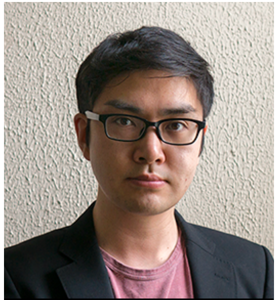
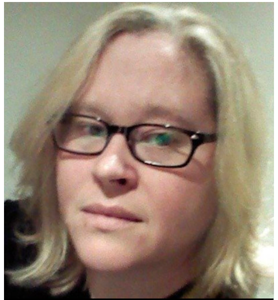
A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage
Chris Callison-Burch, Jeffrey P. Bigham

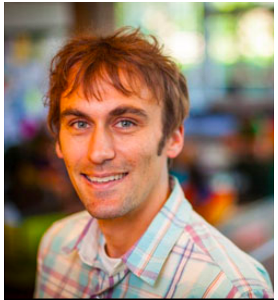

Kotaro Hara


Abigail Adams


Kristy Milland


Saiph Savage


Chris Callison-Burch


Jeffrey P. Bigham

SINGAPORE MANAGEMENT UNIVERSITY
Carnegie Mellon University
MANNERS MAKYTH MAN
McMaster University
West Virginia University
Penn UNIVERSITY of PENNSYLVANIA

A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

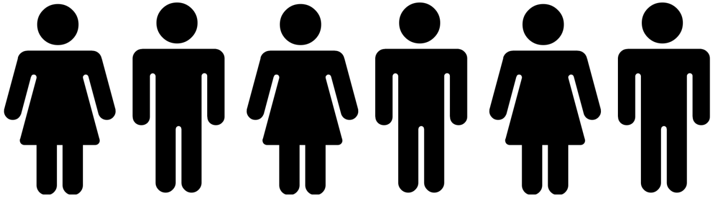
Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage
Chris Callison-Burch, Jeffrey P. Bigham

SINGAPORE MANAGEMENT UNIVERSITY
Carnegie Mellon University
McMaster University
West Virginia University
Penn University of Pennsylvania

CHI 2018
Engage with CHI

600k

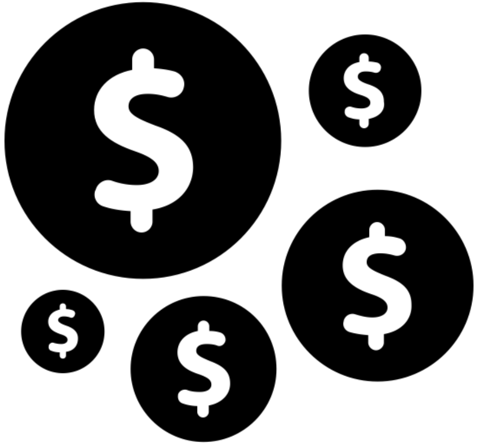
online workers and counting



The Hamilton Project (2015)

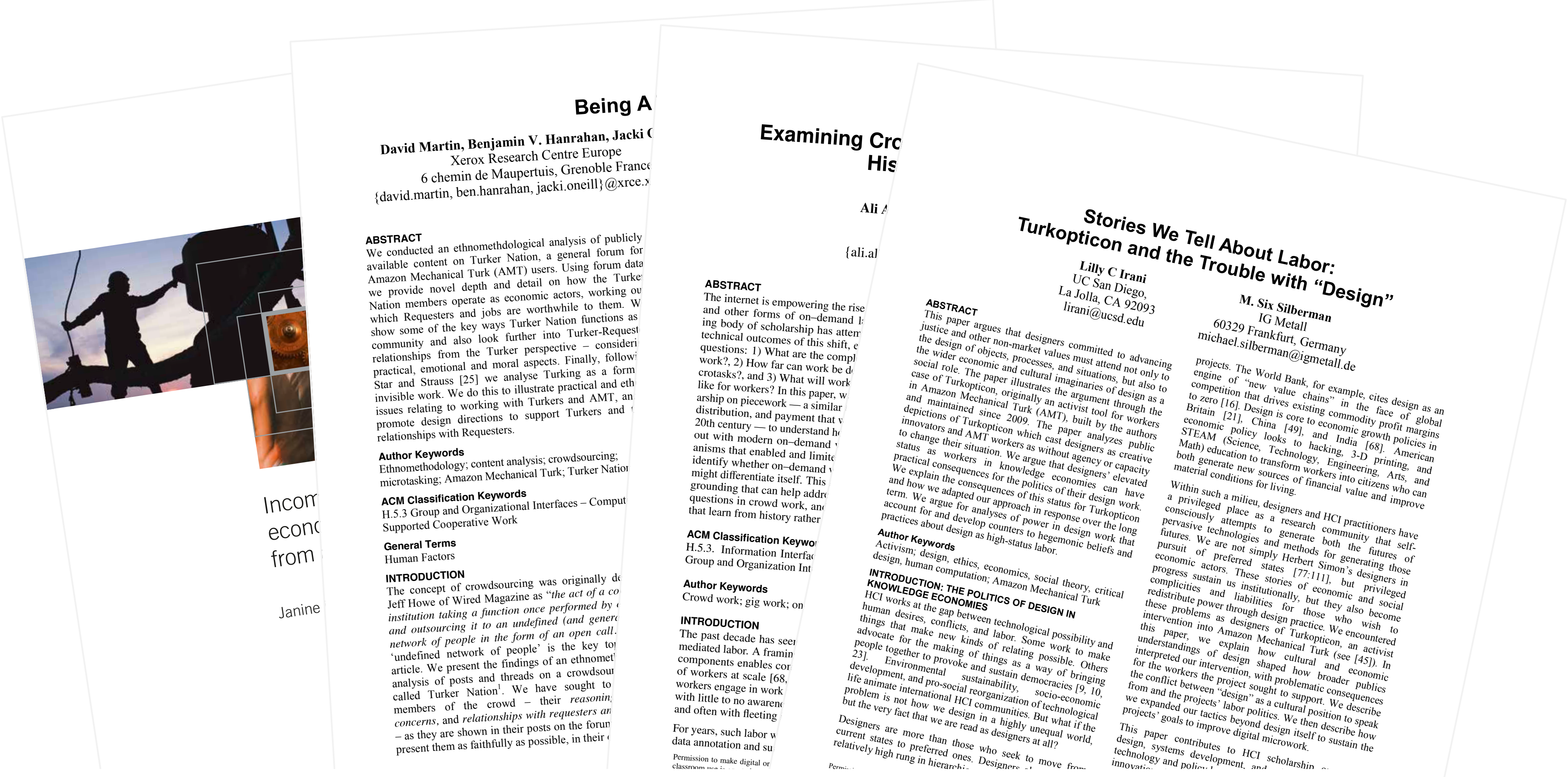
Online outsourcing industry generated

$2b



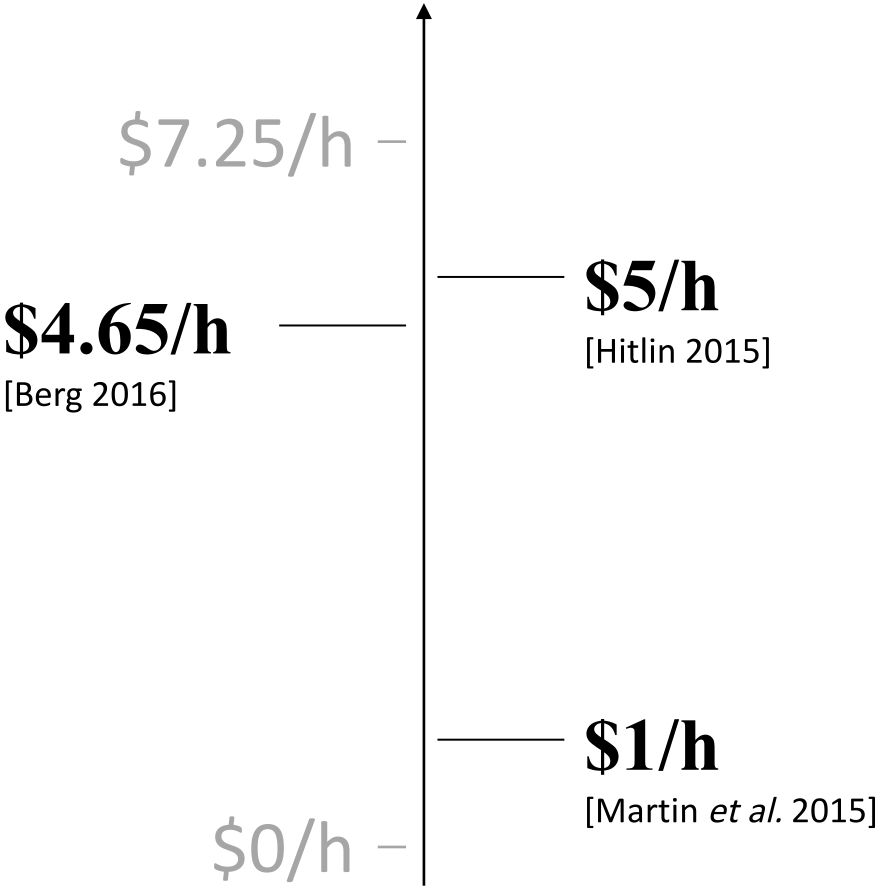
Hitlin (2016), WorldBank (2015)

Are workers treated fairly? How does this new work style affect their lives?



Martin *et al.* 2014; Berg 2016; Irani and Silberman 2016; Alkhatib *et al.* 2017

Workers earn a fraction of the U.S. minimum wage ($7.25/h)



These figures are subjective data based on workers' opinions on an online forum and survey responses

The lack of **reward and** task duration data has prevented us from objectively analysing workers' hourly wage

Crowd Workers

3.8m

task records

https://crowd-workers.com/

Research Questions

How much are workers earning on Amazon Mechanical Turk?

What contributes to the low wage?

Research Questions

How much are workers earning on Amazon Mechanical Turk?

What contributes to the low wage?

Data

- N=2,676 workers
- Task description
    - title, keywords, description, task IDs, requester IDs, **reward ($)**
- Task status
    - submitted *vs.* returned
    - **Timestamps (task start, task end, task return)**

Data

- N=2,676 workers
- Task description
  - title, keywords, description, task IDs, requester IDs, **reward ($)**
- Task status
  - submitted *vs.* returned
  - **Timestamps (task start, task end, task return)**

These pieces of information enable us to calculate hourly wage

It is surprisingly hard to get accurate estimation of hourly wage

Hourly Wage Estimation (Naïve)

$$\text{Task Reward (\$)} \,/\, \text{Task Interval} = \text{Per-task Hourly Wage}$$

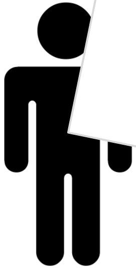
Hourly Wage Estimation (Naïve)

Worker

$$\frac{\sum \text{Task Reward (\$)}}{\sum \text{Task Interval}} = \text{Per-worker Hourly Wage}$$

Naïve method of calculating hourly wage

Hourly Wage Estimation (Naïve)



Worker

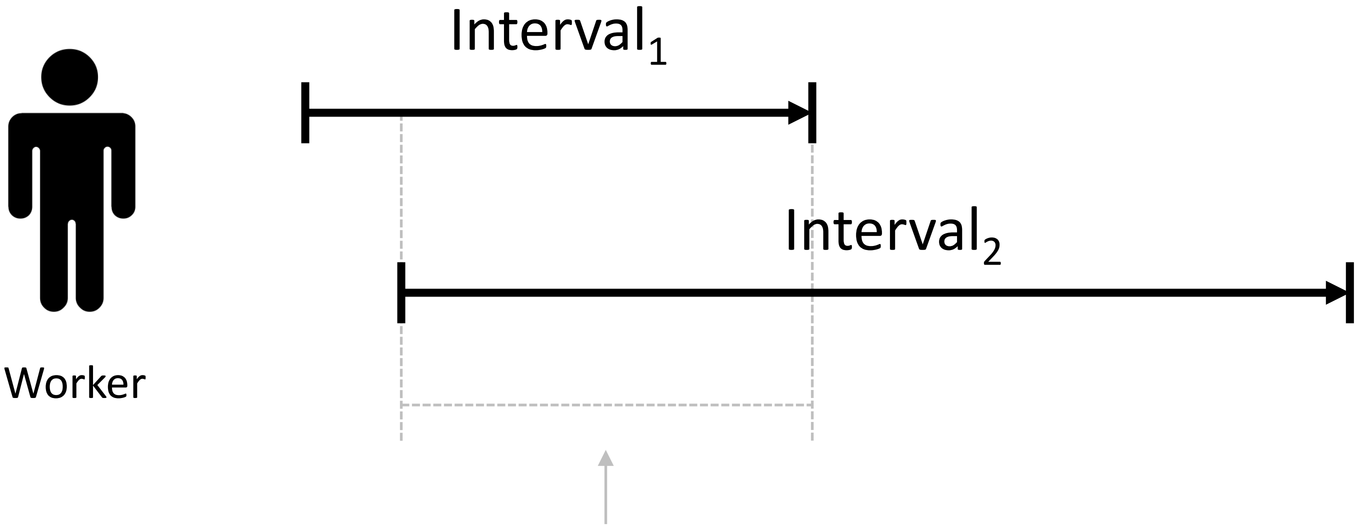
The naïve method may **under- or over-estimate the hourly wage**

Hourly Wage

Σ Task Inter

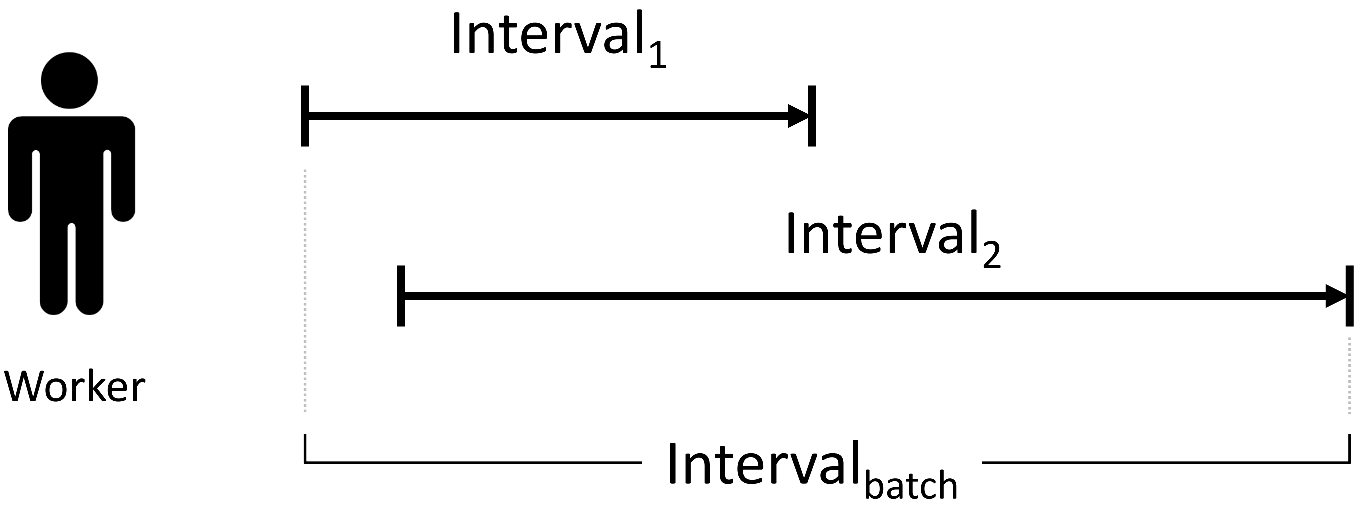
Naïve method of calculating hourly wage

Wage Under-estimation



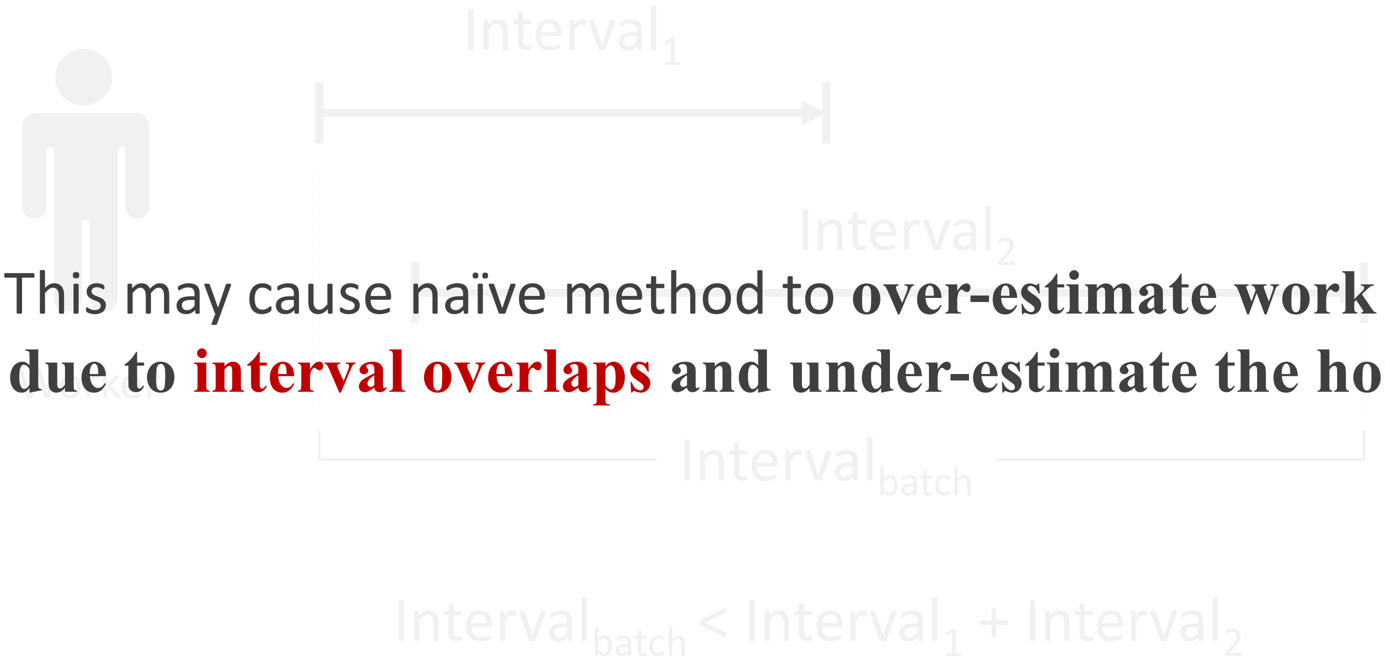
People could work on tasks concurrently

Wage Under-estimation



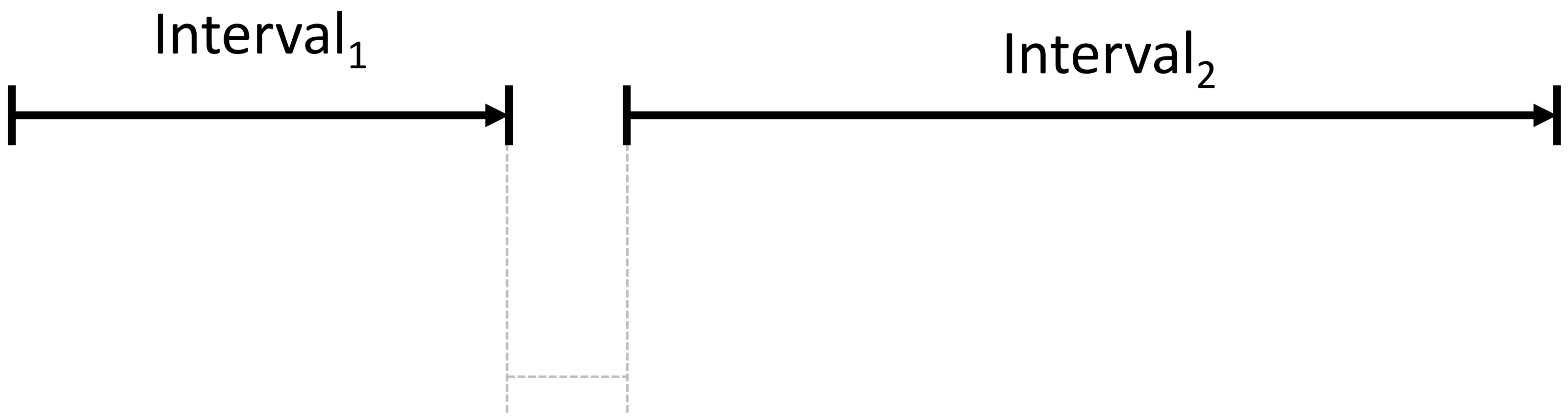
$$\text{Interval}_{\text{batch}} < \text{Interval}_1 + \text{Interval}_2$$

Wage Under-estimation



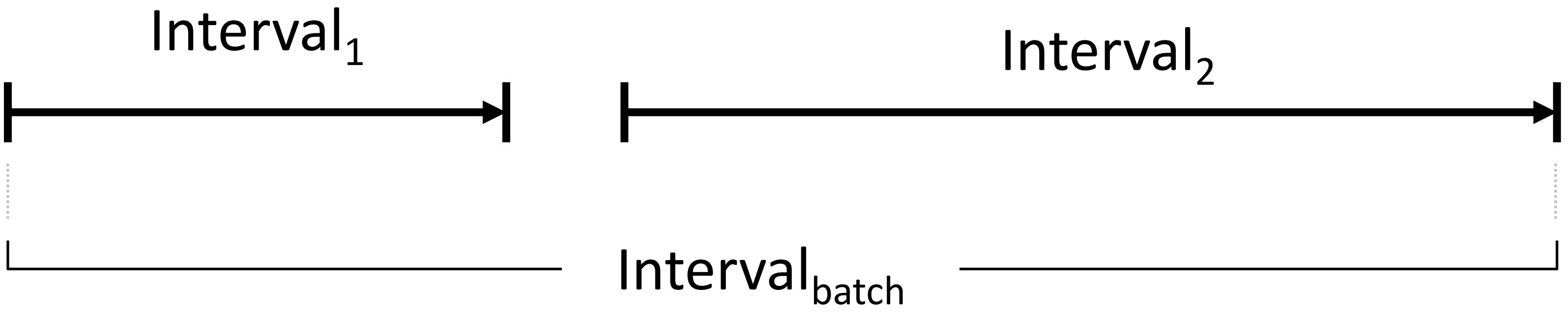
This may cause naïve method to **over-estimate work durations due to interval overlaps and under-estimate the hourly wage**

Wage Over-estimation



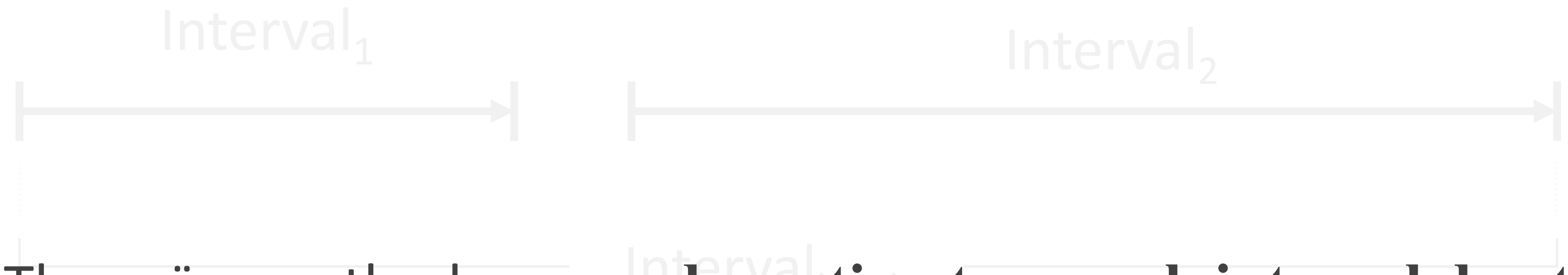
There could be a short gap between two tasks
(*e.g.,* time to search for a task)

Wage Over-estimation



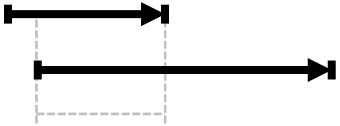
$$Interval_{batch} > Interval_1 + Interval_2$$

Wage Over-estimation

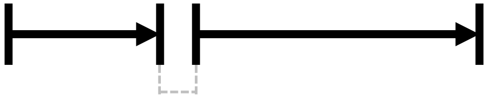


The naïve method may **under-estimate a work interval due to time between tasks and over-estimate the hourly wage**

$\text{Interval}_{batch} > \text{Interval}_1 + \text{Interval}_2$



Interval overlap

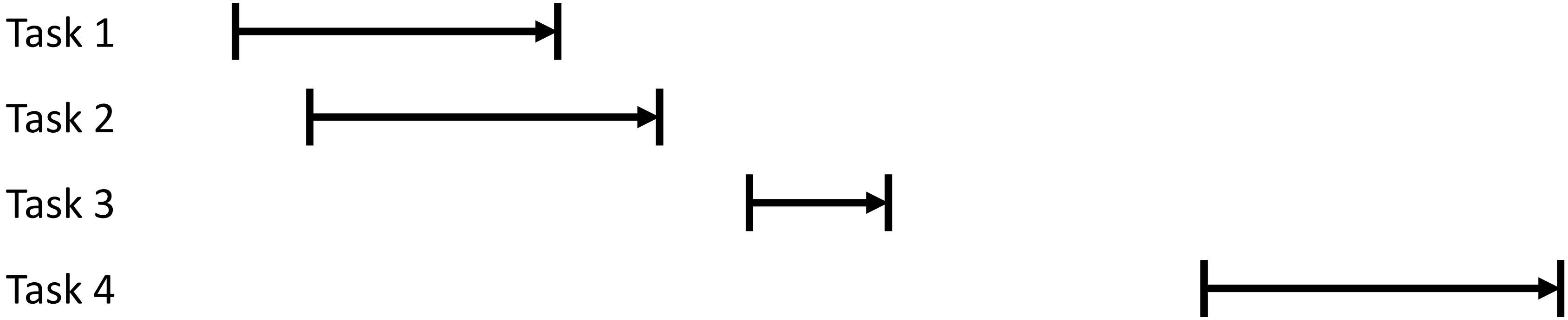


Time between tasks

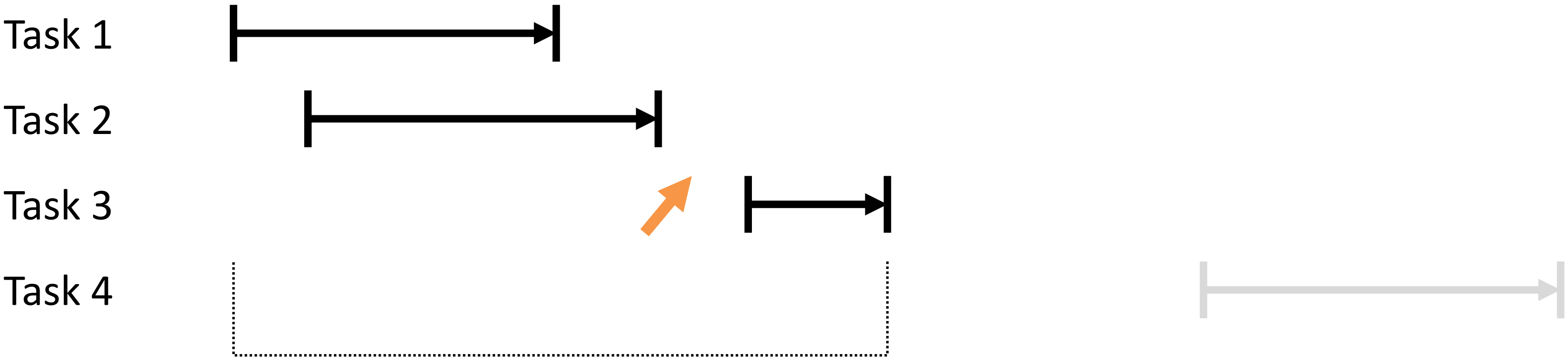
Wage over- and under-estimation may affect the accuracy of hourly wage calculation

To reduce the effects of interval overlaps and time between tasks, we used a **temporal clustering method** to compute hourly wage

Temporal Clustering

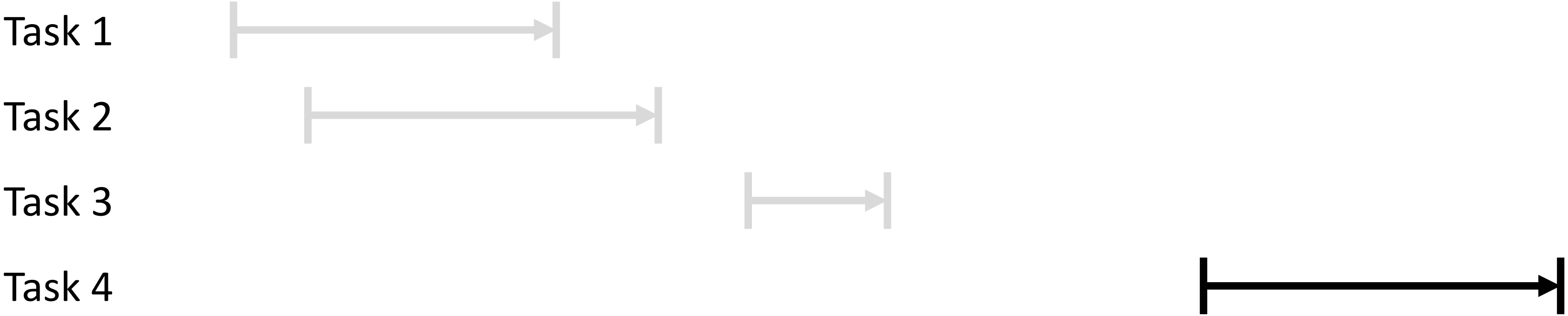


Temporal Clustering



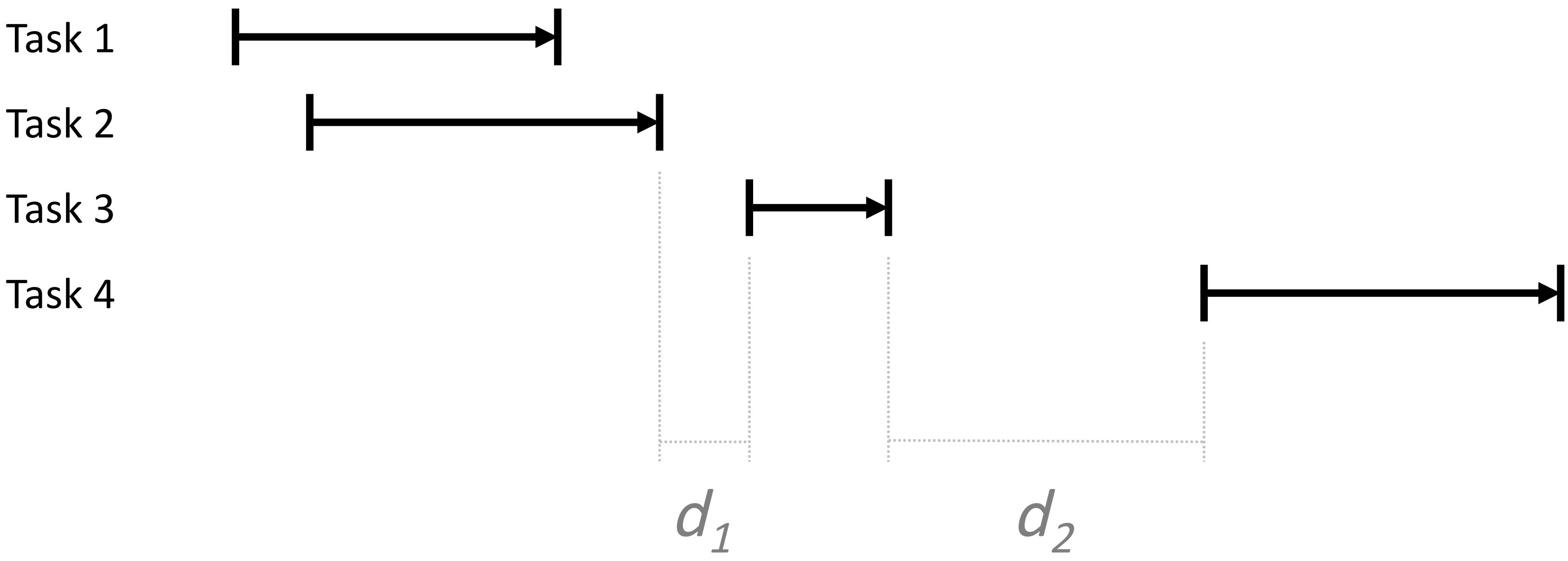
We want to cluster temporally close tasks together to ignore this **small gap**

Temporal Clustering

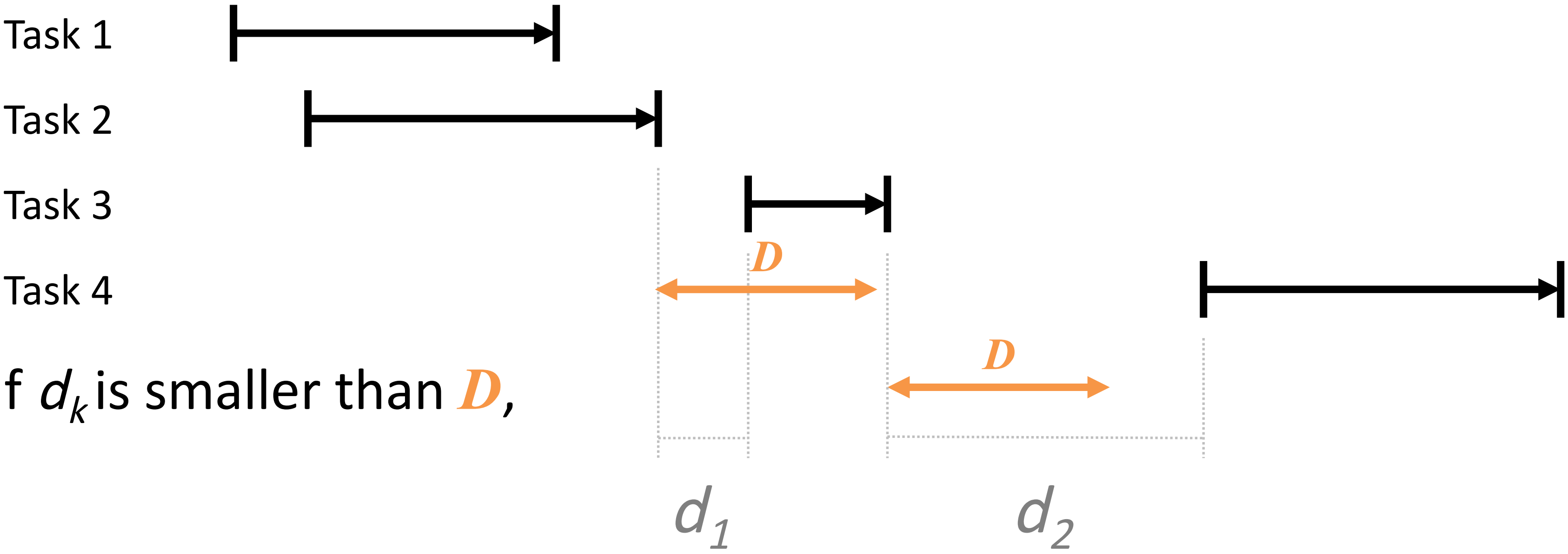


While keeping this isolated task disjoint

Temporal Clustering

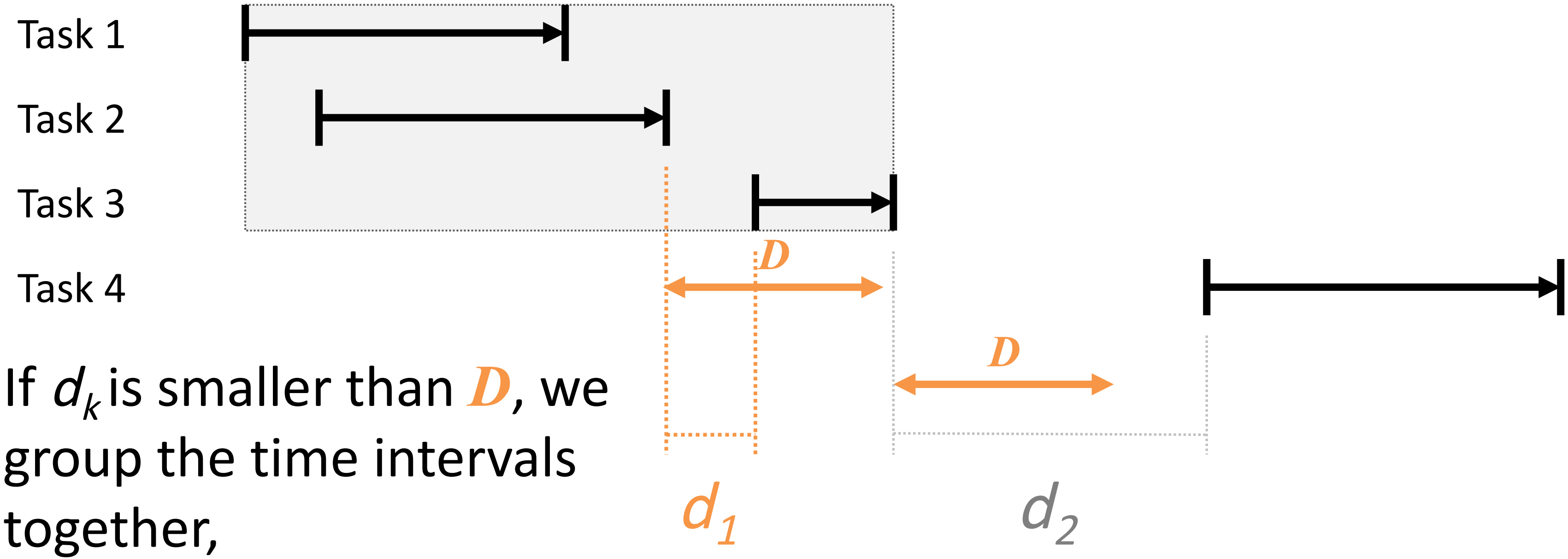


Temporal Clustering



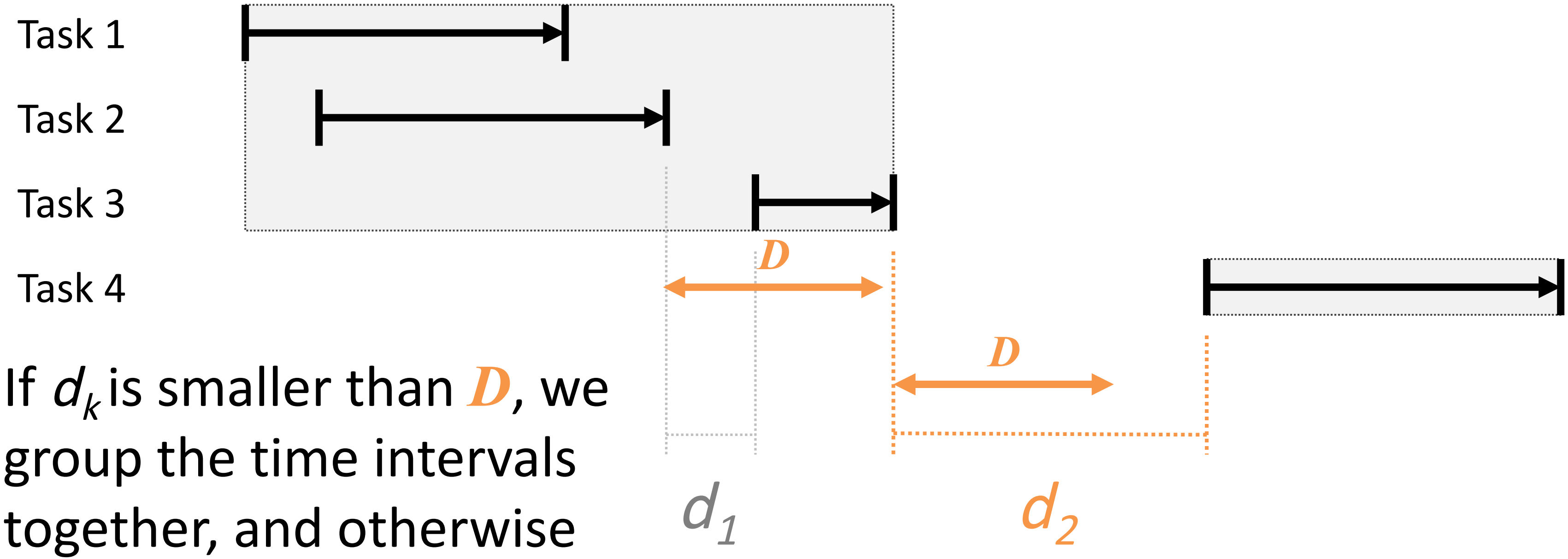
If d_k is smaller than $\boldsymbol{D}$,

Temporal Clustering



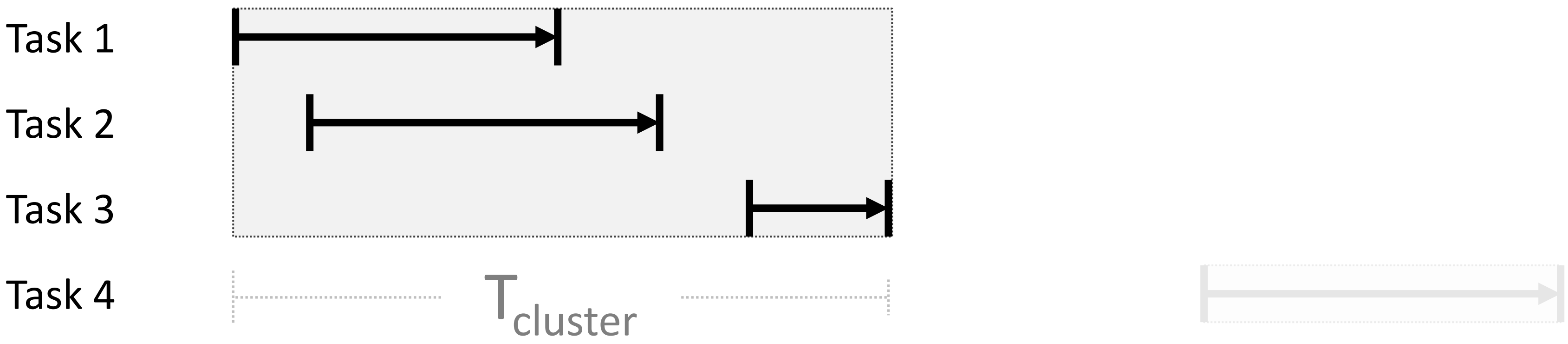
If d_k is smaller than ***D***, we group the time intervals together,

Temporal Clustering

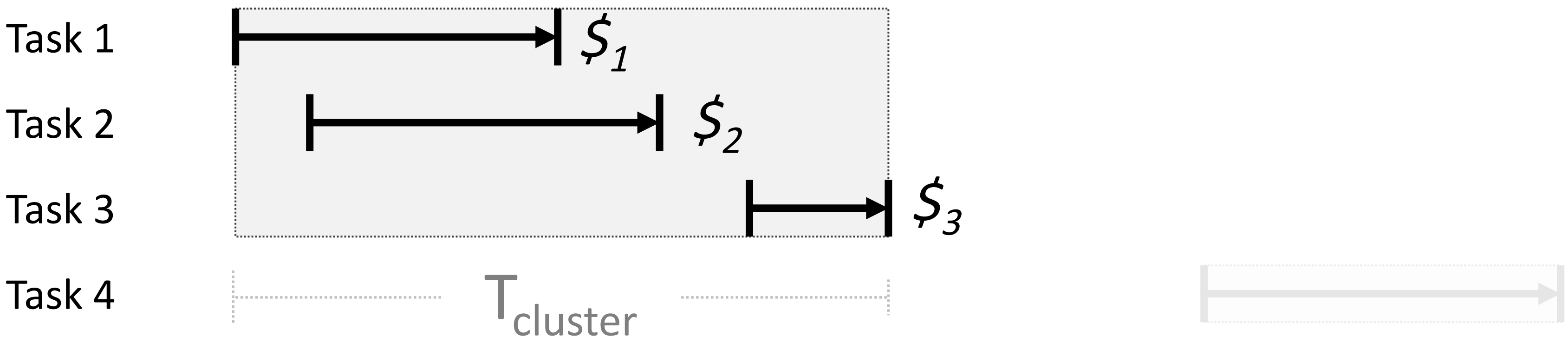


If d_k is smaller than ***D***, we group the time intervals together, and otherwise keep them disjoint

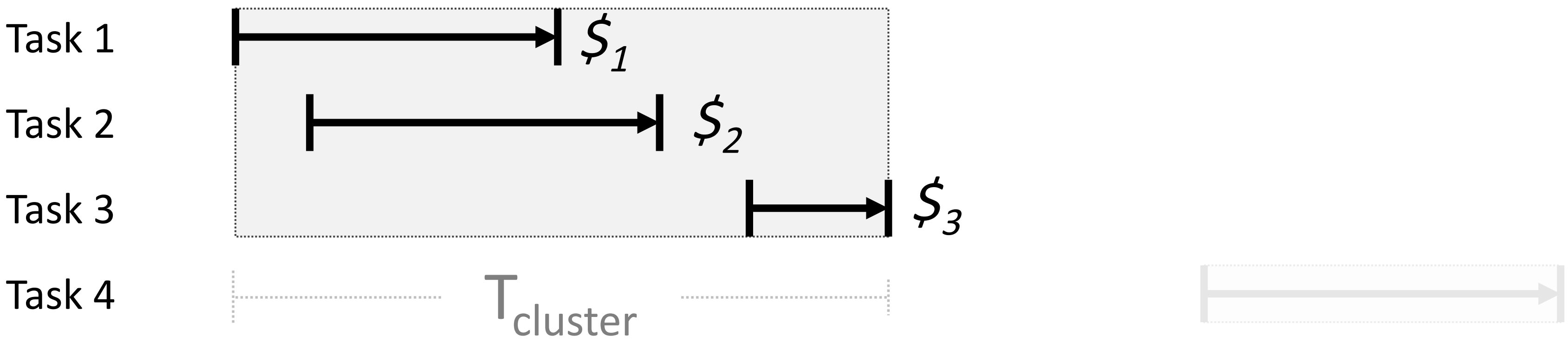
Temporal Clustering: Cluster-based Hourly Wage



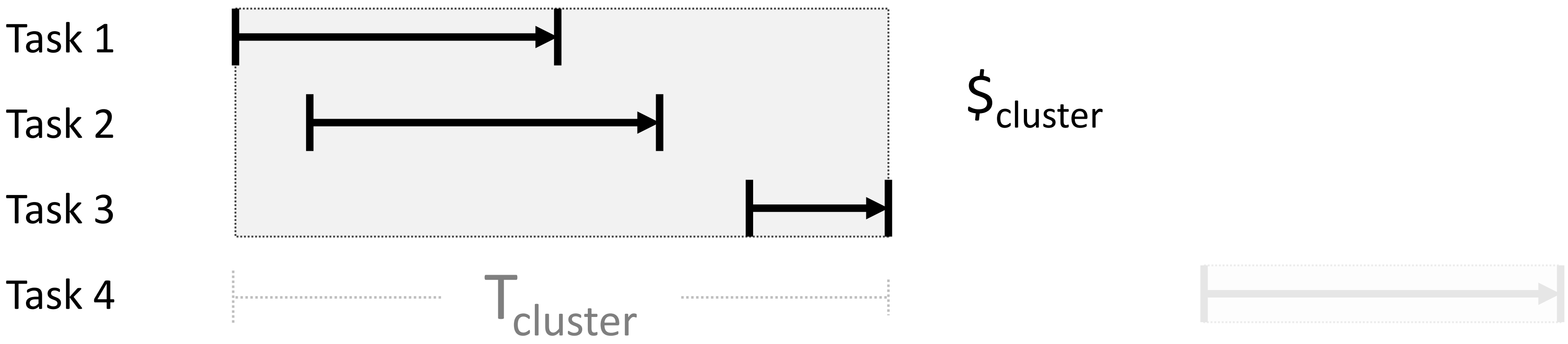
Temporal Clustering: Cluster-based Hourly Wage



Temporal Clustering: Cluster-based Hourly Wage

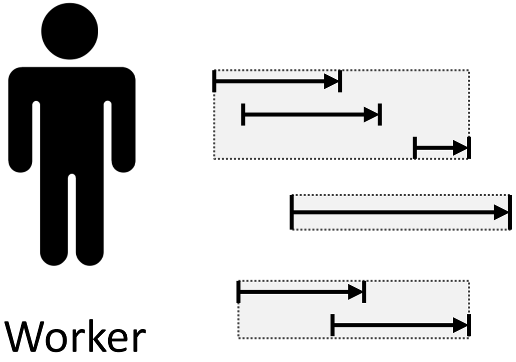


Temporal Clustering: Cluster-based Hourly Wage



We define **per-cluster hourly wage as $\$_{cluster}$ / $T_{cluster}$**

Temporal Clustering: Cluster-based Hourly Wage

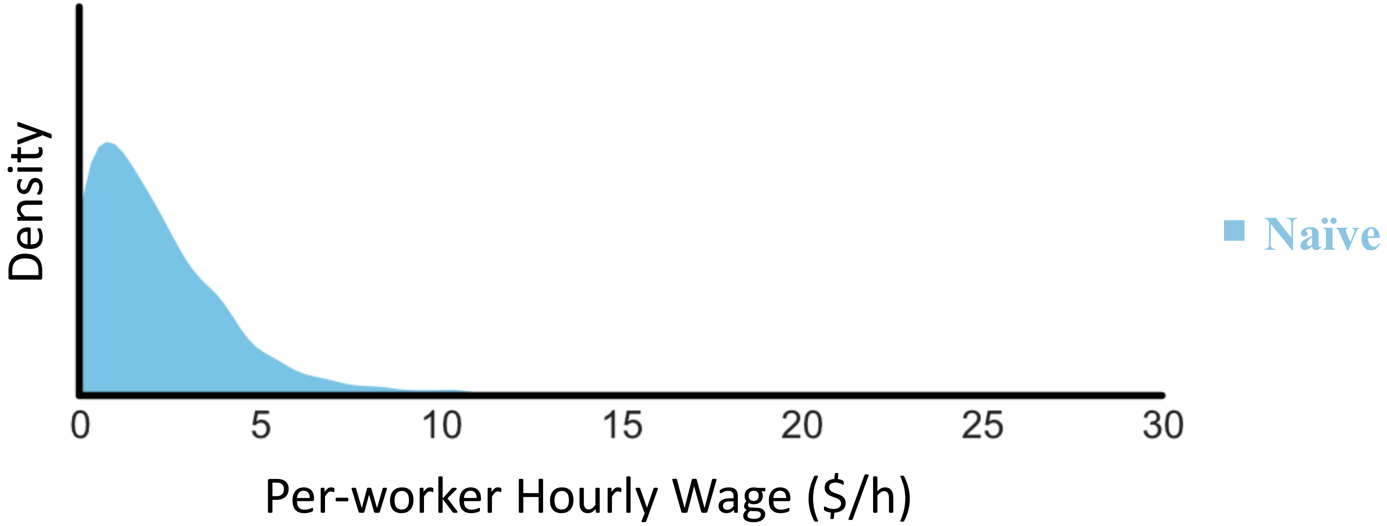


$$\frac{\sum \text{Cluster Reward (\$)}}{\sum \text{Cluster Interval}} = \text{Per-worker Hourly Wage with Clustering}$$

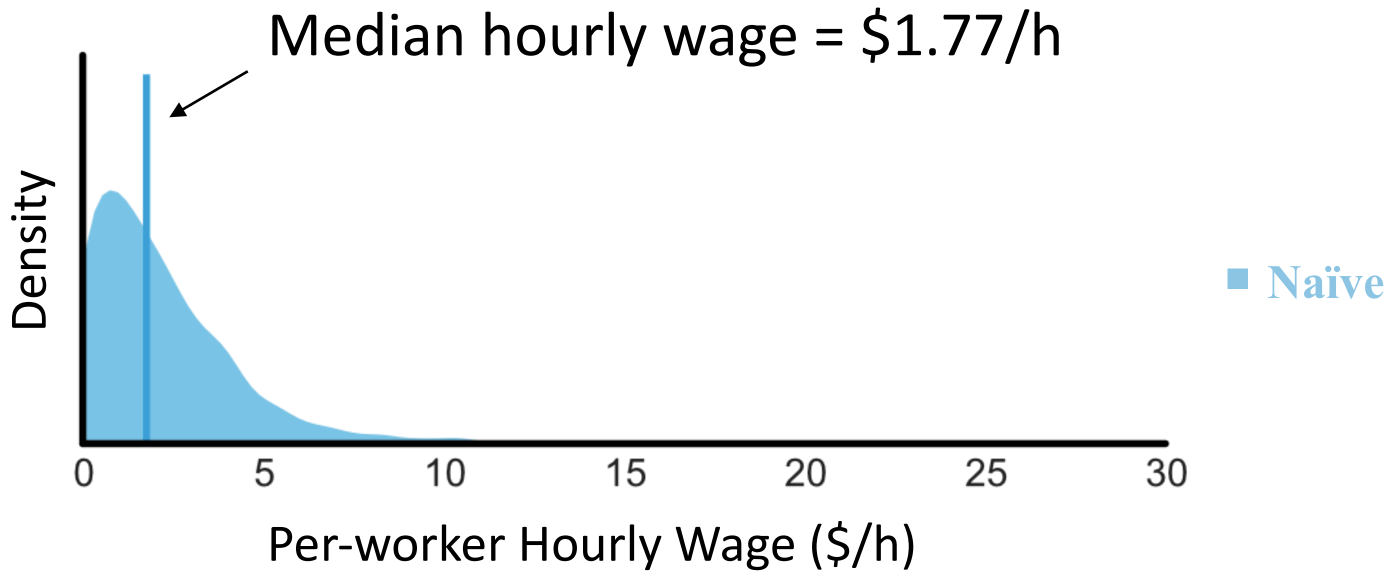
Because different choice of *D* yield different sets of clusters, we use *D*=0 and *D*=1 minute and see their effects on cluster-based hourly wages

Worker Hourly Wage: Result (Naïve)

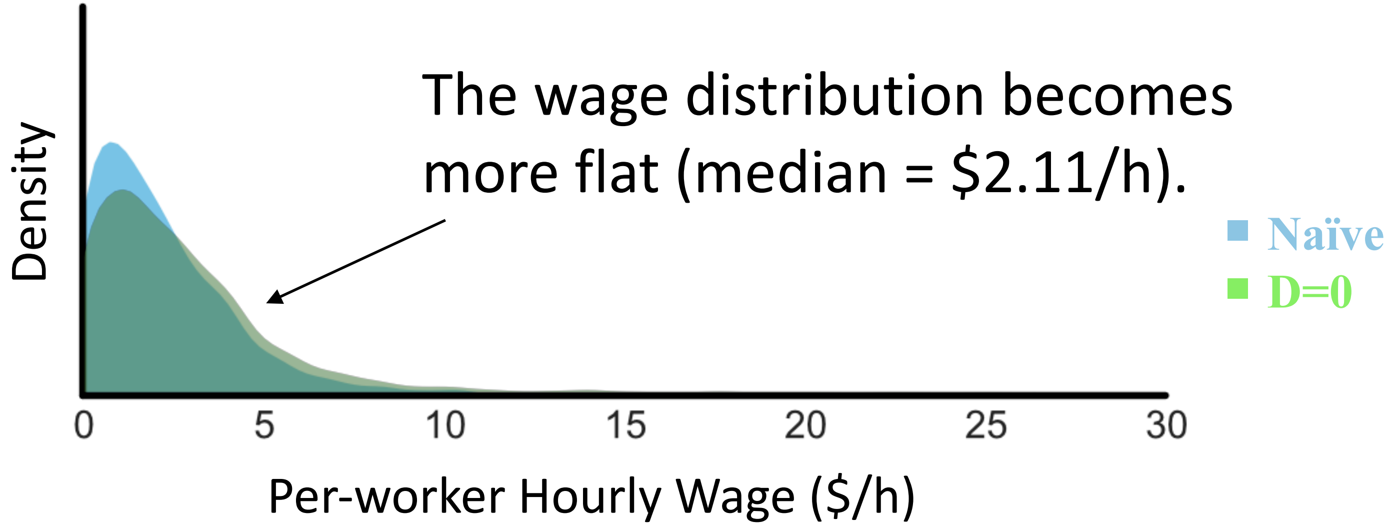
Worker Hourly Wage: Result (Naïve)



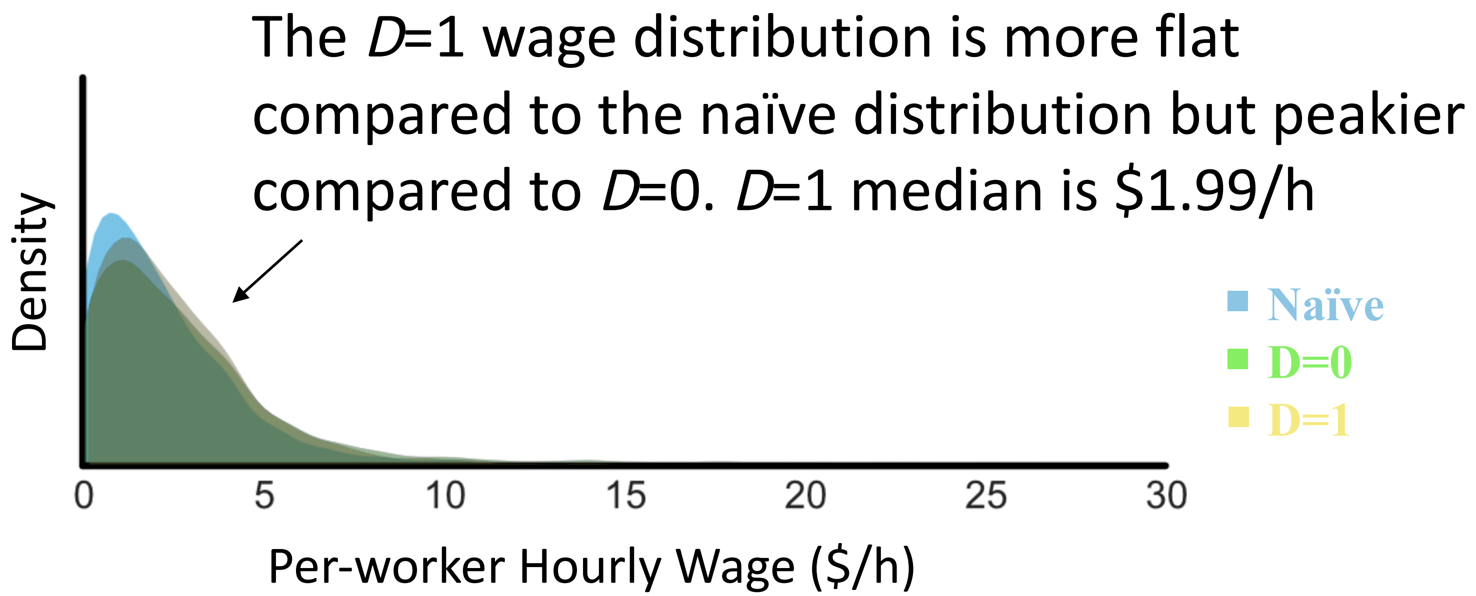
Worker Hourly Wage: Result (Naïve)



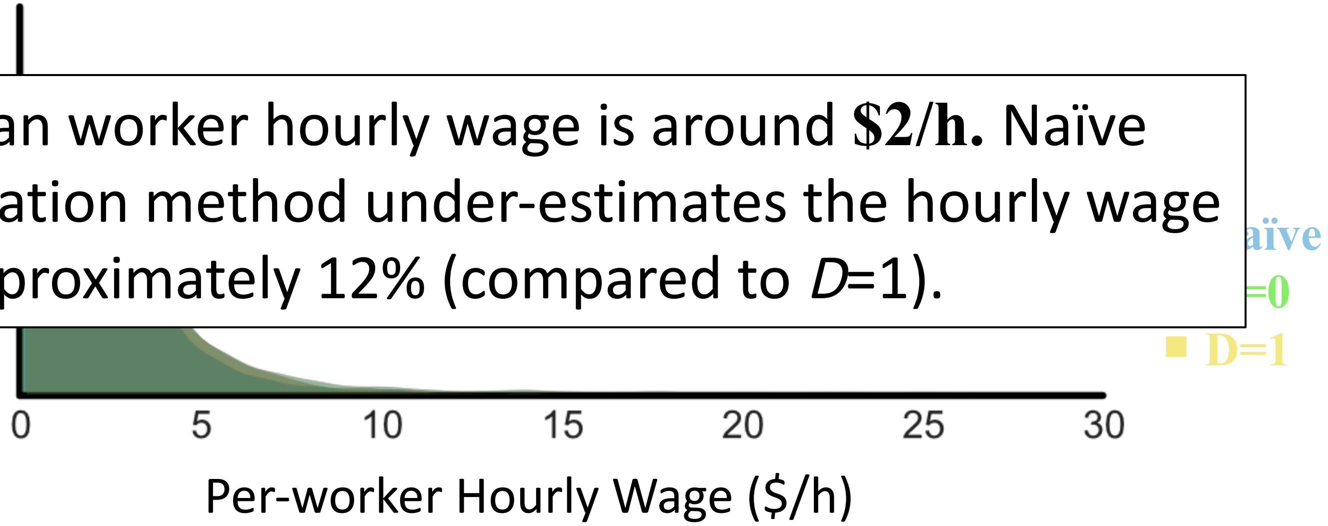
Worker Hourly Wage: Result (Clustered)



Worker Hourly Wage: Result (Clustered)



Worker Hourly Wage: Result



Median worker hourly wage is around **$2/h.** Naïve estimation method under-estimates the hourly wage by approximately 12% (compared to *D*=1).

Takeaway 1

The majority of workers on Amazon Mechanical Turk work with **hourly wage below $2/h**

Research Questions

How much are workers earning on Amazon Mechanical Turk?

What contributes to the low wage?

Research Questions

How much are workers earning on Amazon Mechanical Turk?

What contributes to the low wage?

What contributes to the low wage?

Unpaid work

Low reward

Requesters

Qualifications

Task types

What contributes to the low wage?

Unpaid work Low reward **Requesters** Qualifications Task types

What contributes to the low wage?

Unpaid work Low reward Requesters Qualifications Task types

Being A Turker

David Martin, Benjamin V. Hanrahan, Jacki O'Neill
Xerox Research Centre Europe
6 chemin de Maupertuis, Grenoble France
{david.martin, ben.hanrahan, jacki.oneill}@xrce.xerox.com

Neha Gupta
University of Nottingham
University Park NG7 2TD Nottingham
neha.gupta@xrce.xerox.com

ABSTRACT
We conducted an ethnomethdological analysis of publicly available content on Turker Nation, a general forum for Amazon Mechanical Turk (AMT) users. Using forum data we provide novel depth and detail on how the Turker Nation members operate as economic actors, working out which Requesters and jobs are worthwhile to them. We show some of the key ways Turker Nation functions as a community and also look further into Turker-Requester relationships from the Turker perspective – considering practical, emotional and moral aspects. Finally, following Star and Strauss [25] we analyse Turking as a form of invisible work. We do this to illustrate practical and ethical issues relating to working with Turkers and AMT, and to promote design directions to support Turkers and their relationships with Requesters.

Author Keywords
Ethnomethodology; content analysis; crowdsourcing; microtasking; Amazon Mechanical Turk; Turker Nation.

ACM Classification Keywords
H.5.3 Group and Organizational Interfaces – Computer, Supported Cooperative Work

General Terms
Human Factors

INTRODUCTION
The concept of crowdsourcing was originally defined by Jeff Howe of Wired Magazine as "*the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call*." [8] This 'undefined network of people' is the key topic of this article. We present the findings of an ethnomethodological analysis of posts and threads on a crowdsourcing forum called Turker Nation[1]. We have sought to understand members of the crowd – their *reasoning practices*, *concerns*, and *relationships with requesters and each other* – as they are shown in their posts on the forum. We seek to present them as faithfully as possible, in their own words, in order to provide more definition to this network of people. We believe that this will be beneficial for researchers and businesses working within the crowdsourcing space.

Crowdsourcing encompasses multiple types of activity: invention, project work, creative activities, and microtasking. This latter is our focus here. The most well-known microtask platform is Amazon Mechanical Turk (AMT)[2], and the Turker Nation forum that we studied is dedicated to users of this platform. The basic philosophy of microtasking and AMT is to delegate tasks that are difficult for computers to do to a human workforce. This has been termed 'artificial artificial intelligence'. Tasks like image tagging, duplicate recognition, translation, transcription, object classification, and content generation are common. 'Requesters' (the AMT term for people who have work to be completed) post multiple, similar jobs as Human Intelligence Tasks (HITs), which can then be taken up by registered 'Turkers'. Turkers (termed 'Providers' by AMT) are the users completing the HITs, which typically take seconds or minutes paid at a few cents at a time.

For Amazon, the innovative idea was to have an efficient and cost effective way to curate and manage the quality of content on their vast databases (weeding out duplicates, vulgar content, etc.). While Amazon is still a big Requester, AMT has been deployed as a platform and connects a wide variety of Requesters with up to 500,000 Providers. However, Fort et al. [6] have performed an analysis on the available data and suggest that real number of active Turkers is between 15,059 and 42,912; and that 80% of the tasks are carried out by the 20% most active (3,011–8,582) Turkers. While these numbers are useful, the research community still has little deep qualitative knowledge about this workforce. Questions remain unanswered such as: how and what do they look for in jobs; what are their concerns; and how do they relate to requestors?

LITERATURE REVIEW
To date much of the research on AMT takes the employers' perspective, e.g. [14, 15, 17, 18], and this has in turn been highlighted [6, 16]. Silberman et al. [23] note that this mainstream research looks at how: *"[to] motivate better, cheaper and faster worker performance […] to get good*

[1] http://turkernation.com/forum.php

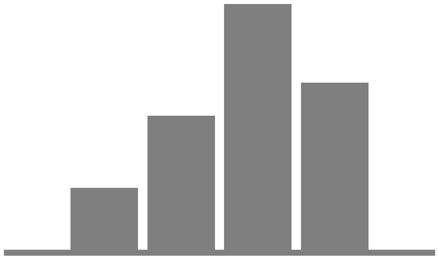
[2] http://www.mturk.com

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

"[…] aspects of turking [(working on Amazon Mechanical Turk)] like simply searching for jobs can take a considerable amount time.

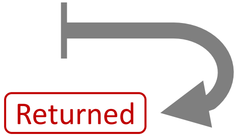
The time spent learning and searching are clear examples of invisible [(unpaid)] work that Turkers must engage in […]."

Martin *et al.*, (2014) *Being a Turker*, CSCW 2014



The issue of unpaid work has been identified in prior work,
but **its effects are not quantified**

We quantify three types of unpaid work

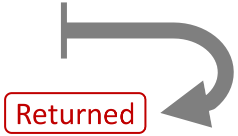


Time spent on
returned tasks

Time spent on
rejected tasks

Time
between tasks

We quantify three types of unpaid work

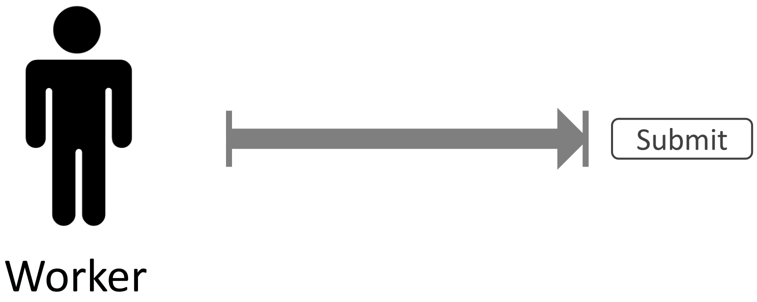


Time spent on
returned tasks

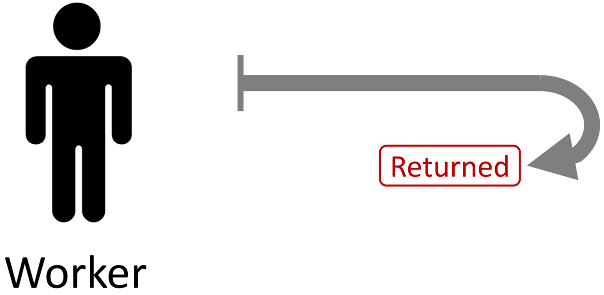
Time spent on
rejected tasks

Time
between tasks

Task Submit and Return



Task Submit and Return



Reward will not be paid and so the time spent is wasted

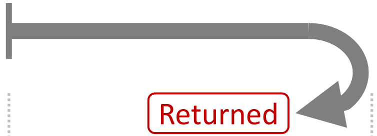
Time Spent on **Returned Tasks**

Submitted

$$\sum \text{Time}_{\text{submitted}}$$

For all tasks from all workers

Returned



$$\sum \text{Time}_{\text{returned}}$$

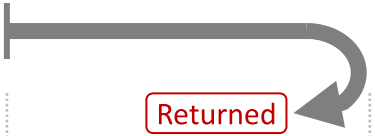
Time Spent on **Returned Tasks: Result**

Submitted

$$\sum \text{Time}_{\text{submitted}}$$

= 98,202 hours (68.2%)

Returned



$$\sum \text{Time}_{\text{returned}}$$

= 45,778 hours (31.8%)

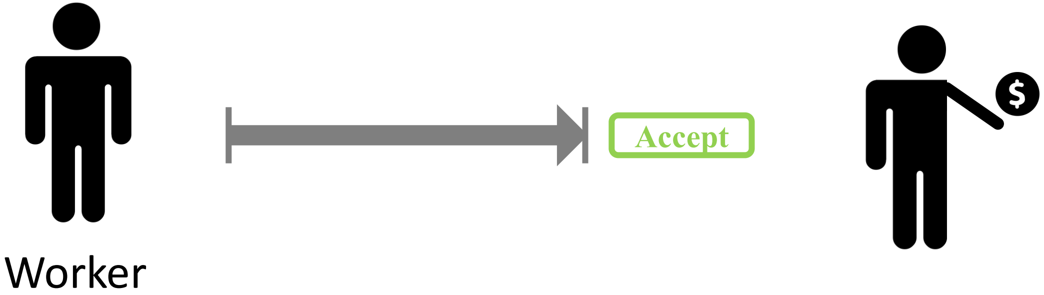
We quantify three types of unpaid work

Time spent on
returned tasks

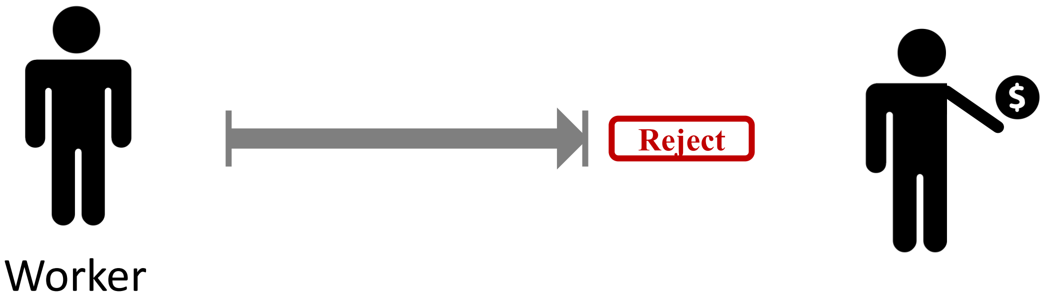
Time spent on
rejected tasks

Time
between tasks

Task **Accept** and Reject

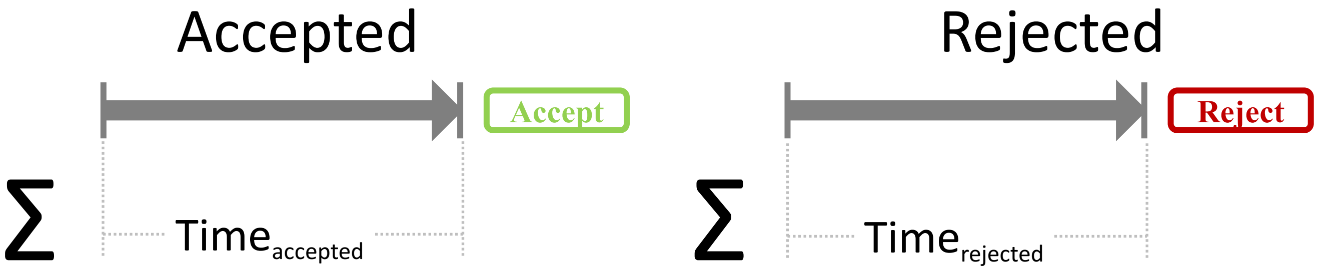


Task Accept and **Reject**



Time Spent on **Rejected Tasks**

We had data on task accept *vs.* reject status for 29.6% of the submitted tasks



Time Spent on **Rejected Tasks: Result**

We had data on task accept *vs.* reject status for 29.6% of the submitted tasks

Accepted

Accept

$$\sum \text{Time}_{\text{accepted}}$$

= 33,130 hours (99.3%)

Rejected

Reject

$$\sum \text{Time}_{\text{rejected}}$$

= 240 hours (0.7%)

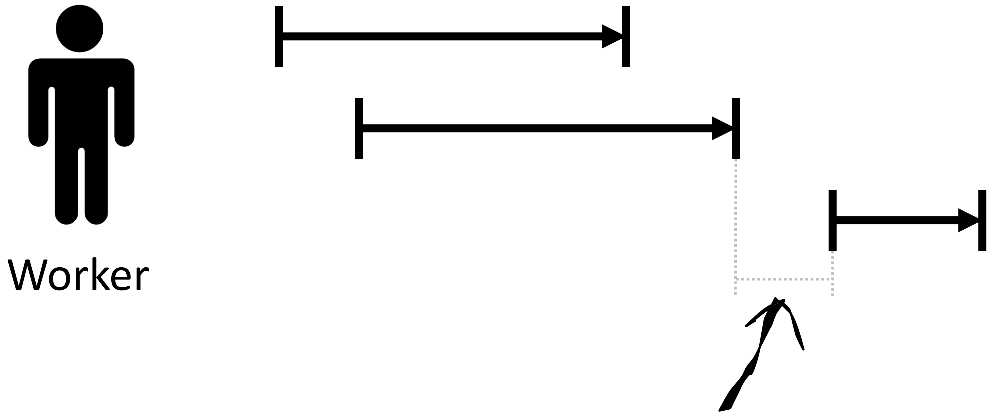
We quantify three types of unpaid work

Time spent on
returned tasks

Time spent on
rejected tasks

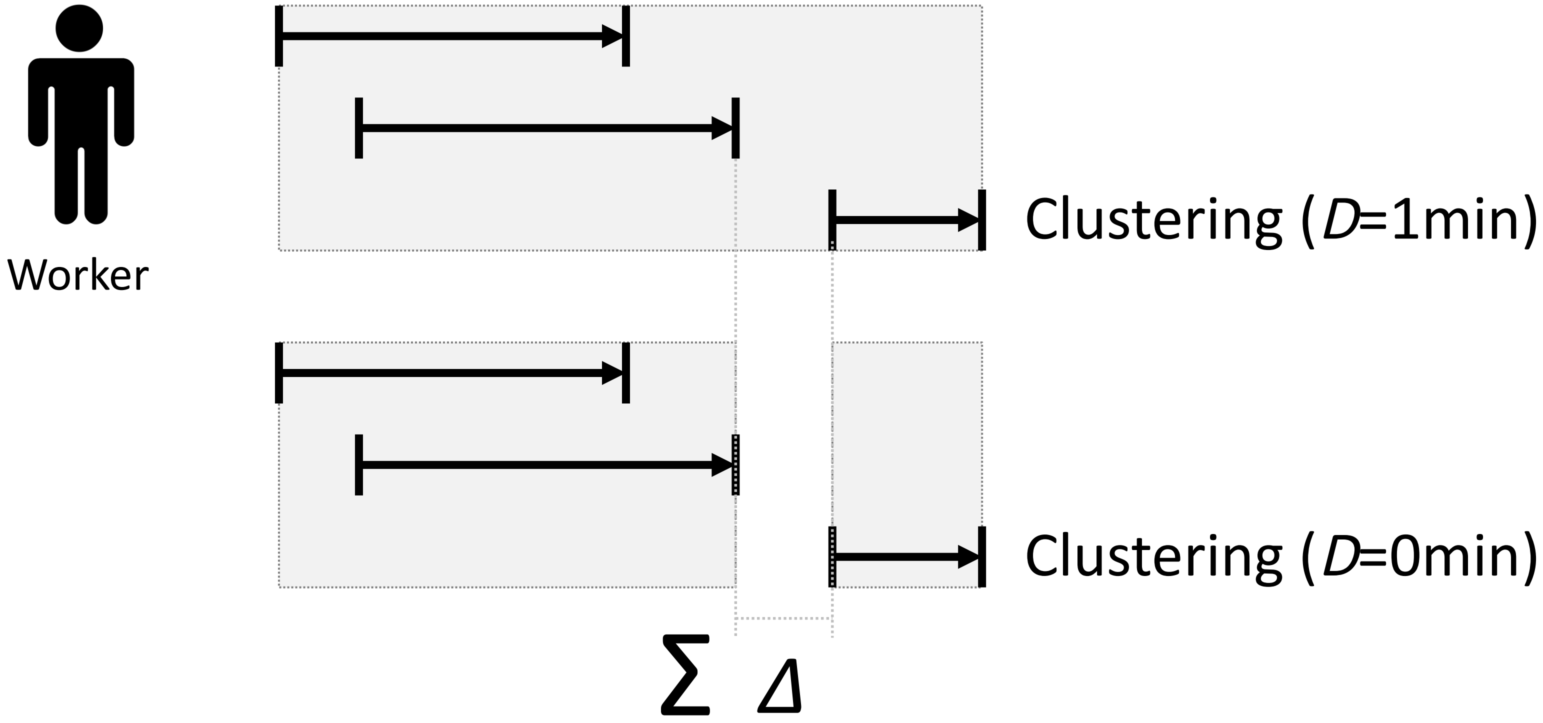
Time
between tasks

Time **Between Tasks**



We want to know the effect of this small gap between tasks (*e.g.,* task search time)

Time **Between Tasks**



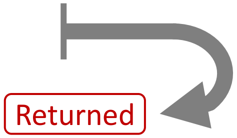
Time **Between Tasks: Result**

$$\sum \Delta = 4{,}603 \text{ hours}$$

Result

45,778 hours

31.8% of work



240 hours

0.7% of work

4,603 hours

4.7% of work

Takeaway 2

Returning tasks has the biggest impact on the hourly wage. In fact, **31.8% of work time is wasted due to this unpaid work.**

What contributes to the low wage?

Unpaid work

Low reward

Requesters

Qualifications

Task types

What contributes to the low wage?

Unpaid work | Low reward | **Requesters** | Qualifications | Task types

Workers are underpaid.

< $2/h

< $2/h

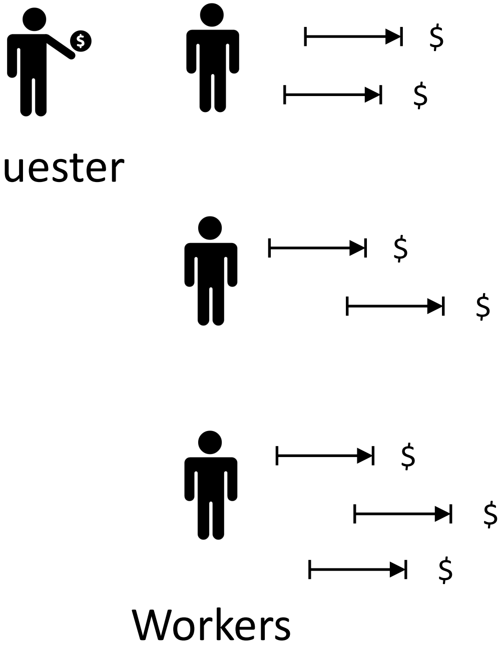
Workers are underpaid. Is this because all requesters treat workers unfairly,

< $2/h

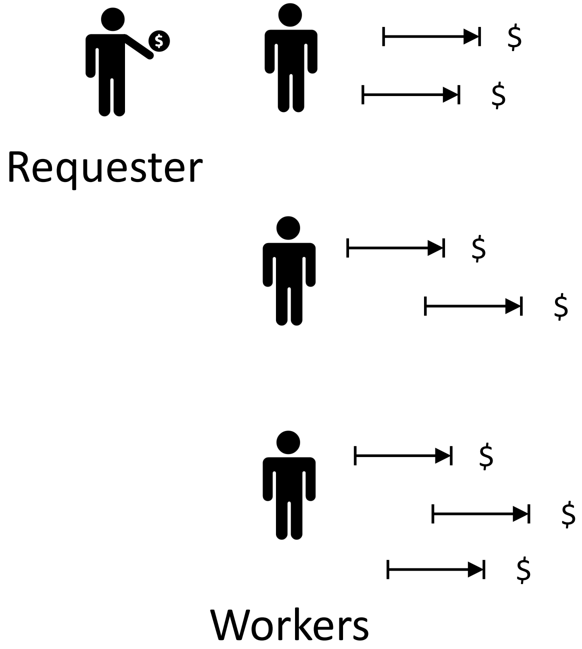
Workers are underpaid. Is this because all requesters treat workers unfairly, or are there a small number of requesters who post many very low paid tasks?

We investigated the distribution of **per-requester hourly payment**

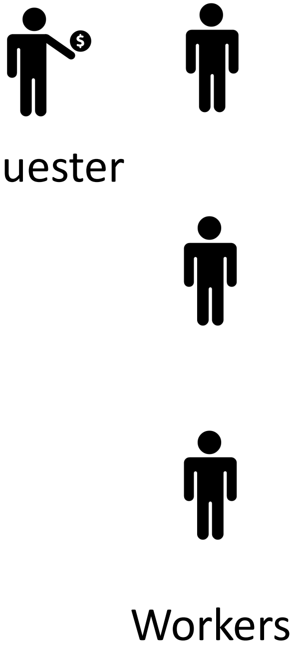
Per-requester Hourly Payment



Per-requester Hourly Payment



Per-requester Hourly Payment

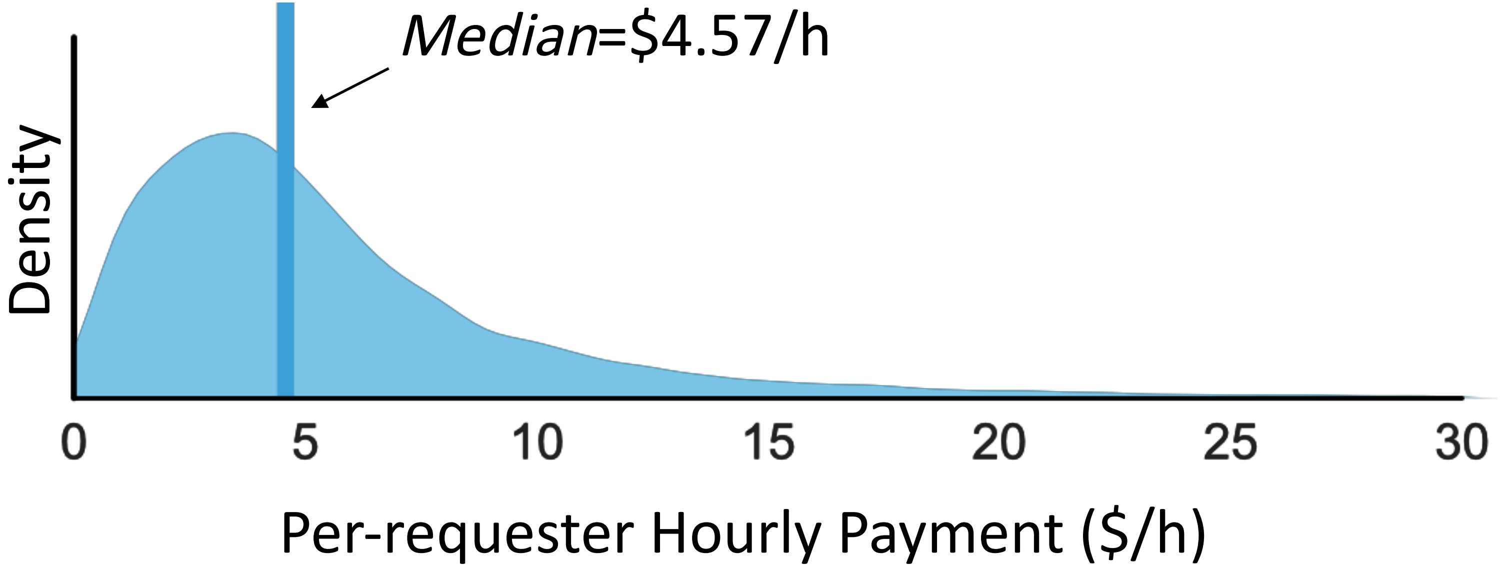


$$\frac{\sum \text{Task Payment (\$)}}{\sum \text{Task Interval}} = \text{Per-requester Hourly Payment}$$

Per-requester Hourly Payment: Result

N=19,598 requesters

Per-requester Hourly Payment: Result



N=19,598 requesters

Takeaway 3

About half of the requesters pay below $5/h, which is below the U.S. federal minimum wage.

Takeaways

< $2/h

Crowd workers are underpaid and they often earn below $2/h

~~$~~

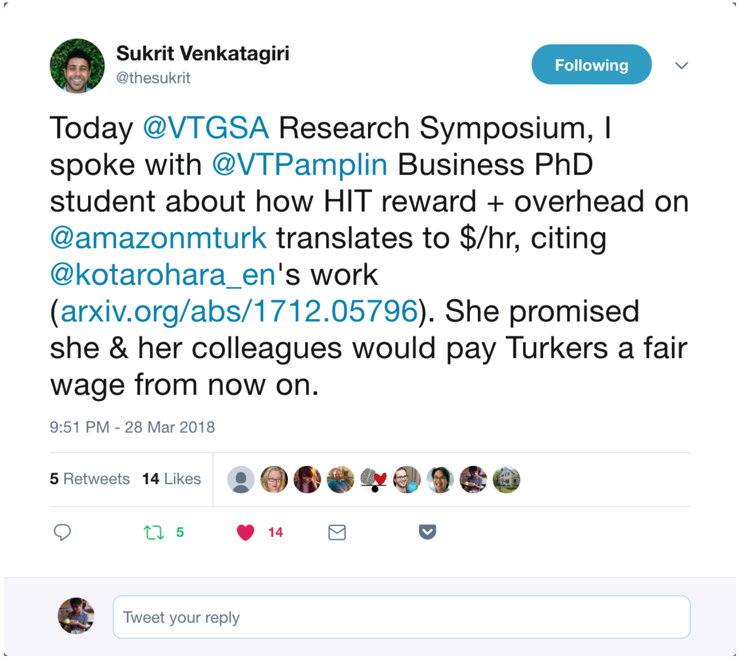
Unpaid work, particularly returning tasks has a large impact on the hourly wage

Majority of the requesters reward workers below $5/h

Discussion and Future Work

- Could we create **tools for workers** to make it easier to search for tasks that give them good wage, avoid tasks that are not completable, and find requesters fair wage?
- Could we create **technologies for requesters** to help them pay fairly?

Discussion and Future Work



How can we nudge more people to pledge and help them keep committing to their promises?

Limitations

- Sampling bias
- Our analyses did not take bonus payment into account
- No information about demographics

A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage
Chris Callison-Burch, Jeffrey P. Bigham

The Economic and Social Research Council, Grant ES/N017099/1

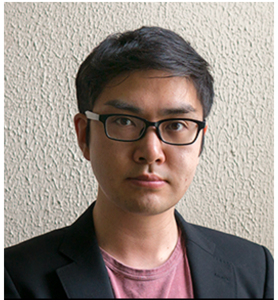
We thank the workers who provided the work log data that enabled this research

Image credits to Saeful Muslim, Kirby Wu from the Noun Project

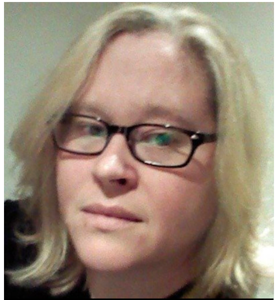
SINGAPORE MANAGEMENT UNIVERSITY   Carnegie Mellon University   MANNERS MAKYTH MAN   McMaster University   West Virginia University   Penn UNIVERSITY of PENNSYLVANIA

A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage
Chris Callison-Burch, Jeffrey P. Bigham

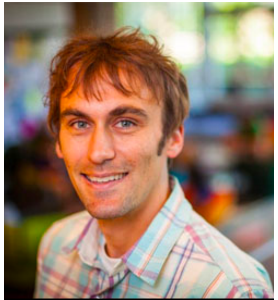

Kotaro Hara


Abigail Adams


Kristy Milland


Saiph Savage


Chris Callison-Burch


Jeffrey P. Bigham

SINGAPORE MANAGEMENT UNIVERSITY
Carnegie Mellon University
MANNERS MAKYTH MAN
McMaster University
West Virginia University
Penn University of Pennsylvania

A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage
Chris Callison-Burch, Jeffrey P. Bigham

Questions?

SINGAPORE MANAGEMENT UNIVERSITY
Carnegie Mellon University
MANNERS MAKYTH MAN
McMaster University
West Virginia University
Penn UNIVERSITY of PENNSYLVANIA

A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage
Chris Callison-Burch, Jeffrey P. Bigham

< $2/h

Crowd workers are underpaid and they often earn below $2/h

Unpaid work, particularly returning tasks has a large impact on the hourly wage

Majority of the requesters reward workers below $5/h

Singapore Management University
Carnegie Mellon University
McMaster University
West Virginia University
Penn University of Pennsylvania

Takeaways

< \$2/h

Crowd workers are underpaid and they often earn below \$2/h

~~\$~~

Unpaid work, particularly returning tasks has a large impact on the hourly wage

Majority of the requesters reward workers below \$5/h

Takeaway 2

Returning tasks has the biggest impact to the hourly wage. In fact, **31.8% of work time is wasted due to returned tasks.**