability. All participants complied with the selected time slots. To obtain the individual valuation of own time availability, we regress the selected remuneration against the selected time slot, obtaining an individual slope coefficient. Naturally, the regression is estimated without a constant.

3.6 Final component of the experiment

After the wait time elapsed, a new screen was visible to participants. The screen contained six additional questions that subjects had to answer. Were subjects not able to complete the survey in the required time, they would still be allowed to answer all the questions, but they will not be given the additional reward. In practice, non-completion was not an issue as every participant complied with stated preferred time to fill the final part of the survey.

The questions in the second part of the survey were of two types. First, we inquired the participants personal characteristics: age, education level and economic situation. Given how sensitive questions on income might be, the last question we adopted the same question as 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). Besides questions on demographic characteristics, subjects were asked to order a list of 16 values according to their importance, adopted from the cyclical study "Modern Polish family" by Bozewicz et al. (2019). The list included in total 16 items to be ranked

from most to least important. We pay particular attention to three items highly relevant for the notion of the *ideal worker*, which include professional success, career and respect of the others, as well as three items opposing this notion – such as family happiness, fulfilling life, and importance of friends. We group the two sets of three items into two indicators. For the *ideal worker* measure, we take the three items and average their importance. Analogously, we average the importance score for the opposing items. For those individuals, who did not rank all 16 options, we impute the importance with each unranked option in a way which preserves mean ranking in the sample. For example, if an individual ranked 10 options and left 6 items unranked, we assign each of them a rank of $10 + 1/(16 - 10) \cdot \sum_{i=11}^{i=16} i = 13.5$. We impute the missing rankings before we average the three items that are consistent with the *ideal worker* and the three items that oppose it.

3.7 Implementation

The experiment was administered online, on a sample of participants pre-registered for experiments and surveys (ANSWEO). Participation was voluntary. Participants were compensated for completing the survey: a flat rate of USD 0.5 for participation and subsequent up to USD 3 depending on individual choices during the experiment. The experiment was implemented between April 23rd and August 9th 2021.

Samples like ours are sometimes criticized on the grounds that the participants are predominantly interested in the pay and thus are not sufficiently careful in filling in the questionnaires (Cheung et al., 2017; Sharpe Wessling et al., 2017). We address this issue extensively in the subsequent section, discussing the implications of the manipulation checks run in this study (Porter et al., 2019).

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.

4 The sample

In total, we observe judgment on 963 vignettes with 321 subjects. Among them 49.84% were women. The average age is 38.5 years. Almost 50% of the sample were subjects with tertiary education, 40% with secondary education, and around 10% with lower education levels. 46% of the sample claim that they never had any managerial experience. The top panel of Table 1 displays the descriptive statistics for individual characteristics such as age, education, income and gender. Given that the four groups do not differ in individual characteristics, we are safe to assume that the randomization across treatments was successful, despite unusual two-step assignment into conditions.

In order to asses the quality of our sample, we provide manipulation check questions. Each individual responded to nine manipulation check questions: three for each vignette. The three questions were related always to the working hours and days in the status quo and the party who initiated the change in discretion over WTAs. Over 40% of the subjects fail at least one of those nine questions. The percentage of the individuals who passed all nine manipulation checks may be systematically related to both treatment conditions (the individual may have been more or less receptive to a vignette) and individual characteristics

as well as norms.

The last row of the upper panel of Table 1 indicates the percentage of observations provided by subjects that have correctly answered all manipulation checks. This is a demanding statistic in a sense that if some individual failed to correctly identify status quo working hours in one of the vignettes and this person is already classified as one who failed at least one manipulation questions. However, the differences across treatment conditions are absent in the case of both gender treatment and initiator treatment.

[Table 1 around here]

To verify the extent to which failure to complete the 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 the Appendices in Table A1. The table presents two specifications. In the first one, we look at the probability of making a mistake in at least one 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, whereas 40% of the subjects who failed record more than one mistaken answer. Mistakes in identifying who requests a discretion over WTAs are less frequent than mistakes in identifying the initiator. 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.

Indeed, failing at least one manipulation checks is not fully random: subjects who passed all the manipulation checks are better educated. They are also somewhat older, though this feature is statistically significant in the vignette-level specification, not in the person-level specification. This suggests that older participants were more likely to do a lower number of mistakes, but not more likely to make a mistake at all. Other individual-level characteristics turn out to be insignificant.

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 the lawyer was the last. 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.

In the remainder of this study, in line with Table 1, we report two sets of the estimates. The first one zooms in on the approx. 60% of the collected sample, where an individual replied correctly to all manipulation check questions. The second one utilizes 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, which they were judging.

5 Results

The main interest of this paper lays on the expected compensation for change from inflexible to flexible working time arrangement and whether this expected compensation differs between treatments. The results of the experiment will be presented three substantive parts. First, we present comparison of means across

treatment conditions. Given that our randomization proved to be successful, this is sufficient to establish the treatment effects. However, the treatment effects could be heterogeneous across individuals: individual characteristics as well as beliefs may mediate the effects of treatment on the provided responses. This gives rise to the second part of this section: we enrich the perspective in a regression model, which permits adjusting the estimates for a variety of the measured individual characteristics. This analysis is complemented by a study of perceived compliance with the social norm. The third part of this section provides an analysis of 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. This section ends with a discussion of our results vis-a-vis the literature.

5.1 Treatment effects at means of responses

Majority of subjects claims that change in working time arrangement (from inflexible to flexible) should not be associated with any change in wage. In total, almost 70% of answers were "wage should remain the same", and 48% of subjects choose this answer in all three contexts. If subjects considered that wages should vary to reflect the new situation, then increases were much more frequent than decreases (26% of answers to only 4%). Table 2 provides a first glance at the results of the experiments

We report several outcome measures of the experiment. First, we provide categorical variables for whether the subjects believed there should be adjustment to wage with the change in WTAs or not. Thus, we report three indicators: % of subject recommending reduction, % or respondents recommending no change, and % of subject recommending an increase in wage subsequent the change in WTAs. In addition, we naturally look at the reported changes in wages in $ per month. Finally, we express the reported changes in wage as a percentage of status quo. This relative change serves to study if characters with higher earnings in are more expected to fulfill the notion of the *ideal worker*. The table reports means for each treatment condition as well as differences between treatments. For the columns reporting the means in each treatment, we also report the $t-statistic$ of the test that this mean is statistically equivalent to zero. For the columns reporting differences between means across treatment conditions, we include the stars to signify whether these differences are statistically equivalent to zero.

The majority of subject in our experiment, approximately 70%, did not see the need for the wage to change. Of the remaining respondents, almost all believed that an increase in wage is in order. We do observe treatment effects, however. When the employer initiates the bargaining process, the expected wage increase is around USD 28 per month, or around 3% of the initial wage. By contrast, whenever the employee wants to modify the wage arrangements, the expected change in wages is much lower, close to USD 6 and is not statistically significant. Much lower average change when employee request a new WTA is reflected in the proportion of people who indicated that no change should be made to wages: almost 80% of answers in that treatment condition versus 60% in the employer treatment.

[Table 2 around here]

We interpret findings reported in Table 2 as a confirmation for the *ideal worker* model. The probability of a fall in wages is almost twice as likely 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*, the subjects were twice as likely to argue for a reduction in wages (5% vs 3% of the

subjects). Analogously, much lower fraction of subjects believes that the wage should be increased when the employee requests discretion over WTAs than when the employer does (16% vs 36 %). However, this finding is not necessarily consistent with the notion of an *ideal worker*: if we assume that the employee should be generally available for work, it is not obvious that the wage should increase at all. A way to think about these treatment effects is the following: clearly, making the employee salient in the role of the party requesting more discretion over WTAs discourages subjects from reporting wage increases as desirable (by 20 percentage points or 55%) and encourages to report wage decreases as desirable (by 2 percentage points or 60%). Moreover, no change in wages is a modal answer across all the conditions, so in majority the subjects to do expect the employers to compensate their workers for requesting discretion over WTAs.

The gender condition reveals lower treatment effects, statistically not different from zero. Differences in the mean wage change fall into a range between -2 and +5.2 USD per month, depending on the sample and are not statistically not different from zero. We find no statistically significant differences for the fraction of respondents who recommend wage reduction, increase or change. These results resonate well with the *ideal worker* model which lacks a 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. The lack of gender differences is also consistent with the conjecture discussed by Goldin (2014).

The inference based on more complex designs than two-sample t-tests yields the same conclusions. This holds for comparing answers within subjects and using non-parametric tests for differences in medians. However, our treatment effects may be mediated by confounding factors such as gender norms, the sense of conflict between work and family life, etc. We move to addressing this issue in the next section.

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

Denote by $y_{i,v}$ the wage change reported by the respondent i in vignette v. This outcome variable can be measured in absolute terms (the adjustment in wage by USD) and in relative terms (the adjustment in USD relative to the base wage described in the vignette). We estimate the following model:

$$y_{i,v} = \beta + \beta_I T : I_{i,v} + \beta_G T : G_{i,v} + \beta_{inter} T : I_{i,v} \times T : G_{i,v} + \gamma_v + \delta_x X_i + e_{i,v} \quad (1)$$

$$y_{i,v} = \beta + \beta_I T : I_{i,v} + \beta_G T : G_{i,v} + \beta_{inter} T : I_{i,v} \times T : G_{i,v} + \gamma_v + \delta_x X_i + \delta_n N_i + e_{i,v} \quad (2)$$

where $T : I_{v,i}$ and $T : G_{v,i}$ 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 γ_v. The components measuring individual social norms, preference for own time availability are denoted by $\delta_n N_i$ and other individual characteristics (age, gender, education, income levels, previous managerial experience) are denoted by the term $\delta_x X_i$. We are going to explicitly verify if the estimates of β's are affected by the inclusion of specific, individual level controls. Explicitly, this series of estimations verifies the potential for heterogeneity in the treatment effects. Note that in order to include the individual level controls, we cannot use the individual fixed effects specifications. We cluster standard errors at the individual-level.

The wage change is measured in quasi-continuous manner: the number of categories in each vignette (above 60 if we consider both positive and negative changes) and the equal spacing make a linear approximation adequate. Similarly, one could be worried that top-codes would restrict possible answers, making

a truncated model more adequate for the analysis. In practice, less than one percent of the answers corresponded to the largest wage changes (in either direction), so alternative models should produce similar coefficients, while putting additional strain on the efficiency of the estimators. In the interest of estimators' efficiency, we estimate equation (1) using OLS. For the relative changes, by design, the outcome variable is constrained to -1 at the lower limit and +1 at the upper limit. However, the linear probability model performs well relative to a nonlinear specifications, we thus continue with a linear estimator for the relative changes as well.

Given our earlier hypothesis, we expect β_I to be greater than zero (Hypotheses 1 and 2). In the *ideal worker* model there is no specific assignment for genders, but Goldin (2014) conjecture suggests that β_G ought to be zero, whereas β_{inter} ought to be of the same sign as β_I (Hypothesis 3). One should bear in mind that β_I represents the *difference* between employers and employees requested discretion over WTAs, likewise for β_G in the context of gender treatment. In other words, the interpretation for β_G 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 employer. The remaining variables in equation (1) are controls for the occupation of the character in the vignettes, which capture also the effect of the order in the vignettes. Finally, we also include controls for respondent characteristics. This is done in two manners: with individual fixed effects and with including the personal characteristics: age, gender, education, previous managerial experience, and household income levels affect the regression. Our preferred specification is the former one, because it allows to fully exploit the within-subject and between-subject variation in our data. With individual fixed effects, the estimators of β''s automatically exploit only the within-subject variation. In all specifications, standard errors are clustered at the individual level to adjust for correlation of responses within subjects.

[Table 3 around here]

Consistent with the earlier results, Table 3 shows significant treatment effects for the initiator condition and insignificant effects for the gender condition. Overall, the wage should increase by roughly USD 17-25 per month or 3% if the employer requests additional discretion over WTAs. These results are remarkably resilient to the inclusion of individual characteristics of the participants. All these results are in line with the hypothesis following the theoretical considerations formulated earlier in this study.

We further observe that the estimates of γ_v are significant only for the salesperson in the case of relative wages. 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. Indeed, one could rationalize this outcome as evidence that subjects value time discretion at a fixed price, independent of income levels. We find no evidence for the role of age, education income or managerial experience. In the preferred subsample we find that women are recommending substantially lower wage rises than men on average: roughly USD 10 or 1%.

[Table 4 around here]

In the next step, we report the results for the augmented specification of equation (2). The results portrayed in Table 4 results help to identify if individual beliefs attenuate or amplify the treatment effects.

Given that the beliefs and norms tend to be correlated, we estimate the model with each one of them separately, to eliminate the risk that the imperfect multicolinearity deflates our t-statistics. One could expect two potential aspects: (i) the inclusion of the beliefs changes the estimated treatment effects and (ii) there is a significant interaction term between beliefs and the treatment effect. In the interest of brevity, we report the results from the former, the results for the latter yield no additional insights and are available upon request.

Surprisingly, individual characteristics of subjects, including measure of individual preference for own time availability, which we obtained in a real consequences question, did not contribute in any way to attenuate or amplify the treatment effects. In fact, most of them prove to be unrelated to the subjects' recommendation about the wage change. The estimates effects are not statistically significant for gender norms, preference for inequality and life values. We find some evidence that subjects who reported higher levels of work-family conflicts and family-work conflicts tend to recommend higher changes in wages (by around USD 8 per month, or 0.6-0.9% relative to status quo). This 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.

The remarkable stability of the estimated treatment effects suggests that the *ideal worker* notion is deeply embedded in the way that our subjects think. Table 2 demonstrates that when the employees request discretion over WTAs (depart from the *ideal worker*), they are penalized and this penalty is gender neutral. We also show that when the employer requests discretion over WTAs, the respondents most frequently recommend no change in wages. Yet, on average, the respondents recommend a rise in wage when the employee is to become more flexibly available to the employer, roughly 3% of the initial wage. In absolute values, this number is higher if the character in the story is a woman. This last result speaks to the generally traditional gender norms of the Polish society. A strong social norm to take care of family and not prioritize career over family, the subjects on average believe that women should be paid more than men, if employer wants them to be more flexibly available.

One potential concern related to these results is that these are individual recommendations. In other words, the subjects may feel that the characters in the vignettes ought not to be compensated for being more available to the employer, neither they should be penalized for requesting greater discretion over own time. However, the individuals may be convinced at the same time that this preference is not shared with the rest of the society. We address this point in the next analysis, where we study the congruence between own recommendations and the individual beliefs about the social norm.

5.3 Social norm

Subsequent each of the vignettes, we asked the subjects to evaluate if their recommendation on the wages are in line with Poles in general. Specifically, the subjects reported if they believe other Poles would agree with them, on average. In case the answer was negative, the subjects reported what they thought would be the overall view of Poles. The subjects were asked this set of questions for each of the vignette.

The answers of what the majority would choose are indicative of what are the social expectations of wage changes following changes in working time arrangements. 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 stated that wages should fall (28%

of responses) and the least frequent among those who claimed that wages should increase following the changes in WTAs (12%).

We use these questions to construct three outcome variables. The first one is a dummy which takes on the value of 1, when the subjects think that their views are in line with the general norm and 0 otherwise. We call this variable *Majority agrees*. The second one is when the subjects report that the overall norm is to change the wage at all, subsequent the changes in WTAs. It takes on the value of 1 if the subjects recommend change and they believe this is the norm, and when the subjects did not recommend the change, but they believed that this is the norm; and zero otherwise. We call this variable *Majority in favor of a change*. Finally, there is a dummy which 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*. The results are presented in Table 5.

[Table 5 around here]

Estimates in Table 5 corroborate our inference from Table 3. Indeed, the subjects appear to expect the norm to be more positive than their own recommendations. When employer requests discretion over WTAs, the subjects are less likely to believe that their recommendation coincides with the social norm, but this effect is small, roughly 6% lower probability of congruence. However, the subjects are more likely to believe that social norm involves changing the wage (21% more likely), in fact increasing the wage (24% more likely). In parallel to the previous estimations, there is no evidence of the gender differences. The key estimate in Table 5 involves the correlation between own recommendation and the social norm. Individuals who recommended wage raise were 20-30% more likely to believe that this is a social norm, relative to those who recommended wage cuts. However, there is no difference for individuals who recommended no change in wages, relative to the base level of recommending wage cuts. This finding confirms that while no wage change is a majority recommendation, consistent with the *ideal worker*, the individuals who recommend a raise believe that to be a social norm.

The norms concerning the three occupations in three vignettes differ, but these differences are minor. Like in the case of Table 3, women are less likely to believe that a wage change is a norm, but equally likely to men to expect a wage raise. Individuals with medium and low income are less likely to believe that others will agree with them, and more likely to recommend a wage raise.

5.4 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, majority of survey respondents find that perfectly acceptable and see no need for changing the wages. When the employee requests discretion, the share of respondents expecting a cut in wages doubles, though respondents reporting no reason for a wage change is still a majority.

Our experiment also partially corroborates the Goldin's conjecture. It is more frequent that when the employers request discretion, the wages of the workers should be raised. 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. Moreover, there is no gender dimension to the compensating differentials.

Admittedly, the estimates in our experiment are small. Our approach is similar to that taken by Gimenez-Nadal et al. (2021) using observational data from time use data for the United States. Our estimates are in line with the existing literature: Mas and Pallais (2017) find slightly larger estimates, but in their design they operated on hourly wages, and the smallest observed by the subjects difference in wages (USD 0.25) accounted for app. 3% of total hourly wage. In our study, the lowest possible answer was equal to 0.8% of total monthly wage (USD 30). Indeed, among those who recommend wage change, a half of wage raises falls short of USD 50 per month. A vast majority of wage cuts falls short of USD 30 per month. Hence, the estimates in Tables 3 and 4 are not a consequence of polarized large and small answers. Even those respondents who believe the change in working time arrangements should be followed by a change in wages select small adjustment in wages.

It could be that we find small effects because our respondents expected the characters in the vignettes to be shifted a few hours earlier or later, but they did not expect that the characters should expect a major disruption to their lives schedules. Vignettes stating more strongly that discretion by the employer implies potentially important disturbance to private life, could encourage the respondents to more frequently report that the wages should change and that the wages should adjust by more. However, this way we would artificially inflate the initiator treatment effect, while size effects are sufficiently large for verifying Hypotheses 1 and 2. Meanwhile, the gender treatment effects would not likely be affected by making the costs associated with employers' discretion more tacit.

Analogously, with reference to the gender dimension and Hypothesis 3, the reported 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. Specifically, we could include in the vignettes the family angle or the education angle (Vandello et al., 2013). While potentially we could boost the gender treatment effect, the downside of such choice would be that we would not identify the effects of gender *per se*, but rather a combination of gender and other reasons. 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. In fact, there are no statistically significant or economically meaningful differences in reported wage changes even in a traditional society in terms of social role of women, such as Polish. It remains a fruitful area of further research to disentangle the contexts which quantitatively tilt the discussions regarding WTAs in the direction of discretion by either employers or employees.

It seems that when asked about wage changes, respondents report based only on productivity. If respondents do not expect the productivity to suffer from workers having more discretion over WTAs, then there is not harm done to the employer. Consequently, there is no need to renegotiate wages. On the other hand, if the employer asks for more discretion over workers time, the worker might anticipate that this change is motivated by productivity increases (otherwise the employer would not engage in costly negotiations). Hence, respondents recommend adjustment in wages in order allow the workers' participation in proceeds from the anticipated productivity increase. This explanation is (in principle) gender neutral and suggests that Goldin (2014) hypothesis is perhaps not universal.

6 Conclusions

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

The *ideal worker* model does not provide clear insights in terms of compensation: should flexibility be rewarded? The *Goldin's conjecture* states implicitly that it should be the case. Further, she argues that women are less frequently able to agree to working time arrangements at the discretion of the employers and that this difference stands behind observed differences in wages between men and women. In other words, employers reward the flexibility of the employers and penalize lack thereof, and this wage differential explains gender wage gaps in their entirety.

Our results lend partial support to the *ideal worker* model. For our subjects, employers request regarding change to flexible WTA do not raise any objections. But we also show that employees requesting discretion over their working time arrangements are not expected to be penalized, whereas employers requesting such discretion ought to pay, at the margin, for having workers available to them. Thus, our results support the mechanism of the Goldin's conjecture, but not its implications for gender inequality. Specifically, the flexibility premium is too small to explain away gender wage inequality and we find no evidence of gender-specific inflexibility penalty.

Our study provides a novel insights on how flexible working time arrangements are socially perceived in the context of the *ideal worker* norm. The fact that introducing change in already existing working arrangements for majority of subjects does not raise an expectation towards change in pay (if it does it is only marginal and only when employer asks for discretion over working times) can be interpreted as conviction that wage should be related to productivity rather then other aspects of work. And productivity seems not to be considered related to WTAs per se.

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, should also be more sensitive to the gender dimension. Non-standard working time arrangements are prevalent, 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.

There are two elephants in the room: external validity of our approach and the gender dimension. For the former, in this paper we prioritized causality, but the ability to implement a randomized experiment comes at the expense of not working with nationally representative samples. Working with labor force surveys or similar samples at this point makes it impossible to identify a causal answer to the main research

question of this study. Also, note that the size effects are simply very small – when the null hypothesis is not rejected, it is not because our estimates lack precision, but rather due to minor differences between the null and the alternative hypothesis.

As to to the gender dimension, the existing empirical literature cannot justify the extent of gender wage gaps with the lower participation of women in flexible working time arrangements. Our study corroborates this established result. It thus appears that there are other mediating and intervening processes, which relate to the mechanisms of both wage setting and the working time arrangements. Further theorizing in this field is thus necessary.

References

Ala-Mursula, L., Vahtera, J., Pentti, J., and Kivimäki, M. (2004). Effect of employee worktime control on health: a prospective cohort study. *Occupational and Environmental Medicine*, 61(3):254–261.

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.

Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2015). Does working from home work? evidence from a chinese experiment. *Quarterly Journal of Economics*, 130(1):165–218.

Boeri, T., Giupponi, G., Krueger, A. B., and Machin, S. (2020). Solo self-employment and alternative work arrangements: a cross-country perspective on the changing composition of jobs. *Journal of Economic Perspectives*, 34(1):170–95.

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.

Goldschmidt, D. and Schmieder, J. F. (2017). The rise of domestic outsourcing and the evolution of the german wage structure. *Quarterly Journal of Economics*, 132(3):1165–1217.

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.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4):1009–1055.

Hayman, J. R. (2009). Flexible work arrangements: Exploring the linkages between perceived usability of flexible work schedules and work/life balance. *Community, work & family*, 12(3):327–338.

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.

Kässi, O. and Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting and Social Change*, 137:241–248.

Katz, L. F. and Krueger, A. B. (2017). The role of unemployment in the rise in alternative work arrangements. *American Economic Review*, 107(5):388–92.

Katz, L. F. and Krueger, A. B. (2019). The rise and nature of alternative work arrangements in the United States, 1995–2015. *ILR Review*, 72(2):382–416.

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.

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.

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., 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.

Peters, P., Den Dulk, L., and Van Der Lippe, T. (2009). The effects of time-spatial flexibility and new working conditions on employees' work–life balance: The dutch case. *Community, Work & Family*, 12(3):279–297.

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.

Shagvaliyeva, S. and Yazdanifard, R. (2014). Impact of flexible working hours on work-life balance. *American Journal of Industrial and Business Management*, 2014.

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

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

[Table A1 around here]

Notes

[1]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), career advancement (e.g. Guillaume and Pochic, 2009). This rich literature has been extensively discussed by Davies and Frink (2014) and we refer the interested reader to this source. 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.

[2]Some evidence suggests that autonomy over working hours is positively correlated with workers' health as well (Ala-Mursula et al., 2004).

[3]Goldin (2014) argues that employers reward workers for the ability to provide hours at employers' discretion (see Goldin and Katz, 2016, for empirical evidence from an occupation where technological innovation relaxed the working hours constraint and simultaneously women's employment shares and wages increased). In fact, using FWAs can reflect constraints resulting from household responsibilities rather than preference as such. As women's home production time endowment becomes more elastic, such as reducing the alternative cost of providing household work through markets, their use of FWAs declines (Cortes and Pan, 2019).

[4]Note that flexible working arrangements (FWAs) may involve a variety of issues, including working from home or other forms of tele-work. Moreover, if the employer has full discretion over setting the working hours, this arrangement would be considered inflexible from the perspective of the worker and flexible from the perspective of the employee. 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.

[5]The questions of interest are available in wave six of EWCS, collected in 2015.

[6]The prevalence of non-standard working time arrangements is rising across developed economies (Katz and Krueger, 2017; Goldschmidt and Schmieder, 2017; Kässi and Lehdonvirta, 2018), and this rise is only marginally driven by the growth of platform jobs (a.k.a gig economy, see Katz and Krueger, 2019; Boeri et al., 2020).

[7]One of the reason standing behind high prevalence of shift work is relatively high share of manufacturing in employment. Another reason is relatively strict labor code.

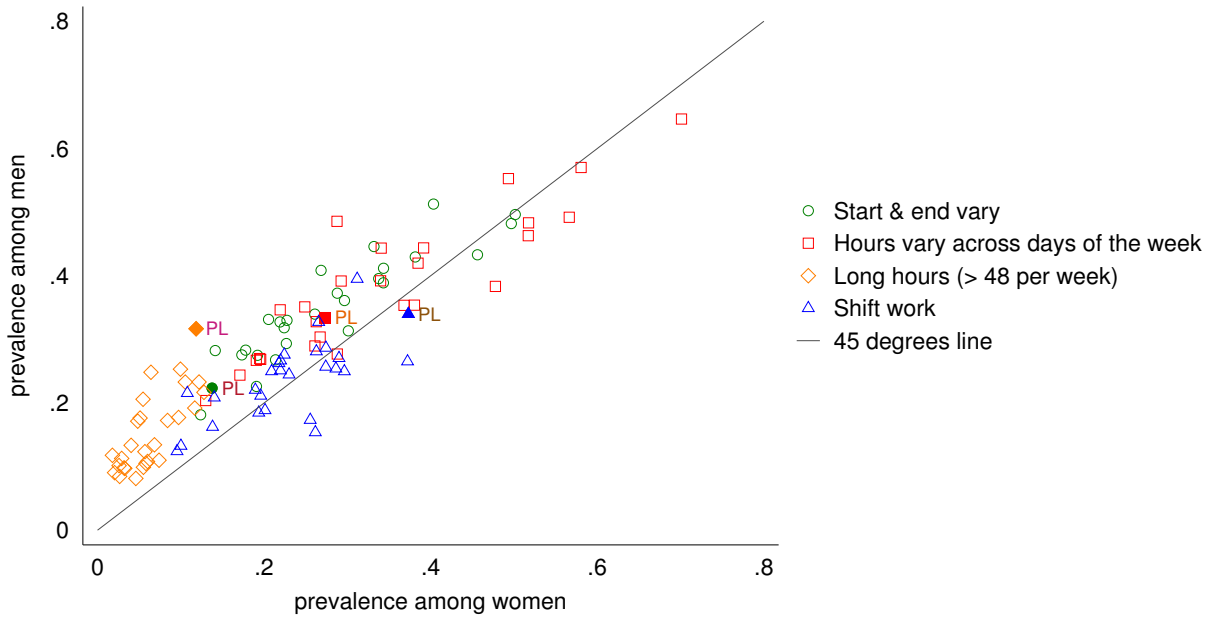
[8]Note that the figures from LFS for Poland and from EWCS do not match exactly, which is to be expected given the methodological differences between these two types of data. However, both sources imply the same ballpark for long hours.

[9]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.

[10]These measures of consistency are slightly higher than those reported in Netemeyer et al. (1996).

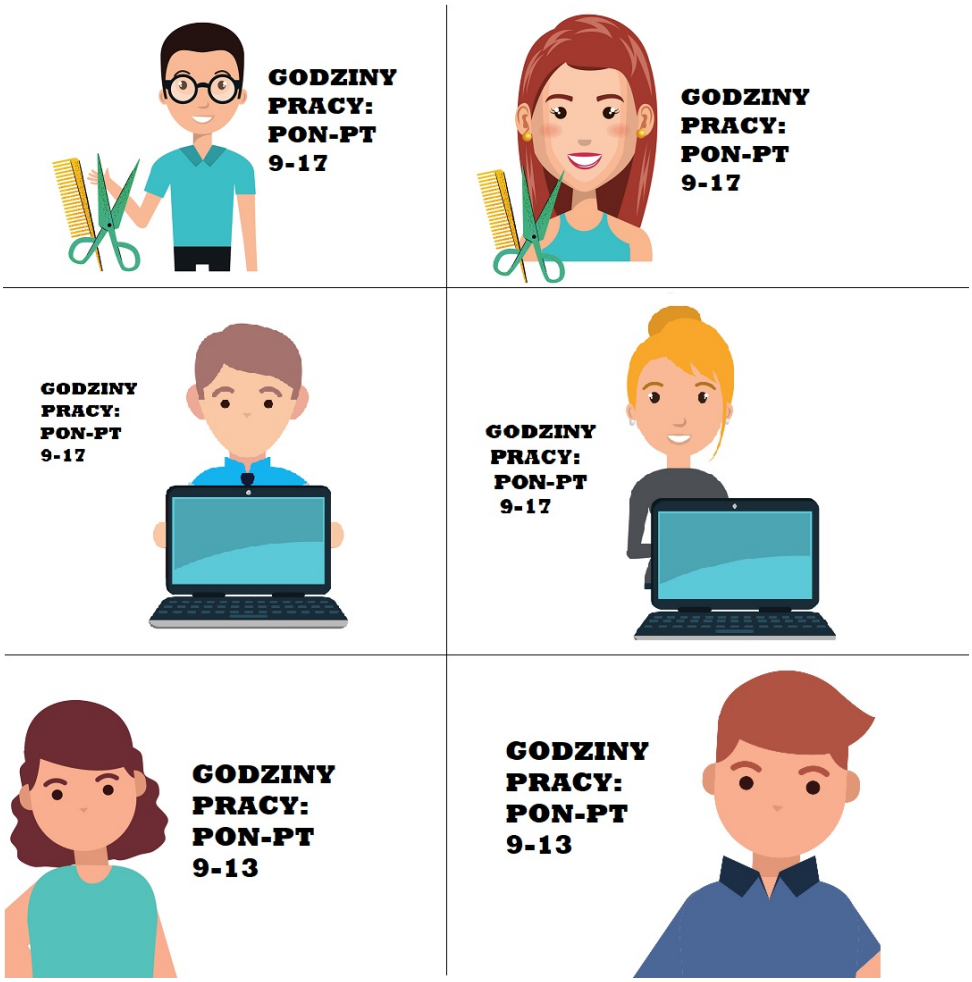
Tables and Figures

Figure 1: Flexible working time arrangements in Europe



Notes: Estimates based on European Working Condition Survey for the year 2015 with country population weights included. Graphs include estimates for: Austria, Belgium, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Greece, Germany, Hungary, Iceland, Italy, Lithuania, Latvia, Netherlands, Norway, Poland (marked with labels and full markers), Portugal, Spain, Sweden, Slovakia and Slovenia.

Figure 2: 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).

Table 1: Sample descriptive statistics by treatment and manipulation check

	Gender of employee		Initiator	
	Men	Women	Employee	Employer
Subjects who passed all manipulation checks				
Age	38.57	38.45	38.62	38.40
% of women participants	0.50	0.50	0.51	0.49
Managerial experience	0.54	0.55	0.53	0.55
Education				
% with primary education	0.10	0.10	0.10	0.10
% with secondary education	0.42	0.40	0.41	0.41
% with tertiary education	0.49	0.50	0.50	0.49
Income level				
% can afford some luxury	0.42	0.43	0.43	0.42
% can make ends meet	0.46	0.46	0.45	0.47
% cannot afford living	0.12	0.11	0.12	0.11
Passed all manipulation checks	0.59	0.60	0.59	0.60
Subjects who failed at least one manipulation check				
Age	38.88	38.65	38.87	38.64
% of women participants	0.53	0.51	0.52	0.51
Managerial experience	0.54	0.54	0.53	0.56
Education				
% with primary education	0.06	0.07	0.05	0.07
% with secondary education	0.39	0.37	0.38	0.38
% with tertiary education	0.55	0.56	0.56	0.54
Income level				
% can afford some luxury	0.41	0.42	0.43	0.40
% can make ends meet	0.46	0.45	0.44	0.48
% cannot afford living	0.12	0.13	0.13	0.12

Notes: Table reports the socio-economic characteristics for 321 individuals participating in the experiment. Note that subjects reported answers to these questions *after* the experimental part. The top panel reports averages for all the participants. The bottom panel reports the averages for those participants who answered correctly all nine manipulation check questions.

Table 2: 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*	$\bar{y}$	$t-stat$		$\bar{y}$	$t-stat$	$\bar{y}$	$t-stat$	
All subjects										
Negative change	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
Positive change	0.363	16.51***	0.159	9.56***	0.20***	0.258	12.76***	0.263	13.27***	-0.01
in USD / month	28.60	4.75***	5.895	1.71*	22.70***	19.86	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
Subjects who passed all manipulation checks										
Negative change	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
Positive change	0.385	13.35***	0.106	5.78***	0.28***	0.239	9.30***	0.252	9.93***	-0.01
in USD / month	29.79	10.31***	7.441	3.15***	22.35***	17.61	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

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 to 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 3: Change in wage subsequent the change in WTAs: experimental results

	in USD per month		in %		in USD per month		in %	
	(1)	(2)	(3)	(4)	(1a)	(2a)	(3a)	(4a)
	All subjects				Subjects who passed all manipulation checks			
T: employer = 1	25.35***	22.11***	3.40***	3.12***	18.05***	17.27***	2.87***	2.94***
	(8.16)	(6.82)	(1.00)	(0.93)	(6.54)	(4.87)	(0.69)	(0.56)
T: woman= 1	-2.16	-3.93	0.35	0.03	-0.11	-2.97	0.44	0.14
	(8.60)	(6.57)	(1.05)	(0.82)	(5.30)	(4.87)	(0.54)	(0.37)
T: employer × woman	-3.02	0.17	-0.75	-0.32	8.96	10.23*	-0.02	0.11
	(19.74)	(13.46)	(1.98)	(1.44)	(7.43)	(6.16)	(0.93)	(0.62)
V: lawyer = 1	14.80	14.96	0.19	0.20	8.84	8.83*	-0.45	-0.45
	(11.78)	(9.56)	(0.93)	(0.76)	(6.24)	(5.13)	(0.48)	(0.40)
V: salesperson = 1	4.15	4.24	2.27**	2.28**	0.32	0.15	1.88***	1.86***
	(6.99)	(5.75)	(1.12)	(0.92)	(3.43)	(2.80)	(0.64)	(0.52)
Age		0.44		0.04		0.04		0.01
		(0.30)		(0.03)		(0.24)		(0.03)
Female subject		-3.51		-0.73		-10.27**		-0.98*
		(9.53)		(1.03)		(4.90)		(0.58)
No managerial experience		5.22		0.38		5.79		0.72
		(9.84)		(1.06)		(5.39)		(0.65)
Educ: secondary		2.62		-0.35		1.49		-0.22
		(7.03)		(1.10)		(6.71)		(1.01)
Educ: tertiary		6.80		0.16		11.52		0.89
		(10.98)		(1.28)		(7.09)		(1.02)
Income: enough		-1.39		-0.31		0.32		0.25
		(9.66)		(1.07)		(5.31)		(0.64)
Income: not enough		10.25		1.99		9.88		1.34
		(11.94)		(1.53)		(8.46)		(1.12)
Intercept	0.10	-21.24	-0.40	-1.66	4.34	-1.85	0.22	-0.74
	(7.27)	(18.95)	(0.89)	(1.94)	(4.30)	(11.28)	(0.51)	(1.58)
Observations	963	963	963	963	570	570	570	570
R^2	0.46	0.02	0.50	0.03	0.54	0.10	0.58	0.13

Notes: Table presents results of OLS 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 secondary 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 USD per month, whereas in Columns (3) and (4) the dependent variable is percentage change relative to status quo wages in each vignette. Columns (1) and (3) include subject fixed effects. Columns denoted by (a) have the analogous structure for the sample of individuals who passed all the 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 4: Change in wage subsequent the change in WTAs: augmented approach

	Dependent variable: Wage change in USD per month						Dependent variable: Wage change in % of status quo					
	(1)	(2)	(3)	(4)	(5)	(6)	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
T: employer = 1	17.22***	17.27***	16.97***	17.65***	17.30***	17.36***	2.93***	2.94***	2.89***	2.98***	2.94***	2.95***
	(4.92)	(4.86)	(4.88)	(4.88)	(4.87)	(4.91)	(0.56)	(0.56)	(0.56)	(0.56)	(0.56)	(0.56)
T: woman= 1	-2.98	-2.98	-3.17	-3.11	-3.05	-3.13	0.14	0.14	0.11	0.13	0.13	0.12
	(4.89)	(4.95)	(4.89)	(4.89)	(4.88)	(4.88)	(0.37)	(0.38)	(0.37)	(0.38)	(0.37)	(0.37)
T: employer × woman	10.34*	10.25	10.51*	10.06	10.15	10.46*	0.14	0.12	0.16	0.09	0.10	0.15
	(6.15)	(6.25)	(6.13)	(6.17)	(6.20)	(6.01)	(0.62)	(0.63)	(0.63)	(0.63)	(0.63)	(0.62)
Own-time availability	4.38						1.15					
	(19.50)						(2.51)					
Gender norm		0.27						0.11				
		(3.17)						(0.36)				
Work-family conflict			3.47						0.54*			
			(2.47)						(0.30)			
Family-work conflict				7.80***						0.88**		
				(2.81)						(0.36)		
Preference for inequality					7.14						1.09	
					(5.86)						(0.80)	
Importance: away from work						-1.28						-0.19
						(1.38)						(0.17)
Importance: towards work						0.39						0.01
						(1.23)						(0.14)
Intercept	18.79	19.43	16.85	14.98	16.98	22.51	1.29	1.44	1.06	0.97	1.09	2.34
	(15.85)	(14.89)	(15.32)	(14.64)	(15.12)	(21.33)	(2.10)	(1.98)	(2.04)	(2.00)	(1.99)	(2.60)
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 OLS regressions subjects' recommendation about the change in wages portrayed by equation (2). 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 3. The estimates reported for individuals who passed all the manipulation checks. Results for the full sample available upon request. Hairdresser story is the base level for the reported coefficients on lawyer and salesperson. 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 5: Beliefs about the social norm (marginal effects)

Majority ...	... agrees		... in favor of a change		in favor of a raise	
T: employer = 1	-0.06**	-0.06	0.21***	0.25***	0.22***	0.27***
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
T: woman= 1	0.01	0.08	0.03	-0.10	-0.03	-0.13
	(0.06)	(0.08)	(0.08)	(0.10)	(0.08)	(0.10)
T: employer = 1 × T: woman= 1	-0.01	-0.04	0.00	0.07	0.02	0.09
	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)
V: lawyer = 1	0.08***	0.08**	-0.14***	-0.09**	-0.12***	-0.08**
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
V: salesperson = 1	0.13***	0.12***	-0.14***	-0.08**	-0.10***	-0.04
	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Age	0.00	0.00*	-0.00	-0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female subject	0.02	0.02	-0.09**	-0.07	-0.06	-0.06
	(0.03)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Managerial experience	0.00	-0.01	-0.01	0.03	-0.01	0.03
	(0.03)	(0.05)	(0.04)	(0.06)	(0.04)	(0.05)
Educ: secondary	0.03	0.07	-0.08	-0.04	-0.06	-0.01
	(0.06)	(0.10)	(0.07)	(0.09)	(0.07)	(0.09)
Educ: tertiary	-0.02	-0.00	-0.02	0.04	0.01	0.05
	(0.06)	(0.10)	(0.08)	(0.09)	(0.07)	(0.09)
Income: enough	-0.06*	-0.05	-0.00	0.04	0.00	0.04
	(0.03)	(0.05)	(0.04)	(0.06)	(0.04)	(0.05)
Income: not enough	-0.11**	-0.14*	0.13*	0.16*	0.07	0.12
	(0.06)	(0.08)	(0.07)	(0.09)	(0.07)	(0.09)
No change in wages	0.12	0.22				
	(0.09)	(0.20)				
Positive change	0.18**	0.27				
	(0.09)	(0.20)				
Observations	963	570	963	570	963	570
P(y=1)	0.830	0.802	0.382	0.360	0.328	0.326

Notes: Table reproduces specifications from Table 3 but with different outcome variables. All specifications include the full set of X's (available upon request). Columns (1) and (1a) report the estimates where the dependent variable is the belief about congruence between own response and social norm. This variable takes on the value of 1 when the subject reported congruence and 0 otherwise. 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 change the wage of the character in the given vignette, and 0 otherwise. Finally, in columns (3) and (3) 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 sample of subjects who passed all the manipulation checks, otherwise the estimates are 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.

Table A1: Determinants of passing the manipulation checks (marginal effects)

	By individual (1)	By vignette (2)
Age	0.00	0.00
	(0.00)	(0.00)
Female subject	0.01	0.01
	(0.06)	(0.03)
Managerial experience	0.03	0.04
	(0.06)	(0.04)
Educ: secondary	0.21**	0.10
	(0.10)	(0.07)
Educ: tertiary	0.34***	0.21***
	(0.10)	(0.07)
Income: enough	0.03	0.06*
	(0.06)	(0.04)
Income: not enough	0.12	0.09*
	(0.09)	(0.05)
T: employer = 1		0.03
		(0.03)
T: woman= 1		-0.01
		(0.03)
treatment_int=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 then secondary education is a base level. Income sufficient to cover some luxury is the base level. Hairdresser is the base level. Column (1) presents the restriction used in the preferred specifications, i.e. whether subject missed 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.