

Discussion Paper Series

IZA DP No. 18976

September 2026

The Price of Unpredictability: Scheduling Demands, Gender and Selection in the Gig Economy

Natasha Jha

Centre for Advanced Financial
Research and Learning

Martin Abell

Bowdoin College
and IZA@LISER

Patrizio Piraino

University of Notre Dame

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.



The Price of Unpredictability: Scheduling Demands, Gender and Selection in the Gig Economy^{*}

Abstract

We study how scheduling demands shape gig-work supply and its composition in an experiment with U.S. gig workers, eliciting reservation wages for contracts varying in advance notice, timing, and duration. No-notice contracts raise reservation wages by 13–14 percent, while outside-hours work with notice carries little penalty. Women require 5–6 percent more than men for outside-hours or no-notice work, and 8–9 percent more when combined. These gaps are largest among workers with heavy household-chore burdens and inflexible schedules. Higher-productivity workers also require larger no-notice premiums. At a given wage, demanding contracts therefore reduce labor supply and select a lower-productivity labor pool.

JEL classification

J22, J16, J31, J41, C93

Keywords

scheduling demands, gig economy, gender gap, reservation wages, worker selection

Corresponding author

Martin Abel

m.abel@bowdoin.edu

^{*} Our experiment was pre-registered, including a pre-analysis plan, at AEARCTR-0016581 and approved by the Bowdoin College Institutional Review Board. The authors have no relevant conflicts or financial interests to disclose. Vharun Boopathy and Prakriti Shakya provided superb research assistance.

1 Introduction

The gig economy is often characterized by flexibility: workers can choose when to work, which tasks to accept, and how much labor to supply. Yet, the most valuable gig opportunities may require workers to give up precisely this flexibility. Higher-paying contracts often demand immediate availability, responsiveness to client needs, work outside standard hours, or sustained commitment over time. Workers who differ in their capacity to absorb these scheduling demands can earn systematically different returns from the same gig work.

This tension connects gig work to a long-standing literature on “greedy” work. [Goldin \(2014\)](#) argues that jobs rewarding continuous availability, on-call responsiveness, and long-hours commitment pay disproportionately more, and that these premia account for a substantial share of persistent gender earnings gaps. Workers with greater non-market obligations face higher costs of supplying these features, and these obligations remain unevenly distributed by gender. Recent evidence suggests that similar forces operate in gig labor markets: gender earnings gaps persist even in relatively standardized platform settings such as Uber and Amazon Mechanical Turk ([Cook et al., 2021](#); [Litman et al., 2020](#)). The existing evidence, however, does not attempt to estimate how much workers value specific dimensions of scheduling flexibility. It also leaves open whether these demands screen out higher-ability workers, altering the pool of talent available to firms.

We study these questions in an online experiment with 969 U.S.-based gig workers. Participants complete a baseline task and then make incentive-compatible choices on follow-up contracts. These contracts randomly vary along three dimensions: whether workers receive advance notice of the task date, whether the task must be completed during standard working hours (or outside them), and whether the contract consists of a one-time task or a repeated engagement. Because the task itself is held fixed, differences in reservation wages identify the compensation premium workers require to accept specific scheduling demands. This design allows us to price two central components of greedy work within the growing gig economy: (i) unpredictability and (ii) non-standard timing. We test whether flexibility penalties differ by gender and use measures of household responsibilities, childcare, and baseline schedule flexibility to examine whether any gap is explained by parental status or the broader domestic workload. We also measure several dimensions of worker productivity to assess how contractual terms shape the talent composition of the labor pool available to firms.

The empirical analysis delivers four main results. First, workers place substantial value on advance notice: contracts that do not specify the task date raise reservation wages by approximately 13–14 percent relative to otherwise comparable contracts performed

during standard hours with notice. Outside-hours work, by contrast, imposes little average penalty when notice is provided, and adding outside-hours scheduling to no-notice work does not compound the penalty. In this setting, unpredictability is therefore more costly than non-standard timing itself.

Second, flexibility penalties are systematically larger for women. Women have lower reservation wages than men for the least demanding contracts but require relatively more compensation as scheduling conditions become more demanding. The additional female penalty is 5–6 percent for contracts that involve either outside-hours work or no advance notice, and it rises to 8–9 percent when no-notice work is combined with outside-hours scheduling. This gender difference is concentrated at low and intermediate wage levels, where demanding scheduling is most likely to push women below their reservation wages.

Third, the gender gap in flexibility penalties is more consistent with binding constraints than with stronger intrinsic preferences against demanding schedules. Three types of evidence support this reading. In post-elicitation surveys, women are no more likely than men to report disliking demanding schedules. They are more likely than men to cite household responsibilities as a barrier to outside-hours and no-notice work. The household-responsibilities gap persists after conditioning on marital and parental status, whereas the raw gender gap in citing childcare largely disappears. The reservation-wage estimates point the same way: the female flexibility penalty is largest among workers with high household-chore intensity and less flexible baseline schedules. Behavioral evidence from the follow-up task corroborates this pattern. Women assigned to the most demanding contracts are less likely to complete them ($p < 0.10$) and, conditional on completion, adjust their routines more often to accommodate the work (though this estimate is imprecise given the small completion subsample).

Fourth, flexibility penalties have implications beyond worker-side costs. If demanding contracts disproportionately deter more productive workers, firms face a smaller and less productive labor pool. We find that they do. Using a pre-specified index combining standardized test scores, performance on a Raven fluid-reasoning task, and quality ratings of a baseline writing task, a one-standard-deviation increase in productivity raises the penalty for no-notice contracts by about 5 percent for within-hours work and 9 percent for outside-hours work. The gradient is concentrated in writing quality and standardized test performance, and does not differ significantly by gender. Because women face larger flexibility penalties overall, however, the female share of workers willing to accept demanding contracts falls relative to less demanding arrangements, and this shift extends across the productivity distribution.

The paper contributes to four strands of research. First, it builds on work valuing job amenities

and the costs of demanding work arrangements. In the canonical framework of compensating differentials, workers trade off wages against non-wage job attributes. Credible estimates of these tradeoffs require variation in job characteristics that is uncontaminated by worker sorting or unobserved productivity. Recent research has made substantial progress in estimating the value of workplace flexibility—including work from home, schedule control, flexible hours, and employer discretion over scheduling—using field experiments and stated-preference designs. (Maestas et al., 2023; Wiswall and Zafar, 2018; He et al., 2021). Most closely related to our setting, Mas and Pallais (2017) estimate that workers forgo roughly 20 percent of wages to avoid employer-controlled schedules.¹ Our contribution is to decompose demanding work into distinct scheduling attributes and estimate their separate prices within an incentivized common-task design. We isolate two dimensions central to platform-mediated work but rarely priced separately: outside-hours timing and advance notice of the task date. Our results show that predictability and timing impose distinct costs, with implications for platform designs that reduce scheduling burdens without necessarily eliminating non-standard hours.

A second contribution is to the literature on gender gaps in gig and online labor markets. Prior work shows that gender earnings gaps persist even in platforms with standardized tasks and limited formal wage-setting. In ride-hailing, Cook et al. (2021) document that a gender earnings gap emerges despite gender-neutral pay algorithms, driven by differences in experience, location choices, and driving speed. In online labor markets, women are often concentrated in lower-paid project types and underrepresented in higher-paying technical categories (Litman et al., 2020; Digital Future Society, 2022). We contribute by pricing a worker-side mechanism that may help explain these gaps: the gender differential in the compensation workers require to accept demanding schedules, holding the task fixed.

We also speak to research on household constraints, motherhood, and gendered labor supply. A large literature shows that children and household production shape gender gaps in labor-market outcomes, including earnings, hours, occupational choice, and the demand for flexible work (Kleven et al., 2019b,a; Cortés and Pan, 2019). Related work emphasizes that gender norms and household specialization make women disproportionately responsible for absorbing non-market demands, even when men and women have similar labor-market attachment (Bertrand et al., 2015; Bursztyn et al., 2020). We contribute by distinguishing childcare from the broader domestic workload as sources of schedule constraints. In our setting, the gender gap in flexibility penalties is not explained by parental status or childcare

¹Related work shows that flexibility is often bundled with other job attributes. Felfe (2012) shows that women adjust working hours and other job amenities around motherhood, explaining part of the motherhood wage gap. Kesternich and De Schouwer (2025) document that women and parents value flexibility more but do not necessarily work in more flexible jobs, partly because meaningful work is less flexible.

intensity, and is largest among workers with high household-chore burdens. Routine domestic labor, rather than motherhood per se, appears to be the binding constraint that makes unpredictable work costly.

Finally, the paper contributes to research on worker sorting, job amenities, and the misallocation of talent. Existing work shows that non-wage job attributes shape sorting across firms, occupations, and work arrangements, with important implications for gender gaps and aggregate productivity (Card et al., 2016; Sorkin, 2018; Morchio and Moser, 2026).² A closely related literature shows that demanding work arrangements affect selection into high-paying occupations. Wasserman (2023) finds that capping work hours in medical residency increased women’s entry into affected specialties. Other evidence shows that the productivity effects of flexible arrangements depend on selection into the arrangement itself (Bloom et al., 2015), and recent work links worker preferences over job attributes to objective productivity measures (Johnston, 2025). We add to this literature by testing whether scheduling demands select workers along different dimensions of productivity within an otherwise identical task. The focus shifts from sorting across firms or occupations to whether specific schedule terms alter the productivity composition of the workers willing to accept a given contract. Firms may therefore face a tradeoff between operational flexibility and the productivity of their available labor pool.

Our results point to a concrete design margin for platforms and firms. The largest costs in our setting arise from unpredictability rather than non-standard timing. This distinction matters because some forms of demanding work may be difficult to eliminate, while notice periods, task windows, and deadline structures are often design choices. Greater predictability may therefore expand the pool of workers willing to accept demanding contracts and reduce the gendered cost of taking them on.

2 Gig Work and the Cost of Demanding Schedules

The gig economy has grown into a substantial segment of the US labor market. As of 2023, over a third of the US workforce earned some income through gig arrangements, up from roughly 27 percent in 2016, with freelance workers contributing an estimated \$1.27 trillion to the US economy (Upwork, 2023; McKinsey & Company, 2022). This growth has

²Related work shows that other labor-market frictions can also generate gendered misallocation, including information frictions in hiring, evaluation, and recognition of worker performance (Sarsons et al., 2021; Exley and Nielsen, 2024; Abel et al., 2020). Our setting focuses on a different margin: whether scheduling demands themselves change the composition of workers willing to accept a given contract.

been sustained across economic cycles rather than peaking with pandemic-era disruptions, reflecting a durable shift in how work is organized.

This sector of the economy spans three broad segments. In-person gigs (such as ride-hailing, delivery, and manual tasks) are location-bound and governed by algorithmic dispatch: workers accept or reject individual assignments in real time, with no meaningful advance notice. Online freelancing platforms facilitate skilled, project-based remote work in software development, design, writing, and consulting, with longer engagements and contracts negotiated between worker and client. Microtask platforms occupy a distinct niche, hosting short, discrete tasks that require human judgment and are completed entirely online, typically within minutes to an hour.

2.1 Greedy Work in the Gig Economy

A growing body of work documents that occupations rewarding continuous availability, on-call responsiveness, and long-hour commitments pay disproportionately more than otherwise similar jobs without these features. [Goldin \(2014\)](#) characterizes such occupations as “greedy” and argues that the convex returns to availability and commitment account for a substantial share of persistent gender earnings gaps in traditional labor markets, since women’s greater non-market obligations make these features more costly to supply. Subsequent work confirms that the wage premium for demanding schedules is large and that women sort away from it ([Wasserman, 2023](#)).

Many higher-paying gig arrangements reward a similar set of features: availability, responsiveness, and commitment. The relevant form of “greediness,” however, varies across segments. In-person dispatch gigs, such as ride-hailing and delivery, reward workers who can accept assignments at peak times and on short notice; [Cook et al. \(2021\)](#) document that peak-hour earnings premiums account for a meaningful share of male drivers’ earnings advantage on Uber. Microtask platforms reward temporal responsiveness without sustained client relationships: workers must monitor task availability and complete assignments within short windows after they appear. Online freelancing, by contrast, often requires more sustained commitment. Higher-return projects may involve repeated client interaction, responsiveness over multiple days or weeks, binding deadlines, and reputational penalties for non-completion. Consistent with this distinction, Upwork reports that a large share of freelancer earnings comes from extended client relationships rather than one-off projects ([Upwork, 2023](#)).

Across these segments, compensation is therefore tied not only to skill or task content but

also to workers’ ability to supply availability and commitment. This raises the question of who can supply greedy work and at what cost. The same features that command higher pay impose costs on workers whose non-market obligations limit their capacity to absorb scheduling demands. If those obligations are distributed unequally, compensation for greedy work can generate systematic earnings differences even under identical contract terms.

2.2 Gender Differences in the Costs of Demanding Work

Women are underrepresented in many of the higher-paying segments of the online gig economy. Technical and software-related freelance categories remain male-dominated, while women are more concentrated in lower-paid categories such as writing, translation, and administrative support ([International Labour Organization, 2021](#); [Digital Future Society, 2022](#)). To the extent that higher-paying categories require greater availability and commitment, this sorting may itself contribute to gender earnings gaps.

Gender earnings gaps also appear within relatively standardized gig settings. In ride-hail work, [Cook et al. \(2021\)](#) show that men earn about 7 percent more than women on Uber despite common platform rules and no employer wage-setting, with a substantial part of the gap attributable to when and where drivers choose to work. In online microtask work, [Litman et al. \(2020\)](#) document a 10.5 percent gender earnings gap on MTurk that persists after accounting for observable differences in task selection and human capital.³

One channel running through this evidence is that women disproportionately bear household and care responsibilities, which raises the cost of supplying long hours, employer-controlled schedules, and continuous availability ([Goldin, 2014](#); [Cubas et al., 2023](#); [Berniell et al., 2021](#)). This evidence points to scheduling as an important margin in gig-work earnings gaps. What remains unclear is how much compensation workers require for specific scheduling attributes, and whether women require more.

³Beyond gig settings, [Cortés et al. \(2023\)](#) combine field evidence from MBA students with laboratory job-choice experiments to document that women accept jobs at lower wages than men and require greater compensation for jobs offering less workplace flexibility.

3 Study Design

3.1 Worker Recruitment and Participant Characteristics

To isolate the price of flexibility, we ran an online task-market experiment in which the work itself was held fixed and only the scheduling terms varied across contracts. Between August and September 2025, we recruited 1,289 U.S.-based gig workers aged 25 to 42 from Amazon’s Mechanical Turk (MTurk).⁴ MTurk suits our design because its workers are experienced with paid online tasks. To improve data quality, we recruited through CloudResearch, which screens MTurk participants and reduces the risk of bots and low-quality responses (Hauser et al., 2023).⁵ Of the 1,289 recruited workers, 174 (13.5%) failed our attention check, and 145 of the remaining 1,114 declined the follow-up work described below. This leaves an analysis sample of 969 participants.⁶

Table 1 reports summary statistics and pairwise gender comparisons. The sample is roughly gender-balanced, with a mean age of 35 and high educational attainment (89 percent hold a tertiary degree). Educational attainment and performance on the Raven-fluid reasoning task are balanced across genders, as are online work experience (13 years on average) and weekly hours on their internet job (roughly 16). Women are less likely than men to hold a non-internet job (52% versus 61%, $p < 0.01$), more likely to be married (49% versus 35%, $p < 0.01$) and more likely to have children (53% versus 33%, $p < 0.01$). Domestic responsibilities differ significantly. Women report 2.7 hours per day on home chores versus 1.9 hours for men, and 2.6 hours per day on other household tasks versus 1.8 ($p < 0.01$). The childcare gap is larger still: 4.2 hours per day for women compared to 1.3 for men. Women in the sample therefore bear a larger share of domestic labor, which may reduce their capacity to absorb unexpected schedule shocks.

3.2 Experimental Design

The study proceeds in three stages, summarized in Figure 1. First, workers complete a baseline task that measures performance and indicate whether they are interested in follow-up

⁴Our initial plan was to have a minimum age of 28. To help reach our target sample size, we lowered it to 25.

⁵Recruitment was further restricted to participants with an MTurk approval rate of 97% or more, and at least 1,000 approved HITs. As specified in our pre-analysis plan, we also excluded participants who expressed interest in a follow-up contract paying only 0.30 per hour, which we treat as an indicator of inattention.

⁶Data collection took place across five waves. Summary statistics for each wave are reported in Table A1.

Table 1: Gender Balance - Demographic and Work

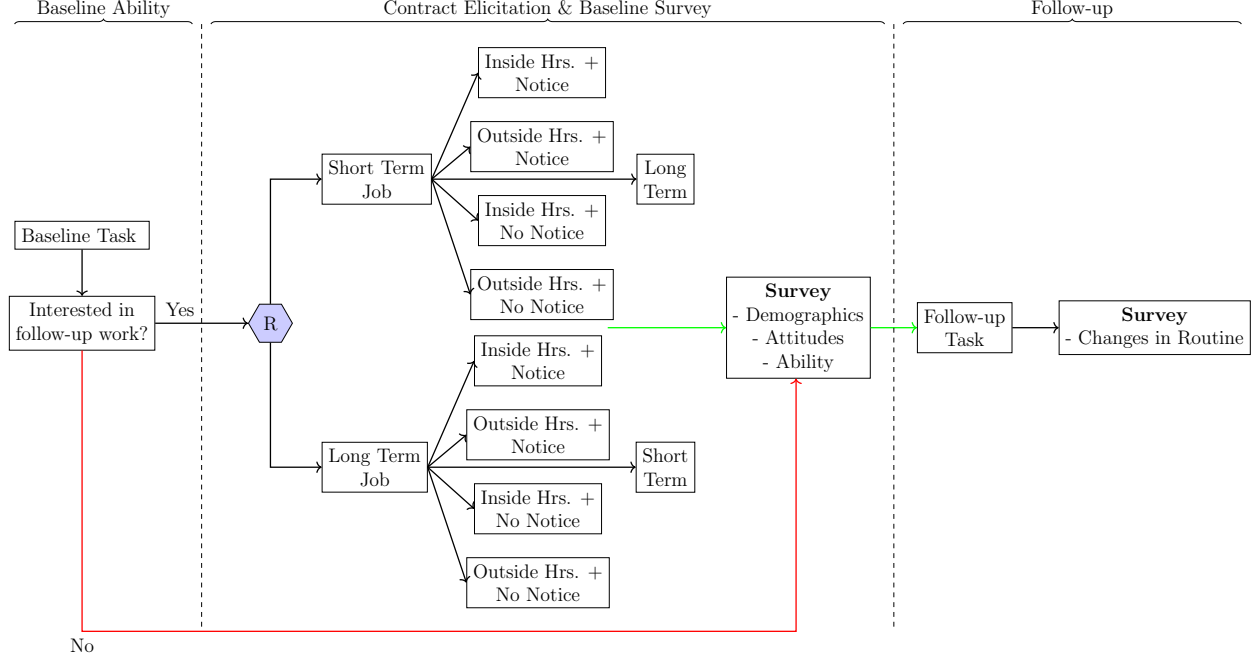
Variable	(1) Total Mean/(SE)	(2) Male Mean/(SE)	(3) Female Mean/(SE)	(2)-(3) Pairwise t-test Mean difference
Age	35.299 (0.161)	35.146 (0.228)	35.446 (0.229)	-0.301
White	0.784 (0.013)	0.755 (0.020)	0.812 (0.018)	-0.057**
Married	0.419 (0.016)	0.348 (0.022)	0.487 (0.022)	-0.139***
Has Child	0.430 (0.016)	0.331 (0.022)	0.525 (0.022)	-0.194***
Tertiary Edu	0.893 (0.010)	0.888 (0.014)	0.897 (0.014)	-0.009
Raven Score	4.655 (0.082)	4.730 (0.118)	4.584 (0.114)	0.146
Yrs Experience	13.256 (0.206)	13.332 (0.294)	13.183 (0.289)	0.149
Job Non-Internet	0.562 (0.016)	0.605 (0.022)	0.521 (0.022)	0.084***
Hours working Internet	15.598 (0.483)	16.039 (0.695)	15.174 (0.671)	0.865
Regular Hours Intern	0.419 (0.016)	0.437 (0.023)	0.402 (0.022)	0.035
Home chores (h)	2.319 (0.062)	1.910 (0.069)	2.709 (0.098)	-0.798***
Childcare (h)	2.781 (0.157)	1.297 (0.117)	4.207 (0.271)	-2.910***
Other (h)	2.229 (0.104)	1.831 (0.109)	2.613 (0.174)	-0.781***
F-test of joint significance (F-stat)				16.760***
F-test, number of observations				934

Notes: This table reports the mean, standard error and t-test (difference in the means across groups) for the overall sample and by gender groups. Column (1) reports the mean and standard error for the overall sample while Columns (2) and (3) report the mean and standard error for Males and Females respectively. Column (4) reports the difference in means between males and females ((2)-(3)). The table also reports the F-stat for joint significance across all balance variables. The number of observations used in the F-test is lower than the size of the analytical sample due to missing values in some of the variables. ***, **, * mean statistical significance at 1, 5, 10 percent levels.

work. Second, interested workers make a series of binary choices over contracts varying in scheduling terms and wages, and then complete a survey covering demographics, attitudes, domestic responsibilities, and self-reported barriers to accepting demanding contracts. Third, each worker is randomly assigned one of 56 contracts and invited to complete it only if their earlier choices revealed a willingness to accept it. An endline survey then measures behavioral

adjustments, perceived burden, and retention. Each stage is described below.

Figure 1: Experimental Design



3.2.1 Baseline Task

Participants first complete a baseline task: they watch a short news video clip and write a 4–5 sentence summary highlighting a key message. The task and accompanying survey take approximately 15 minutes and pay a fixed \$3.50, an hourly rate of \$14. Written summaries are rated on a 0–10 scale by a separate sample of evaluators, following the procedure in Appendix B.2. We also record completion times for the writing component. Participants then answer two factual questions about the video and indicate whether they are interested in follow-up work. Those who decline proceed to the remaining survey modules; those who express interest advance to the contract elicitation, where they make incentivized choices over follow-up contracts.

3.2.2 Contract Variations

The follow-up contracts vary along three dimensions: scheduling terms, wage, and duration. Scheduling terms have two components. The first is timing: tasks are completed either during

standard hours (9am–5pm) or outside them (7pm–midnight). The second is advance notice: workers either learn the task date three days in advance, or they know only that the task will arrive sometime within the next week. Cross-randomizing the two components yields four treatment arms: within hours with advance notice (T1), outside hours with advance notice (T2), within hours with no advance notice (T3), and outside hours with no advance notice (T4).

Contracts are offered at seven hourly wage levels, from \$4 to \$10 in \$1 increments.⁷ Duration takes one of two forms: a one-off contract with a single 10-minute task, or a five-week contract with a two-hour weekly commitment. The task itself is identical across durations; the number of hours, engagement length and total compensation differ.⁸

3.2.3 Elicitation Procedure

Before making contract decisions, participants are told that one of their choices will be randomly selected and implemented. This follows the logic of BDM-style elicitation: because each decision is potentially payoff-relevant, participants have an incentive to evaluate every contract as if it could determine their follow-up offer. They are further informed that some contracts are available in larger quantities and are therefore more likely to be selected, that rejecting one contract does not increase the probability of being offered another, and that their fixed payment for the current survey does not depend on their contract choices.

During elicitation, participants learn each contract’s timing window, advance-notice condition, duration, and wage, but not the exact task date or implementation details, which are revealed only after assignment. The design elicits valuations of scheduling attributes rather than responses to known calendar conflicts.

Each participant is randomly assigned to either the one-off or five-week duration (see Figure 1) and completes the main elicitation within it, with the order of treatment arms randomized. Within each arm, elicitation begins at \$7 per hour and adjusts upward after a rejection and downward after an acceptance, yielding three binary decisions that bracket the switching point.⁹ Participants then report their exact reservation wage at the switching point. After completing all four arms in their assigned duration frame, participants evaluate one contract

⁷These rates correspond to payments of approximately \$0.67, \$0.83, \$1.00, \$1.17, \$1.33, \$1.50, and \$1.67 for a 10-minute task.

⁸Workers also see a single contract under T1 conditions paying \$0.30 per hour, which serves as a consistency check.

⁹Compared to a full price list, this adaptive procedure reduces respondent burden and limits mechanical response patterns while preserving incentive compatibility, as each decision may still be implemented. It recovers the reservation wage without presenting every wage-treatment combination.

from each arm at \$7 under the alternative duration (Figure 1). These choices provide a common-wage comparison across durations but do not enter the main reservation-wage measure.

Following the contract elicitation, participants complete the remaining survey modules, which collect demographic information, household characteristics, domestic responsibilities, self-reported barriers to accepting demanding contracts, and attitudes toward gender norms around flexible work. They also complete Raven Progressive Matrices and a verbal fluency task, and report standardized test scores.¹⁰ These measures enter the productivity index used in Section 6.

3.2.4 Contract Implementation and Follow-up

Following the elicitation stage, participants are randomly assigned to one of the 56 contracts defined by the interaction of duration, treatment arm, and wage level, with assignment stratified by gender and baseline schedule flexibility.¹¹ Participants receive a follow-up offer only if they accepted their assigned contract during elicitation. Of the 985 participants assigned a follow-up contract, 628 (63.8 percent) accepted it and were therefore offered follow-up work (Table A1).¹² Baseline covariates are balanced across the four treatment arms (Appendix Table A2).

Participants offered follow-up work receive two messages before the task. The first confirms the contract and reveals the implementation details of the assigned scheduling terms. Participants in the advance-notice arms (T1 and T2) receive the specific date on which the task will be sent, while those in the no-notice arms (T3 and T4) are told only that it will arrive within the next few days. The second message launches the task, at which point workers have until 5:00 PM local time in the within-hours arms, or midnight in the outside-hours arms, to complete it. Appendix Table B1 reports the exact wording. The follow-up task is identical to the baseline task. After submitting it, participants complete an endline survey measuring behavioral adjustments, perceived task burden, and interest in future work.

¹⁰The verbal fluency task asks participants to produce as many words as possible beginning with the letter “Q” in thirty seconds, excluding proper nouns and repetitions; the score is the number of valid words. Raven Progressive Matrices measure fluid reasoning through abstract pattern recognition. Appendix B.2 describes both tasks.

¹¹A small share of participants (0.6 percent) was assigned to five-week contracts for operational reasons.

¹²The follow-up pool is slightly larger than the 969-participant analysis sample because some Wave-1 participants were offered a contract despite not meeting inclusion requirements.

3.3 Empirical Design

3.3.1 Outcomes

We study two families of outcomes: (i) stated preferences elicited during the contract-choice stage; and (ii) realized behavior among workers who received a follow-up offer.

Our primary stated-preference outcome is the reservation wage, which is the point at which a worker’s decision switches from rejection to acceptance (or vice versa), elicited once per treatment arm. Wages are measured in dollars per hour, and all main specifications use a log transformation. For robustness, we replace the elicited reservation wage with the midpoint between the two wage levels bracketing the switch, and alternatively use that interval as the outcome in an interval regression.

Our second stated-preference outcome is a binary indicator for contract acceptance at each wage level. The adaptive procedure yields accept/reject decisions at three wage levels per arm; we recover the remaining levels by imposing monotonicity, so that a worker who accepts at wage w is assumed to accept the same contract at any $w' > w$, and a worker who rejects at w is assumed to reject at any $w' < w$.

Among workers who received a follow-up offer, we observe two outcomes. The first is a binary indicator for task submission, which measures whether workers follow through on contracts they indicated a willingness to accept. The second, defined only for submitters, is whether the worker adjusted their usual routine to complete the task, capturing the realized logistical cost of accommodating scheduling demands.¹³

3.3.2 Empirical Strategy

All primary specifications follow the pre-analysis plan.¹⁴ We first estimate the flexibility penalty associated with demanding contracts:

$$y_{ic} = \alpha + \beta_1 \text{Female}_i + \beta_2 T2_{ic} + \beta_3 T3_{ic} + \beta_4 T4_{ic} + \delta X_i + r_i + \varepsilon_{ic}, \quad (1)$$

where y_{ic} is the log reservation wage of worker i for contract c . The omitted category is T1 (within-hours work with advance notice). The coefficients β_2 , β_3 , and β_4 estimate the penalties for outside-hours work with notice, within-hours work without notice, and outside-hours work

¹³Elicited via a single Yes/No item in the endline survey: “To accomplish this task did you need to make any changes to your usual daily routine?”

¹⁴See [urlhttps://www.socialscisearch.org/trials/16581](https://www.socialscisearch.org/trials/16581).

without notice, respectively. The vector X_i controls for age, household income, and indicators for being white and unmarried. r_i denotes fixed effects for the order in which treatment arms were presented (e.g., “T1–T2–T3–T4” or “T2–T3–T1–T4”), absorbing any systematic effects of the elicitation sequence. Standard errors are clustered at the participant level. The main specification pools one-off and five-week contracts; we also report estimates separately by duration.

To test whether penalties differ by gender, we add interactions between the treatment indicators and a female indicator:

$$\begin{aligned}
y_{ic} = & \alpha + \beta_1 \text{Female}_i + \beta_2 T2_{ic} + \beta_3 T3_{ic} + \beta_4 T4_{ic} \\
& + \gamma_1 (T2_{ic} \times \text{Female}_i) + \gamma_2 (T3_{ic} \times \text{Female}_i) + \gamma_3 (T4_{ic} \times \text{Female}_i) \\
& + \delta X_i + r_i + \varepsilon_{ic}.
\end{aligned} \tag{2}$$

The coefficients γ_1 , γ_2 , and γ_3 capture the additional compensation women require relative to men for each contract type.

We also examine whether these gender differences vary across the wage distribution. Because reservation wages are observed only at the switching point, this analysis uses the binary contract-acceptance outcome described above. We group the seven hourly wage levels into three tiers: low (\$4–\$5), mid (\$6–\$7), and high (\$8–\$10). We estimate equation (2) separately for each tier and test whether the treatment and treatment-by-female coefficients differ across them.

3.3.3 Heterogeneity, Mechanisms, and Selection

We use the same interaction framework (equation 2) to examine mechanisms and selection. First, we ask whether the gender differential in flexibility penalties is concentrated among workers for whom scheduling constraints are more likely to bind. We focus on four pre-specified moderators: baseline schedule flexibility, parental status, childcare intensity, and household-chore intensity. For each moderator, we estimate two sets of models. The first estimates the gender-interaction specification separately within moderator-defined subgroups. The second interacts treatment indicators with the moderator, separately by gender. Together, these distinguish whether a constraint predicts larger flexibility penalties within gender from whether the gender gap itself is larger among workers facing the constraint.

Second, we examine whether scheduling demands affect the productivity composition of the willing labor pool, estimating analogous models in which the moderator is baseline productivity.

Our primary measure is a pre-specified index combining standardized test scores, Raven matrices, and baseline writing quality. We also report estimates for each component separately and for secondary measures of verbal fluency, comprehension, and baseline task time. Positive treatment-by-productivity interactions indicate that higher-productivity workers require larger compensation premia, so that demanding contracts draw a less productive pool at a given wage. Negative interactions indicate the opposite.

4 Results

4.1 Pooled Results

We begin with a pooled analysis of 3,864 contract decisions across treatment arms. In the control condition (T1: within-hours with advance notice), the mean reservation wage is \$8.27 per hour (1.94 in logs, Table 2)), and the overall acceptance rate across wage levels is 53%. These serve as benchmarks for assessing the flexibility penalties reported below.

Contracts requiring outside-hours work but providing advance notice (T2) do not significantly raise reservation wages relative to T1, in the pooled sample or separately for one-off and five-week contracts (Table 2, columns 1-4). Confidence intervals are tight enough to rule out economically meaningful effects. The absence of a timing penalty is notable given how central non-standard hours are to existing accounts of scheduling barriers, and may reflect a distinctive feature of the gig-economy workforce.

By contrast, the absence of advance notice generates a large and precisely estimated flexibility penalty. Relative to T1, contracts requiring within-hours work with no advance notice (T3) raise reservation wages by approximately 14 percent, or around \$1.30 per hour (Table 2, Col. 1-2, $p < 0.01$). Contracts combining no advance notice with outside-hours scheduling (T4) have a nearly identical effect of 13 percent, and we cannot reject equality of the two coefficients ($p = 0.75$). With the near-zero coefficient on T2, these estimates indicate that the two dimensions are not additive: outside-hours scheduling imposes no additional cost once advance notice has been removed.

Splitting the sample by contract duration reverses the relative magnitude of the two no-notice penalties (Col. 3-4). For one-off contracts, the T4 penalty slightly exceeds the T3 penalty (16 versus 13 log points); for five-week contracts, the ordering reverses (10 versus 14 log points, $p = 0.02$). One interpretation is that outside-hours work involves fixed adjustment costs that

Table 2: Flexibility Penalty

	Log(Reported Reservation Wage) (USD/hr)			
	Pooled	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.00 (0.02)	-0.00 (0.01)	0.00 (0.02)	-0.01 (0.02)
No Notice / Within (T3)	0.14*** (0.01)	0.14*** (0.01)	0.13*** (0.02)	0.14*** (0.01)
No Notice / Outside (T4)	0.13*** (0.02)	0.13*** (0.02)	0.16*** (0.02)	0.10*** (0.02)
Female (=1)		-0.11*** (0.04)	-0.08 (0.05)	-0.13*** (0.05)
Participant F.E	Yes	No	No	No
Participant Controls	No	Yes	Yes	Yes
Control Mean	1.94	1.94	1.76	2.13
Observations	3864	3864	1952	1912
R^2	0.82	0.05	0.04	0.09
P-val: T2 = T3	0.00	0.00	0.00	0.00
P-val: T2 = T4	0.00	0.00	0.00	0.00
P-val: T3 = T4	0.75	0.73	0.17	0.02

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. Columns (1) and (2) show pooled regressions, while columns (3) and (4) present results for one-off and five-week job sub-samples respectively. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include age, income, race, and marital status. The bottom three rows report p-values from pairwise tests of equal coefficients within each regression. ***, **, * indicate statistical significance at 1, 5, 10 percent levels.

Source: Waves 1 - 5.

pay off only over a longer engagement, dampening the T4 penalty when the commitment is sustained.

4.2 Treatment Effects by Gender

We next examine heterogeneity by gender. Women's reservation wages for T1 contracts are approximately 16 percent lower than men's in the pooled sample (Table 3, Col. 2), consistent with established evidence of lower reservation wages among women in the United States (Cortés et al., 2023), France (Le Barbanchon et al., 2021), and Germany (Caliendo et al.,

2017).¹⁵

Interacting the treatment indicators with a female dummy shows that women’s reservation wages are not only lower on average but also more sensitive to scheduling demands. All interaction terms are positive and significant in the pooled sample (Table 3). This matters most for T2: outside-hours work with advance notice generates no significant penalty on average, but the positive $T2 \times \text{Female}$ interaction shows the average masks a gender difference, with the penalty about 6 percent larger for women (column 2, $p < 0.05$). The no-notice penalties are also larger for women: 5 percent for T3 and 8 to 9 percent for T4 (columns 1–2). These gender-interaction patterns are robust to alternative outcome specifications: replacing the elicited reservation wage with the midpoint of the switching bracket, estimating an interval regression on that bracket (Table A3), and using a binary acceptance indicator at each wage level (Table A4) yield qualitatively similar gender differences.

The gender interactions are larger and statistically significant for five-week contracts (column 4), while the corresponding coefficients for one-off contracts are smaller and insignificant (column 3). The difference between them is not itself significant, so we treat the contrast as suggestive rather than conclusive. One reading is that the cumulative costs of demanding schedules weigh more heavily on women over repeated engagements.

¹⁵A similar gap appears in online gig markets: women’s earnings on Mechanical Turk are around 10 percent lower than men’s, even absent overt discrimination or labor market segregation (Litman et al., 2020). McGee and McGee (2023) find a gender gap in reservation wages of 14 to 18 percent (in the same range as our estimate) in a laboratory search experiment with North American participants.

Table 3: Flexibility Penalty by Gender

	Log(Reported Reservation Wage) (USD/hr)			
	Pooled	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.03 (0.02)	-0.03 (0.02)	-0.01 (0.03)	-0.05* (0.03)
No Notice / Within (T3)	0.11*** (0.02)	0.11*** (0.01)	0.11*** (0.02)	0.11*** (0.02)
No Notice / Outside (T4)	0.09*** (0.02)	0.09*** (0.02)	0.12*** (0.03)	0.05** (0.03)
Female (=1)		-0.16*** (0.04)	-0.12** (0.05)	-0.19*** (0.05)
T2 $\times$ Female	0.05* (0.03)	0.06** (0.03)	0.03 (0.04)	0.08** (0.04)
T3 $\times$ Female	0.05** (0.03)	0.05** (0.02)	0.04 (0.04)	0.06** (0.03)
T4 $\times$ Female	0.09** (0.04)	0.08*** (0.03)	0.07 (0.05)	0.10** (0.04)
Participant F.E	Yes	No	No	No
Participant Controls	No	Yes	Yes	Yes
Control Mean	1.94	1.94	1.76	2.13
Observations	3864	3864	1952	1912
R-squared	0.82	0.05	0.04	0.09
P-val: T2 = T3	0.00	0.00	0.00	0.00
P-val: T2 = T4	0.00	0.00	0.00	0.00
P-val: T3 = T4	0.31	0.26	0.61	0.01
P-val: T2 $\times$ Fem = T3 $\times$ Fem	0.94	0.85	0.91	0.65
P-val: T2 $\times$ Fem = T4 $\times$ Fem	0.23	0.23	0.34	0.47
P-val: T3 $\times$ Fem = T4 $\times$ Fem	0.30	0.25	0.53	0.29

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and female dummy to test for heterogeneous treatment effects by gender. Columns (1) and (2) show pooled regressions, while columns (3) and (4) present results for one-off and five-week contract sub-samples respectively. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression.

***, **, * indicate statistical significance at 1, 5, 10 percent levels.

Source: Waves 1 - 5.

4.3 Heterogeneity by Wage Level

Because reservation wages are observed only at the switching point, the analysis of gender interactions across the wage distribution uses the binary acceptance outcome, where negative coefficients indicate that women are less likely than men to accept. Table 4 shows that at high wages the coefficients are small and insignificant across all contract types: men and women respond similarly to demanding contracts when compensation is sufficiently high. At low and mid wages, however, the interactions are negative and significant, particularly for T3 and T4: the gender gap in contract acceptance widens as compensation falls. These patterns hold across one-off and five-week contracts (Table B2). The gender gap in flexibility penalties is therefore concentrated at lower wages, where demanding contracts more often push women below their reservation wages.

4.4 Follow-through on Accepted Contracts

The reservation-wage estimates measure ex ante willingness to accept. We next ask whether these elicited penalties predict follow-through. Because workers receive an offer only for a contract they accepted during elicitation, any shortfall in completion arises among workers who had already agreed to the terms. We define completion as a binary indicator for the submission of the follow-up task (non-submission coded as zero) and estimate treatment effects with and without interactions between treatment and gender.

Table 5 reports the coefficient estimates on task completion. Completion is lower in outside-hours contracts: T2 reduces completion by 16 percentage points and T4 by 20, while the within-hours no-notice arm (T3) shows no significant effect. The T2 estimate runs against the reservation-wage results, where outside-hours work with notice carries no measurable penalty. Workers thus appear to price unpredictability more accurately than the cost of non-standard timing.

The clearest gender pattern appears in the most demanding contract. The $T4 \times \text{Female}$ interaction of 0.19 ($p < 0.10$) implies that the gender gap in completion for T4 exceeds the gap for T1 by roughly 19 percentage points. The estimate is large but only marginally significant. While suggestive, it aligns with the reservation-wage estimates, where T4 also generates the largest gender gap in flexibility penalties. Columns 3–4 report treatment effects on whether workers reported adjusting their usual routines to complete the task, which we discuss in Section 5.2.

Table 4: Analysis by Wage Level

	Contract Acceptance (0/1)					
	Individual FE			Controls		
	Low	Mid	High	Low	Mid	High
Notice / Outside (T2)	0.03** (0.02)	0.03 (0.02)	0.02 (0.02)	0.03** (0.01)	0.03* (0.02)	0.02 (0.02)
No Notice / Within (T3)	-0.04*** (0.01)	-0.06*** (0.01)	-0.09*** (0.01)	-0.04*** (0.01)	-0.06*** (0.01)	-0.09*** (0.01)
No Notice / Outside (T4)	-0.02 (0.01)	-0.04** (0.02)	-0.07*** (0.02)	-0.03* (0.01)	-0.04** (0.02)	-0.08*** (0.02)
Female (=1)				0.12*** (0.03)	0.12*** (0.03)	0.10*** (0.03)
T2 X Female	-0.06*** (0.02)	-0.02 (0.02)	-0.02 (0.02)	-0.06*** (0.02)	-0.02 (0.02)	-0.02 (0.02)
T3 X Female	-0.06*** (0.02)	-0.05** (0.02)	-0.02 (0.02)	-0.05*** (0.02)	-0.04** (0.02)	-0.02 (0.02)
T4 X Female	-0.07*** (0.02)	-0.06** (0.03)	-0.02 (0.03)	-0.07*** (0.02)	-0.06** (0.02)	-0.02 (0.02)
Participant F.E	Yes	Yes	Yes	No	No	No
Participant Control	No	No	No	Yes	Yes	Yes
Control Mean	0.282	0.525	0.702	0.282	0.525	0.702
Observations	7736	7737	11607	7736	7737	11607
R^2	0.63	0.72	0.66	0.03	0.03	0.04

Notes: The outcome variable is a binary variable indicating the Y/N preference for the contract. Columns (1), (2) and (3) show regressions of low, mid and high wage groups while including participant fixed effects. Columns (4), (5) and (6) show regressions of low, mid and high wage groups while controlling for age, income, race, marital status, and contract order effects. Standard errors are clustered at the participant level and reported in parentheses.

***, **, * indicate statistical significance at 1, 5, 10 percent levels.

Source: Waves 1 - 5.

5 Mechanisms

5.1 Heterogeneity by Baseline Flexibility

We begin by testing whether the gender differential in flexibility penalties varies with workers' baseline scheduling autonomy. The distinction matters in gig and online labor markets, where setting one's own hours is a primary draw and the workforce may therefore differ

Table 5: Work Completion, and Routine Changes

	Followup Task Completion (0/1)		Changes in Routine (0/1)	
Notice / Outside (T2)	-0.16*** (0.05)	-0.13* (0.08)	0.03 (0.07)	-0.02 (0.10)
No Notice / Within (T3)	0.03 (0.05)	0.07 (0.08)	0.03 (0.06)	-0.02 (0.09)
No Notice / Outside (T4)	-0.20*** (0.05)	-0.09 (0.08)	0.11 (0.07)	0.02 (0.10)
Female (=1)	-0.11*** (0.04)	-0.03 (0.08)	0.06 (0.05)	-0.02 (0.09)
T2 $\times$ Female		-0.04 (0.10)		0.08 (0.14)
T3 $\times$ Female		-0.07 (0.11)		0.10 (0.13)
T4 $\times$ Female		-0.19* (0.11)		0.22 (0.16)
Observations	626	626	222	222
R-squared	0.05	0.06	0.04	0.05
Control Mean	0.44	0.44	0.12	0.12
P-val: T2 = T3	0.00	0.01	0.91	0.97
P-val: T2 = T4	0.47	0.61	0.27	0.68
P-val: T3 = T4	0.00	0.04	0.29	0.63
P-val: T2 $\times$ Fem = T3 $\times$ Fem		0.80		0.85
P-val: T2 $\times$ Fem = T4 $\times$ Fem		0.16		0.38
P-val: T3 $\times$ Fem = T4 $\times$ Fem		0.28		0.45

Notes: Columns (2) and (4) include interaction terms between treatment indicators and a female dummy to test for heterogeneous treatment effects by gender. The outcome variable is a binary indicator for completion of the followup task in columns (1)-(2), and a binary indicator for the participant having to change their routine in order to accommodate the followup task in columns (3)-(4). Standard errors are clustered at the participant level and reported in parentheses. The regressions control for the contract's wage level. The table also reports p-values from pairwise tests of equal coefficients within each regression.

***, **, * indicate statistical significance at 1, 5, 10 percent levels.

Source: Waves 1-5.

systematically from traditional labor markets in its scheduling flexibility. Following the pre-analysis plan, we split the sample by a binary indicator for whether the participant has no regular working hours in their current internet job, yielding 2,244 observations in the flexible group and 1,620 in the non-flexible group. Table 6 reports estimates of equation (4) separately for the two groups.

Table 6: Effects by Baseline Work Flexibility

	Log(Reported Reservation Wage) (USD/hr)					
	Flexible Schedule			Not Flexible Schedule		
	Pooled	Short Term	Long Term	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.04 (0.03)	-0.02 (0.04)	-0.06 (0.04)	-0.02 (0.03)	-0.00 (0.05)	-0.04 (0.04)
No Notice / Within (T3)	0.15*** (0.02)	0.18*** (0.03)	0.13*** (0.03)	0.05** (0.02)	0.02 (0.04)	0.09*** (0.03)
No Notice / Outside (T4)	0.11*** (0.03)	0.15*** (0.04)	0.07** (0.03)	0.06* (0.03)	0.09* (0.05)	0.02 (0.04)
Female	-0.18*** (0.05)	-0.15** (0.07)	-0.17** (0.07)	-0.14** (0.06)	-0.05 (0.08)	-0.24*** (0.07)
T2 $\times$ Female	0.04 (0.04)	0.02 (0.06)	0.06 (0.05)	0.08** (0.04)	0.07 (0.07)	0.10* (0.05)
T3 $\times$ Female	0.01 (0.03)	-0.00 (0.04)	0.02 (0.03)	0.10*** (0.04)	0.09 (0.06)	0.12*** (0.04)
T4 $\times$ Female	0.06 (0.04)	0.07 (0.06)	0.04 (0.05)	0.12** (0.05)	0.06 (0.08)	0.18*** (0.06)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	1.96	1.78	2.14	1.92	1.72	2.12
Observations	2244	1143	1101	1620	809	811
R-squared	0.07	0.06	0.13	0.04	0.06	0.11
p-val: T2 = T3	0.00	0.00	0.00	0.02	0.63	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.02
p-val: T3 = T4	0.09	0.50	0.06	0.90	0.15	0.06
p-val: T2 $\times$ Fem = T3 $\times$ Fem	0.45	0.72	0.42	0.68	0.77	0.80
p-val: T2 $\times$ Fem = T4 $\times$ Fem	0.61	0.35	0.59	0.23	0.93	0.06
p-val: T3 $\times$ Fem = T4 $\times$ Fem	0.21	0.20	0.69	0.71	0.73	0.21

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. The sample is split by current work schedule flexibility, where flexible schedule is measured as a binary indicator for participants who do not have regular working hours in their current internet job. This specification includes interaction terms between treatment indicators and a female dummy to test for heterogeneous treatment effects by gender within each flexibility group. Columns (1)-(3) show results for participants with flexible schedules (pooled, one-off job, five-week job) and columns (4)-(6) show results for participants without flexible schedules. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * indicate statistical significance at 1, 5, 10 percent levels.

Source: Waves 1-5

Among male workers with flexible baseline schedules, the no-notice penalty is large and precisely estimated for both one-off and five-week contracts, replicating the pooled results.

T3 contracts (within hours, no notice) raise reservation wages by approximately 15 percent, and T4 contracts (outside hours, no notice) by around 11 percent (Table 6, col. 1). The T2 null also holds in this group. The gender interactions, however, are small and insignificant across all contract types, indicating that men and women with flexible schedules respond similarly to demanding terms.

The pattern differs among workers without flexible baseline schedules. In the pooled sample, the gender interactions are positive and significant across all contract types—8 log points for T2, 10 for T3, and 12 for T4 (column 4). The T2 interaction is itself significant: outside-hours work with advance notice carries a larger penalty for women than men in this group. Splitting by contract duration reveals an asymmetry. The interactions are small and insignificant for one-off contracts (column 5) but grow for five-week contracts, where the T2, T3, and T4 reach 10, 12, and 18 log points (columns 6). Differences across durations are not always significant, so we treat the contrast as suggestive rather than conclusive.

These results locate the pooled gender differential documented in Section 4.2 within a specific group. Among workers with flexible baseline schedules, men and women respond similarly to demanding contracts. Among those with fixed schedules, however, the gender differential is substantial and the estimates are largest for five-week contracts. The pattern points to constraints on women with fixed schedules that do not bind equally on men in the same position. The next section asks whether household responsibilities and childcare obligations account for this pattern, and examines attitudinal evidence on gender differences in preferences.

5.2 Household Work, Constraints and Preferences

Section 3.1 documents large baseline gender differences in domestic labor. Women in our sample spend more time than men on both childcare and household chores (Table 1). These differences are not fully explained by marital or parental status (controlling for both narrows the chore gap, but women still report more time on domestic tasks, Appendix Table B4). The gender gap in flexibility penalties may reflect this asymmetry: when a no-notice or outside-hours assignment arrives, men and women do not have the same slack to absorb it.

Self-reported reasons, collected after contract decisions were finalized, point to household chores rather than childcare. Across the full sample, household responsibilities are cited nearly twice as often as childcare as a reason for declining work outside regular hours or without advance notice (Table 9, Panel A and B). Women are 6.3 percentage points more

likely than men to cite household responsibilities as a barrier to outside-hours work and 7.5 p.p. more likely for no-notice work (roughly 10 and 14 percent). The gap persists after conditioning on marital and parental status.

Reservation-wage estimates by subgroup sharpen this distinction. In the female-only sample, the no-notice penalty is not significantly larger for mothers than for non-mothers, and results for men show no parental gradient either (Table A5, Column 1).¹⁶ Across genders, the female interaction is similar and statistically indistinguishable across childcare-intensity subgroups (Table 7). Household chores show a different pattern. Among women, heavier burdens are again not significantly associated with larger penalties, but in the pooled sample the female interaction is concentrated in the top tercile of chore intensity for both no-notice arms, most so for T4 (Table 8). Combined with the self-reported barriers, this points to general household burdens rather than childcare as a more likely source of the gender gap in the flexibility penalties.

One interpretation is that reported childcare hours capture direct, schedulable care time, whereas broader household work involves a more diffuse (and less predictable) set of responsibilities. Daminger (2019) argues that household labor has a cognitive dimension distinct from its physical tasks: anticipating needs, identifying options, making decisions, and monitoring whether things get done. Two features of cognitive labor bear on the capacity to absorb schedule shocks. It is ongoing and open-ended rather than confined to discrete blocks, so it competes with work demands even when reported hours are similar. It is also harder to hand off on short notice. A partner can take over a specific task, but transferring responsibility for tracking and coordinating the household requires context that does not transfer easily. Conditional on reported hours, women who carry this load have less slack to absorb an unpredictable assignment.¹⁷ Gendered expectations may also be reinforced by external institutions. Buzard et al. (2025) find that in a large-scale field experiment with U.S. schools, mothers are more likely than fathers to be contacted about child-related matters even when both parents are presented as available.

Two pieces of behavioral evidence are consistent with this interpretation. First, as discussed

¹⁶Although not pre-specified, we also test whether the age of the child matters. Table B3 shows that among mothers, the flexibility penalty is largely indistinguishable for children above and below age six, with the exception of a significant negative $T3 \times \text{Young Child}$ coefficient in the five-week subsample, pointing, if anything, in the opposite direction from a childcare-constraint mechanism.

¹⁷This matters for how our measures should be interpreted. When respondents are asked about hours spent on childcare, they plausibly report direct care time (supervising, feeding, transporting) rather than the anticipation and coordination around it. Much of the cognitive load of running a household with children falls outside the childcare category and likely shows up in the broader household-work measure instead. Read this way, the contrast is not between parenting and chores, but between care time that can be scheduled and household responsibility that cannot.

in section 4.4, women are less likely to complete the most demanding contracts, consistent with binding constraints and a lack of temporal buffer. Second, among workers who complete the task, the gender interaction in reported routine adjustment rises from near zero in T2 to about 22 percentage points in T4 ($p = 0.162$; Table 5, column 4), though the estimate is imprecise.¹⁸

An alternative to the constraint interpretation is that women require higher compensation because they have stronger preferences against outside-hours or no-notice work. Table 9 does not support this. If anything, women are less likely than men to report disliking outside-hours work (30.3 versus 39.4 percent), and the same holds for no-notice work (46.0 versus 53.9 percent). Nor does a broader preference-based account find support: when asked to choose between an additional hour of market work and an additional hour of household work, gender differences are small and shrink further once marital and parental status are held fixed. Women therefore report no greater distaste for demanding schedules yet require more compensation to accept them, which points to constraints rather than preferences.

The post-elicitation survey offers descriptive support for the role of gendered expectations. Asked which gender is more likely to accept jobs requiring outside-hours work or no advance notice, respondents disproportionately name men (Panel D, Table 9). Asked why women in the United States spend more time on unpaid domestic work, the most common response is that women are socially expected to perform these tasks—cited by 84.9 percent of women and 68.9 percent of men (Panel G). These expectations plausibly leave women with less slack when an unpredictable assignment arrives, consistent with broader evidence that perceived gender norms shape labor-market behavior in ways that need not reflect underlying preferences (Bursztyn et al., 2020).

In sum, what separates men and women in this setting is not how they feel about unpredictable work but how much slack they have to absorb it. The gender gap in flexibility penalties is concentrated among workers with the heaviest domestic burdens, where the demands of unpredictable work may bind most tightly. Women are more likely to report that household responsibilities are the reason they turn down demanding work offers and they are less likely to actually complete the most demanding contracts. At the same time, women report no stronger distaste for outside-hours or no-notice conditions than men. The pattern aligns with

¹⁸Because the sample is conditional on completion, the result reflects intensive-margin behavior, and we cannot rule out that workers with the largest adjustment costs declined the contract entirely. We therefore read the routine-change results as supportive evidence of realized schedule disruption.

Table 7: Effects by Childcare Levels

	Log(Reported Reservation Wage) (USD/hr)					
	Low Childcare			High Childcare		
	Pooled	Short Term	Long Term	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.02 (0.02)	-0.01 (0.03)	-0.03 (0.03)	-0.05 (0.04)	-0.02 (0.07)	-0.08 (0.05)
No Notice / Within (T3)	0.14*** (0.02)	0.14*** (0.03)	0.14*** (0.02)	0.05* (0.03)	0.04 (0.05)	0.06* (0.04)
No Notice / Outside (T4)	0.09*** (0.02)	0.12*** (0.04)	0.06* (0.03)	0.08* (0.04)	0.12* (0.07)	0.04 (0.05)
Female (=1)	-0.18*** (0.05)	-0.10 (0.07)	-0.24*** (0.07)	-0.13** (0.06)	-0.15* (0.08)	-0.15** (0.07)
T2 × Female	0.02 (0.04)	0.01 (0.05)	0.04 (0.04)	0.10* (0.05)	0.08 (0.08)	0.12** (0.06)
T3 × Female	0.05 (0.03)	0.03 (0.06)	0.08** (0.03)	0.09** (0.04)	0.09* (0.06)	0.07* (0.04)
T4 × Female	0.11*** (0.04)	0.10 (0.06)	0.12** (0.05)	0.07 (0.05)	0.06 (0.09)	0.08 (0.07)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2143	1063	1080	1625	825	800
R-squared	0.05	0.05	0.13	0.07	0.06	0.11
Control Mean	1.96	1.78	2.14	1.91	1.73	2.11
p-val: T2 = T3	0.00	0.00	0.00	0.02	0.40	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.01	0.00
p-val: T3 = T4	0.03	0.47	0.01	0.36	0.15	0.54
p-val: T2×Fem = T3×Fem	0.54	0.83	0.43	0.80	0.82	0.43
p-val: T2×Fem = T4×Fem	0.00	0.07	0.02	0.50	0.77	0.41
p-val: T3×Fem = T4×Fem	0.14	0.26	0.36	0.76	0.62	0.87

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and female dummy to test for heterogeneous treatment effects by gender. The sample is split by childcare intensity, where childcare intensity is measures as an indicator for above-median hours spent on childcare. Columns (1)-(3) show results for individuals with low childcare responsibility (pooled, short term job, long term job) and columns (4)-(6) show results for individuals with high childcare responsibility. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5

gendered expectations about who absorbs domestic shocks and who can take advantage of more demanding work opportunities.

Table 8: Effects by Household Work Levels

	Log(Reported Reservation Wage) (USD/hr)								
	Low HH Work			Mid HH Work			High HH Work		
	Pooled	Short	Long	Pooled	Short	Long	Pooled	Short	Long
Notice / Outside (T2)	-0.06* (0.03)	-0.03 (0.04)	-0.08** (0.04)	-0.01 (0.04)	0.01 (0.05)	-0.03 (0.05)	-0.01 (0.04)	-0.01 (0.07)	-0.00 (0.04)
No Notice / Within (T3)	0.12*** (0.02)	0.13*** (0.03)	0.10*** (0.03)	0.11*** (0.03)	0.12*** (0.05)	0.11*** (0.03)	0.08*** (0.02)	0.01 (0.03)	0.14*** (0.03)
No Notice / Outside (T4)	0.11*** (0.03)	0.18*** (0.05)	0.04 (0.03)	0.09** (0.04)	0.10* (0.06)	0.07 (0.05)	0.03 (0.05)	0.01 (0.07)	0.04 (0.06)
Female (=1)	-0.13** (0.06)	-0.08 (0.09)	-0.18** (0.08)	-0.14** (0.07)	-0.06 (0.09)	-0.22** (0.09)	-0.16* (0.08)	-0.16 (0.12)	-0.18* (0.10)
T2 × Female	0.02 (0.05)	-0.02 (0.07)	0.07 (0.06)	0.04 (0.05)	0.01 (0.08)	0.08 (0.07)	0.06 (0.05)	0.08 (0.09)	0.04 (0.06)
T3 × Female	0.05 (0.04)	0.03 (0.06)	0.08* (0.04)	0.04 (0.04)	0.00 (0.07)	0.08 (0.05)	0.08** (0.04)	0.15*** (0.05)	0.02 (0.05)
T4 × Female	0.03 (0.05)	-0.02 (0.08)	0.08 (0.06)	0.11** (0.05)	0.10 (0.08)	0.12* (0.07)	0.15** (0.06)	0.19** (0.09)	0.10 (0.08)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	2.03	1.86	2.18	1.91	1.73	2.12	1.88	1.68	2.08
Observations	1382	659	723	1346	727	619	1132	562	570
R-squared	0.05	0.05	0.13	0.06	0.06	0.11	0.07	0.08	0.13
p-val: T2 = T3	0.00	0.00	0.00	0.00	0.04	0.00	0.04	0.72	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.76	0.45
p-val: T3 = T4	0.79	0.34	0.08	0.38	0.69	0.32	0.16	0.99	0.06
p-val: T2×Fem = T3×Fem	0.60	0.56	0.90	0.95	0.92	0.98	0.71	0.45	0.74
p-val: T2×Fem = T4×Fem	0.92	0.97	0.89	0.04	0.08	0.36	0.14	0.25	0.41
p-val: T3×Fem = T4×Fem	0.61	0.51	0.98	0.17	0.24	0.48	0.19	0.57	0.26

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and female dummy to test for heterogeneous treatment effects by gender. The sample is split by household work intensity, where household work intensity is measured as terciles of hours spent on home chores. Columns (1)-(3) show results for individuals with low household work responsibility (pooled, short term job, long term job) and columns (4)-(6) show results for individuals with high household work responsibility. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5

6 Selection Effects of Demanding Work

We turn now to whether the worker-side constraints documented so far have consequences for the productivity composition of the labor pool available to firms. Demanding contracts can affect not only the size of the applicant pool but also its composition, and the direction depends on how worker productivity correlates with the cost of accommodating demanding schedules. If higher-productivity workers have better outside options or face higher marginal costs, they will raise their reservation wages more sharply, and demanding contracts will

Table 9: Barriers, Preferences, and Gender-Norm Attitudes toward Demanding Work

	Women	Men	Difference	Adjusted Difference
	(1)	(2)	(3)	(4)
Panel A: Barriers to outside hours				
Household Responsibilities	67.2	60.9	6.3**	6.4**
Childcare Responsibilities	36.9	21.6	15.3	2.9
Dislike Outside Hours Jobs	30.3	39.5	-9.2	-4.7
Observations	488	458	946	946
Panel B: Barriers to no advance notice				
Household Responsibilities	62.9	55.4	7.5**	6.8**
Childcare Responsibilities	37.5	22.1	15.4	2.9
Dislike no Advance Notice Jobs	45.8	54.1	-8.3**	-3.5
Observations	491	462	953	953
Panel C: Choice b/w Work and HH Responsibilities				
HH Responsibilities	34.5	28.1	6.5**	4.3
Job Task	32.9	39.2	-6.3**	-4.6
Indifferent	32.5	32.7	-0.2	0.3
Observations	495	474	969	969
Panel D: More likely to accept demanding contracts				
Men (Outside Hours)	27.3	42.2	-14.9	-16.7***
Men (Advance Notice)	34.6	40.9	-6.3**	-7.7**
Observations (Outside Hours)	495	474	969	969
Observations (Advance Notice)	491	474	965	965
Panel E: Reasons for disparity in domestic work				
Women are expected to	85.5	69.6	15.8	16.9***
Women are better	19.2	23.8	-4.6*	-6.4**
Women care more	43.6	42.8	0.8	-1.0
Observations	495	474	969	969

Notes: Column (1) reports the share of women selecting the corresponding response, Column (2) reports the share of men selecting the corresponding response, Column (3) reports the difference in shares (Women - Men) derived as the coefficient on an indicator variable for female from regressing the responses on the indicator variable, and Column (4) reports the coefficient on the indicator variable for female while controlling for marital status and parental status. Standard errors are clustered at the participant level. ***, **, * mean statistical significance at 1, 5, 10 percent levels.

deter the most productive talent. If, by contrast, lower-productivity workers are more readily deterred, these same contracts select positively on productivity. We test these possibilities by examining whether the flexibility penalty varies with our measures of worker productivity.

The task-based literature in labor economics emphasizes that jobs reward different bundles of skills — cognitive, task-specific, communication, and effort-related — rather than a single

measure of productivity (Autor, 2013; Deming, 2017). We therefore measure three dimensions: (i) standardized test scores (SAT/ACT), capturing prior academic skills; (ii) Raven Progressive Matrices, capturing fluid reasoning and abstract problem-solving; and (iii) quality ratings of the baseline writing task, capturing written communication and information synthesis. Our pre-specified primary measure is an index of the three, constructed following Kling et al. (2007).¹⁹ We also report selection effects for each component separately, and for three (related but distinct) secondary measures of worker performance: verbal fluency, performance on simple factual-question scores (designed to capture attention to the video), and time spent completing the baseline summary (as a coarse proxy for effort).

6.1 Selection on Productivity

We first estimate whether flexibility penalties vary with baseline productivity, interacting each treatment indicator with the productivity measures (Table 10). We find evidence of negative selection on productivity in contracts without advance notice. A one-unit increase in the productivity index – equivalent to a one-standard-deviation improvement in each of its three standardized components – raises the flexibility penalty by approximately 5 percent for T3 (within hours, no notice) and 9 percent for T4 (outside hours, no notice), or 36 and 69 percent relative to the underlying T3 and T4 penalties. These effects are statistically significant and large in magnitude.²⁰ The corresponding interaction for T2 (outside hours with notice) is small and insignificant, indicating that non-standard timing alone does not deter higher-productivity workers when notice is provided.

Columns 3–8 of Table 10, show which dimensions of productivity drive the composite result. The gradient is concentrated in task quality and SAT/ACT scores, while the Raven interactions are close to zero and statistically insignificant. Appendix Table A7 extends the analysis to the secondary measures of worker productivity and engagement. Verbal fluency shows the same negative-selection pattern, particularly for T3. On the other hand, the interactions for factual-question scores and baseline task time – proxies for attention and effort rather than productivity – are small, and where significant they are negative.

¹⁹Writing ratings, SAT/ACT scores, and Raven scores are winsorized at the 99th percentile before standardization to mean zero and unit variance within the sample. Following Kling et al. (2007), the productivity index is the equally-weighted average of the three standardized components; for participants missing one component, the index is the average of the two non-missing components.

²⁰Results are robust to alternative treatments of missing values in SAT/ACT scores and writing ratings. Coefficients on the productivity-index interactions are very similar when we impute missing standardized values as zero and interact treatment indicators with a missing-value indicator, allowing treatment effects to differ for participants with imputed values (Online Appendix Table B5).

Table 10: Selection: Flexibility Penalty by Productivity

	Log(Reported Reservation Wage) (USD/hr)							
	Productivity Index		Rating		Raven		SAT/ACT	
Notice / Outside (T2)	-0.00 (0.02)	-0.00 (0.01)	-0.00 (0.02)	-0.00 (0.01)	-0.00 (0.02)	-0.00 (0.01)	-0.01 (0.02)	-0.02 (0.02)
No Notice / Within (T3)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.02)	0.14*** (0.01)
No Notice / Outside (T4)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)
Ability		0.03 (0.03)		-0.01 (0.02)		0.05** (0.02)		0.00 (0.03)
T2 $\times$ Ability	0.00 (0.03)	0.01 (0.02)	0.00 (0.02)	0.00 (0.01)	0.00 (0.02)	0.01 (0.01)	-0.00 (0.03)	0.00 (0.03)
T3 $\times$ Ability	0.05** (0.02)	0.05*** (0.02)	0.03*** (0.01)	0.03*** (0.01)	0.00 (0.01)	0.00 (0.01)	0.03* (0.02)	0.03** (0.01)
T4 $\times$ Ability	0.07** (0.03)	0.07*** (0.03)	0.04** (0.02)	0.04*** (0.01)	0.01 (0.02)	0.01 (0.02)	0.05* (0.03)	0.06** (0.03)
Participant F.E	Yes	No	Yes	No	Yes	No	Yes	No
Participant Controls	No	Yes	No	Yes	No	Yes	No	Yes
Control Mean	1.94	1.94	1.94	1.94	1.94	1.94	1.96	1.96
Observations	3864	3864	3848	3848	3864	3864	2675	2675
R-squared	0.82	0.04	0.82	0.04	0.82	0.05	0.82	0.05
p-val: T2 = T3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T3 = T4	0.78	0.76	0.78	0.75	0.76	0.73	0.71	0.66
p-val: T2 $\times$ Ability = T3 $\times$ Ability	0.13	0.09	0.06	0.04	0.94	0.88	0.38	0.29
p-val: T2 $\times$ Ability = T4 $\times$ Ability	0.00	0.00	0.00	0.00	0.53	0.62	0.00	0.00
p-val: T3 $\times$ Ability = T4 $\times$ Ability	0.31	0.30	0.55	0.56	0.55	0.56	0.38	0.31

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and worker productivity measures to test for heterogeneous treatment effects by productivity level. 'Productivity Index' is measured using a standardized index combining three baseline measures: (1) test scores (SAT/ACT), (2) fluid intelligence scores from Raven matrices performance, and (3) task quality ratings of baseline writing performance on a 0-10 scale. Each component is standardized to have mean zero and unit variance within the baseline sample before taking an average to arrive at the productivity index. Columns (1)-(2) show regressions wherein 'Ability' refers to the productivity index. Columns (3)-(4), (5)-(6), and (7)-(8) show regressions wherein 'Ability' refers to the task quality ratings, the raven scores, and the test scores respectively. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5.

6.2 Selection on Productivity by Gender

We next ask whether the productivity gradient differs by gender. Two potential mechanisms imply different answers. The first is market-based: higher productivity workers tend to have better outside options, raising their reservation wage more for contracts they find costly. Because this channel operates through outside options rather than gender, it implies similar gradients for men and women. The second is constraint-based and gendered: if high-productivity women are disproportionately the ones absorbing unpredictable work demands within their household (Goldin, 2014), the cost of unpredictability falls especially heavily on them, producing a steeper gradient for women.²¹

Table A8 interacts treatment assignment, gender, and each productivity measure. The triple-interaction coefficients are small and statistically insignificant for the index and for every component. Demanding contracts therefore appear to deter higher-productivity workers similarly regardless of gender, consistent with the outside-options mechanism.

Read alongside the gender gap documented in Section 5, this null separates a level effect from a slope. Women require more compensation for demanding contracts at every point in the productivity distribution, but the productivity gradient itself is no steeper for them. Even so, firm-side consequences are not gender-neutral: the level penalty implies that the female share of workers willing to accept demanding contracts falls relative to less demanding arrangements, and the shift runs across the productivity distribution rather than concentrating among the most productive. That is, firms lose access to female talent through a differential reduction in labor supply rather than through differential selection by productivity.

7 Conclusions

This paper studies a central tension in gig work: the same arrangements that give workers flexibility also reward those who can surrender it when clients or platforms demand responsiveness. By experimentally varying advance notice, timing, and duration while holding the task fixed, we show that demanding work is not a single job attribute. Unpredictability and non-standard timing impose distinct costs: contracts without advance notice raise reservation wages by 13–14 percent, while outside-hours work carries little average penalty when notice is provided. The capacity to absorb these schedule shocks is unequally distributed across workers.

²¹This could arise if assortative matching places high-productivity women in households with partners in equally demanding jobs (Eika et al., 2019).

The worker-side results point to a gendered cost of demanding schedules. Women face 5–6 percent larger flexibility penalties than men for either outside-hours or no-notice work, and 8–9 percent when both are combined. The gap is not explained by a gendered preference for avoiding these schedules, nor by motherhood or childcare intensity. Rather, it is largest among workers with high household-chore burdens and among those with less flexible baseline schedules. The binding constraint therefore appears broader than childcare: domestic responsibilities erode the temporal buffer workers need to absorb unexpected work demands. Women are more likely than men to cite household responsibilities as a barrier to demanding contracts, with suggestive evidence of higher routine adjustment among those who complete the most demanding contracts. The absence of this buffer translates unequal domestic workloads into unequal labor-market costs.

The firm-side results show that scheduling demands affect more than the level of labor supply. Higher-productivity workers require more compensation for no-notice work: a one-standard-deviation increase in our productivity index raises the flexibility penalty by about 5 percent for within-hours no-notice contracts and 9 percent when outside-hours no-notice contracts. The gradient does not differ significantly by gender, but because women face larger penalties at every productivity level, the female share of the willing pool falls as contracts become more demanding. At a given wage, operational flexibility therefore comes at the cost of a smaller and less productive labor pool. This connects the worker-side cost of demanding schedules to a broader misallocation concern: the organization of work determines not only how much labor firms attract, but also which workers they attract.

The scope of the findings is shaped by three features of the study. First, the estimates come from U.S.-based online gig workers completing short remote tasks. While this is a fast-growing segment of platform-mediated work, a natural next step would be to extend the analysis to in-person shift work, commuting-intensive jobs, or longer-term employment relationships. Second, our reservation wages come from an incentivized elicitation rather than equilibrium wage-setting. This is by design: the BDM structure and follow-up offer isolate worker-side valuations of specific scheduling attributes, holding fixed the bargaining and sorting processes that shape observed wages. How these valuations translate into equilibrium compensation when firms and workers bargain jointly over scheduling and pay remains a separate question. Third, the design randomizes contract terms but not household circumstances, so the evidence on mechanisms rests on subgroup comparisons and self-reported barriers. These identify where the gender gap concentrates rather than the causal effect of the constraints themselves. Designs that exploit shocks to household constraints could identify that effect directly.

Within these limits, the gender and productivity results reflect a common logic: demanding

schedules are costliest for workers with the least temporal slack or the highest opportunity cost of accommodating schedule shocks – conditions that disproportionately describe women with heavy domestic workloads and higher-productivity workers with better outside options. These forces shift the composition of the labor pool, so that relative to less demanding arrangements, the female share and the average productivity of workers willing to accept a contract both fall. As platform-mediated work, on-demand staffing, and just-in-time contracting expand, this selection is likely to grow in importance. The design of work therefore matters not only for worker welfare, but also for the composition of the workforce firms attract to greedy work.

References

- ABEL, M., R. BURGER, AND P. PIRAINO (2020): “The Value of Reference Letters: Experimental Evidence from South Africa,” *American Economic Journal: Applied Economics*, 12, 40–71.
- AUTOR, D. H. (2013): “The “Task Approach” to Labor Markets: An Overview,” *Journal for Labour Market Research*, 46, 185–199.
- BERNIELL, I., L. BERNIELL, D. D. L. MATA, M. EDO, AND M. MARCHIONNI (2021): “Gender gaps in labor informality: The motherhood effect,” *Journal of Development Economics*, 150, 102599.
- BERTRAND, M., E. KAMENICA, AND J. PAN (2015): “Gender Identity and Relative Income within Households,” *Quarterly Journal of Economics*, 130, 571–614.
- BLOOM, N., J. LIANG, J. ROBERTS, AND Z. J. YING (2015): “Does Working from Home Work? Evidence from a Chinese Experiment,” *Quarterly Journal of Economics*, 130, 165–218.
- BURSZTYN, L., A. L. GONZÁLEZ, AND D. YANAGIZAWA-DROTT (2020): “Misperceived social norms: Women working outside the home in Saudi Arabia,” *American economic review*, 110, 2997–3029.
- BUZARD, K., L. K. GEE, AND O. STODDARD (2025): “Who you gonna call? gender inequality in external demands for parental involvement,” *The Quarterly Journal of Economics*, 140, 2805–2849.
- CALIENDO, M., W.-S. LEE, AND R. MAHLSTEDT (2017): “The Gender Wage Gap and the Role of Reservation Wages: New Evidence for Unemployed Workers,” *Journal of Economic Behavior & Organization*, 136, 161–173.
- CARD, D., A. R. CARDOSO, AND P. KLINE (2016): “Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women,” *Quarterly Journal of Economics*, 131, 633–686.

- COOK, C., R. DIAMOND, J. V. HALL, J. A. LIST, AND P. OYER (2021): “The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers,” *The Review of Economic Studies*, 88, 2210–2238.
- CORTÉS, P. AND J. PAN (2019): “When Time Binds: Substitutes for Household Production, Returns to Working Long Hours, and the Skilled Gender Wage Gap,” *Journal of Labor Economics*, 37, 351–398.
- CORTÉS, P., J. PAN, L. PILOSSOPH, E. REUBEN, AND B. ZAFAR (2023): “Gender Differences in Job Search and the Earnings Gap: Evidence from the Field and Lab,” *Quarterly Journal of Economics*, 138, 2069–2126.
- CUBAS, G., C. JUHN, AND P. SILOS (2023): “Coordinated Work Schedules and the Gender Wage Gap,” *The Economic Journal*, 133, 1036–1066.
- DAMINGER, A. (2019): “The cognitive dimension of household labor,” *American sociological review*, 84, 609–633.
- DEMING, D. J. (2017): “The growing importance of social skills in the labor market,” *The quarterly journal of economics*, 132, 1593–1640.
- DIGITAL FUTURE SOCIETY (2022): “Global Perspectives on Women, Work and Digital Labour Platforms,” Report, Digital Future Society, Barcelona.
- EIKA, L., M. MOGSTAD, AND B. ZAFAR (2019): “Educational assortative mating and household income inequality,” *Journal of Political Economy*, 127, 2795–2835.
- EXLEY, C. L. AND K. NIELSEN (2024): “The Gender Gap in Confidence: Expected but Not Accounted For,” *American Economic Review*, 114, 851–885.
- FELFE, C. (2012): “The Motherhood Wage Gap: What about Job Amenities?” *Labour Economics*, 19, 59–67.
- GOLDIN, C. (2014): “A Grand Gender Convergence: Its Last Chapter,” *American Economic Review*, 104, 1091–1119.
- HAUSER, D. J., A. J. MOSS, C. ROSENZWEIG, S. N. JAFFE, J. ROBINSON, AND L. LITMAN (2023): “Evaluating CloudResearch’s Approved Group as a solution for problematic data quality on MTurk,” *Behavior Research Methods*, 55, 3953–3964.
- HE, H., D. NEUMARK, AND Q. WENG (2021): “Do workers value flexible jobs? A field experiment,” *Journal of Labor Economics*, 39, 709–738.
- INTERNATIONAL LABOUR ORGANIZATION (2021): “World Employment and Social Outlook 2021: The Role of Digital Labour Platforms in Transforming the World of Work,” Flagship report, International Labour Office, Geneva.
- JOHNSTON, A. C. (2025): “Preferences, Selection, and the Structure of Teacher Pay,” *American Economic Journal: Applied Economics*, 17, 310–346.

- KESTERNICH, I. AND T. DE SCHOUWER (2025): “Work Meaning and the Flexibility Puzzle,” *Journal of Labor Economics*.
- KLEVEN, H., C. LANDAIS, J. POSCH, A. STEINHAEUER, AND J. ZWEIMÜLLER (2019a): “Child Penalties across Countries: Evidence and Explanations,” *AEA Papers and Proceedings*, 109, 122–126.
- KLEVEN, H., C. LANDAIS, AND J. E. SØGAARD (2019b): “Children and Gender Inequality: Evidence from Denmark,” *American Economic Journal: Applied Economics*, 11, 181–209.
- KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): “Experimental analysis of neighborhood effects,” *Econometrica*, 75, 83–119.
- LE BARBANCHON, T., R. RATHELOT, AND A. ROULET (2021): “Gender Differences in Job Search: Trading off Commute against Wage,” *Quarterly Journal of Economics*, 136, 381–426.
- LITMAN, L., J. ROBINSON, Z. ROSEN, C. ROSENZWEIG, J. WAXMAN, AND L. M. BATES (2020): “The Persistence of Pay Inequality: The Gender Pay Gap in an Anonymous Online Labor Market,” *PLOS ONE*, 15, e0229383.
- MAESTAS, N., K. J. MULLEN, D. POWELL, T. VON WACHTER, AND J. B. WENGER (2023): “The Value of Working Conditions in the United States and Implications for the Structure of Wages,” *American Economic Review*, 113, 2007–2047.
- MAS, A. AND A. PALLAIS (2017): “Valuing Alternative Work Arrangements,” *American Economic Review*, 107, 3722–3759.
- MCGEE, A. AND P. MCGEE (2023): “Gender Differences in Reservation Wages in Search Experiments,” .
- MCKINSEY & COMPANY (2022): “Freelance, Side Hustles, and Gigs: Independent Workers in America,” Tech. rep., McKinsey Global Institute.
- MORCHIO, I. AND C. MOSER (2026): “The Gender Pay Gap: Micro Sources and Macro Consequences,” *American Economic Review*, 116, 1765–1810.
- SARSONS, H., K. GÖRKHANI, E. REUBEN, AND A. SCHRAM (2021): “Gender Differences in Recognition for Group Work,” *Journal of Political Economy*, 129, 101–147.
- SORKIN, I. (2018): “Ranking Firms Using Revealed Preference,” *Quarterly Journal of Economics*, 133, 1331–1393.
- UPWORK (2023): “Freelance Forward 2023: The Annual Study of the US Independent Workforce,” Tech. rep., Upwork Inc.
- WASSERMAN, M. (2023): “Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents,” *The Review of Economic Studies*, 90, 1535–1568.
- WISWALL, M. AND B. ZAFAR (2018): “Preference for the Workplace, Investment in Human Capital, and Gender*,” *The Quarterly Journal of Economics*, 133, 457–507.

A Appendix

Appendix Tables

Table A1: Participant Flow and Exclusions by Wave

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Total
Survey Timeline						
Survey Date	25th Aug	8th Sep	15th Sep	22nd Sep	29th Sep	-
Follow-up Date	2nd Sep	19th Sep	26th Sep	3rd Oct	10th Oct	-
Baseline Duration (Days)	2	7	6	6	7	28
Recruitment and Exclusions						
Initial Participants	201	402	216	325	145	1289
Female/Male Ratio	0.72	1.04	1.06	1.11	1.54	1.05
Less: No Consent	0	0	1	0	0	1
Less: Failed Attention	0	3	3	1	0	7
Less: Terrible Contract	16	60	36	38	17	167
Attention Pass Rate (%)	92.0	84.3	81.9	88.0	88.3	86.5
Less: Not Interested in Follow-up	24	56	19	26	20	145
Analytical Sample	161	283	157	260	108	969
Follow-up Process						
Eligible for Follow-up	177	283	157	260	108	985
Selected for Follow-up	114	187	97	167	63	628
Selection Rate (%)	64.4	66.1	61.8	64.2	58.3	63.8

Notes: This table presents the participant flow across five experimental waves of data collection. Survey Date indicates when each wave was launched, Follow-up Date shows when follow-up surveys were sent, and Survey Duration shows how many days each wave remained open for recruitment. The Recruitment and Exclusions section shows initial participant counts and reasons for exclusion: participants who did not provide consent, failed attention checks, accepted an unreasonably low wage contract (\$0.3/hour), or declined follow-up participation. Female/Male Ratio indicates the ratio of female to male participants in the initial sample. Attention Pass Rate represents the percentage of initial participants who neither failed the attention check nor accepted the low-paying contract. Analytical Sample represents participants who completed the survey without any of these exclusions and were interested in follow-up. The Follow-up Process section shows those eligible for follow-up (same as Analytical Sample), those ultimately selected for the follow-up study, and the selection rate.

Table A2: Randomization Balance Test

Variable	(1) T1	(2) T2	(3) T3	(4) T4	Pairwise t-test		
	Mean/(SE)	Mean/(SE)	Mean/(SE)	Mean/(SE)	Mean diff	Mean diff	Mean diff
Female	0.54 (0.03)	0.56 (0.03)	0.50 (0.03)	0.44 (0.03)	0.02	-0.04	-0.10**
White	0.76 (0.03)	0.77 (0.03)	0.83 (0.02)	0.79 (0.03)	0.01	0.07*	0.02
Age	35.08 (0.35)	35.16 (0.33)	35.49 (0.29)	35.35 (0.32)	0.09	0.42	0.27
Married	0.39 (0.03)	0.42 (0.03)	0.46 (0.03)	0.42 (0.03)	0.03	0.07	0.03
Has Child	0.44 (0.03)	0.43 (0.03)	0.44 (0.03)	0.42 (0.03)	-0.01	0.00	-0.02
Tertiary Edu	0.88 (0.02)	0.90 (0.02)	0.90 (0.02)	0.89 (0.02)	0.01	0.02	0.01
Yrs Experience	12.70 (0.45)	13.14 (0.42)	13.51 (0.37)	13.41 (0.40)	0.44	0.81	0.71
Hours working	38.04 (1.31)	39.11 (1.03)	39.72 (0.95)	38.89 (0.93)	1.07	1.68	0.85
Job Non-Internet	0.56 (0.03)	0.51 (0.03)	0.58 (0.03)	0.60 (0.03)	-0.05	0.03	0.04
Hours Working Intern	14.92 (0.97)	16.96 (0.98)	15.00 (0.95)	15.21 (0.93)	2.04	0.08	0.29
Regular Hours Intern	0.38 (0.03)	0.45 (0.03)	0.47 (0.03)	0.39 (0.03)	0.06	0.08*	0.01
School Score	0.77 (0.01)	0.78 (0.01)	0.77 (0.01)	0.76 (0.01)	0.01	-0.01	-0.01
Raven Score	4.42 (0.18)	4.83 (0.16)	4.58 (0.16)	4.73 (0.16)	0.41*	0.16	0.31
Home chores (h)	2.29 (0.11)	2.27 (0.13)	2.28 (0.11)	2.45 (0.14)	-0.02	-0.01	0.16
Childcare (h)	2.95 (0.34)	2.96 (0.32)	2.51 (0.27)	2.75 (0.31)	0.01	-0.44	-0.20
Other (h)	2.20 (0.23)	2.31 (0.18)	2.32 (0.23)	2.11 (0.18)	0.12	0.12	-0.09

Notes: This table presents the results of a balance test, illustrating the means and differences in means among the various treatment groups and the control group. Columns (1) through (4) provide the means for the T1, T2, T3, and T4. Columns (5) through (7) display the differences in means between pairs of treatment types. Robust standard errors are reported (in parentheses). ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Randomization till Wave 5. The total 981 excludes 4 participants who were eligible and assigned to the long-run treatment.

Table A3: Flexibility Penalties - Robustness Test

	Log(Imputed RW) (USD/hr)			Intervals of Imputed RW		
	Pooled	Short Term	Long Term	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.03** (0.01)	-0.01 (0.02)	-0.04** (0.02)	-0.04** (0.02)	-0.02 (0.03)	-0.08** (0.03)
No Notice / Within (T3)	0.07*** (0.01)	0.07*** (0.02)	0.06*** (0.01)	0.12*** (0.02)	0.11*** (0.02)	0.14*** (0.02)
No Notice / Outside (T4)	0.05*** (0.01)	0.07*** (0.02)	0.02 (0.02)	0.09*** (0.02)	0.13*** (0.03)	0.05* (0.03)
Female (=1)	-0.12*** (0.03)	-0.11*** (0.04)	-0.13*** (0.04)	-0.19*** (0.04)	-0.16*** (0.05)	-0.21*** (0.06)
T2 × Female	0.04** (0.02)	0.02 (0.03)	0.07*** (0.03)	0.07** (0.03)	0.04 (0.04)	0.10** (0.04)
T3 × Female	0.05*** (0.01)	0.04* (0.02)	0.05*** (0.02)	0.07*** (0.02)	0.07** (0.03)	0.06* (0.03)
T4 × Female	0.06*** (0.02)	0.05 (0.03)	0.07** (0.03)	0.08** (0.03)	0.07 (0.05)	0.09** (0.04)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	1.83	1.83	1.83			
Observations	3868	1952	1916	3868	1952	1916
R-squared	0.05	0.05	0.08			
P-val: T2 = T3	0.00	0.00	0.00	0.00	0.00	0.00
P-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00
P-val: T3 = T4	0.12	1.00	0.02	0.18	0.62	0.00
P-val: T2×Fem = T3×Fem	0.81	0.45	0.62	0.97	0.48	0.43
P-val: T2×Fem = T4×Fem	0.30	0.21	0.97	0.62	0.34	0.82
P-val: T3×Fem = T4×Fem	0.52	0.76	0.55	0.70	0.93	0.50

Notes: The outcome variable is the log of imputed reservation wage (\$/hr) elicited at contract switch points in Columns (1)-(3) and intervals of reservation wage (\$/hr) in Columns (4)-(6). This specification includes interaction terms between treatment indicators and female dummy to test for heterogeneous treatment effects by gender. Columns (1) and (2) show pooled regressions, while columns (3) and (4) present results for short run and long run sub-samples respectively. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, indicator for race (participant being white), and participant being unmarried. ***, **, * mean statistical significance at 1, 5, 10 percent levels. The tables also reports p-values from pairwise tests of equal coefficients within each regression. Source: Waves 1 - 5.

Table A4: Flexibility Penalties - Contract Acceptance

	Contract Acceptance (0/1)			
	Pooled	Pooled	Short Term	Long Term
Notice / Outside (T2)	0.02* (0.01)	0.02** (0.01)	0.01 (0.02)	0.04** (0.02)
No Notice / Within (T3)	-0.07*** (0.01)	-0.07*** (0.01)	-0.07*** (0.01)	-0.06*** (0.01)
No Notice / Outside (T4)	-0.05*** (0.01)	-0.05*** (0.01)	-0.08*** (0.02)	-0.03 (0.02)
Female (=1)		0.11*** (0.02)	0.09*** (0.03)	0.12*** (0.03)
T2 × Female	-0.03* (0.02)	-0.03* (0.02)	-0.01 (0.03)	-0.05** (0.02)
T3 × Female	-0.04*** (0.01)	-0.04*** (0.01)	-0.03* (0.02)	-0.04** (0.02)
T4 × Female	-0.05** (0.02)	-0.05** (0.02)	-0.03 (0.03)	-0.06** (0.03)
Participant F.E	Yes	No	No	No
Participant Controls	No	Yes	Yes	Yes
Wage Rate Controls	Yes	Yes	Yes	Yes
Control Mean	0.53	0.53	0.53	0.53
Observations	27080	27080	13664	13416
R-squared	0.62	0.16	0.18	0.17
P-val: T2 = T3	0.00	0.00	0.00	0.00
P-val: T2 = T4	0.00	0.00	0.00	0.00
P-val: T3 = T4	0.16	0.18	0.78	0.01
P-val: T2×Fem = T3×Fem	0.67	0.74	0.44	0.75
P-val: T2×Fem = T4×Fem	0.25	0.33	0.30	0.78
P-val: T3×Fem = T4×Fem	0.62	0.65	0.95	0.55

Notes: The outcome variable is a binary variable indicating the Y/N preference for the contract. This specification includes interaction terms between treatment indicators and female dummy to test for heterogeneous treatment effects by gender. Columns (1) and (2) show pooled regressions, while columns (3) and (4) present results for short run and long run sub-samples respectively. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, indicator for race (participant being white), and participant being unmarried. ***, **, * mean statistical significance at 1, 5, 10 percent levels. The tables also reports p-values from pairwise tests of equal coefficients within each regression. Source: Waves 1 - 5.

Table A5: Effects of Childcare: Within-Gender Analysis

	Log(Reported Reservation Wage) (USD/hr)					
	Women			Men		
	Pooled	Short Term	Long Term	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.01 (0.02)	-0.02 (0.04)	0.01 (0.03)	-0.01 (0.02)	0.01 (0.03)	-0.04 (0.03)
No Notice / Within (T3)	0.19*** (0.03)	0.16*** (0.04)	0.21*** (0.03)	0.13*** (0.02)	0.14*** (0.03)	0.12*** (0.02)
No Notice / Outside (T4)	0.19*** (0.03)	0.19*** (0.04)	0.18*** (0.04)	0.09*** (0.02)	0.13*** (0.04)	0.05* (0.03)
Has Children	-0.13** (0.06)	-0.24*** (0.09)	-0.04 (0.07)	0.01 (0.07)	0.11 (0.10)	-0.05 (0.09)
T2 × Has Children	0.06 (0.04)	0.09 (0.06)	0.04 (0.05)	-0.05 (0.05)	-0.07 (0.07)	-0.03 (0.06)
T3 × Has Children	-0.05 (0.03)	-0.03 (0.06)	-0.07* (0.04)	-0.06* (0.03)	-0.09 (0.05)	-0.03 (0.04)
T4 × Has Children	-0.03 (0.04)	-0.00 (0.07)	-0.06 (0.06)	-0.01 (0.05)	-0.00 (0.08)	-0.01 (0.06)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1977	1011	966	1887	941	946
R-squared	0.07	0.10	0.09	0.04	0.04	0.12
Control Mean	1.98	1.82	2.14	2.09	1.90	2.28
p-val: T2 = T3	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T3 = T4	0.99	0.54	0.41	0.09	0.75	0.02
p-val: T2×Has Children = T3×Has Children	0.01	0.10	0.05	0.78	0.77	1.00
p-val: T2×Has Children = T4×Has Children	0.01	0.08	0.03	0.22	0.26	0.57
p-val: T3×Has Children = T4×Has Children	0.60	0.64	0.83	0.19	0.22	0.61

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and parental status to test for heterogeneous treatment effects by childcare responsibilities. Children is measured as a binary indicator for whether the participant has any children. Columns (1)-(3) show results for women (pooled, short term job, long term job) and columns (4)-(6) show results for men. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5

Table A6: Effects of Household Work: Within-Gender Analysis

	Log(Reported Reservation Wage) (USD/hr)					
	Women			Men		
	Pooled	Short Term	Long Term	Pooled	Short Term	Long Term
Notice / Outside (T2)	-0.04 (0.04)	-0.07 (0.06)	-0.02 (0.04)	-0.06* (0.03)	-0.03 (0.04)	-0.08** (0.04)
No Notice / Within (T3)	0.16*** (0.03)	0.16*** (0.05)	0.17*** (0.03)	0.12*** (0.02)	0.13*** (0.03)	0.11*** (0.03)
No Notice / Outside (T4)	0.14*** (0.04)	0.16*** (0.06)	0.11** (0.05)	0.11*** (0.03)	0.18*** (0.05)	0.04 (0.03)
High Household Work	-0.13** (0.06)	-0.13 (0.09)	-0.11 (0.07)	-0.09 (0.06)	-0.12 (0.08)	0.00 (0.07)
T2 $\times$ High HH Work	0.09** (0.04)	0.12* (0.07)	0.06 (0.05)	0.05 (0.04)	0.03 (0.06)	0.06 (0.05)
T3 $\times$ High HH Work	-0.01 (0.03)	-0.01 (0.06)	0.00 (0.04)	-0.02 (0.03)	-0.04 (0.05)	0.01 (0.04)
T4 $\times$ High HH Work	0.05 (0.05)	0.04 (0.07)	0.05 (0.06)	-0.04 (0.04)	-0.10 (0.07)	0.02 (0.05)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1977	1011	966	1883	937	946
R-squared	0.06	0.08	0.09	0.04	0.05	0.12
Control Mean	1.86	1.70	2.03	2.03	1.83	2.24
p-val: T2 = T3	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T3 = T4	0.43	0.94	0.26	0.70	0.35	0.06
p-val: T2 $\times$ High HH Work = T3 $\times$ High HH Work	0.06	0.11	0.31	0.13	0.27	0.38
p-val: T2 $\times$ High HH Work = T4 $\times$ High HH Work	0.27	0.19	0.82	0.01	0.01	0.25
p-val: T3 $\times$ High HH Work = T4 $\times$ High HH Work	0.23	0.41	0.42	0.56	0.37	0.92

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and household work intensity to test for heterogeneous treatment effects by household work responsibilities. 'High HH Work' is measured as a binary indicator for whether the participant has above-median time spent on home chores. Columns (1)-(3) show results for women (pooled, short term job, long term job) and columns (4)-(6) show results for men. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5

Table A7: Flexibility Penalties by Productivity

	Log(Reported Reservation Wage) (USD/hr)					
	Fluency		Factual		Time (Effort)	
Notice / Outside (T2)	-0.00 (0.02)	-0.00 (0.01)	-0.00 (0.02)	-0.00 (0.01)	-0.00 (0.02)	-0.00 (0.01)
No Notice / Within (T3)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)
No Notice / Outside (T4)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)
Ability		0.01 (0.02)		0.01 (0.02)		-0.06*** (0.02)
T2 $\times$ Ability	-0.00 (0.02)	-0.00 (0.02)	-0.04** (0.02)	-0.04** (0.02)	0.01 (0.01)	0.01 (0.01)
T3 $\times$ Ability	0.03** (0.01)	0.03** (0.01)	0.00 (0.01)	0.00 (0.01)	-0.02* (0.01)	-0.02* (0.01)
T4 $\times$ Ability	0.03 (0.02)	0.03* (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.01)
Participant F.E	Yes	No	Yes	No	Yes	No
Participant Controls	No	Yes	No	Yes	No	Yes
Control Mean	1.94	1.94	1.94	1.94	1.94	1.94
Observations	3864	3864	3864	3864	3864	3864
R-squared	0.82	0.04	0.82	0.04	0.82	0.05
p-val: T2 = T3	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T3 = T4	0.75	0.72	0.76	0.72	0.75	0.72
p-val: T2 $\times$ Ability = T3 $\times$ Ability	0.10	0.06	0.03	0.01	0.06	0.03
p-val: T2 $\times$ Ability = T4 $\times$ Ability	0.01	0.00	0.04	0.01	0.65	0.62
p-val: T3 $\times$ Ability = T4 $\times$ Ability	0.89	0.85	0.43	0.37	0.15	0.10

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and worker productivity measures to test for heterogeneous treatment effects by productivity level. Columns (1)-(2), (3)-(4), and (5)-(6) show regressions wherein 'Ability' refers to the verbal fluency scores, the factual question scores, and the time taken on the baseline task respectively. Each productivity measure is standardized to have mean zero and unit variance within the baseline sample. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5.

Table A8: Productivity Selection Effects by Gender

	Log(Reported Reservation Wage) (USD/hr)						
	Productivity Index	Rating	Raven	SAT/ACT	Fluency	Factual	Time (Effort)
Notice / Outside (T2)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.04 (0.03)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)
No Notice / Within (T3)	0.11*** (0.01)	0.11*** (0.01)	0.11*** (0.01)	0.11*** (0.02)	0.11*** (0.02)	0.11*** (0.02)	0.11*** (0.01)
No Notice / Outside (T4)	0.09*** (0.02)	0.09*** (0.02)	0.09*** (0.02)	0.09*** (0.03)	0.09*** (0.02)	0.09*** (0.02)	0.09*** (0.02)
Female	-0.16*** (0.04)	-0.17*** (0.04)	-0.16*** (0.04)	-0.14*** (0.05)	-0.17*** (0.04)	-0.16*** (0.04)	-0.16*** (0.04)
Ability	0.10** (0.04)	0.03 (0.02)	0.07** (0.03)	0.05 (0.04)	0.05* (0.03)	0.01 (0.03)	-0.05 (0.04)
T2 x Ability	0.02 (0.03)	-0.00 (0.02)	0.02 (0.02)	0.00 (0.04)	-0.02 (0.03)	-0.04 (0.02)	0.01 (0.02)
T3 x Ability	0.07*** (0.02)	0.04*** (0.01)	0.02 (0.01)	0.03 (0.02)	0.01 (0.02)	0.00 (0.02)	-0.02** (0.01)
T4 x Ability	0.07* (0.04)	0.02 (0.02)	0.02 (0.02)	0.05 (0.04)	0.01 (0.03)	-0.02 (0.02)	-0.01 (0.02)
Female x Ability	-0.14** (0.06)	-0.10*** (0.04)	-0.04 (0.04)	-0.09* (0.05)	-0.05 (0.04)	-0.01 (0.04)	-0.01 (0.04)
T2 x Female	0.06** (0.03)	0.06** (0.03)	0.06** (0.03)	0.05 (0.03)	0.06** (0.03)	0.05* (0.03)	0.06** (0.03)
T3 x Female	0.05** (0.02)	0.05** (0.02)	0.05** (0.02)	0.07** (0.03)	0.05** (0.02)	0.05** (0.02)	0.05** (0.02)
T4 x Female	0.08*** (0.03)	0.08*** (0.03)	0.08*** (0.03)	0.09** (0.04)	0.08*** (0.03)	0.08*** (0.03)	0.08*** (0.03)
T2 x Female x Ability	-0.02 (0.05)	0.01 (0.03)	-0.03 (0.03)	0.00 (0.05)	0.03 (0.03)	-0.00 (0.03)	-0.00 (0.02)
T3 x Female x Ability	-0.05 (0.03)	-0.02 (0.02)	-0.04 (0.02)	0.01 (0.03)	0.02 (0.02)	0.01 (0.03)	-0.00 (0.02)
T4 x Female x Ability	0.01 (0.05)	0.03 (0.03)	-0.02 (0.03)	0.01 (0.05)	0.03 (0.03)	0.03 (0.03)	0.02 (0.02)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	1.94	1.94	1.94	1.96	1.94	1.94	1.94
Observations	3864	3848	3864	2675	3864	3864	3864
R-squared	0.06	0.05	0.06	0.06	0.05	0.05	0.06

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. ‘Productivity Index’ is measured using a standardized index combining three baseline measures: (1) test scores (SAT/ACT), (2) Raven matrices performance, and (3) task quality ratings. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for age, income, race, marital status, and contract order effects. ***, **, * mean statistical significance at 1, 5, 10 percent levels.

B Online Appendix

B.1 Experimental Protocol

B.1.1 Eligibility Screening and Consent

Participants completed a standard attention check embedded in a short paragraph about internet-based work. The paragraph instructed participants to select “Prefer not to answer” from a list of response options that also included “Yes,” “No,” “Increase productivity,” and “Decrease productivity.” Participants who selected any option other than “Prefer not to answer” were exited from the survey.

B.1.2 Baseline Task

Participants were asked to write a brief message of 4–5 sentences highlighting the benefits of hand washing, using details from the video.²² They were informed that an anonymous version of their message would be shared with others to promote safe hand washing behaviors.²³ This framing was chosen to encourage effortful engagement with the writing task by giving it a prosocial purpose, without revealing the study’s focus on scheduling flexibility or gender. Additionally, the participants were asked two questions to capture attention to the video:

- “In which country is the school discussed in the video located?”
- “What is the disease mentioned in the video?”

B.1.3 Contract Elicitation

Before making contract decisions, participants were presented with the following introductory text:

Jobs vary regarding their demands for being flexible. Some industries, like media, often require work at odd hours or on short notice to meet tight or time-sensitive deadlines. Next you will be asked about your interest in different types of jobs.

²²This is a 125 second video from UNICEF: <https://www.youtube.com/watch?v=KHPkZTYKUPg>

²³The exact prompt for the writing task is as follows: “Please watch the following video carefully. After watching, write a brief message (4-5 sentences) highlighting the benefits of hand washing, using details from the video. Note that an anonymous version of this message on how to practice safe hand washing behaviors will be shared with others.”

For each type, we will ask whether you would accept or reject the job. These jobs may differ in: (1) Timing — when the work is scheduled; (2) Notice period — how far in advance you are informed about the time to complete the job; (3) Pay — how much the job pays per hour.

The following instruction was displayed before participants began making decisions:

Since only one of your choices will be randomly selected and used, rejecting a contract will not increase the chance of being offered a different contract — this means you should answer each question as if that one job might be the only one you get. Therefore, it’s important to express your true preferences.

Each treatment arm was introduced with a brief description of its scheduling conditions, as follows:

- **T1 (Within Hours, Advance Notice):** “Next, you will see contracts for a [10-minute / 2-hour per week for 5 weeks] task where you are told about the job 3 days in advance, so you have time to plan ahead. The task must be completed within typical working hours on the day we specify (between 9 AM and 5 PM).”
- **T2 (Outside Hours, Advance Notice):** “Next, you will see contracts for a [10-minute / 2-hour per week for 5 weeks] task where you are told about the job 3 days in advance, so you have time to plan ahead. The task must be completed outside typical working hours on the day we specify (between 7 PM and midnight).”
- **T3 (Within Hours, No Advance Notice):** “Next, you will see contracts for a [10-minute / 2-hour per week for 5 weeks] task with no advance notice, so you don’t have time to plan ahead. The task must be completed within typical working hours (between 9 AM and 5 PM).”
- **T4 (Outside Hours, No Advance Notice):** “Next, you will see contracts for a [10-minute / 2-hour per week for 5 weeks] task with no advance notice, so you don’t have time to plan ahead. The task must be completed outside typical working hours (between 7 PM and midnight).”

Figure B1 shows a representative contract screen as presented to participants, illustrating how the three contract dimensions were displayed for each offer.

Figure B1: Representative Contract Screen

Next, you will see contracts for a **2-hour task per week for 5 weeks** where you are told about the job **3 days in advance**, so you have time to plan ahead. The task must be completed **outside typical working hours** on the day we specify (between 7 pm and midnight).

Decide whether you would commit to accepting this job.

Offer -

- 1) Wage - pays an hourly wage of **\$7**
- 2) Timing - **occurs outside typical work hours** (after 7 PM)
- 3) Advance notice - **with advance notice**

- ☐ Yes
☐ No

Notes: Screenshot of the T2 (Outside Hours, Advance Notice) [2-hour per week] contract screen as presented to participants. Each offer listed the wage, timing condition, and advance notice condition. Participants indicated their interest by selecting Yes or No.

In addition to the four treatment arms, participants were presented with a dominated contract under T1 conditions paying \$0.30 per hour (\$0.05 for a 10-minute task). Figure B2 shows how this contract was presented to participants.

Figure B2: Dominated Contract Screen

Next, you will see contracts for a **10-minute task** where you are told about the job **3 days in advance**, so you have time to plan ahead. The task must be completed **within typical working hours** (between 9 AM and 5 PM).

Decide whether you would accept this job.

Offer -

- 1) Wage - pays an hourly wage of **\$0.3** (i.e. \$0.05 for a 10-minute task)
- 2) Timing - **occurs within typical work hours**
- 3) Advance notice - **with advance notice**

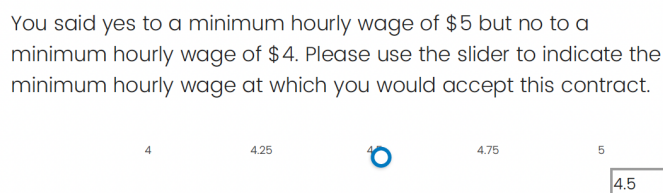
- ☐ Yes
☐ No

Notes: Screenshot of the dominated contract as presented to participants. Participants who accepted this offer were excluded from the analytical sample on the grounds of inattention.

B.1.4 Reservation Wage Elicitation

At the identified switch point — the wage level at which the participant’s decision changed from rejection to acceptance or from acceptance to rejection — participants were asked to report their exact reservation wage using a slider bounded by the two wage levels bracketing the switch point, in increments of \$0.25. Figure B3 shows the slider as presented to participants. Participants who accepted all wage levels including the lowest (\$4 per hour) or rejected all wage levels including the highest (\$10 per hour) were asked to enter their reservation wage directly in a free-text box without a bounded range.

Figure B3: Reservation Wage Slider



Notes: Screenshot of the reservation wage slider as presented to participants at the switch point. The slider was bounded by the two wage levels bracketing the switch, in increments of \$0.25.

After completing elicitation for all four treatment arms in their assigned contract length, participants were presented with four additional contracts at the medium wage of \$7 per hour under the alternative contract length, one for each treatment arm, allowing preferences over contract duration to be elicited alongside preferences over scheduling conditions. Participants who indicated a preference for the alternative contract length were asked to select from a list of reasons for their preference, covering household obligations, scheduling flexibility, income stability, and other factors.

B.1.5 Barriers, Preferences, and Gender-Norm Attitudes

Upon completion of contract decisions, participants are asked a series of questions on the barriers they face when considering demanding contracts (Figure B4), a hypothetical capturing their preference between work and household responsibilities (Figure B5, and their attitudes on gender-norms (Figures B6 and B7).

Figure B4: Barriers to Accepting Demand Contracts

What **barriers** do you face when considering work-related tasks that require you to **work outside your usual work hours**?
(select all that apply)

- ☐ I have other household responsibilities (e.g., elder care, chores) that demand attention
- ☐ I have childcare responsibilities that demand attention
- ☐ I dislike work outside my usual hours
- ☐ Family members will object
- ☐ I lack reliable transportation to work

Notes: Screenshot of the question on barriers to accepting an “Outside Hours” contract as presented to participants. The same question is rephrased as “What barriers do you face when considering work-related tasks that require you to work with no advance notice ?” to capture the barriers participants face when considering a “No Advance Notice” contract.

Figure B5: Preference for Work vs. Household Responsibilities

Imagine you had the **choice** between work and household responsibilities, what would you choose?

- ☐ Indifferent between the two
- ☐ Household responsibilities
- ☐ Job task

Notes: Screenshot of the question the choice between work and household responsibilities as presented to participants.

Figure B6: Gender Norm-Attitudes: Contract Decisions

Which of these groups is more likely to **accept** a job that requires **working on short notice**?

- ☐ Men
- ☐ Women
- ☐ Both men and women equally

Notes: Screenshot of the question on gender-norm attitudes as presented to the participants. The question captures participants’ beliefs on which gender is more likely to accept a “No Advance Notice” contract. The question is also rephrased as “Which of these groups is more likely to accept a job that requires working outside regular hours?” in order to capture beliefs on which gender is more likely to accept a “Outside Hours” contract.

Figure B7: Gender Norm-Attitudes: Domestic Work

As per a recent report, **women** in the United States spend **double the amount of time** as **men** on unpaid domestic work such as cleaning, childcare, preparing meals and caregiving. What do you think could be reasons for this statistic? (Select all that apply)

- ☐ Women are expected to take care of domestic work
- ☐ Women are better than men at domestic work
- ☐ Women care more about domestic work than men
- ☐ Other

Notes: Screenshot of the question on gender-norm attitudes as presented to the participants. The question captures participants' beliefs on the reasons behind the gender differences in unpaid domestic work.

B.2 Ability Measures

This appendix describes all the ability measures used in the analyses. All the measures are winsorized at the 99th percentile before standardization to mean zero and unit variation within the analytical sample.

SAT/ACT. Participants are asked two questions each for the SAT and the ACT scores: the marks obtained by the participant and the maximum marks that could be obtained in the year of examination. The proportion of marks obtained by the participant is calculated as the marks obtained by the participant normalized by the maximum marks. This exercise is done for both the SAT and the ACT scores. Further, these 2 proportions are aggregated into one "SAT/ACT" variable wherein a participant is assigned their ACT proportion as a priority, and are assigned their SAT proportion in cases wherein the ACT scores are not reported.²⁴

²⁴In the analytical sample of 969 individuals, 299 individuals have unreported scores for both the SAT and the ACT. Of these, 207 individuals state that they have not taken the SAT or the ACT. The remaining 92 state that they have taken at least one of the two tests, however, they do not report either of the two scores.

Figure B8: SAT/ACT Scores

Please enter the SAT score you obtained as well as the maximum total points in the year of examination

SAT score	Maximum Total Points

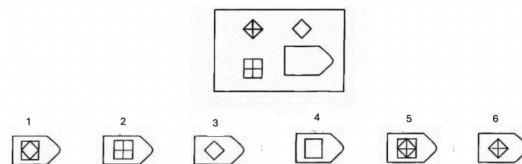
Notes: Screenshot of the questions on the SAT scores as presented to the participants. Additionally, the same set of questions are asked for the ACT scores.

Raven Progressive Matrices. The Raven Progressive Matrices task measures fluid intelligence through abstract pattern recognition. Each item presents a matrix of geometric patterns with one cell missing; participants identify the correct completion from a set of options. We implement an adaptive version of the task in which all participants begin at the same starting item. Correct responses route participants to a more difficult item; incorrect responses route participants to an easier item. Each participant completes at most three items. The Raven scores range from 0 to 7 based on the number of correct responses and the difficulty of the route taken by the participants. Figure B9 shows a representative item as presented to participants.

Figure B9: Raven Progressive Matrices: Example Item

Raven 4 (B11)

Next, you will complete a series of puzzles. In each puzzle, you will see a pattern with one piece missing, and your task is to select the correct piece that completes the pattern from the given options.



Enter the correct digit of the missing piece.

Notes: Screenshot of a representative Raven Progressive Matrices item as presented to participants. Participants enter the number corresponding to the correct missing piece.

Verbal Fluency Task. Participants are presented with a thirty-second timed task in which they are asked to generate as many words as possible beginning with the letter “Q,” excluding proper nouns and repeated words. The score is the total number of valid words produced within the time limit. Figure B10 shows the task as presented to participants.

Figure B10: Verbal Fluency Task

Self-Assessment

Now we will ask you a few questions. Please note that your answers will not affect your approval rate for this HIT.

29

Generate as many words as possible starting with the letter "Q" within thirty seconds. Please avoid proper nouns and repeating words.

https://nd.yull1.qualtrics.com/Q/EditSection/Blocks/Ajsa/GetSurveyPrintPreview?ContextSurveyID=5V_elpQ3CKiaBSKaR0&ContextLibraryID=UR_658PG23eS... 4/64

9/14/25, 9:25 PM Qualtrics Survey Software

words.

Notes: Screenshot of the verbal fluency task as presented to participants. Participants had thirty seconds to generate words beginning with the letter “Q,” excluding proper nouns and repetitions.

Writing Quality Ratings. The task quality ratings used as a component of the productivity index in Section 4 are constructed from a separate evaluation survey. We recruited a separate sample of evaluators on Prolific. After completing an attention check, evaluators were shown the same one-minute video presented to baseline participants and instructed: “Please watch the following 1min video carefully. Next, you will read short summaries of this video and assess how well these capture the content.” Each evaluator was then presented with a random subset of ten baseline summaries in random order and asked: “What is the quality of this summary (1 = very poor, 10 = very strong)?” Each baseline summary was independently rated by multiple evaluators, and we use the average rating across evaluators as our measure of writing quality.

Factual Question Scores. As described in Appendix B.1.2, upon completion of the baseline writing task, participants were asked two questions to capture attention to the video. The responses to these questions are coded as a binary variable, with valid responses taking

the value 1 and invalid responses taking the value 0. The average of the two binary variables is considered the factual question scores of the participants.

Time Taken (Baseline Task). The time taken is measured as the time each participant spends on the baseline writing task. This time is captured in seconds and is used as a proxy for effort.

B.3 Tables

Table B1: Follow-up Messages

Group	Contract	Message 1	Message 2
Short Run	Within Work Hours, Advance Notice (T1)	Congratulations! You have qualified for an upcoming follow-up task that pays \$1.17 for approximately 10 minutes of work. An invitation with the HIT will be sent on 02sep2025 by noon (local time). You will have until 5pm to complete it. Thank you for your participation.	The follow-up task you qualified for is now live. It pays \$1.17 for approximately 10 mins of work. Once the task is sent, you will have until 5pm local time to complete it
	Outside Work Hours, Advance Notice (T2)	Congratulations! You have qualified for an upcoming follow-up task that pays \$1.17 for approximately 10 minutes of work. An invitation with the HIT will be sent on 02sep2025 by 10 pm (local time). You will have until midnight to complete it. Thank you for your participation.	The follow-up task you qualified for is now live. It pays \$1.17 for approximately 10 mins of work. Once the task is sent, you will have until midnight local time to complete it
	Within Work Hours, No Advance Notice (T3)	Congratulations! You have qualified for an upcoming follow-up task that pays \$1.17 for approximately 10 minutes of work. An invitation with the HIT will be sent within the next couple of days, at some point during regular work hours. You will have until 5pm to complete it. Thank you for your participation.	The follow-up task you qualified for is now live. It pays \$1.17 for approximately 10 mins of work. Once the task is sent, you will have until 5pm local time to complete it
	Outside Work Hours, No Advance Notice (T4)	Congratulations! You have qualified for an upcoming follow-up task that pays \$1.17 for approximately 10 minutes of work. An invitation with the HIT will be sent within the next couple of days, at some point after regular work hours. You will have until midnight to complete it. Thank you for your participation.	The follow-up task you qualified for is now live. It pays \$1.17 for approximately 10 mins of work. Once the task is sent, you will have until midnight local time to complete it

Notes: The table reports the messages that the workers receive when they are selected for a Short-Run Medium Wage job. Message 1 informs the workers that they are selected for a follow-up job, and Message 2 informs the workers that the job they are selected for is now live. Workers assigned to contracts with advance notice (T1 and T2) are given information on the date and time when the job would go live while workers assigned to contracts with no advance notice (T3 and T4) are not given this information.

Table B2: Analysis by wage level

	Contract Acceptance (0/1)					
	Short Term			Long Term		
	Low	Mid	High	Low	Mid	High
Notice / Outside (T2)	0.02 (0.02)	0.02 (0.02)	-0.01 (0.02)	0.04** (0.02)	0.03 (0.02)	0.04** (0.02)
No Notice / Within (T3)	-0.06*** (0.02)	-0.06*** (0.02)	-0.09*** (0.02)	-0.03*** (0.01)	-0.06*** (0.02)	-0.09*** (0.02)
No Notice / Outside (T4)	-0.05** (0.02)	-0.05** (0.03)	-0.11*** (0.03)	-0.00 (0.02)	-0.03 (0.02)	-0.04* (0.02)
Female (=1)	0.14*** (0.04)	0.11*** (0.04)	0.04 (0.03)	0.09*** (0.03)	0.12*** (0.04)	0.14*** (0.04)
T2 X Female	-0.03 (0.03)	-0.01 (0.03)	0.00 (0.03)	-0.09*** (0.03)	-0.04 (0.03)	-0.04 (0.03)
T3 X Female	-0.05* (0.03)	-0.05** (0.03)	-0.01 (0.03)	-0.06*** (0.02)	-0.04 (0.02)	-0.04 (0.03)
T4 X Female	-0.07** (0.03)	-0.07* (0.04)	0.01 (0.04)	-0.07*** (0.03)	-0.06* (0.03)	-0.05 (0.03)
Participant Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	0.282	0.525	0.702	0.282	0.525	0.702
Observations	3904	3904	5856	3832	3833	5751
R^2	0.03	0.03	0.03	0.05	0.06	0.07

Notes: The outcome variable is a binary variable indicating the Y/N preference for the contract. Columns (1), (2) and (3) show regressions of low, mid and high wage groups for the short term job sub-sample. Columns (4), (5) and (6) show regressions of low, mid and high wage groups for the long term job sub-sample. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. Standard errors are clustered at the participant level and reported in parentheses. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1 - 5.

Table B3: Within-Gender Analysis: Young Child

	Log(Reported Reservation Wage) (USD/hr)					
	Women			Men		
	Pooled	Short Term	Long Term	Pooled	Short Term	Long Term
Notice / Outside (T2)	0.05 (0.04)	0.08 (0.08)	0.02 (0.05)	-0.07 (0.08)	-0.09 (0.12)	-0.06 (0.10)
No Notice / Within (T3)	0.16*** (0.03)	0.12*** (0.04)	0.20*** (0.04)	0.08 (0.07)	0.10 (0.11)	0.06 (0.08)
No Notice / Outside (T4)	0.18*** (0.05)	0.23*** (0.09)	0.13** (0.06)	0.01 (0.09)	0.03 (0.16)	-0.00 (0.09)
Young Child	0.07 (0.07)	0.03 (0.11)	0.10 (0.10)	0.16 (0.10)	0.08 (0.10)	0.19 (0.15)
T2 $\times$ Young Child	0.00 (0.06)	-0.06 (0.10)	0.06 (0.07)	0.00 (0.09)	0.04 (0.14)	-0.04 (0.11)
T3 $\times$ Young Child	-0.06 (0.04)	-0.02 (0.06)	-0.10** (0.05)	-0.03 (0.07)	-0.10 (0.12)	0.03 (0.09)
T4 $\times$ Young Child	-0.04 (0.07)	-0.09 (0.11)	-0.00 (0.10)	0.09 (0.10)	0.12 (0.17)	0.05 (0.11)
Participant Controls	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	1.92	1.92	1.92	1.92	1.92	1.92
Observations	973	495	478	588	294	294
R-squared	0.08	0.10	0.11	0.14	0.24	0.22
p-val: T2 = T3	0.02	0.57	0.00	0.01	0.08	0.06
p-val: T2 = T4	0.00	0.00	0.02	0.11	0.19	0.29
p-val: T3 = T4	0.65	0.12	0.12	0.28	0.55	0.21
p-val: T2 $\times$ Young Child = T3 $\times$ Young Child	0.32	0.66	0.04	0.65	0.32	0.42
p-val: T2 $\times$ Young Child = T4 $\times$ Young Child	0.33	0.67	0.31	0.18	0.46	0.18
p-val: T3 $\times$ Young Child = T4 $\times$ Young Child	0.80	0.44	0.25	0.10	0.11	0.72

Notes: The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and parental status to test for heterogeneous treatment effects by childcare responsibilities. 'Young Child' is measured as a binary indicator for whether the participant has a child less than 6 years of age. Columns (1)-(3) show results for women (pooled, short term job, long term job) and columns (4)-(6) show results for men. All regressions are limited to participants who report having a child. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5

Table B4: Gender Differences in Household Responsibilities

	Women		Men		Difference		Adjusted Difference	
	Mean	N	Mean	N	Coeff.	N	Coeff.	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Home chores (h)	2.709	495	1.910	473	0.798***	968	0.679***	968
Childcare (h)	4.207	482	1.297	463	2.910***	945	1.788***	945
Other (h)	2.613	482	1.831	465	0.781***	947	0.757***	947

Notes: Columns (1) and (3) report the mean of the corresponding variable for Males and Females respectively, Column (5) reports the coefficients from regressing the variables on an indicator variable for female, and Column (7) reports the coefficient from the same regression while controlling for marital status and parental status. Standard errors are clustered at the participant level. ***, **, * mean statistical significance at 1, 5, 10 percent levels. **Source:** Waves 1-5

Table B5: Selection - Robustness: Missing values imputed as 0 & include indicator variables for missing values

	Log(Reported Reservation Wage) (USD/hr)							
	Productivity Index		Rating		Raven		SAT/ACT	
Notice / Outside (T2)	-0.01 (0.02)	-0.02 (0.02)	-0.00 (0.02)	-0.00 (0.01)	-0.00 (0.02)	-0.00 (0.01)	-0.01 (0.02)	-0.02 (0.02)
No Notice / Within (T3)	0.14*** (0.02)	0.13*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	0.14*** (0.02)	0.14*** (0.01)
No Notice / Outside (T4)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)
Ability		0.04 (0.03)		-0.01 (0.02)		0.05** (0.02)		-0.00 (0.02)
T2 × Ability	0.01 (0.03)	0.01 (0.03)	0.00 (0.02)	0.00 (0.01)	0.00 (0.02)	0.01 (0.01)	-0.00 (0.03)	0.00 (0.03)
T3 × Ability	0.05** (0.02)	0.05*** (0.02)	0.03*** (0.01)	0.03*** (0.01)	0.00 (0.01)	0.00 (0.01)	0.03* (0.02)	0.03** (0.01)
T4 × Ability	0.09** (0.03)	0.09*** (0.03)	0.04** (0.02)	0.04*** (0.01)	0.01 (0.02)	0.01 (0.02)	0.05* (0.03)	0.06** (0.03)
Missing		0.00 (0.04)		0.02 (0.26)				-0.01 (0.04)
T2 × Missing	0.04 (0.03)	0.05* (0.03)	0.15 (0.15)	0.15 (0.13)			0.04 (0.03)	0.04 (0.03)
T3 × Missing	0.00 (0.03)	0.01 (0.02)	0.12 (0.32)	0.12 (0.28)			-0.01 (0.03)	-0.01 (0.02)
T4 × Missing	0.02 (0.04)	0.02 (0.03)	0.03 (0.16)	0.03 (0.14)			0.00 (0.04)	0.00 (0.03)
Participant F.E	Yes	No	Yes	No	Yes	No	Yes	No
Participant Controls	No	Yes	No	Yes	No	Yes	No	Yes
Control Mean	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94
Observations	3864	3864	3864	3864	3864	3864	3864	3864
R-squared	0.82	0.04	0.82	0.04	0.82	0.05	0.82	0.04
p-val: T2 = T3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T2 = T4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p-val: T3 = T4	0.66	0.62	0.78	0.75	0.76	0.73	0.71	0.66
p-val: T2×Ability = T3×Ability	0.20	0.15	0.06	0.04	0.94	0.88	0.38	0.29
p-val: T2×Ability = T4×Ability	0.00	0.00	0.00	0.00	0.53	0.62	0.00	0.00
p-val: T3×Ability = T4×Ability	0.27	0.23	0.55	0.55	0.55	0.56	0.38	0.31

Notes: This table reports results from Table 10 while imputing missing values of test scores (SAT/ACT) and task quality ratings as 0, and controlling for interactions between an indicator variable for if the individual has a missing value (in the corresponding ability metric) and the treatment indicators. The outcome variable is the log of reported reservation wage (\$/hr) elicited at contract switch points. This specification includes interaction terms between treatment indicators and worker productivity measures to test for heterogeneous treatment effects by productivity level. 'Productivity Index' is measured using a standardized index combining three baseline measures: (1) test scores (SAT/ACT), (2) fluid intelligence scores from Raven matrices performance, and (3) task quality ratings of baseline writing performance on a 0-10 scale. Each component is standardized to have mean zero and unit variance within the baseline sample before taking an average to arrive at the productivity index. Columns (1)-(2) show regressions wherein 'Ability' refers to the productivity index. Columns (3)-(4), (5)-(6), and (7)-(8) show regressions wherein 'Ability' refers to the task quality ratings, the raven scores, and the test scores respectively. Standard errors are clustered at the participant level and reported in parentheses. The regressions control for contract order effects. Participant controls include controls for age, income, race, and marital status. The table also reports p-values from pairwise tests of equal coefficients within each regression. ***, **, * mean statistical significance at 1, 5, 10 percent levels. Source: Waves 1-5.