Department of Economics Discussion Paper Series

The Gender Wage Gap in an Online Labour Market: The Cost of Interruptions

Abi Adams-Prassl

Number 944
June, 2021

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

The Gender Wage Gap in an Online Labour Market: The Cost of Interruptions

Abi Adams-Prassl*

June 30, 2021

Abstract

This paper analyses gender differences in working patterns and wages on Amazon Mechanical Turk, a popular online labour platform. Using information on 2 million tasks, I find no gender differences in task selection nor experience. Nonetheless, women earn 20% less per hour on average. Gender differences in working patterns are a statistically and economically significant driver of this wage gap. Women are more likely to interrupt their working time on the platform with consequences for their task completion speed. A follow up survey shows that the gender differences in working patterns and hourly wages are concentrated amongst workers with children.

*University of Oxford and Institute for Fiscal Studies. Email: abi.adams@economics.ox.ac.uk. I gratefully acknowledge funding from the Economic and Social Research Council (ESRC) under grant ES/I024808 and the Turing Foundation under grant TEDSA2100046. I am incredibly grateful to Kotaro Hara, Saiph Savage, Kristy Milland, Chris Callison Burch, Jeffrey Bingham for access to the data and an understanding of the platform. I would like to thank Jeremias Adams-Prassl, Joseph Altonji, Alison Andrew, Oriana Bandiera, Janine Berg, Giulia Giupponi, Hamish Low, Karl Schlag, and Johanna Rickne for helpful comments.

1 Introduction

There is a growing body of literature documenting gender differences in pay and work in "gender blind" workplaces. Even in the face of identical remuneration and promotion structures, men and women make different choices over when, where, and how much to work away from home. In the face of returns to experience (Bertrand, Goldin, and Katz 2010; Blau and Kahn 2017), convexity in the hours-earnings relationship (Goldin and Katz 2016), and monetary incentives to work at specific times in particular geographic areas (Bolotnyy and Emanuel 2018; Cook et al. 2018), these choices have consequences for women's earnings and the gender wage gap.

These gender differences in choices are hypothesised to arise from inequality in the division of household duties and variation in "the value of time not at paid work" that influence women's ability and preference for working outside the home (Cook et al. 2018).[1] On this basis, it is natural to think that innovations making it easier to work from home could facilitate a further convergence of working patterns and pay. However, as the Covid-19 pandemic has made salient, domestic care responsibilities might differentially affect men and women's ability to work productively from home. During the 2020 lockdowns, women were more likely to report engaging in paid work and caregiving simultaneously (Andrew et al. 2020). Alon et al. (2021) hypothesise that working such a "double shift" might have undermined mothers' productivity with consequences for their career progression. However, data limitations have held back the ability of researchers to test directly for productivity losses arising from any blurring of the boundary between work and care. Indeed, Alon et al. (2021) conclude that "the productivity losses of working parents during the pandemic will show up in the data only some years down the road".

In this paper, I shed new light on gender differences in working from home in a gender-blind setting using detailed task-level data from a popular online labour market platform, Amazon Mechanical Turk (MTurk). MTurk is a platform on which employers upload small tasks to be completed by workers for a fixed piece-rate. Gender is not visible and workers have full discretion over when they work and what tasks they work on. Crucially, work is performed in the home. This data was collected between 2015 and 2017. It therefore facilitates an assessment of whether gender differences in working from home observed during the Covid-19

[1] Or in the case of the taxi company Uber, the speed at which men and women drive (Cook et al. 2018).

pandemic are likely to persist even when lockdowns ease. The dynamics of online platform work is also an interesting phenomenon in its own right: Kuek et al. (2015) estimate that more than 48 million workers worldwide are registered with online labour platforms, while the number of tasks posted on the five largest online labour platforms rose by 40% between December 2016 and December 2019 (Online Labour Index, 2020).[2]

In this environment, I find no difference in task selection nor the accumulation of experience across men and women; there are no gender differences in the piece rates nor characteristics of tasks selected, and men and women complete a similar number of tasks per month. This suggests that gender differences in preferences for tasks and work are not salient in this setting.

Nonetheless, women earn 20% less per hour on average. This is because women take longer on average to complete similar tasks to men. Drawing on the literature on multi-tasking (Vasilescu et al. 2016; Coviello et al. 2014), I examine the role of gender differences in working patterns in generating this phenomenon. I find that completing a large number of tasks in quick succession ("batch working") is associated with faster task completion times for both men and women. However, women are more likely to interrupt their working time on the platform, taking longer breaks between completing one task and starting the next, and they are less likely to complete a large number of tasks in the same sitting. This has consequences for women's average task completion speed, reducing their earnings per hour.

A follow up survey of the workers in my task-level data set shows that the gender difference in working patterns and hourly wages is confined to workers with children; there is no significant gender difference in wages for individuals without children. This is consistent with childcare and domestic production constraints continuing to influence women's ability to schedule paid work, even when this work is performed within the home (Tremblay 2002; Kossek, Lautsch, and Eaton 2006). Thus, my results suggest that even in this idealised setting is not possible to consider market work independently of domestic production and childcare for women.

[2] See Pelletier and Thomas (2018) for a review of the literature on online labour markets.

2 Context & Data

MTurk is one of the largest online micro-task platforms in the world. Workers registered with MTurk browse the platform for tasks and self-select into those which they wish to work on. Workers choose tasks from a list that provides a short description of the work to be completed, the employer's name, the expiration date for the task, the time frame within which the task must be completed, and the reward for successful completion.[3] Employers usually post a large number of similar tasks on the site, which are often referred to as "batches". For example, under the same batch identifier, an employer might post 1,000 similar images to be classified. After accepting, completing, and submitting one task from such a batch, a worker is immediately redirected to the next available task of the same type. When a worker completes a task, the employer receives the output, along with information on the time elapsed between accepting and submitting the task, and the worker's identification code. The employer then decides whether to accept the work and remunerate the worker, or to reject the output if the quality is not considered sufficiently high (Irani 2015).

Relevance for Studying Gender Differences in Work & Wages There are three particularly important features of the MTurk workplace that make it an interesting setting to study the gender wage gap. First, workers are anonymous and gender is not directly visible to employers. A worker's ID is simply a collection of letters and numbers that has no connection to their sex. While requesters can restrict the type of worker to whom a task is made available according to certain characteristics (e.g. their approval rating, whether they have acquired certain AMT-specific qualifications, and their geographical location), ex ante restrictions on the basis of gender are not possible nor are requesters informed about a worker's sex when choosing to accept or reject output unless this information has been directly collected as part of the task (Irani and Silberman 2013). These features imply that direct discrimination on the platform is highly unlikely to be the primary reason for any gender differentials uncovered on the platform.

Second, there are no explicit returns to tenure built into the payment structure. Tasks are remunerated on a piece rate basis and thus earnings are proportional to the number of jobs completed. This suggests a limited role for any "job flexibility penalty" arising from a convex hours-earning relationship (Goldin and Katz 2016). Further, the nature of MTurk's payment structure rules out a significant role for negotiation

[3] Figure A.1 in the Appendix shows a screenshot of the interface seen by workers.

and competition in governing rates of pay (Card, Cardoso, and Kline 2015).

Finally, work on MTurk can be done without leaving the home. If prior findings of gender differences in behaviour in "gender blind" contexts are partly driven by women having to accommodate domestic care constraints, one might expect fewer differences in choices in a setting with minimal fixed costs of work that permits more flexibility for combining domestic and market production. Furthermore, as all work on MTurk is performed online, with no direct relationship to the employer or consumers, harassment is unlikely to be present in this setting. For example, women may accumulate less experience on Uber because of consumer behaviour (e.g. sexual advances late at night) which results in women facing higher costs to work in the industry (Westmarland and Anderson 2001).[4]

2.1 Data

For my primary analysis, I use unique task-level data collected between 2015 and 2017 from a third-party MTurk plug-in, *CrowdWorkers*.[5] The plugin tracks what tasks workers complete and records timestamps for when workers accept and submit tasks, allowing a panel of task level effective hourly wages to be constructed. The plug-in is used by workers on an opt-in basis and was designed to disclose the effective hourly wage rates of tasks (Hara et al. 2018).

Following the approach of Larivière et al. (2013), I predict gender from the name an individual used to sign-up to the plugin. A worker is considered "female" or "male" when their first name occurred at least ten times as frequently for one gender than the other in the 1990 US Census name files. As many first names are common across both sexes and some names do not appear at all in the census files, not all workers' gender can be established in this way. I am able to match the gender of 1,805 of the 2,683 workers in the dataset. Table A.1 gives summary statistics for the gender-matched versus unmatched samples. Workers in the gender-matched sample completed 65.5% of all tasks in the dataset: 1,619,463 out of my full sample of 2,473,679 tasks.[6] Workers whose gender could not be established complete lower value tasks on average

[4]Cook et al. (2018) find a 4-7% gender wage gap on Uber in the US, 36% of which is explained by the fact that men accumulate more experience than women and 28% by gender differences in where to drive.

[5]Previous studies on MTurk have used either surveys or information scraped directly from the MTurk interface (Dube, Jacobs, Naidu, and Suri 2018; Adams and Berg 2017; Ipeirotis 2010). However, these sources do not capture worker behaviour at the task level.

[6]Gender established by this method and self-identified gender for the subset of workers who complete the demographic survey in Section 5 is the same for all but 25 individuals: 11 (14) who were classified as female (male) according to their name but self-identified as male (female) in the survey. Table A.7 shows the robustness of results to dropping these workers or coding

and thus have lower hourly wages on average.[7] However, there are few other differences in the observables of gender matched versus unmatched workers. While, by definition, I cannot examine whether there is any differential selection by gender into my gender-matched versus unmatched sample, I show in Appendix Table A.8 that there are no significant differences when running the various regression specifications reported in the main text (excluding the gender dummy) on the sample of matched workers compared to the full dataset. In the main text, the sample is restricted to the gender matched sample in the analysis that follows.

The Crowdworker log data is the only source on MTurk that I know of that contains detailed information on task completion at the worker-task level.[8] However, it is unlikely to be a random selection of MTurk workers that use the plugin. The plugin is designed for less experienced workers as its functionalities can largely be replicated using worker-written software for those with sufficient technical expertise. However, there is no evidence of gender differences in selection into the plug-in. Within workers whose gender can be determined, 52% are women. This is consistent with a range of survey evidence that approximately 50% of the MTurk population are women (Boas, Christenson, and Glick 2020; Difallah, Filatova, and Ipeirotis 2018; Adams and Berg 2017).[9]

3 Task Selection, Experience & Hourly Wages

I first analyse differences in task selection and the accumulation of experience on MTurk. I classify tasks into categories, e.g. data entry, on the basis of their title description following the schema of Hara et al. (2018).[10] I also distinguish between "batch" and "unit" tasks. Employers typically post multiple tasks of the same type on the platform (batch tasks), e.g. classifying 1,000 similar images appears as 1,000 separate tasks under the same group identifier. A task is classified as a batch task if I observe a worker completing at least one other task with the same batch identifier and the task piece-rate is less than $0.50. Unit tasks are

their gender as per their self-identified report.

[7] Closer examination of the unmatched names suggests that the unmatched sample includes three distinct groups: workers who report a comedy name (e.g. cupcake); workers whose name is gender-ambiguous by my classification method (e.g. Alex); workers whose name does not appear in the 1990 US Census name files and often appear to be of South Asian origin (e.g. Cherupally). As my interest is in gender differences in earnings in an OECD context, I do not attempt to include any of these workers in my main sample.

[8] See Hara et al. (2018) for a previous analysis of the same data source.

[9] The demographic characteristics of those who complete the follow-up survey are also consistent with prior evidence: namely, my sample is also disproportionately highly educated and a majority are in paid-work in the "offline" economy. See Table A.5 and, e.g. Adams and Berg (2017).

[10] Full details of the tagging procedure are given in the Appendix.

disproportionately research tasks (Table A.9).

Figures 1 (a) and (b) gives the distribution of task characteristics by gender. They reveal that there are few differences in task selection on the platform by male and female workers. Table A.1 shows that any differences are not statistically significant. Table A.1 also gives information on accumulated experience by gender. There is no significant gender difference in the average number of tasks completed each month in my sample; women and men on average complete 272 and 257 tasks per month respectively. Thus, within my sample, there is no evidence of the common finding in offline labour markets that women to accumulate less experience than men. (Bertrand, Goldin, and Katz 2010; Blau and Kahn 2017).[11]

3.1 Wage Regressions

I now turn to the relationship between gender, tasks characteristics, and variation in hourly wages. I compute effective hourly task-level wages by simply dividing a task's remuneration by the time taken between accepting and submitting the job. The wage for individual i completing task j is then:

$$w_{ij} = \frac{PieceRate_j}{SubmitTime_{ij} - AcceptTime_{ij}} \tag{1}$$

One limitation of this measure is that it will overstate working time, and thus understate hourly wages, if workers do not work continuously between accepting and submitting a task.[12] The majority of tasks on MTurk are "micro"-tasks and are designed to be completed rapidly; the average task completion time in my sample is 2 minutes suggesting that there is little room for lengthy breaks to be taken while the timer on a task is running. Nonetheless, to rule out the possibility that my results are driven by measurement error, I will examine heterogeneity in gender gaps across different types of task to test the validity of my proposed mechanism in Section 4.[13]

Despite no evidence of systematic gender differences in the types of tasks selected on the platform nor

[11] Adams and Berg (2017) also find no significant gender difference in the accumulation of experience on MTurk using survey evidence in which workers self reported the number of tasks they had completed on the platform.

[12] Workers might not work continuously while the clock is running if they take breaks/are interrupted or if they work on multiple tasks at once (often referred to as "hoarding"). I can observe, and control for, hoarding in the data (i.e. if a worker accepts a new task before submitting their previous one).

[13] It is also important to note that this measure of working time is not a perfect measure of productivity: faster work might be of lower quality. For example, (Cook et al. 2018) find that male Uber drivers drive faster but this might also bring associated passenger safety risks. I cannot explore this in my main data set but do consider work quality to survey responses in Section 5.

overall experience, women on average earn 21.6% less per hour than men (Table A.1). This is surprising given MTurk's institutional features. To examine these patterns further, Figure 1 (c) reports the results a set of standard wage regressions to explore the relationship between the characteristics of tasks, accumulated experience, and the gender wage gap:

$$\log w_{ij} = \beta_0 + \beta_1 Female + \rho \mathbf{X}_{ij} + \epsilon_{ij} \tag{2}$$

All regressions include time period fixed effects, have standard errors clustered at the worker level. In this section, the controls of interest are the log piece rate, a set of dummies for task type, and a worker's accumulated experience at the point the task is completed.

As should be expected given the absence of gender differences in task selection, there is little change in the coefficient on $Female$ once characteristics of the task and experience are controlled for: the unconditional log wage difference is -0.201 (p-value: 0.052) and -17.7 (p-value: 0.027) when log task reward, task type, and accumulated experience is included. Given the potential for measurement error on $Female$ (as gender is predicted on the basis of name), this wage gap is, if anything, an attenuation of the true gender gap.[14] These results imply that on average women earn less per hour than men, not because they choose systematically different tasks but because women complete similar tasks more slowly. This can of course be inferred from the regression specifications that control for the piece rate but is reported directly in Table A.2 column (6). On average, women take 2.25 minutes to complete a task compared to 1.95 minutes for men (Table A.1). The remainder of this paper explores why this is the case.

4 Gender Differences in Working Patterns

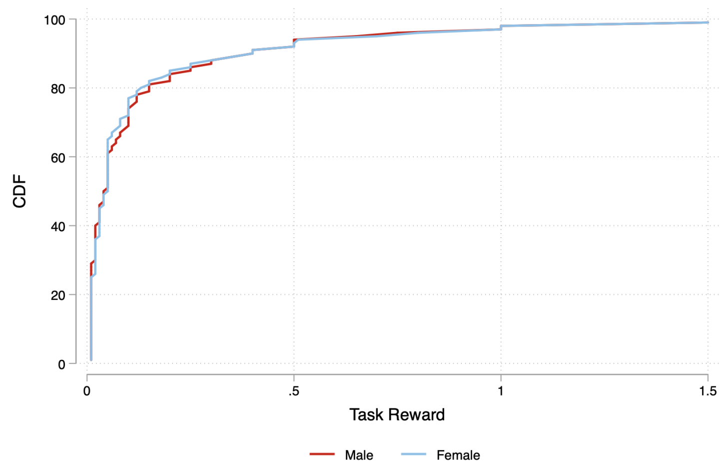
The literature on worker productivity identifies task scheduling and "multi-tasking", in addition to innate worker ability and effort, as important for work completion times (Vasilescu et al. 2016).[15] There are few characteristics available in my core dataset to facilitate a detailed exploration of ability and effort differences

[14] In Appendix Table A.7, we show that all results are robust to excluding the 25 workers for whom there is a discrepancy between their predicted gender and that which they self-report in the demographic survey.

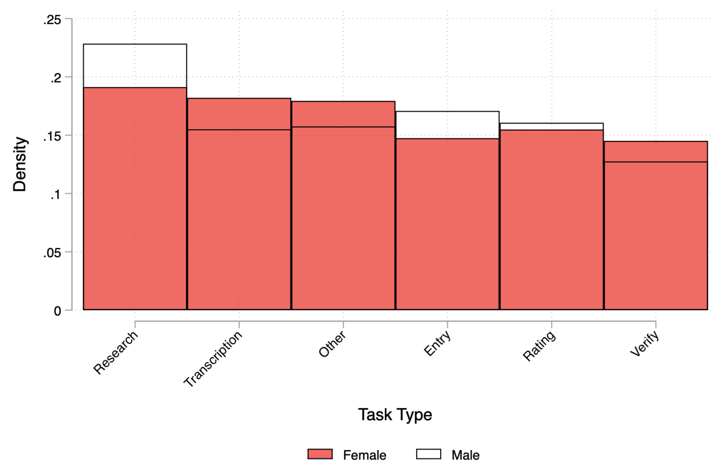
[15] Previous surveys of MTurk workers have found that about a fifth of workers report multi-tasking while on the platform (Necka et al. 2016).

Figure 1: Task Selection: Descriptive Statistics & Wage Gaps

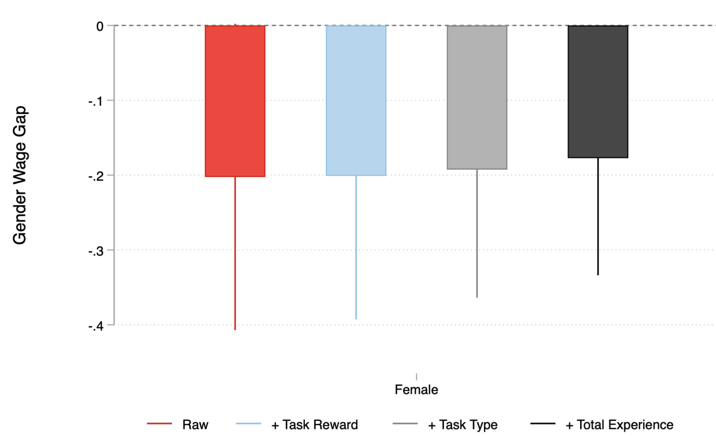
(a) Distribution of Piece Rates Chosen



(b) Distribution of Task Types



(c) Gender Wage Gap with Task & Experience Controls



Notes: Panel (c) gives the coefficients on *Female* for a set of regressions of log wage on task characteristic controls (Equation 2) along with the 95% confidence interval. All regressions include time period fixed effects, have standard errors clustered at the worker level, and the coefficients are also reported in Table A.2. The controls of interest are the log piece rate, a set of dummies for task type, and a worker's accumulated experience at the point the task is completed.

by gender. However, given the detailed time information available, I am able to consider whether gender differences in work scheduling and interruptions have a role to play. There is a rich literature that demonstrates the productivity consequences of simultaneously working on multiple tasks and of interruptions. Across a range of scenarios, it has been confirmed that workers who juggle projects take longer to complete each of them compared to if the tasks were completed sequentially (Buser and Peter 2012; Coviello, Ichino, and Persico 2015; Coviello, Ichino, and Persico 2014).

The grouping of tasks into batches might heighten the importance of work scheduling for measured productivity on MTurk. In qualitative studies of online platform work, workers have described an ability to "hone their practises over successive iterations of a task" and that batch hits allowed them to "get into a rhythm" (Kässi, Lehdonvirta, and Dalle 2019). For these reasons, workers often express a preference for working on batches of repetitive tasks, even if each individual unit is not well paid (Kaplan, Saito, Hara, and Bigham 2018).

MTurk tasks are performed at home and potentially alongside domestic production. The home is an "interdependent workplace", an environment in which other actors can demand immediate attention that might distract a worker from their own task (Perlow 1999). Women are more likely to be "passive carers" (Folbre et al. 2005), ready to be called upon when issues of childcare arise. As described in a survey of MTurk users conducted by Adams and Berg (2017), "I am able to help my disabled husband when he needs me and still bring in money" or "MTurk allows me to stop a HIT, if need be, so that I can care for the baby if she starts coughing, shaking, etc". These additional demands might influence workers' ability to work on batches of tasks without interruption.

4.1 Measuring Working Patterns

These constraints could manifest themselves in two ways. First, the need to take time out to "fire fight"[16] urgent domestic tasks might result in longer intervals between completed tasks and, potentially, a more fragmented work schedule. Juggling activities on and off the platform might directly lower productivity and could also result in higher rates of switching between different types of tasks if other workers complete the remaining jobs available while one is attending to other duties. Second, working time on tasks might be

[16] See Bohn (2000) for a discussion of the productivity impact of fire fighting, i.e. where problems are fixed as they arise.

systematically mis-measured; if workers attend to domestic requirements without submitting a task, measured working time will be longer than the actual time spent completing a task. Controlling for breaks between tasks and features of the work schedule will account for the former set of influences but not the latter.

I construct four types of controls from the worker log data (which only records activity on MTurk):

- *Length of break between adjacent tasks*: The CrowdWorkers app records the time when a task is accepted and submitted, allowing breaks taken between finishing one task and starting the next to be identified. I control flexibly for breaks between tasks with a set of binary variables for the decile of break length between the current and previous task (if the break is less than 24 hours long and positive). It is possible for workers to accept a task before submitting their previous task. I include a separate control for whether a task is "overlapping" and also check the robustness of results to excluding these tasks from the analysis completely.[17]

- *Tasks completed previously in a work session*: Following Hara et al. (2018), I define a work session as a period in which no more than 10 minutes elapses between completing one task and starting the next. To account for the "work flow" effects described by Kässi et al. (2019), I control flexibly for the number of tasks previously completed in a work session to allow for any learning by doing in a given work session over and above that captured by experience on all completed tasks. These are referred to as work session controls.

- *Task switching*: direct switching between different sorts of tasks is captured by a binary variable that equals one if a task is in the same batch as the previous task completed.

- *Time of day*: domestic constraints might affect when men and women work. Unfortunately I am unable to measure time of day with any precision as the time stamp in my dataset gives time at the location of the app's server, not at the worker's location. I include hour of day fixed effects nonetheless but do not consider heterogeneity in time-of-day given this measurement error.

[17]See Columns (1) and (2) Table A.7.

4.2 Work Schedule Results

Table 1 explores the explanatory power of these features of work scheduling for effective hourly wages: including the work scheduling variables reduces the magnitude of the gender gap by 48% to -0.0928, and it becomes statistically significant. The importance of differences in work patterns for explaining the gender wage gap is confirmed by a Gelbach decomposition in Figure 2. Intuitively, a Gelbach decomposition accounts for correlation between regressors when accounting for the extent to which a given variable/set of variables "explains" the gender wage gap and it is invariant to the order in which covariates are introduced (Gelbach 2016). This exercise reveals that work scheduling variables are significant determinants of the observed gender difference in hourly wages, in particular the sets of work session and break controls.

The importance of work scheduling controls derives from two features of the data. First, women are more likely to have a fragmented work schedule. Figures 3 (a) and (b) give the distribution of break length and work session controls by gender. Women are more likely to take longer breaks between adjacent tasks and are less likely to complete a large number of tasks in a work session. Second, working patterns are systematically related to task completion speed: task completion speed increases with the number of tasks completed in a work session and decreases in the length of time taken between adjacent tasks. To show this, Figure 3 (c) and (d) gives the coefficients on the the break length and work session position controls when wage regressions that include worker fixed effects are performed separately for men and women.[18] Note that these results are driven by within worker variation in working patterns, limiting concerns that the relationship between work patterns and effective hourly wages is driven by time-invariant unobserved worker characteristics (e.g. ability).[19] This analysis reveals that a fragmented work schedule reduces worker productivity (i.e. is associated with lower hourly wages). Furthermore, men and women are equally penalized for a fragmented work schedule.[20] However, as women are more likely to work in a fragmented pattern, they complete tasks more slowly than men on average.

[18]The regressions underlying these figures also control for all other task characteristics as in Table 1 and standard errors are clustered at the worker level.

[19]See Appendix Table A.3 for the equivalent graphs without worker fixed effects which correspond to the specification reported in Table 1 (2) that includes a gender dummy.

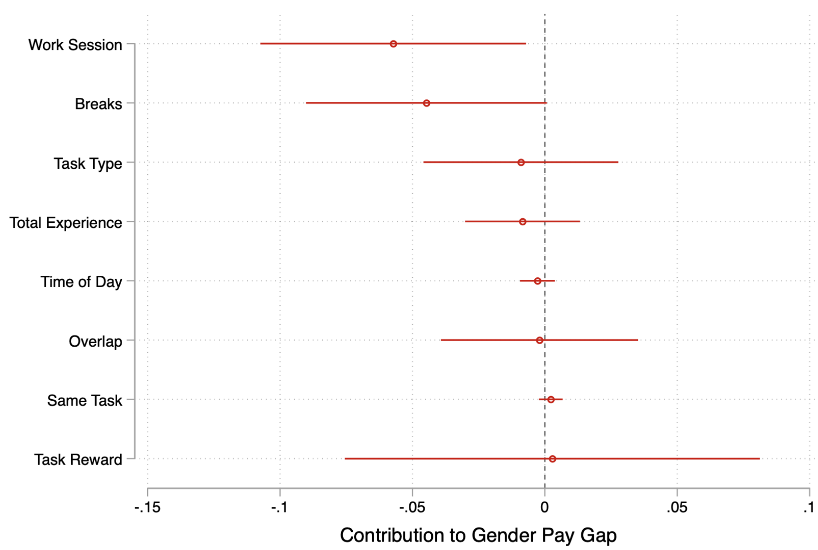
[20]There is no significant gender difference in the coefficients on work schedule controls (p-value: 0.5621) nor breaks (p-value: 0.2582) reported in Figures 3 (c) and (d).

Table 1: Task-Level Log Wage Regressions: Working Pattern Controls

	All Tasks		Batch v. Unit Tasks			
			Batch	Unit	Batch	Unit
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.1772**	-0.0928	-0.1798**	-0.0257	-0.0940	-0.0217
	(0.0267)	(0.1361)	(0.0269)	(0.3119)	(0.1384)	(0.3629)
Test for Equality *Female*						
p-value:	0.0287		0.0514		0.2690	
Task Controls	yes	yes	yes	yes	yes	yes
Work Pattern Controls	no	yes	no	no	yes	yes
Observations	1616850	1616850	1582730	34120	1582730	34120

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * p < 0.10, ** p < 0.05, *** p < 0.01. All specifications include year-quarter fixed effects. Task characteristics refers to all controls in column (6) of Table 1. Dependent variable is log wage in all specifications.

Figure 2: Gelbach Decomposition of *Female* in Wage Regressions



Notes: Figure uses the method described in Gelbach (2016) to plot the share of the gender wage gap that can be attributed to each set of controls and the 95% confidence interval.

Heterogeneity by Task Type The returns to continuous work are likely to differ across different types of tasks. Thus, I consider heterogeneity in the gender gap across "batch" versus "unit" tasks as an indirect check of the importance of work fragmentation for observed gender differences. Unit tasks (e.g. surveys, see Table A.9) are often designed to be completed only once, limiting any gains from learning by doing within a work session.[21] The design of batch tasks, however, generates more potential for productivity spillovers across tasks (Kässi et al. 2019). Systematic variation in the wage gap across these different types of task supports the hypothesis that my results are not simply due to mis-measurement of working time as this would affect both batch and unit tasks.[22]

Figure 3 (e) and (f) show the coefficients on work schedule controls when the sample is split into batch and unit tasks and separate wage regressions including worker fixed effects are performed. They confirm that the returns to continuous work are significantly larger for batch tasks.[23] In line with this, I find no gender difference in hourly wage amongst unit tasks. Columns (3)-(6) of Table 1 give the standard wage regression results with the sample split into batch and unit tasks. The coefficient on $Female$ is close to zero for unit tasks, both including and excluding work scheduling controls. However, work scheduling controls significantly reduce the magnitude of the $Female$ coefficient for batch tasks such that, once work patterns are controlled for, there is no significant difference in the magnitude of the gender gap across batch and unit tasks.[24] This heterogeneity helps to provide further reassurance that differences in work patterns are driving the main results in Table 1 (1) and Figure 2.

5 Survey Evidence

The question that next arises is *why* men and women working on the platform have different working patterns. I complement the task level log-data with linked survey data on worker demographics and their self-reported motivations and experiences for working on MTurk. 711 of the workers present in my task-level data set completed a demographic survey administered in December 2017 and January 2018. The survey collected

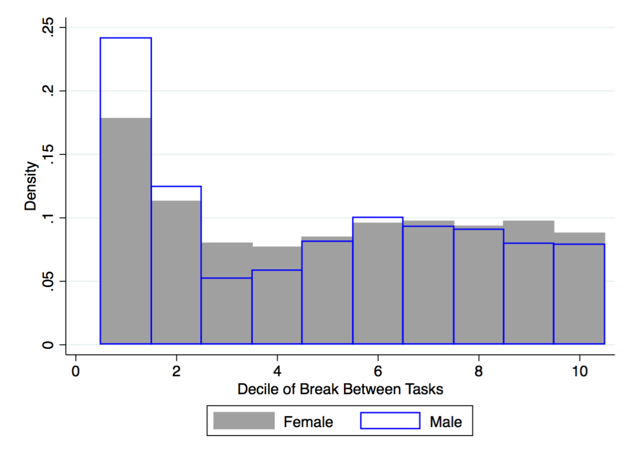
[21] The majority of tasks are Batch tasks, which is to be expected given the micro-task nature of the platform.

[22] Indeed, as unit tasks take longer to perform on average (6.67 minutes versus 1.97 minutes), one might expect that there is more potential for taking breaks within these tasks compared to batch tasks.

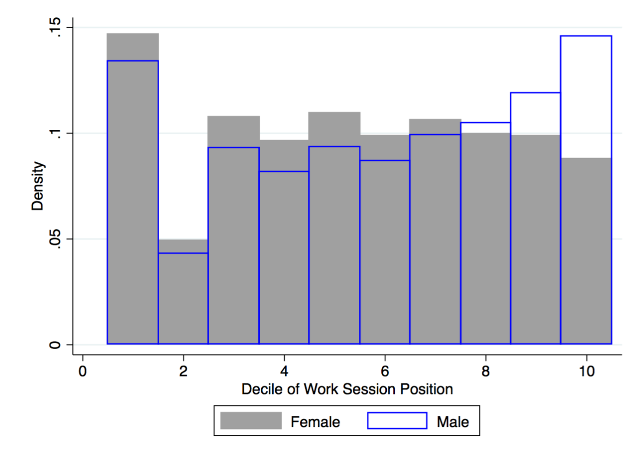
[23] The coefficients on work schedule and break controls are significantly different across the two samples, i.e. p-value ≈ 0.0000.

[24] Appendix Figure A.2 gives the results for specifications where the sample is split by sex and task type.

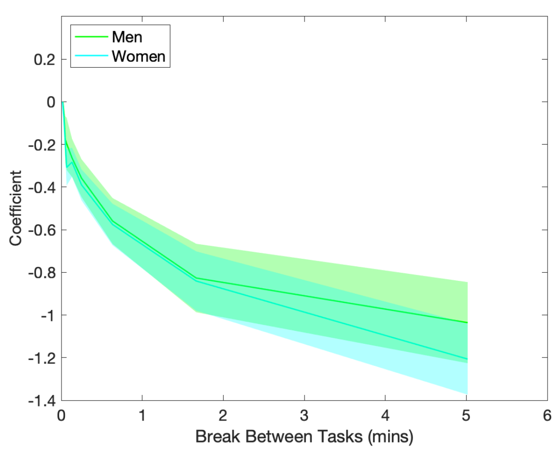
Figure 3: Work Schedule Controls & Wages: Worker Fixed Effect Specifications



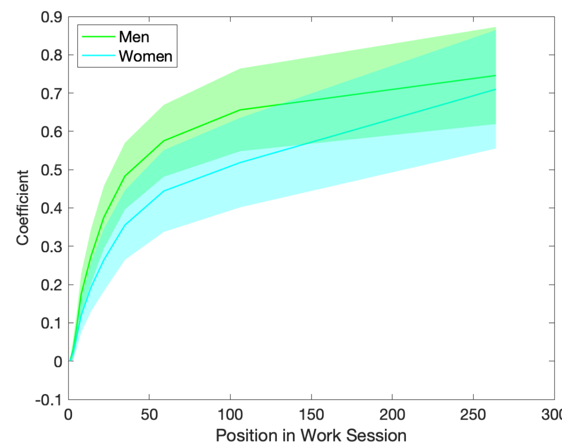
(a) Density of Break Controls



(b) Density of Work Session Controls

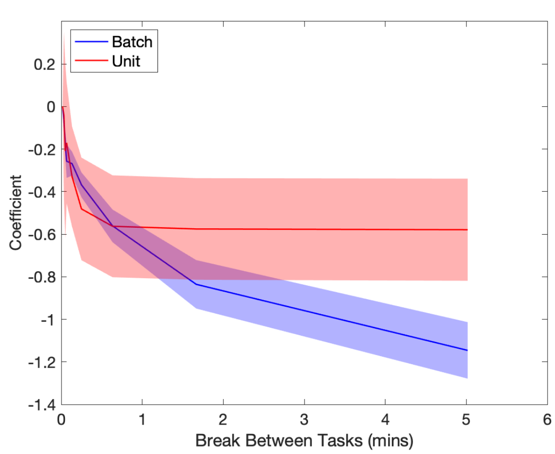


(c) Break Between Tasks (p= 0.2582)

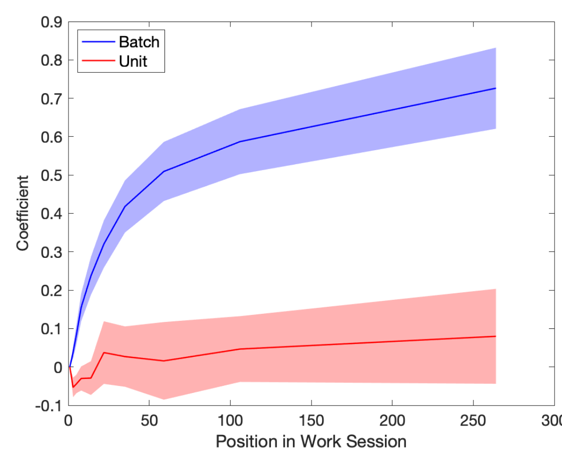


(d) Position in Work Session (p=0.5621)

Batch versus Unit Tasks



(e) Break Between Tasks (p=0.0000)



(f) Position in Work Session (p=0.0000)

Notes: Panel (a) and (b) gives the distribution of break length and work session position deciles by men and women. Panels (c) - (f) give the coefficients in regression of log wage on controls corresponding to (2), (4) and (6) of Table 1 along with the 95% confidence interval with worker fixed effects.

information on age, education, health, and the reasons why an individual worked on MTurk. Descriptive statistics and a short commentary are given in Appendix Table A.5.

While it is not a random selection of workers into my survey,[25] the same gap in hourly earnings and pattern with respect to the additional of work scheduling controls emerges (Table A.3). Amongst the sample of workers completing the survey, women again earn 20% less than men per hour on average. Controlling for the same task and work session features as described in Section 4 explains approximately half of this gap, with the coefficient on $Female$ shrinking to -0.1251.

Crucially, the survey responses enable an investigation of heterogeneity in the gender wage gap by household structure. It is reasonable to expect that the spillover effects of domestic production on productivity in paid work would be greatest for women with children (Tremblay 2002; Kossek, Lautsch, and Eaton 2006).[26] Consistent with this hypothesis, I find no significant gender difference in the length of work sessions and in the length of breaks between tasks between workers without children (Appendix Figure A.5). However, there are significant gender differences in work patterns for those with children. On average, mothers complete 80 fewer tasks in a work session and take 1.33 minutes longer between adjacent HITs compared to fathers.[27]

Appendix Table A.3 provides the regression results for the same specifications as reported in Table 1 with the sample split into workers with and without children. It demonstrates that the gender wage gap is significantly larger amongst workers with children. Indeed, there is no significant gender difference in wages amongst childless workers (columns (3) and (5)). Figure 4 gives the Gelbach decomposition of the gender wage gap amongst childless workers and those with children. Work session controls explain a statistically significant component of the coefficient on $Female$ for workers with children.[28] This evidence is therefore consistent with domestic care responsibilities differentially affecting the productivity of mothers and fathers in paid work from home. It is worth noting that the coefficient on $Female$ remains significant and negative for workers with children, albeit at a much reduced size, even after the work scheduling controls are introduced. It is not the aim of this paper to fully explain the gender gap but rather to show that gender differences in

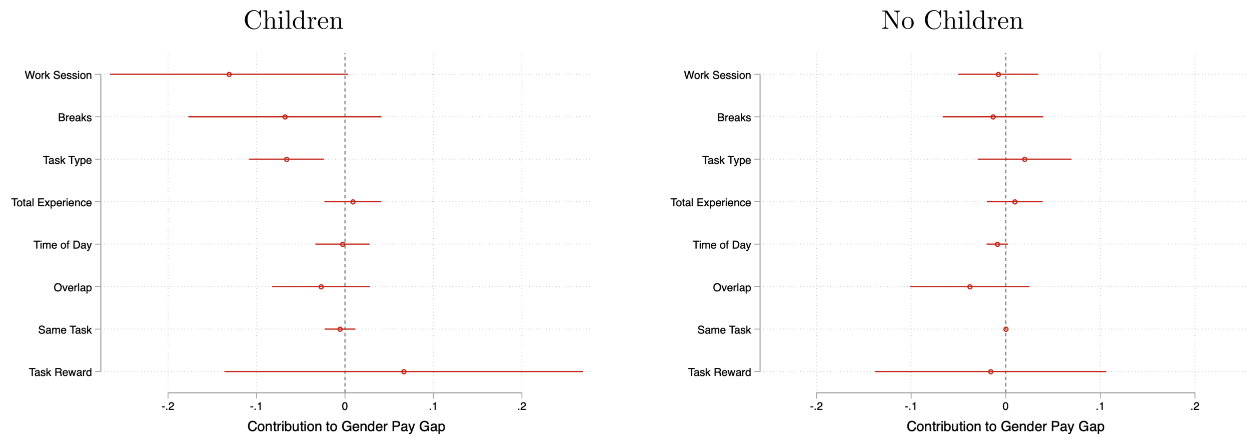
[25] Workers who complete the survey perform more unit tasks and are more experienced. See Table A.6.

[26] Although, there is also inequality in care giving amongst elderly and in housework which could also differentially affect the ability of women to engage in paid work at home.

[27] These differences are statistically significant, p-value≈0.0000. At the median, mothers complete 34 fewer tasks in a work session and take 0.6 minutes longer between adjacent HITs.

[28] Figure shows the coefficients on the work session and break controls for those with and without children.

Figure 4: Gelbach Decomposition of *Female* in Wage Regressions: By Children



Notes: Figure uses the method described in Gelbach (2016) to plot the share of the gender wage gap that can be attributed to each set of controls and the 95% confidence interval.

work scheduling and interruptions have an important role to play.[29]

Self-reported responses to other survey modules also provide indirect support for the importance of child-care and domestic interruptions in explaining gender differences in online work.[30] Women are 10 percentage points more likely than men to report that having a task that was "easy to do alongside caring responsibilities" was "extremely important" in their choice of what to work on. Women are 12 percentage points more likely to report that "I can only work from home" as an "extremely important" factor explaining why they crowd-work. Caring decisions also appear differentially to affect men and women's ability to complete tasks; women are 10 percentage points more likely to report that if they return a task, it is at least common for them to have returned it due to caring responsibilities.[31]

Finally, I am also able to assess gender differences in the quality of survey responses, which could also generate differences in completion time and thus effective hourly wages. To this end, I consider the length of all open text responses to the survey (Figure A.6).[32] One cannot reject the null hypothesis of equality of distribution to the length of open text responses across men and women using a Kolmorgov-Smirnov test (p-value: 0.316).

[29]It is also worth noting that there are no statistically significant differences in the translation of demographic characteristics such as age and education into wages for those with and without children: Table A.5.

[30]See Appendix Figure A.7 for full distribution of results to these variables.

[31]Note that there are no significant gender differences in the self-reported importance of piece-rate for task selection (60.6% of women and 57.4% of men report this as extremely important).

[32]i.e. The length of peoples' answers to questions such as "How do you decide which tasks to complete? How do you search for the best tasks for you?".

6 Conclusion

The findings in this paper demonstrate that family responsibilities differentially affect men and women's ability to engage in paid work, even when this work is performed within the home in a gender-blind setting. Women on MTurk earn 20% less per hour than men despite the absence of any significant gender difference in task selection or total experience. Women are less likely to work in continuous batches on tasks and are more likely to take longer breaks between submitting one task and starting another. These work patterns are associated with significantly slower task completion times for both sexes but are more common amongst women. Heterogeneity in the wage penalty by whether workers have children or not, is consistent with childcare responsibilities placing more constraints on women's ability to schedule paid work without interruption: there is no wage gap for individuals without children.

These results are important for three reasons. First, childcare responsibilities interact with paid work in different ways for men and women, even when work is performed in the home. This has consequences for the extent to which online labour markets and telework might be able to equalise men and women's economic opportunities. This has also taken on a heightened importance with the rise of working from home initiated by the pandemic. Second, the finding of no significant gender difference in task selection and accumulation of experience suggests that gender differences in the nature of work are not given and may not be as salient in online labour markets as the occupational segregation in offline labour markets might suggest. Finally, interruptions can have real effects on productivity. Children are not the only factor producing interruptions; there is increasing evidence that modern technology produces many more ways for our attention to be unconsciously diverted (Duke and Montag 2017; Kushlev, Proulx, and Dunn 2016). These results call for further research into the effectiveness of different strategies that can employed to reduce the influence of family responsibilities on work in the home and how task design can be adjusted to minimise the impact of interruptions.

References

Adams, A. and J. Berg (2017). When home affects pay: An analysis of the gender pay gap among crowdworkers. *Available at SSRN 3048711*.

Alon, T., S. Coskun, M. Doepke, D. Koll, and M. Tertilt (2021). From mancession to shecession: Women's employment in regular and pandemic recessions. Technical report, National Bureau of Economic Research.

Andrew, A., S. Cattan, M. Costa Dias, C. Farquharson, L. Kraftman, S. Krutikova, A. Phimister, and A. Sevilla (2020). The gendered division of paid and domestic work under lockdown.

Berg, J. (2015). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comp. Lab. L. & Pol'y J. 37*, 543.

Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics 2*(3), 228–55.

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

Boas, T. C., D. P. Christenson, and D. M. Glick (2020). Recruiting large online samples in the united states and india: Facebook, mechanical turk, and qualtrics. *Political Science Research and Methods 8*(2), 232–250.

Bohn, R. (2000). Stop fighting the fires. *Harvard Business Review 78*(4), 83–92.

Bolotnyy, V. and N. Emanuel (2018). Why do women earn less than men? evidence from bus and train operators. Technical report, Working Paper.

Buser, T. and N. Peter (2012). Multitasking. *Experimental Economics 15*(4), 641–655.

Card, D., A. R. Cardoso, and P. Kline (2015). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Technical report, National Bureau of Economic Research.

Coviello, D., A. Ichino, and N. Persico (2014). Time allocation and task juggling. *American Economic Review 104*(2), 609–23.

Coviello, D., A. Ichino, and N. Persico (2015). The inefficiency of worker time use. *Journal of the European Economic Association 13*(5), 906–947.

Difallah, D., E. Filatova, and P. Ipeirotis (2018). Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pp. 135–143.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2018). Monopsony in online labor markets. Technical report, National Bureau of Economic Research.

Duke, É. and C. Montag (2017). Smartphone addiction, daily interruptions and self-reported productivity. *Addictive behaviors reports 6*, 90–95.

Folbre, N., J. Yoon, K. Finnoff, and A. S. Fuligni (2005). By what measure? family time devoted to children in the united states. *Demography 42*(2), 373–390.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics 34*(2), 509–543.

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

Hara, K., A. Adams, K. Milland, S. Savage, C. Callison-Burch, and J. P. Bigham (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 449. ACM.

Ipeirotis, P. G. (2010). Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students, Forthcoming*.

Irani, L. (2015). The cultural work of microwork. *New Media & Society 17*(5), 720–739.

Irani, L. C. and M. Silberman (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 611–620. ACM.

Kaplan, T., S. Saito, K. Hara, and J. P. Bigham (2018). Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

Kässi, O., V. Lehdonvirta, and J.-M. Dalle (2019). Workers' task choice heuristics as a source of emergent structure in digital microwork.

Kossek, E. E., B. A. Lautsch, and S. C. Eaton (2006). Telecommuting, control, and boundary management: Correlates of policy use and practice, job control, and work–family effectiveness. *Journal of Vocational Behavior 68*(2), 347–367.

Kuek, S. C., C. Paradi-Guilford, T. Fayomi, S. Imaizumi, P. Ipeirotis, P. Pina, and M. Singh (2015). The global opportunity in online outsourcing.

Kushlev, K., J. Proulx, and E. W. Dunn (2016). Silence your phones: Smartphone notifications increase inattention and hyperactivity symptoms. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 1011–1020. ACM.

Larivière, V., C. Ni, Y. Gingras, B. Cronin, and C. R. Sugimoto (2013). Bibliometrics: Global gender disparities in science. *Nature News 504*(7479), 211.

Necka, E. A., S. Cacioppo, G. J. Norman, and J. T. Cacioppo (2016). Measuring the prevalence of problematic respondent behaviors among mturk, campus, and community participants. *PloS one 11*(6), e0157732.

Pelletier, A. and C. Thomas (2018). Information in online labour markets. *Oxford Review of Economic Policy 34*(3), 376–392.

Perlow, L. A. (1999). The time famine: Toward a sociology of work time. *Administrative science quarterly 44*(1), 57–81.

Tremblay, D.-G. (2002). Balancing work and family with telework? organizational issues and challenges for women and managers. *Women in Management Review 17*(3/4), 157–170.

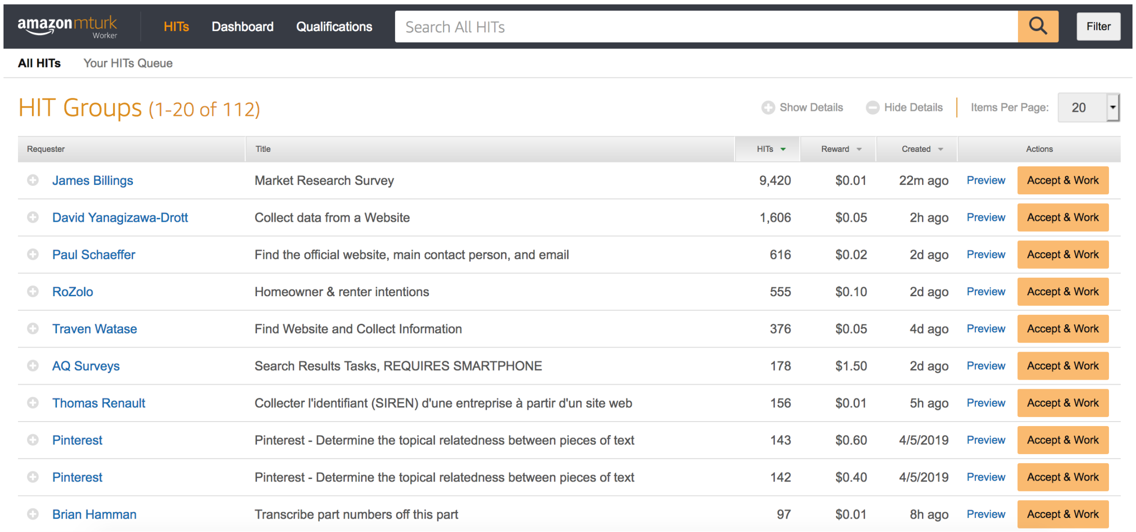
Vasilescu, B., K. Blincoe, Q. Xuan, C. Casalnuovo, D. Damian, P. Devanbu, and V. Filkov (2016). The sky is not the limit: multitasking across github projects. In *2016 IEEE/ACM 38th International Conference* *on Software Engineering (ICSE)*, pp. 994–1005. IEEE.

Westmarland, N. and J. Anderson (2001). Safe at the wheel? security issues for female taxi drivers. *Security Journal 14*(2), 29–40.

A Appendix

A.1 Data

Figure A.1: Screenshot of MTurk Interface



Requester	Title	HITs	Reward	Created	Actions	
James Billings	Market Research Survey	9,420	$0.01	22m ago	Preview	Accept & Work
David Yanagizawa-Drott	Collect data from a Website	1,606	$0.05	2h ago	Preview	Accept & Work
Paul Schaeffer	Find the official website, main contact person, and email	616	$0.02	2d ago	Preview	Accept & Work
RoZolo	Homeowner & renter intentions	555	$0.10	2d ago	Preview	Accept & Work
Traven Watase	Find Website and Collect Information	376	$0.05	4d ago	Preview	Accept & Work
AQ Surveys	Search Results Tasks, REQUIRES SMARTPHONE	178	$1.50	2d ago	Preview	Accept & Work
Thomas Renault	Collecter l'identifiant (SIREN) d'une entreprise à partir d'un site web	156	$0.01	5h ago	Preview	Accept & Work
Pinterest	Pinterest - Determine the topical relatedness between pieces of text	143	$0.60	4/5/2019	Preview	Accept & Work
Pinterest	Pinterest - Determine the topical relatedness between pieces of text	142	$0.40	4/5/2019	Preview	Accept & Work
Brian Hamman	Transcribe part numbers off this part	97	$0.01	8h ago	Preview	Accept & Work

Notes: Screenshot of MTurk worker interface taken on 8th December 2019.

Table A.1: Summary Statistics: All & Gender Matched Sample

	Full Sample			Matched Sample		
	Matched	Unmatched	Diff	Female	Male	Diff
Task Level						
Piece rate ($)	0.105	0.086	0.020***	0.103	0.109	-0.006
	(0.005)	(0.005)	(0.007)	(0.006)	(0.010)	(0.011)
Completion time (min)	2.118	2.214	-0.096	2.252	1.952	0.300
	(0.099)	(0.118)	(0.154)	(0.120)	(0.152)	(0.193)
Task Type:						
Data Entry	0.157	0.208	-0.051**	0.147	0.170	- 0.023
	(0.013)	(0.020)	(0.024)	(0.011)	(0.026)	(0.029)
Verification	0.136	0.140	-0.004	0.144	0.127	0.018
	(0.012)	(0.012)	(0.017)	(0.017)	(0.017)	(0.024)
Viewing	0.006	0.008	-0.002	0.005	0.007	-0.002
	(0.001)	(0.004)	(0.004)	(0.001)	(0.002)	(0.003)
Rating	0.156	0.149	0.008	0.154	0.160	-0.006
	(0.014)	(0.019)	(0.024)	(0.014)	(0.026)	(0.030)
Transcription	0.169	0.191	-0.022	0.181	0.154	0.027
	(0.014)	(0.024)	(0.028)	(0.015)	(0.024)	(0.028)
Research	0.207	0.160	0.047*	0.190	0.227	-0.037
	(0.019)	(0.015)	(0.024)	(0.012)	(0.039)	(0.041)
Other	0.169	0.144	0.024	0.178	0.156	0.022
	(0.011)	(0.011)	(0.015)	(0.015)	(0.012)	(0.021)
Wage ($ p/h)	6.972	6.483	0.487	6.174	7.966	-1.791*
	(0.466)	(0.439)	(0.635)	(0.287)	(0.928)	(0.970)
log Wage	1.221	1.052	0.169*	1.125	1.341	-0.216**
	(0.055)	(0.071)	(0.090)	(0.057)	(0.095)	(0.111)
Tasks per Day	43.11	45.03	-1.918	41.651	45.091	-3.439
	(2.204)	(2.988)	(3.712)	(2.433)	(4.003)	(4.684)
N Tasks	1,619,463	854,216	2,473,679	898,302	721,161	1,616,850
Worker Level						
Mean Tasks per Month	264.7	290.2	-25.42	272.2	256.9	15.32
	(14.02)	(20.53)	(24.86)	(17.02)	(22.52)	(28.23)
Mean Earnings per Month ($)	39.74	34.91	4.836*	41.42	37.98	3.444
	(1.703)	(2.011)	(2.636)	(2.428)	(2.385)	(3.405)
Mean Completion Time (min)	3.111	3.241	-0.128	3.188	3.029	0.162*
	(0.047)	(0.073)	(0.087)	(0.064)	(0.070)	(0.094)
N Workers	1,805	878	2,683	936	869	1,805

Notes: Standard errors (clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Task-Level Wage Regressions: Task Characteristic & Experience Controls

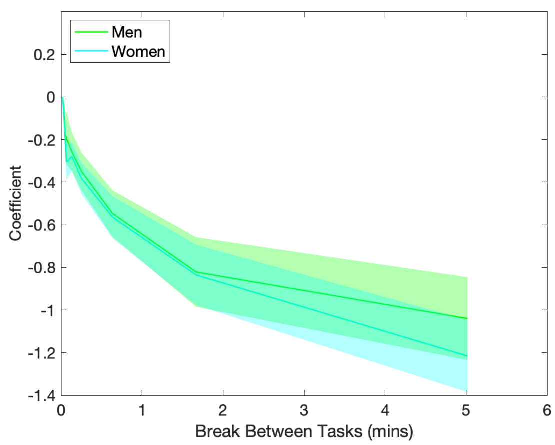
	(1)	(2)	(3)	(4)	(5)
Female	-0.2024*	-0.2010**	-0.1925**	-0.1772**	0.1772**
	(0.0525)	(0.0399)	(0.0277)	(0.0267)	(0.0267)
Log Piece Rate		0.2468***	0.2236***	0.2462***	0.7538***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Entry & Discovery			0.0000	0.0000	0.0000
			(.)	(.)	(.)
Verification			0.2821*	0.2762*	-0.2762*
			(0.0553)	(0.0676)	(0.0676)
Viewing			-0.0417	-0.0064	0.0064
			(0.8908)	(0.9832)	(0.9832)
Rating			0.8211***	0.8005***	-0.8005***
			(0.0000)	(0.0000)	(0.0000)
Transcription			-0.2506**	-0.2179**	0.2179**
			(0.0147)	(0.0349)	(0.0349)
Research			0.4741***	0.4559***	-0.4559***
			(0.0001)	(0.0001)	(0.0001)
Other			0.4929***	0.4960***	-0.4960***
			(0.0000)	(0.0000)	(0.0000)
Log Experience				0.0964***	-0.0964***
				(0.0000)	(0.0000)
Observations	1616850	1616850	1616850	1616850	1616850

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)-(5) give standard mean regressions of log wages on the controls indicated. Column (5) regresses log working time on controls. All specifications include year-quarter fixed effects. Experience refers to the total tasks completed at the point at which task j is performed.

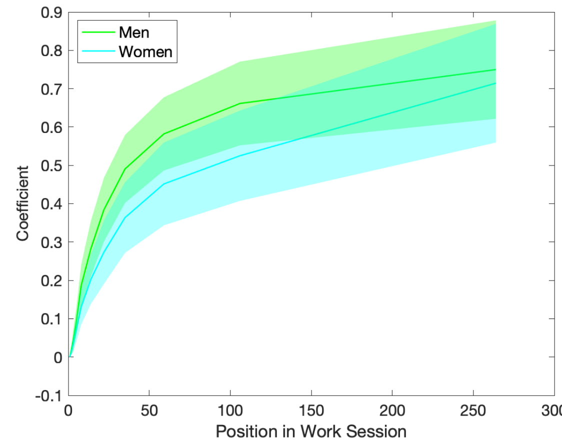
A.2 Work Pattern

Figure A.2: Work Schedule Controls & Wages: By Gender & Task Type including Worker Fixed Effects

Batch Tasks

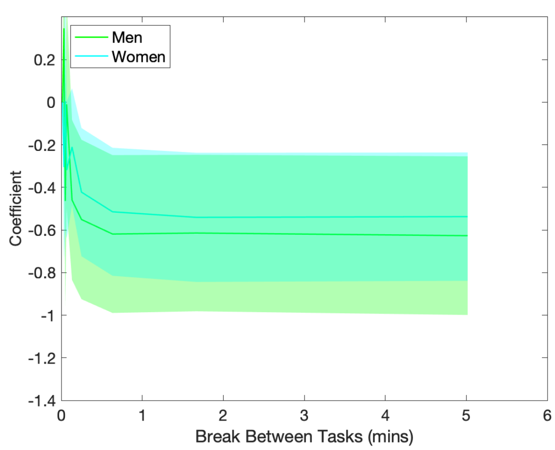


(a) Coefficients on Decile of Break Controls

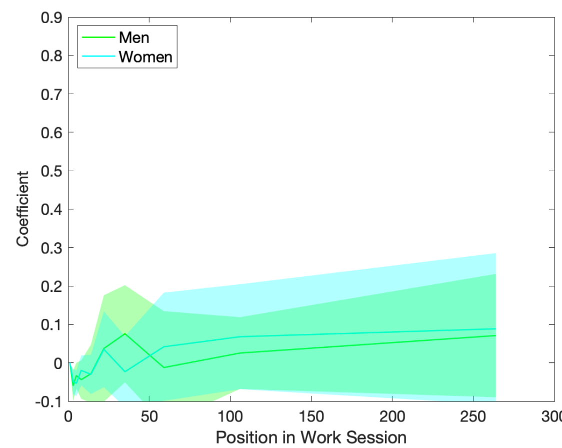


(b) Coefficients on Decile of Work Session Controls

Unit Tasks



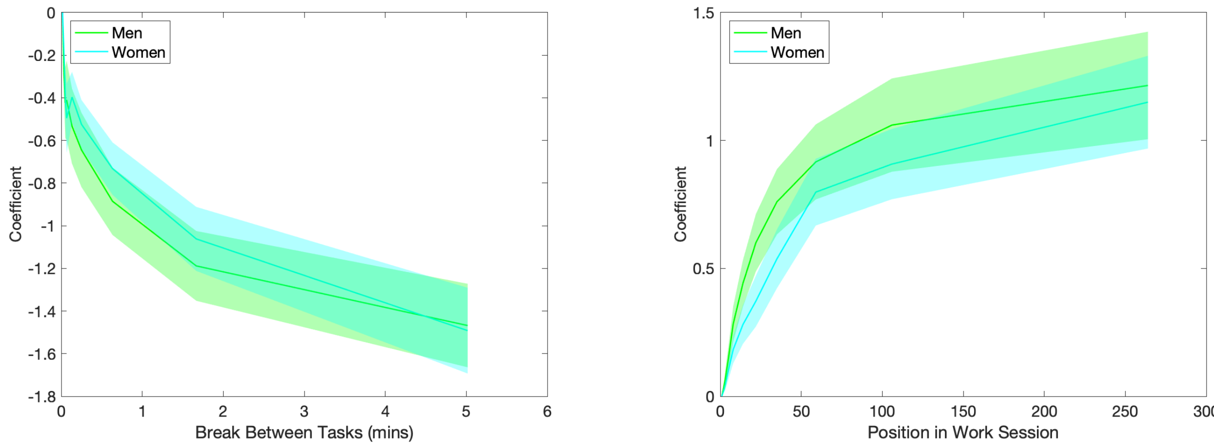
(c) Coefficients on Decile of Break Controls



(d) Coefficients on Decile of Work Session Controls

Notes: Figure gives the coefficients on the break and work session controls in separate wage regressions by split by batch and unit tasks and by gender. All specifications include worker fixed effects and standard errors are clustered at the worker level.

Figure A.3: Work Schedule Controls & Wages: Pooled OLS Specification

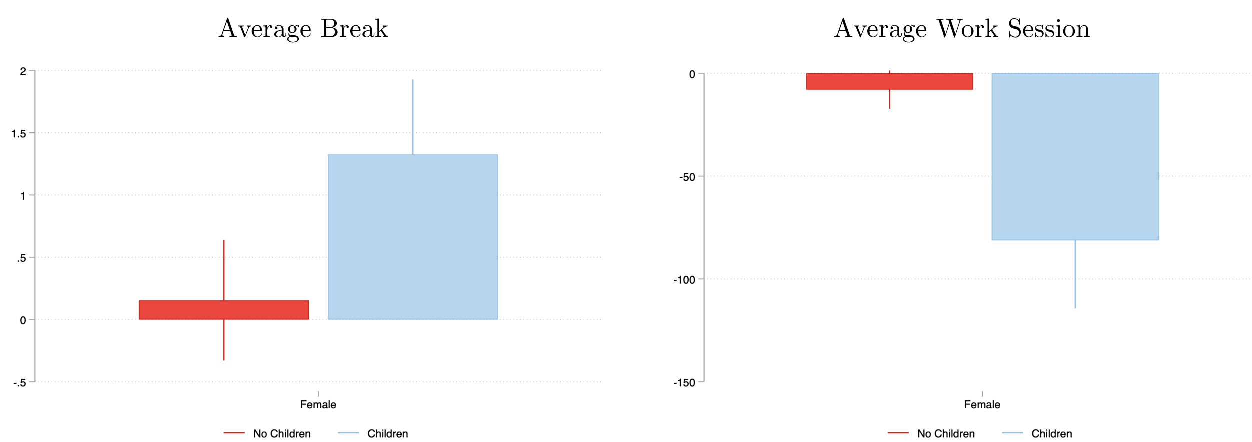


(a) Coefficients on Decile of Break Controls (b) Coefficients on Decile of Work Session Controls

Notes: Panels (a) and (b) give the coefficients on break length and work session position deciles by men and women for wage regressions including the additional task characteristic and experience controls at in Table 1 along with the 95% confidence interval. Standard errors are clustered at the worker level.

A.3 Survey Evidence

Figure A.4: Difference in Work Patterns: By Children



Notes: Mean difference in work session length and break between tasks for workers who completed the survey by whether they report having children in the household or not.

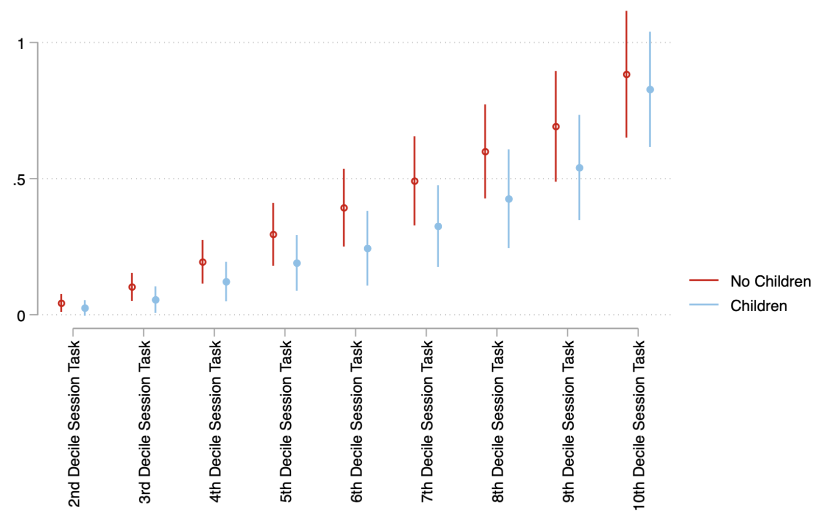
Table A.3: Task-Level Wage Regressions: By Children

	Pooled		Children v. No Children			
			No Children	Children	No Children	Children
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.2025***	-0.1251*	-0.0966	-0.3416***	0.0061	-0.1962**
	(0.0089)	(0.0569)	(0.3009)	(0.0046)	(0.9244)	(0.0487)
p-value:	0.0932		0.1059		0.0870	
Task Controls	no	yes	no	no	yes	yes
Work Pattern Controls	no	yes	no	no	yes	yes
Observations	833075	833075	342288	490787	342288	490787

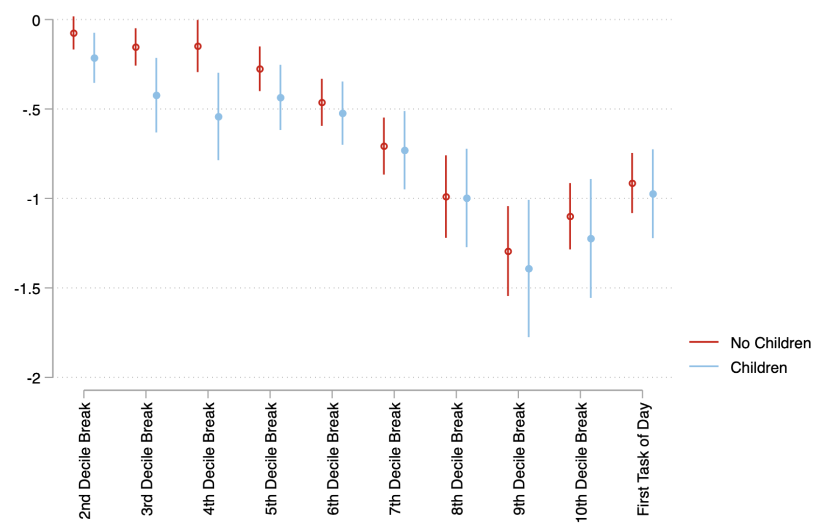
Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include year-quarter fixed effects. Task characteristics refers to all controls in column (4) of Table A.2. Work pattern controls are those described in Section 4: deciles of position in work session; deciles of break between adjacent tasks; whether a task overlaps with an adjacent task; whether the task is the same HIT as the previous; hour of day. p-value of test of equality of $Female$ coefficient across specifications (implied by standard errors clustered at the worker level) given in main table.

Figure A.5: Work Pattern Coefficients: By Children

Break

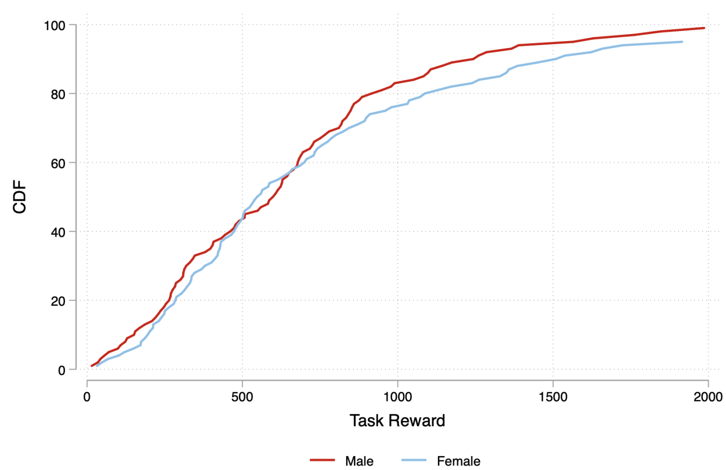


Work Session



Notes: Panels (a) and (b) give the coefficients on break length and work session position deciles by whether a worker has children for wage regressions including the additional task characteristics and experience controls at in Table 1 along with the 95% confidence interval. Standard errors are clustered at the worker level.

Figure A.6: Total Length to Open Text Responses



A.4 Additional Survey Results

Note that consistent with prior studies, my sample is disproportionately highly educated. My follow-up survey reveals that 40% of workers have a College degree or higher. Berg (2015) report that 45% of MTurk workers have at least College level education. Boas et al (2018) do not report Education in a directly comparable manner. However, in their sample of US MTurk workers, average education is 3.88 compared to 2.87 in a probability sample (the American National Election Study). Other studies have also found a similar proportion of respondents who work in the offline economy to those in my sample (Adams and Berg (2017): 74% to 64%).

Figure A.7: Self-Reported Importance of Caring Responsibilities for Task Choice

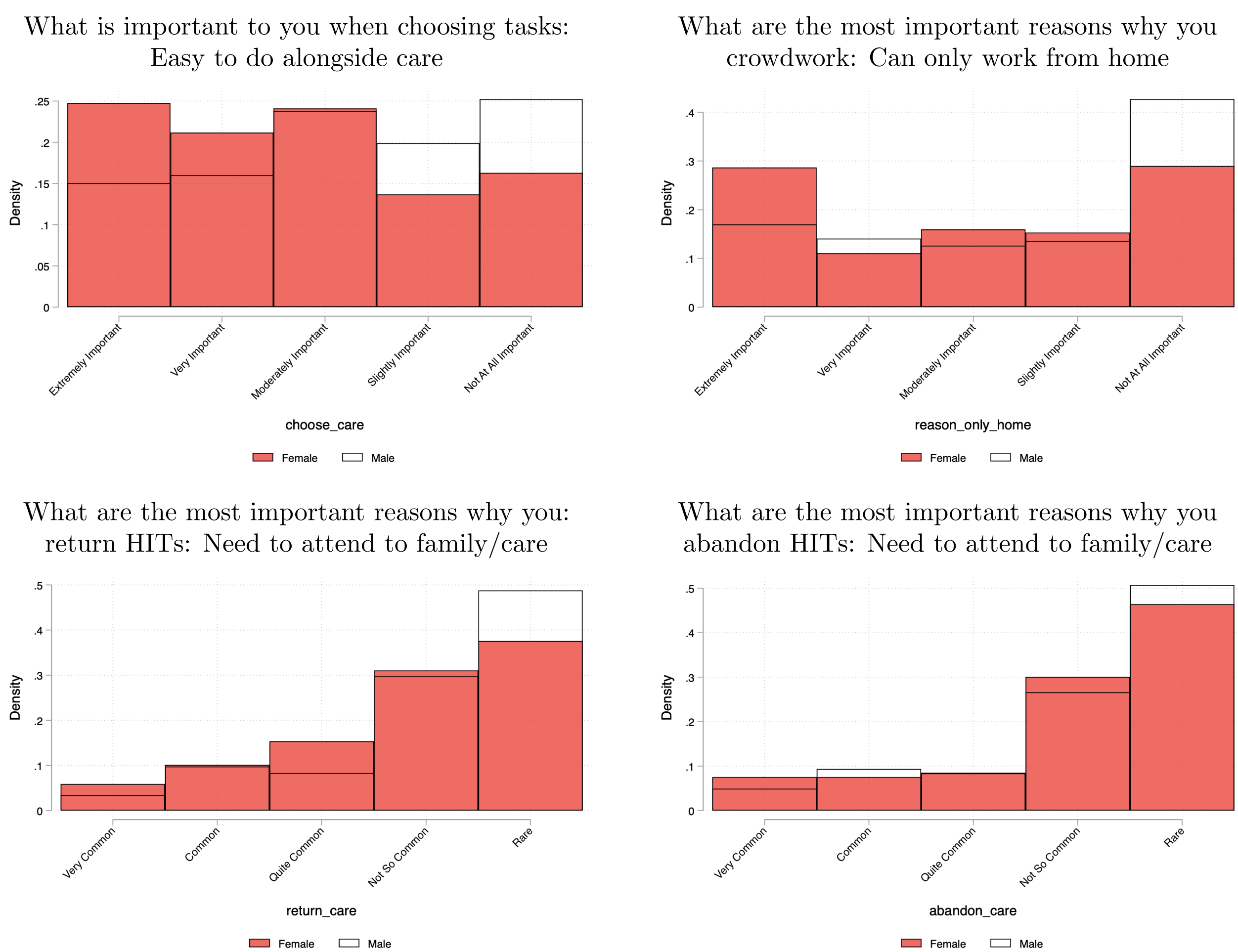


Table A.4: Summary Statistics: Survey Responses

	Female	Male	Difference
Demographics			
Children	0.6635	0.4152	0.2483***
	(0.0232)	(0.0297)	(0.0373)
College Education	0.3534	0.4513	-0.0979**
	(0.0235)	(0.0300)	(0.0377)
Bad Health	0.0697	0.0433	0.0264
	(0.0125)	(0.0122)	(0.0183)
Age	40.27	36.89	3.379***
	(0.5583)	(0.6512)	(0.8664)
Employment			
Full Time	0.3582	0.5523	-0.1942***
	(0.2353)	(0.0299)	(0.0377)
Part Time	0.1250	0.0830	0.0420*
	(0.0162)	(0.0166)	(0.0241)
Homemaker	0.1490	0.0361	0.1129***
	(0.0175)	(0.0112)	(0.0233)
Retired	0.0601	0.0361	0.0240
	(0.0112)	(0.0112)	(0.0170)
Self Employed	0.1034	0.0578	0.0456**
	(0.0149)	(0.0140)	(0.0216)
Unemployed	0.0529	0.1083	-0.0554***
	(0.0110)	(0.0187)	(0.0203)

Notes: Standard errors given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Wage Regressions: Demographics

	All	No Children	Children	p-value
Female	-0.1375*	-0.0731	-0.3258**	0.1973
	(0.0990)	(0.4911)	(0.0492)	
Children	-0.0272			
	(0.7707)			
College +	-0.0387	-0.1031	0.0379	0.4554
	(0.7123)	(0.3435)	(0.8066)	
Age	-0.0106*	-0.0064	-0.0082	0.8722
	(0.0861)	(0.1562)	(0.4321)	
Bad Health	-0.1028	0.1278	-0.3521	0.1015
	(0.4715)	(0.3699)	(0.1707)	
Full Time	0.0000	0.0000	0.0000	-
	(.)	(.)	(.)	
Part Time	-0.3285***	-0.2829**	-0.3755**	0.6550
	(0.0020)	(0.0310)	(0.0205)	
Homemaker	-0.3940**	-0.3950	-0.3220*	0.8136
	(0.0127)	(0.1244)	(0.0650)	
Retired	-0.4021	-0.5562*	-0.1935	0.5358
	(0.2101)	(0.0653)	(0.7008)	
Self-Employed	-0.3508**	-0.2781**	-0.3431	0.8201
	(0.0263)	(0.0457)	(0.1717)	
Unemployed	-0.1197	-0.1135	-0.0903	0.9257
	(0.3182)	(0.5467)	(0.5793)	
Observations	832752	341965	490787	

Notes: Standard errors given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Selection into Survey

	(1)	(2)	(3)
Average Wage	-0.0002	-0.0003	-0.0004
	(0.4247)	(0.3338)	(0.3017)
Average Not-Batch	0.4114***	0.2820***	0.3816***
	(0.0000)	(0.0009)	(0.0000)
1st Quartile Experience	0.0000	0.0000	0.0000
	(.)	(.)	(.)
2nd Quartile Experience	0.0501**	0.0429	0.0638**
	(0.0294)	(0.1296)	(0.0314)
3rd Quartile Experience	0.1300***	0.1162***	0.1529***
	(0.0000)	(0.0000)	(0.0000)
4th Quartile Experience	0.2217***	0.2457***	0.2808***
	(0.0000)	(0.0000)	(0.0000)
Months since Last Task	-0.0061***	-0.0090***	-0.0078***
	(0.0010)	(0.0001)	(0.0008)
Female		0.0751***	
		(0.0001)	
Female (Inc Survey)			0.0914***
			(0.0000)
Observations	2626	1762	1932
R^2	0.0484	0.0653	0.0705

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table gives the results of a linear probability model in which a dummy for where a worker completed the survey is regressed on characteristics derived from the task log information. Column (1) gives results on the full sample, column (2) gives the gender matched sample based only on the names given to sign up for the CrowdWorkers app, and column (3) gives the gender matched sample based on survey responses also.

A.5 Robustness

Table A.7: Excluding Overlapping Tasks & Known Gender Measurement Error

	Exclude Overlap		Drop Mis-Measured Gender		Use Survey Gender	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.1220*	-0.0356	-0.1937**	-0.0950		
	(0.1000)	(0.5215)	(0.0291)	(0.1320)		
Female (Survey)					-0.1719**	-0.1021
					(0.0381)	(0.1041)
Observations	1374981	1374981	1592435	1592435	1854140	1854140
R^2	0.1772	0.3107	0.1503	0.3490	0.1357	0.3314
Task Characteristics	yes	yes	yes	yes	yes	yes
Work Pattern Characteristics	no	yes	no	yes	no	yes

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * p < 0.10, ** p < 0.05, *** p < 0.01. Year-quarter fixed effects in all specifications. Columns (1)-(2) drop observations where task accept and submit times overlap. Columns (3)-(4) exclude workers where there is an inconsistency between the gender predicted from their first name and the survey response. Columns (5)-(6) use self-reported gender from the survey.

Table A.8: Wage Regressions: Matched versus Unmatched Samples

	Task Characteristics		+ Work Session		p-value
	(1)	(2)	(3)	(4)	
Log Piece Rate	0.2462***	0.2924***	0.4079***	0.4132***	0.9125
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
Entry & Discovery	0.0000	0.0000	0.0000	0.0000	
	(.)	(.)	(.)	(.)	
Verification	0.2645*	0.3232**	0.2174**	0.3464***	0.4092
	(0.0892)	(0.0285)	(0.0397)	(0.0027)	
Viewing	-0.0026	-0.7913	-0.0041	-0.3904*	0.2217
	(0.9931)	(0.1271)	(0.9859)	(0.0675)	
Rating	0.7953***	0.6209***	0.6221***	0.5329***	0.6180
	(0.0000)	(0.0015)	(0.0000)	(0.0007)	
Transcription	-0.2288**	-0.3092*	-0.1802**	-0.1826	0.9854
	(0.0239)	(0.0641)	(0.0133)	(0.1096)	
Research	0.4555***	0.5830***	0.3602***	0.4548***	0.4752
	(0.0001)	(0.0001)	(0.0000)	(0.0000)	
Other	0.4844***	0.3696***	0.3789***	0.2933***	0.4990
	(0.0000)	(0.0066)	(0.0000)	(0.0038)	
Log Experience	0.0989***	0.0932***	0.0528***	0.0511*	0.9598
	(0.0001)	(0.0013)	(0.0020)	(0.0868)	
Observations	1616850	853063	1616850	853063	
Task Characteristics	yes	yes	yes	yes	
Work Pattern Characteristics	no	no	yes	yes	

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * p < 0.10, ** p < 0.05, *** p < 0.01. Year-quarter fixed effects in all specifications. Column (5) reports the p-value on a test of equality of coefficients between columns (3) and (4).

A.6 Task Type Classification

Task Type To classify tasks into broader themes, I largely follow the approach of Hara et al (2018). Using the same underlying data as this paper, these authors used the task title, description, and keywords to develop a classification schema for MTurk task type. A K-Means algorithm was used to cluster tasks that were close to each other in the latent space of descriptive words used to describe. Using their methodology, the following types were identified: surveys; research; verification; viewing; data entry; rating; content creation. As only 433 tasks were tagged as a research task, I merge surveys and research into one category. Based on the most common words used to describe tasks in each category as reported by Hara et al (2018), I tag tasks according to a key word search of the task title and description. Table A.9 shows the proportion of unit tasks by each task type.

Table A.9: Proportion of Unit Tasks by Task Type

	Unit Tasks	Std. Error
Data Entry	0.0029	0.0000
Verification	0.0078	0.0002
Viewing	0.0269	0.0013
Rating	0.0084	0.0002
Transcription	0.0029	0.0000
Research	0.1109	0.0004
Other	0.0326	0.0003