DISCUSSION PAPER SERIES

DP14294

THE GENDER WAGE GAP ON AN ONLINE LABOUR MARKET: THE COST OF INTERRUPTIONS

Abigail Adams

LABOUR ECONOMICS

ISSN 0265-8003

THE GENDER WAGE GAP ON AN ONLINE LABOUR MARKET: THE COST OF INTERRUPTIONS

Abigail Adams

Discussion Paper DP14294
Published 09 January 2020
Submitted 08 January 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE GENDER WAGE GAP ON AN ONLINE LABOUR MARKET: THE COST OF INTERRUPTIONS

Abstract

This paper analyses gender differences in working patterns and wages on Amazon Mechanical Turk. Using information on 2 million tasks, I find no gender difference in task selection nor experience on the platform. Nonetheless, women earn 20% less per hour on average. Half of this gap is explained by differences in the scheduling of work; women have more fragmented work patterns with consequences for their task completion speed. A follow up survey shows that the wage gap is concentrated amongst women with young children, who also report that domestic responsibilities affect their ability to plan and complete work online.

JEL Classification: J16, J24, J31

Keywords: Gender Difference, Flexibility, Multitasking, Telework, Gig economy

Abigail Adams - abi.adams@economics.ox.ac.uk
University of Oxford and CEPR

Acknowledgements
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.

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

Abi Adams-Prassl*

January 6, 2020

Abstract

This paper analyses gender differences in working patterns and wages on Amazon Mechanical Turk. Using information on 2 million tasks, I find no gender difference in task selection nor experience on the platform. Nonetheless, women earn 20% less per hour on average. Half of this gap is explained by differences in the scheduling of work; women have more fragmented work patterns with consequences for their task completion speed. A follow up survey shows that the wage gap is concentrated amongst women with young children, who also report that domestic responsibilities affect their ability to plan and complete work online.

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

*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, Oriana Bandiera, Janine Berg, Giulia Giupponi, Hamish Low, Karl Schlag, and Johanna Rickne for helpful comments.

and Emanuel 2018; Cook et al. 2018), these choices have consequences for women's earnings and the gender wage gap. For example, when offered the same rota and remuneration opportunities, Bolotnyy and Emanuel (2018) find that women select lower paying work schedules to avoid work times when children are more likely to be at home and they are less likely to accept better remunerated overtime when it is offered at last minute. 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.

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. The recent rise of online labour market platforms, in which work is both advertised and performed online, provides promise in this regard. 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). Online platform work can often provide workers with greater flexibility to substitute between market and household production as it typically does not require face-to-face interaction nor colocation with employers, customers or other workers (Pelletier and Thomas 2018). Online work might further be associated with lower fixed costs of employment if additional childcare is not required while work is performed at home.

In this paper, I analyse whether there exist gender differences in work patterns and hourly wages on a popular online labour market, Amazon Mechanical Turk (MTurk), using a unique task-level data set. 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 typically performed in the home. There is thus a lower cost of substituting between domestic production and market work than has been the case in the related literature.

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 or characteristics of tasks selected, and men and women complete a similar number of tasks per month. There is no significant gender difference in monthly

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

earnings and I also find no difference in the (albeit imperfect) measure of work quality that I am able to observe. This suggests that gender differences in preferences for tasks and work are not salient in this setting.

Nonetheless, women earn 21% less per hour on average. This is because women take longer on average to complete similar tasks to men. There is no gender difference in monthly earnings because women work for longer overall. Drawing on the literature on multi-tasking (Vasilescu et al. 2016; Coviello et al. 2014), I am able to explain 50% of this hourly wage gap by differences in the work scheduling patterns of men and women. Women have more fragmented work patterns when working online with consequences for their productivity. I find that completing a large number of tasks in quick succession ("batch working") is associated with faster task completion times. 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 with consequences for women's average task completion speed.

A follow up survey of the workers in my task-level data set shows that the gender difference in hourly wages is confined to women with children; there is no gender difference in wages for individuals without children. Examining heterogeneity by the age of the youngest child in the household, the gender wage gap is largest for women with small children (aged under five) in the household. These stylized facts are 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). Supporting this hypothesis, I find that women with children are more likely to report facing interruptions and having to abandon work on a task because of a need to "attend to caring responsibilities" and that they self-report that these responsibilities affect their working on the platform. 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 Context & Data

MTurk is one of the largest online micro-task platforms in the world. It is popular amongst economists as a forum for posting survey and experimental tasks as well as being a labour market of interest in its own right (Dube, Jacobs, Naidu, and Suri 2018). 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.[2] 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 tasks and its payment structure rules out a significant role for negotiation and competition in governing rates of pay, which have also been found to be drivers of gender differences in wages and working patterns in the labour market (Card, Cardoso, and Kline 2015).

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

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 other potential 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 with intoxicated consumers) which results in women facing higher costs to work in the industry (Westmarland and Anderson 2001).

2.1 Data

Most prior studies of MTurk have relied on survey responses or on information scraped directly from the MTurk interface (Dube, Jacobs, Naidu, and Suri 2018; Adams and Berg 2017; Ipeirotis 2010). While both types of data source have some key advantages, they are of limited use for a detailed study of worker behaviour. Data scraped directly from MTurk, for example, does not capture which workers are completing the tasks nor task completion times. While survey evidence allows remuneration to be understood at the worker level, collection of detailed task information and work diaries has proved difficult through this channel.

I use unique task-level data collected between 2015 and 2017 from a third-party MTurk plug-in, *CrowdWorkers*, to study gender differences in work patterns on the site. 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). On the basis of responses to a survey of CrowdWorker users as part of this study (discussed below), I exclude users if they state that their country of residence is not the US or UK.

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 datasets. These workers completed 65.5% of all tasks in the dataset: 1,619,463 out of my full sample of 2,473,679 tasks.[3]

This is the only data source on MTurk that I know of that contains detailed information on task completion at the worker level, enabling a detailed analysis of gender differences in work patterns at a granular level.[4] 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 little a priori reason to expect any significant gender differences in selection into the plug-in and a benefit of perhaps being disproportionately used by inexperienced workers is that a more complete account of these workers' task histories will be captured by the plugin.

Survey I conducted a short survey of *Crowd Workers* users to provide additional demographic information to complement the task-level log data. The survey was posted as a task on MTurk with a piece rate of $2.50 in December 2017 and took approximately 10 minutes to complete. I match survey responses to individual work histories on the basis of a worker's unique ID. The piece rate was high relative to the average piece-rate on MTurk (Table 1) to persuade high-wage users to complete the survey. 711 of the 1,805 workers in my sample completed the survey.[5] As discussed in the Appendix, there are some differences in the task selection patterns of workers who select into my survey (see Table A.7); they are more likely to complete survey and research tasks and complete more tasks on average. Thus, in Sections 3 and 4, I analyse the full sample of workers in the CrowdWorkers log data and in Section 5, I demonstrate that the same patterns hold in the sample of workers who completed my survey and analyse the relationship of gender differences in wages and working patterns with demographic characteristics.

[3]Gender established by this method and self-identified gender for the subset of workers who complete the demographic survey is the same for all but 25 individuals.

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

[5]Workers who identified their location as outside of the US and Europe were excluded from my primary sample given the very different labour market context.

3 Task Selection & Experience

Table 1 gives summary statistics on the characteristics of tasks completed by all workers in my dataset (the "full sample") and for the subset who could be classified as either male or female (the "matched sample"). All standard errors are clustered at the worker level. My primary unit of analysis is the worker-task level as this allows me to control for the detailed characteristics of the task and, later, detailed controls for working patterns (e.g. breaks between tasks, whether a worker switches between different sorts of tasks, and the number of tasks completed in a work session).

I classify tasks into categories, e.g. data entry, on the basis of their title description following the schema of Hara et al. (2018).[6] 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 competing at least one other task with the same batch identifier and the task piece-rate is less than \$0.50. Unit tasks are disproportionately research tasks (Table A.9).[7]

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:

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

The average effective hourly wage is computed as a worker's total earnings on MTurk divided by their total work time as recorded by the difference between submitting and accepting tasks:

$$\bar{w}_{ij} = \frac{\sum_{j \in S_i} PieceRate_j}{\sum_{j \in S_i} SubmitTime_{ij} - AcceptTime_{ij}} \tag{2}$$

where S_i is the set of tasks completed by worker i.

[6] Tasks were tagged according to a key word search of the title. Full details of the tagging procedure are given in the Appendix.

[7] Some research tasks are batch tasks, however. For example, "We want your reactions to injustice" which was completed 101 times by users who did the task more than once despite it being categorised as a research task.

Gender Matched v. Not Matched Workers who's gender could not be established complete lower value tasks on average than those in the gender matched sample, and thus have lower wages on average.[8] There are few other differences in the observables of these groups. 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 A.2 that there are no significant or qualitative 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 (Table A.3 to Table A.6).

Female v. Male Within the gender matched sample, there are no significant differences in task selection nor experience accumulated on the platform by male and female workers. 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). There are also no significant differences in the broad categories of tasks (e.g. transcription, research) nor the piece rates of tasks that the genders select.

Despite no evidence of systematic gender differences in the types of tasks selected on the platform nor overall experience, Table 1 highlights that, nonetheless, there is a significant gender difference in effective hourly pay. At the task level, women on average earn 21.6% less per hour than men. Aggregating remuneration and total working time to the worker level, women earn 10% less per hour. This is surprising given the institutional features of the platform described in Section 1 and no significant gender differences in task selection.

To examine these patterns further, Table 2 gives a set of standard wage regressions to explore the relationship between the characteristics of tasks, accumulated experience, and the gender wage gap:

$$\log w_{ij} = \beta_0 + \beta_1 Female + \rho \mathbf{X}_{ij} + \epsilon_{ij} \tag{3}$$

[8] 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.

Table 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 ($)	39.74	34.91	4.836*	41.42	37.98	3.444
	(1.703)	(2.011)	(2.636)	(2.428)	(2.385)	(3.405)
Mean Completion Time (min)	3.111	3.241	-0.128	3.188	3.029	0.162*
	(0.047)	(0.073)	(0.087)	(0.064)	(0.070)	(0.094)
Mean Hourly Wage ($ p/h)	6.256	5.355	0.894	5.994	6.499	-0.512
	(0.724)	(0.724)	(1.024)	(1.315)	(0.504)	(1.408)
Mean log Hourly Wage	1.346	1.095	0.251***	1.299	1.396	-0.097**
	(0.021)	(0.034)	(0.040)	(0.029)	(0.030)	(0.042)
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: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include time period fixed effects.

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 mean and median wage difference are approximately 21.6% and 24.0% respectively (columns (1) and (2) in Table 2).[9] Controlling for the piece rate and broad task category do little to explain the wage difference. In column (5), the total tasks completed at the point at which task j is performed (i.e. experience) is controlled for, again with little consequence for the estimated wage gap. Specifically, a set of dummy variables for the decile of tasks completed before task j are controlled for. There is no sensitivity of the coefficient on $Female$ to instead allowing for a linear or quadratic trend in experience (see Table A.1 in the Appendix).[10]

Women earn less per hour than men on average, not because they choose poorly remunerated tasks but because women complete similar tasks more slowly. This can of course be inferred from the specifications that control for the piece rate but is reported directly in column (6). On average, women take 2.25 minutes to complete a task compared to 1.95 minutes for men (Table 1). The remainder of this paper explores why this is the case.

4 Gender Differences in Working Patterns

Gender differences in working time explain the difference in effective hourly wages; on average women complete similar tasks more slowly. The literature on worker productivity (for independent work) identifies task scheduling and "multi-tasking", in addition to innate worker ability and effort, as important for completion times (Vasilescu et al. 2016). There are few characteristics available in my core dataset to facilitate a detailed exploration of ability and effort differences by gender (although I am able to explore this more fully in the auxiliary survey data described in Section 5). However, given the detailed time information available in the worker log data, I am able to consider whether gender differences in work scheduling and interruptions have a role to play in rationalising differences in effective hourly pay. There is a rich literature that demonstrates the productivity consequences of simultaneously working on multiple tasks and of interruptions. Across a

[9] The deviation between column 1 of Table 2 and 21.6% is because all regressions include time fixed effects.

[10] There are no qualitative differences in the coefficients on task characteristics when separate regressions are performed on the subsamples of female and male tasks including worker fixed effects.

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

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.2024*	-0.2397***	-0.2010**	-0.1925**	-0.1724**	0.1724**
	(0.0525)	(0.0080)	(0.0399)	(0.0277)	(0.0320)	(0.0320)
Log Piece Rate			0.2468***	0.2236***	0.2463***	0.7537***
			(0.0000)	(0.0000)	(0.0000)	(0.0000)
Task Type						
Entry & Discovery				0.0000	0.0000	0.0000
				(.)	(.)	(.)
Verification				0.2821*	0.2819*	-0.2819*
				(0.0553)	(0.0652)	(0.0652)
Viewing				-0.0417	-0.0007	0.0007
				(0.8908)	(0.9983)	(0.9983)
Rating				0.8211***	0.8086***	-0.8086***
				(0.0000)	(0.0000)	(0.0000)
Transcription				-0.2506**	-0.2120**	0.2120**
				(0.0147)	(0.0413)	(0.0413)
Research				0.4741***	0.4492***	-0.4492***
				(0.0001)	(0.0001)	(0.0001)
Other				0.4929***	0.5010***	-0.5010***
				(0.0000)	(0.0000)	(0.0000)
Constant	1.3690***	1.5041***	2.1413***	1.7174***	1.4752***	6.7135***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	1616850	1616850	1616850	1616850	1616850	1616850
Experience	no	no	no	no	yes	yes

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (3)-(5) give standard mean regressions of log wages on the controls indicated. Column (2) reports a quantile regression at the 0.5 quantile. Column (6) 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 set of dummy variables for the decile of tasks completed before task j are controlled for. See Table A.1 for alternative experience controls.

range of scenarios, it has been confirmed that workers who juggle projects take longer to complete each of them than if the tasks were completed sequentially (Buser and Peter 2012; Coviello, Ichino, and Persico 2015; Coviello, Ichino, and Persico 2014). This might be due to a mechanical effect or because workers quickly forget small details of tasks when taking time out (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 to joint projects that might distract a worker from their own task (Perlow 1999). As described by respondents to 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. Women are more likely to be "passive carers" (Folbre, Yoon, Finnoff, and Fuligni 2005), ready to be called upon when issues of childcare arise. Thus, women might be more constrained in their ability to work on continuous batches of tasks, potentially with a consequence for task completion time.

4.1 Measuring Working Patterns

These constraints could manifest themselves in two ways. First, the need to take time out to "fire fight"[11] urgent domestic tasks might result in longer intervals between completed tasks and, potentially, a more fragmented work schedule overall, i.e. a given number of tasks are less likely to be completed in succession. 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

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

while one is attending to other duties. Second, working time on tasks might be systematically mismeasured; 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 control for whether a task is the first to have been completed in a 24 hour period; 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).

- *Tasks completed previously in a work session*: 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 (Bolotnyy and Emanuel 2018). This is controlled for with hour-of-day fixed effects.

Work Schedule Results Panel (a) of Table 3 explores the explanatory power of these features of work scheduling for effective hourly earnings. While the inclusion of task characteristic controls does little to change the coefficient on $Female$, including all work scheduling variables reduces the magnitude of the gender gap by more than half. The importance of these characteristics for explaining the gender wage gap is confirmed by a Gelbach decomposition, the results of which are given in Appendix Figure A.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 explain 47.6% of the observed gender wage gap; the controls for length of break and tasks completed previously in the same work session explain 26.0% and 21.9% of the total gap respectively.

Figures 1 (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. Figure 1 (c) and (d) reveal why this reduces women's average earnings per hour; these figures show the coefficients on the break length and work session position controls in the wage regressions. Productivity increases with the number of tasks completed in a work session and decreases in the length of time taken between adjacent tasks. As women are more likely to work in a fragmented pattern, these features imply that on average they complete tasks more slowly than men. In Appendix Table A.2, I show that these patterns are unchanged when worker fixed effects are included and thus only within-worker variation is relied upon. Further, there are no qualitative differences in the coefficients on break length and work session position when separate fixed effect regressions are run by gender.

The returns to continuous work are likely to differ across different types of tasks. Unit tasks (e.g. surveys and research tasks, see Table A.9) are often designed to be completed only once limiting any gains from learning by doing within a work session. The design of batch tasks, however, generates more potential for productivity spillovers across tasks (Kässi et al. 2019). Panels (b) and (c) of Table 3 split the sample into batch and unit tasks to examine heterogeneity in the magnitude of the gender pay gap and the contribution of work pattern controls across these categories.[12] Indeed, Figure 1 (e) and (f) shows that the returns to short breaks and completing a large number of tasks per work session are much smaller for unit tasks. In line with this, I find no gender difference in hourly wage amongst unit tasks; the coefficient on $Female$ is close to zero in all specifications albeit imprecisely estimated. In other words, I find no gender difference in hourly earnings amongst tasks for which the productivity penalty to fragmented work is low.

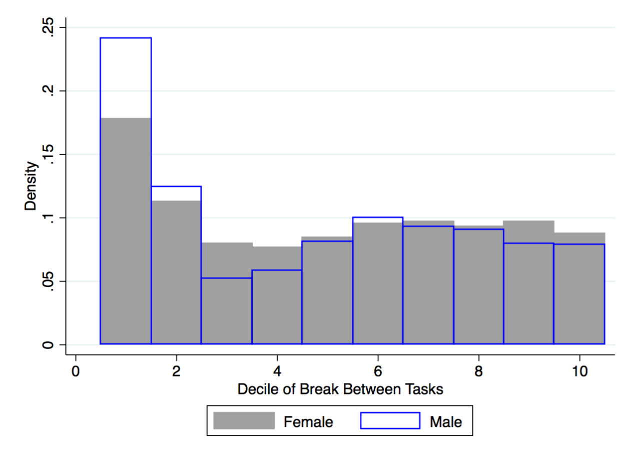
[12]The majority of tasks are Batch tasks, which is to be expected given the micro-task nature of the platform.

Table 3: Task-Level Wage Regressions: Working Pattern Controls

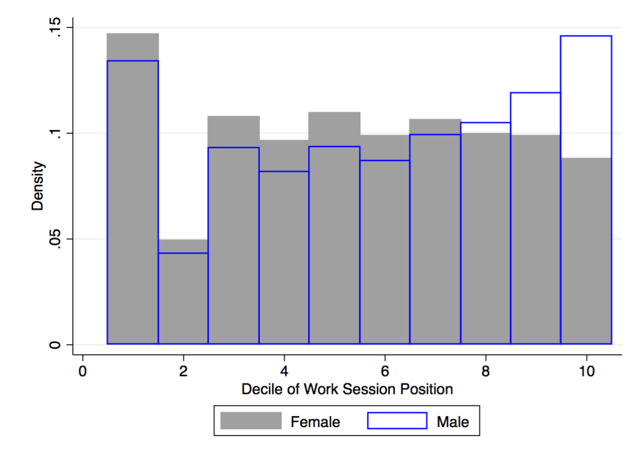
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All Tasks							
Female	-0.2024*	-0.1724**	-0.1719**	-0.1631**	-0.1289*	-0.1243**	-0.0970*
	(0.0525)	(0.0320)	(0.0288)	(0.0351)	(0.0618)	(0.0496)	(0.0843)
Observations	1616850	1616850	1616850	1616850	1616850	1616850	1616850
R^2	0.0154	0.1667	0.1698	0.1840	0.2277	0.2964	0.3270
Batch Tasks							
Female	-0.2080*	-0.1753**	-0.1749**	-0.1660**	-0.1312*	-0.1265**	-0.0986*
	(0.0514)	(0.0321)	(0.0287)	(0.0355)	(0.0619)	(0.0500)	(0.0847)
Observations	1582730	1582730	1582730	1582730	1582730	1582730	1582730
R^2	0.0159	0.1610	0.1642	0.1784	0.2229	0.2935	0.3244
Unit Tasks							
Female	-0.0115	-0.0270	-0.0276	-0.0268	-0.0253	-0.0232	-0.0224
	(0.6857)	(0.2996)	(0.2897)	(0.3027)	(0.3098)	(0.3508)	(0.3534)
Observations	34120	34120	34120	34120	34120	34120	34120
R^2	0.0032	0.1565	0.1583	0.1567	0.1584	0.1656	0.1690
Task Characteristics	no	yes	yes	yes	yes	yes	yes
Work Pattern Controls							
Time of the Day	no	no	yes	no	no	no	yes
Task Switching	no	no	no	yes	no	no	yes
Work Session Decile	no	no	no	no	yes	no	yes
Break Length Decile	no	no	no	no	no	yes	yes

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include year-quarter fixed effects. Task characteristics refers to all controls in column (6) of Table 2.

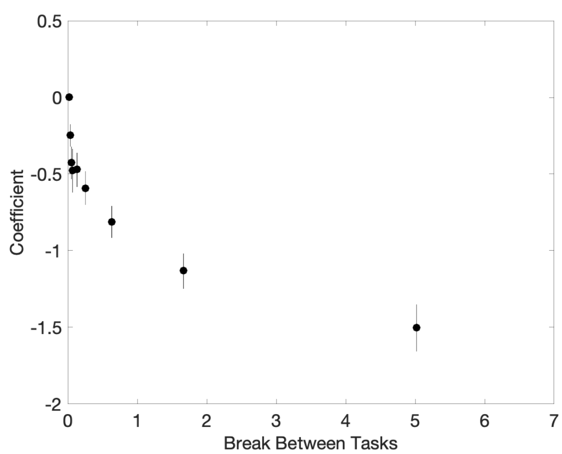
Figure 1: Work Schedule Controls & Wages



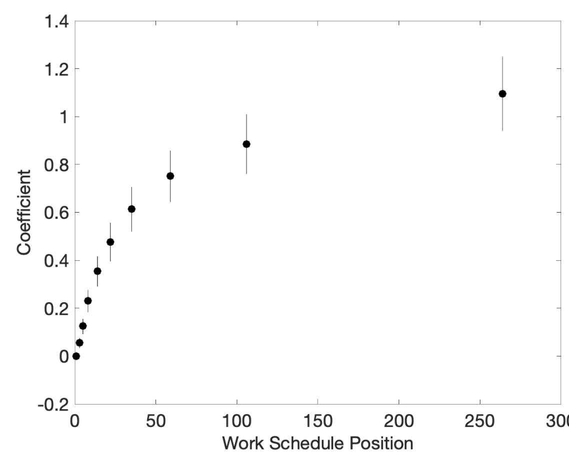
(a) Density of Break Controls



(b) Density of Work Session Controls

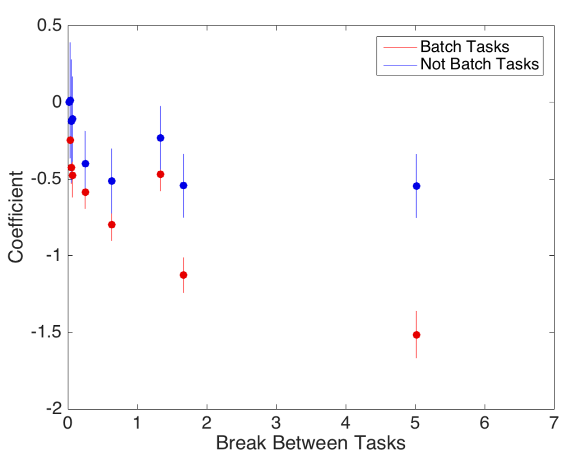


(c) Coefficients on Decile of Break Controls

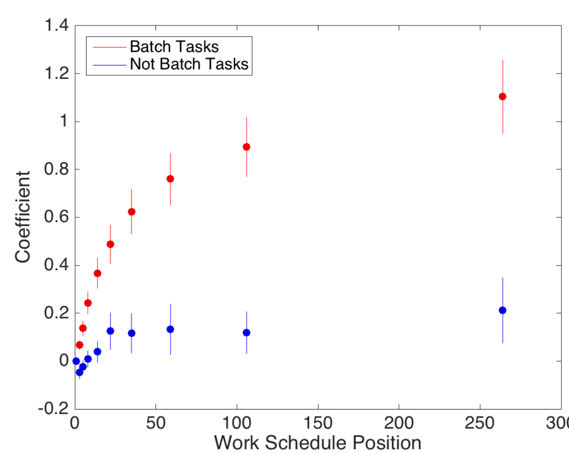


(d) Coefficients on Decile of Work Session Controls

Batch versus Unit Tasks



(e) Coefficients on Decile of Break Controls



(f) Coefficients on Decile of Work Session Controls

Notes: Panel (a) and (b) gives the distribution of break length and work session position deciles by men and women. Note that the decile groups are formed with respect to the whole data-set including unmatched workers. Panels (c) and (d) give the coefficients on these controls corresponding to columns (5) and (6) of Table 3 along with the 95% confidence interval.

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 (see Section 2). The survey collected information on age, education, health, and the reasons why an individual worked on MTurk. Descriptive statistics are given in Appendix Table A.8. Self-reported gender was collected to validate the measure based off of the first name an individual used to sign up for the CrowdWorkers app. Predicted gender is correct for 90% of those taking the survey; eleven individuals that I classify as a woman based on their first name self-identify as a man, and fourteen individuals that I classify as a man self-identify as a woman. All results throughout this paper are robust to excluding these individuals from the sample.

While it is not a random selection of workers into my survey,[13] the same gap in hourly earnings and pattern with respect to the additional of work scheduling controls emerges (Table 4). Amongst the sample of workers completing the survey, women earn approximately 22.7% 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 -0.107.

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 those with children, especially young children who's care needs are less predictable and who are less likely to respect boundaries between paid and domestic work (Tremblay 2002; Kossek, Lautsch, and Eaton 2006). 20% of female and male workers report having at least one child aged five years old or younger within the household, and 39% of women and 18% of men report having children older than this.[14] Consistent with this hypothesis, the gender wage gap is confined to women with children (column 4) and largest amongst those with young children (column 6). The unexplained component of the gender wage is especially large for women with young children at approximately 50%. Task level controls

[13] Workers who complete the survey perform more unit tasks and are more experienced. See Table A.7.

[14] In the survey, respondents were asked about the number of children in the household aged between (i) 0-5 years; (ii) 6-10 years; (iii) 11-15 years; (iv) 16+ years.

continue to explain approximately half of the unexplained gap for women with children in line with the results reported in Table 3.

Columns (5)-(7) control for further demographic characteristics and self-reported measures of what factors influence work decisions on MTurk. 25.2% of women compared to 14.8% of men said 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 more likely than men to report that they find it harder to work away from the home; 27.4% of women report "I can only work from home" as an "extremely important" factor explaining why they crowd-work compared to 15.5% of men. Caring decisions and interruptions also appear differentially to affect men and women's ability to complete tasks; 8.2% of women compared to 3.7% of men report that a "need to attend to family/caring responsibilities" is a "very common" factor explaining why they have to abandon work on tasks, while 12% of women compared to 8% of men report that being "interrupted" is a "very common" factor explaining abandonment.[15]

These self-reported measures and other demographic factors enter the wage equation with the expected signs. Individuals who report that interruptions are a "very common" or "common" reason for explaining why they abandon tasks, those who state that preferring to work from home or only being able to work from home is a "very important" or "important" reason for explaining why they work online, and homemakers learn less per hour on average. Older workers and those who state that they have a work condition that reduces their ability to carry out day-to-day activities also earn less per hour. Given the number of workers in the data set, I do not interact these variables with the gender and children variables.

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 within a gender-blind setting. Women on MTurk earn 20% less per hour than men, not because of any gender difference in task selection or total experience but because of differences in the scheduling of work. Women are less likely to work in continuous batches on tasks and are more likely to take longer breaks between submitting one task

[15] 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).

Table 4: Task-Level Wage Regressions & Survey Responses: Demographic Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.2265**	-0.1070***	-0.0357	0.0113	0.0369	0.0443	0.0154
	(0.0178)	(0.0094)	(0.7584)	(0.8219)	(0.4671)	(0.3758)	(0.7535)
Female x Young Kids			-0.5020*	-0.2598***	-0.1883**	-0.1559	-0.1216
			(0.0603)	(0.0085)	(0.0302)	(0.1112)	(0.2009)
Female x Older Kids Only			-0.3049	-0.1722*	-0.1632*	-0.1585*	-0.1258
			(0.2401)	(0.0681)	(0.0797)	(0.0744)	(0.1775)
Young Kids			0.0903	0.1276**	0.1438***	0.0813	0.0527
			(0.5033)	(0.0289)	(0.0072)	(0.2074)	(0.3933)
Older Kids Only			0.1607	0.0306	0.0668	0.1242*	0.0970
			(0.4803)	(0.6946)	(0.3787)	(0.0980)	(0.2204)
Self-Reported Measures							
Abandon: Care					-0.0600	-0.0384	-0.0288
					(0.3247)	(0.5035)	(0.6415)
Abandon: Interrupt					-0.0657	-0.0868**	-0.0999**
					(0.1534)	(0.0424)	(0.0191)
Reason: Home					-0.1380***	-0.1026**	-0.0615
					(0.0007)	(0.0150)	(0.1805)
Demographics							
College +						0.0034	-0.0014
						(0.9330)	(0.9716)
Bad Health						-0.1205*	-0.1034
						(0.0585)	(0.1057)
Age 30-40						-0.0100	-0.0025
						(0.8172)	(0.9502)
Age 40-50						-0.1313**	-0.1164**
						(0.0143)	(0.0257)
Age 50+						-0.2680***	-0.2395***
						(0.0000)	(0.0001)
Part Time							-0.1192**
							(0.0325)
Homemaker							-0.1363**
							(0.0234)
Retired							-0.1133*
							(0.0959)
Self-Emp							-0.0573
							(0.2819)
Unemployed							-0.1241
							(0.1233)
Constant	1.3596***	4.0076***	1.3471***	3.9801***	4.0548***	4.1326***	4.1512***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	833075	833075	833075	833075	833075	833075	832752
R^2	0.0175	0.5847	0.0290	0.5871	0.5907	0.5979	0.5991
Task Log Variables	no	yes	no	yes	yes	yes	yes
Length Quartile	no	no	no	no	no	yes	yes

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Signficance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include year-quarter fixed effects. Task-log controls are those reported in Table 3. The length control is a set of dummy variables for the quartile of the total number of words written in response to the open text questions in the survey.

and starting another. These work patterns are associated with slower task completion times.

Heterogeneity in the wage penalty by family structure, and self-reported reasons for task abandonment and completion, are consistent with childcare responsibilities placing more constraints on women's ability to schedule paid work without interruption within the home. There is no wage gap for individuals without children and the difference in hourly earnings is highest for women with young children, who are often less respectful of boundaries placed between market and care (Tremblay 2002; Kossek, Lautsch, and Eaton 2006). Women are also much more likely to self-report that domestic responsibilities affect their ability to complete task successfully and without interruption.

These results are important for three reasons. First, 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 a given and may not be as salient in online labour markets as the occupational segregation in offline labour markets might suggest. Second, 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. 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). My results suggest these distractions could have damaging effects on individual productivity. These results call for further research into the effectiveness of different strategies that can employed to reduce the influence of family responsibilities on online work and ways in which 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 crowd-workers. *Available at SSRN 3048711*.

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.

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.

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.

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.

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.

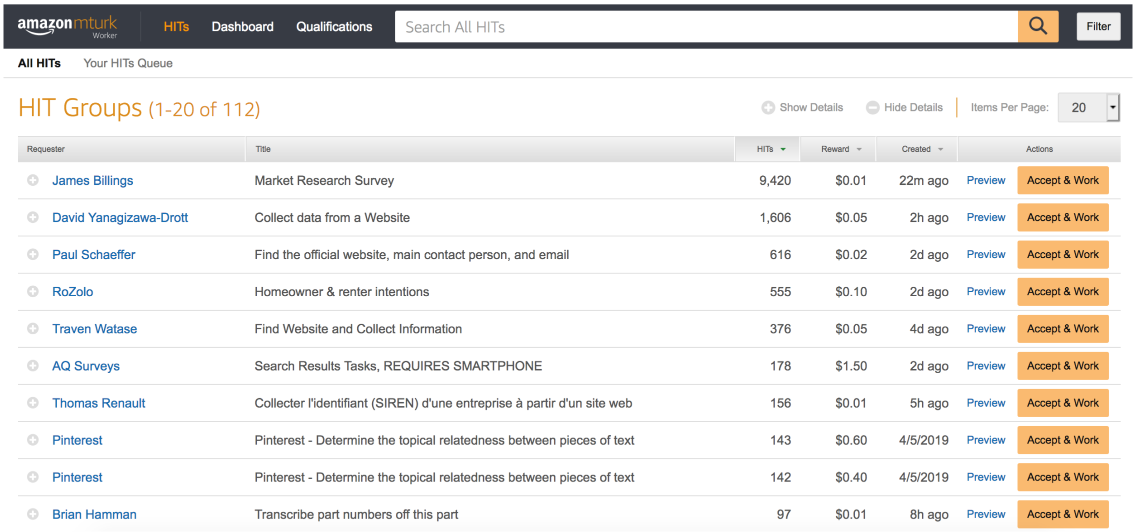
Vasilescu, B., K. Blincoe, Q. Xuan, C. Casalnuovo, D. Damian, P. Devanbu, and V. Filkov (2016). The sky is not the limit: multitasking across github projects. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pp. 994–1005. IEEE.

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 Additional Results

Figure A.1: Screenshot of MTurk Interface



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

Table A.1: Task-Level Wage Regressions: Alternative Experience Controls

	(1)	(2)	(3)
Female	-0.1608**	-0.1662**	-0.1724**
	(0.0473)	(0.0395)	(0.0320)
Experience	0.0197***	0.0379**	
	(0.0032)	(0.0164)	
Experience × Experience		-0.0006	
		(0.1001)	
1st Decile			0.0000
			(.)
2nd Decile			0.1522***
			(0.0000)
3rd Decile			0.2050***
			(0.0000)
4th Decile			0.3141***
			(0.0000)
5th Decile			0.3170***
			(0.0000)
6th Decile			0.3529***
			(0.0000)
7th Decile			0.3173***
			(0.0003)
8th Decile			0.4443***
			(0.0003)
9th Decile			0.5489***
			(0.0036)
10th Decile			0.6623***
			(0.0000)
Observations	1616850	1616850	1616850

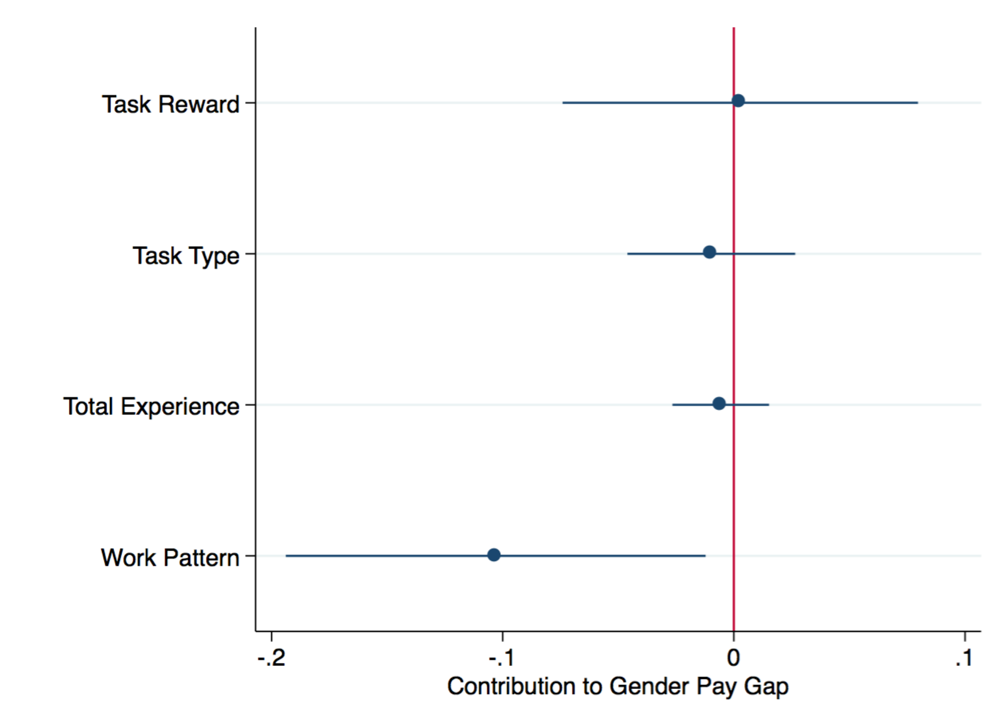
Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Experience is the number of tasks completed before a given task divided by 1000. Decile refers to the decile of the experience variable. Year-quarter fixed effects in all specifications.

Table A.2: Inclusion of Worker Fixed Effects: Pooled and by Gender

	(1)		(2)		(3)	
1st Decile Break	0.0000	(.)	0.0000	(.)	0.0000	(.)
2nd Decile Break	-0.0510**	(0.0187)	-0.0703**	(0.0141)	-0.0397	(0.1703)
3rd Decile Break	-0.1573***	(0.0000)	-0.1769***	(0.0000)	-0.1481***	(0.0039)
4th Decile Break	-0.2262***	(0.0000)	-0.2705***	(0.0000)	-0.1672***	(0.0025)
5th Decile Break	-0.2500***	(0.0000)	-0.2605***	(0.0000)	-0.2357***	(0.0000)
6th Decile Break	-0.3561***	(0.0000)	-0.3703***	(0.0000)	-0.3272***	(0.0000)
7th Decile Break	-0.5534***	(0.0000)	-0.5629***	(0.0000)	-0.5168***	(0.0000)
8th Decile Break	-0.8284***	(0.0000)	-0.8384***	(0.0000)	-0.7851***	(0.0000)
9th Decile Break	-1.1294***	(0.0000)	-1.2143***	(0.0000)	-0.9855***	(0.0000)
10th Decile Break	-1.0316***	(0.0000)	-1.0631***	(0.0000)	-0.9679***	(0.0000)
First Task of Day	-0.9009***	(0.0000)	-0.9578***	(0.0000)	-0.7908***	(0.0000)
1st Decile Session Task	0.0000	(.)	0.0000	(.)	0.0000	(.)
2nd Decile Session Task	-0.2016***	(0.0000)	-0.2084***	(0.0000)	-0.1952***	(0.0001)
3rd Decile Session Task	-0.1924***	(0.0000)	-0.2068***	(0.0000)	-0.1749***	(0.0012)
4th Decile Session Task	-0.1544***	(0.0000)	-0.1792***	(0.0000)	-0.1214**	(0.0446)
5th Decile Session Task	-0.0959**	(0.0137)	-0.1275***	(0.0044)	-0.0511	(0.4512)
6th Decile Session Task	-0.0286	(0.5118)	-0.0688	(0.1911)	0.0292	(0.6823)
7th Decile Session Task	0.0544	(0.2236)	0.0071	(0.9007)	0.1191*	(0.0791)
8th Decile Session Task	0.1318***	(0.0069)	0.0834	(0.1919)	0.1960***	(0.0056)
9th Decile Session Task	0.2083***	(0.0001)	0.1563**	(0.0243)	0.2745***	(0.0003)
10th Decile Session Task	0.3340***	(0.0000)	0.3161***	(0.0002)	0.3595***	(0.0000)
Same Task at t-1	-0.1310***	(0.0018)	-0.1534***	(0.0043)	-0.0733	(0.1931)
Observations	1616850		896694		720156	

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Worker and year-quarter fixed effects in all specifications. Column (1) gives the pooled sample, column (2) gives the same regression on the subsample of women, and column (3) gives results on the subsample of men.

Figure A.2: Gelbach Decomposition



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.

A.2 Gender Matched v Unmatched Results

Table A.3: Pooled OLS Task Characteristic Regressions: Matched versus Unmatched Samples

	(1)		(2)		(3)	
Log Piece Rate	0.7416***	(0.0000)	0.7532***	(0.0000)	0.7018***	(0.0000)
Entry & Discovery	0.0000	(.)	0.0000	(.)	0.0000	(.)
Verification	-0.2918**	(0.0118)	-0.2720*	(0.0833)	-0.3355**	(0.0199)
Viewing	0.3218	(0.3522)	-0.0041	(0.9889)	0.7710	(0.1287)
Rating	-0.7496***	(0.0000)	-0.8049***	(0.0000)	-0.6249***	(0.0014)
Transcription	0.2470***	(0.0066)	0.2214**	(0.0304)	0.3191**	(0.0481)
Research	-0.5001***	(0.0000)	-0.4451***	(0.0001)	-0.5855***	(0.0000)
Other	-0.4665***	(0.0000)	-0.4895***	(0.0000)	-0.3755***	(0.0051)
Observations	2469913		1616850		853063	
Experience	yes		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 (1) gives results on the full sample, column (2) gives the gender matched sample, and column (3) gives the gender unmatched sample.

Table A.4: Fixed Effect Task Characteristic Regressions: Matched versus Unmatched Samples

	(1)		(2)		(3)	
Log Piece Rate	0.7628***	(0.0000)	0.7735***	(0.0000)	0.7412***	(0.0000)
Entry & Discovery	0.0000	(.)	0.0000	(.)	0.0000	(.)
Verification	-0.1167**	(0.0241)	-0.1691***	(0.0026)	-0.0549	(0.5501)
Viewing	0.3847	(0.2093)	0.0004	(0.9978)	1.0513*	(0.0742)
Rating	-0.3588***	(0.0000)	-0.4263***	(0.0000)	-0.2528*	(0.0938)
Transcription	0.2477***	(0.0001)	0.1856**	(0.0121)	0.3458***	(0.0006)
Research	-0.2256***	(0.0001)	-0.2261***	(0.0014)	-0.2763***	(0.0036)
Other	-0.1920***	(0.0003)	-0.2212***	(0.0003)	-0.1617*	(0.0786)
Observations	2469913		1616850		853063	
Experience	yes		yes		yes	

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column (1) gives results on the full sample, column (2) gives the gender matched sample, and column (3) gives the gender unmatched sample.

Table A.5: Pooled OLS Working Pattern Regressions: Matched versus Unmatched Samples

	(1)		(2)		(3)	
1st Decile Break	0.0000	(.)	0.0000	(.)	0.0000	(.)
2nd Decile Break	-0.2766***	(0.0000)	-0.2320***	(0.0000)	-0.3597***	(0.0000)
3rd Decile Break	-0.3784***	(0.0000)	-0.4024***	(0.0000)	-0.3415***	(0.0000)
4th Decile Break	-0.3965***	(0.0000)	-0.4324***	(0.0000)	-0.3331***	(0.0000)
5th Decile Break	-0.4535***	(0.0000)	-0.4282***	(0.0000)	-0.4967***	(0.0000)
6th Decile Break	-0.5745***	(0.0000)	-0.5454***	(0.0000)	-0.6178***	(0.0000)
7th Decile Break	-0.8064***	(0.0000)	-0.7688***	(0.0000)	-0.8615***	(0.0000)
8th Decile Break	-1.1262***	(0.0000)	-1.0891***	(0.0000)	-1.1709***	(0.0000)
9th Decile Break	-1.5306***	(0.0000)	-1.4602***	(0.0000)	-1.6258***	(0.0000)
10th Decile Break	-1.4986***	(0.0000)	-1.3564***	(0.0000)	-1.7100***	(0.0000)
First Task of Day	-1.3416***	(0.0000)	-1.2253***	(0.0000)	-1.5143***	(0.0000)
1st Decile Session Task	0.0000	(.)	0.0000	(.)	0.0000	(.)
2nd Decile Session Task	-0.2869***	(0.0000)	-0.2283***	(0.0000)	-0.3762***	(0.0000)
3rd Decile Session Task	-0.2700***	(0.0000)	-0.2062***	(0.0000)	-0.3683***	(0.0000)
4th Decile Session Task	-0.2207***	(0.0000)	-0.1482***	(0.0005)	-0.3339***	(0.0000)
5th Decile Session Task	-0.1437***	(0.0001)	-0.0608	(0.2137)	-0.2735***	(0.0000)
6th Decile Session Task	-0.0593	(0.1640)	0.0347	(0.5357)	-0.2072***	(0.0008)
7th Decile Session Task	0.0423	(0.3762)	0.1478**	(0.0153)	-0.1252*	(0.0840)
8th Decile Session Task	0.1526***	(0.0042)	0.2638***	(0.0001)	-0.0292	(0.7150)
9th Decile Session Task	0.2806***	(0.0000)	0.3883***	(0.0000)	0.0942	(0.2776)
10th Decile Session Task	0.4977***	(0.0000)	0.5633***	(0.0000)	0.3698***	(0.0003)
Same Task at t-1	-0.2102***	(0.0000)	-0.2017***	(0.0000)	-0.2086***	(0.0002)
Observations	2469913		1616850		853063	

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 (1) gives results on the full sample, column (2) gives the gender matched sample, and column (3) gives the gender unmatched sample.

Table A.6: Fixed Effect Working Pattern Regressions: Matched versus Unmatched Samples

	(1)		(2)		(3)	
1st Decile Break	0.0000	(.)	0.0000	(.)	0.0000	(.)
2nd Decile Break	-0.0603***	(0.0013)	-0.0703**	(0.0141)	-0.0397	(0.1703)
3rd Decile Break	-0.1494***	(0.0000)	-0.1769***	(0.0000)	-0.1481***	(0.0039)
4th Decile Break	-0.2146***	(0.0000)	-0.2705***	(0.0000)	-0.1672***	(0.0025)
5th Decile Break	-0.2618***	(0.0000)	-0.2605***	(0.0000)	-0.2357***	(0.0000)
6th Decile Break	-0.3698***	(0.0000)	-0.3703***	(0.0000)	-0.3272***	(0.0000)
7th Decile Break	-0.5618***	(0.0000)	-0.5629***	(0.0000)	-0.5168***	(0.0000)
8th Decile Break	-0.8261***	(0.0000)	-0.8384***	(0.0000)	-0.7851***	(0.0000)
9th Decile Break	-1.1583***	(0.0000)	-1.2143***	(0.0000)	-0.9855***	(0.0000)
10th Decile Break	-1.1167***	(0.0000)	-1.0631***	(0.0000)	-0.9679***	(0.0000)
First Task of Day	-0.9531***	(0.0000)	-0.9578***	(0.0000)	-0.7908***	(0.0000)
1st Decile Session Task	0.0000	(.)	0.0000	(.)	0.0000	(.)
2nd Decile Session Task	-0.2360***	(0.0000)	-0.2084***	(0.0000)	-0.1952***	(0.0001)
3rd Decile Session Task	-0.2260***	(0.0000)	-0.2068***	(0.0000)	-0.1749***	(0.0012)
4th Decile Session Task	-0.1898***	(0.0000)	-0.1792***	(0.0000)	-0.1214**	(0.0446)
5th Decile Session Task	-0.1325***	(0.0000)	-0.1275***	(0.0044)	-0.0511	(0.4512)
6th Decile Session Task	-0.0680**	(0.0351)	-0.0688	(0.1911)	0.0292	(0.6823)
7th Decile Session Task	0.0109	(0.7499)	0.0071	(0.9007)	0.1191*	(0.0791)
8th Decile Session Task	0.0904**	(0.0156)	0.0834	(0.1919)	0.1960***	(0.0056)
9th Decile Session Task	0.1708***	(0.0000)	0.1563**	(0.0243)	0.2745***	(0.0003)
10th Decile Session Task	0.3039***	(0.0000)	0.3161***	(0.0002)	0.3595***	(0.0000)
Same Task at t-1	-0.1307***	(0.0000)	-0.1534***	(0.0043)	-0.0733	(0.1931)
Observations	2469913		896694		720156	

Notes: p-values (implied by standard errors clustered at the worker level) given in parentheses. Significance of differences indicated by: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Worker and year-quarter fixed effects in all specifications. Column (1) gives results on the full sample, column (2) gives the gender matched sample, and column (3) gives the gender unmatched sample.

A.3 Additional Survey Results

Table A.7: 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.

Table A.8: Summary Statistics: Survey Responses

	Female	Male	Difference
Demgraphics			
Children	0.6635	0.4152	0.2483***
	(0.0232)	(0.0297)	(0.0373)
Children $\leq$ 5 years	0.2019	0.1913	0.0106
	(0.0197)	(0.0237)	(0.0309)
Children 6+ years	0.3918	0.1805	0.2113***
	(0.0239)	(0.0232)	(0.0349)
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)
Abandoning Tasks & Why Crowdworking: Very Important			
Abandon: Care	0.0817	0.0361	0.0456**
	(0.0134)	(0.0112)	(0.189)
Abandon: Interrupt	0.1202	0.0830	0.0371
	(0.0160)	(0.0166)	(0.0238)
Reason: Only Home	0.2740	0.1552	0.1188***
	(0.0219)	(0.0218)	(0.0322)

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

A.4 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