The Sixth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2018)

How Do Crowdworker Communities and Microtask Markets Influence Each Other? A Data-Driven Study on Amazon Mechanical Turk

Jie Yang,[1] Carlo van der Valk,[2] Tobias Hoßfeld,[3] Judith Redi,[2,4] Alessandro Bozzon[2]
[1]University of Fribourg, [2]Delft University of Technology, [3]University of Würzburg, [4]Exact B.V.
jie@exascale.info, {c.p.vandervalk, a.bozzon}@tudelft.nl, judith.redi@exact.com, tobias.hossfeld@uni-wuerzburg.de

Abstract

Crowdworker online communities — operating in fora like mTurkForum and TurkerNation — are an important actor in microwork markets. Albeit central to market dynamics, how the behavior of crowdworker communities and the dynamics of online marketplaces influence each other is yet to be understood. To provide quantitative evidence of such influence, we performed an analysis on 6-years worth of mTurk market activities and community discussions in six fora. We investigated the nature of the relationships that exist between activities in fora, tasks published in mTurk, requesters for such tasks, and task completion speed. We validate — and expand upon — results from previous work by showing that (i) there are differences between market demand and community activities that are specific to fora and task types; (ii) the temporal progression of HIT availability in the market is predictive of the upcoming amount of crowdworker discussions, with significant differences across fora and discussion categories; (iii) activities in fora can have a significant positive impact on the completion speed of tasks available in the market.

Introduction

Microtask crowdsourcing has become a highly appealing approach for data collection and augmentation purposes. Microwork markets such as Amazon Mechanical Turk (mTurk) and Crowdflower are still on the rise, providing *requesters* with tools to publish work in the form of microtasks — or HITs (Human Intelligence Tasks); and *crowdworkers* with interfaces to actively seek for HIT groups to complete and to be rewarded for.

Microwork markets are socio-technical systems regulated by complex mechanisms that relate the activities of requesters and crowdworkers. This class of online labor has been widely studied in many aspects, from crowdworker analysis (Bozzon et al. 2013; Difallah, Demartini, and Cudré-Mauroux 2013), to market analysis (Ipeirotis 2010; Difallah et al. 2015); from incentive mechanism design (e.g. pricing schemes) (Gao and Parameswaran 2014), to crowdworker retention (Difallah et al. 2014). Recent work (Gray et al. 2016; Martin et al. 2014) has shown that crowdworkers interact and collaborate outside microwork markets, in online fora such as `mTurkForum` and `Turkopticon`.

These fora are virtual social environments that aim at developing social capital in online labour, by supporting the social and technical needs of their members. Using survey and forum data, Yin et al. (2016) discovered that a high proportion of crowdworkers use at least one forum (59.1% of all crowdworkers in their survey), and that crowdworkers within the same forum are more likely to establish direct interactions with their fellows. These findings clearly point to the need for a better understanding of crowdworker communities and their relationships with microwork markets.

Fora provide a unique vantage point to observe the activities of a large amount of crowdworkers. Previous studies (McInnis et al. 2016; Ross et al. 2010; Yin et al. 2016) mainly analyze crowdworker activities in a limited amount of fora (often a single one) and for a short period of time, thus covering an incomplete set of activities performed by a subset of crowdworker communities. More importantly, they do not address the relationship between such activities and microwork market dynamics. In this paper, we focus on the Amazon Mechanical Turk (mTurk) marketplace, and study how the activities of crowdworkers in fora are influenced by — and can influence — the status of mTurk. Specifically, we seek answers to the following research questions:

- **RQ1:** How is the activity of crowdworker communities in online fora influenced by the content, requesters, and variations of availability of HIT groups in the mTurk market?
- **RQ2:** To what extent does the activity of crowdworker communities affect task completion speed in mTurk?

Answering these questions is of crucial importance for a variety of purposes, including the design of tasks, incentive and task allocation schemes, human computation systems, and microwork markets (Gaikwad et al. 2015; Yang and Bozzon 2016).

In this work, we adopt a data-driven approach, and collect, enrich, and analyze a dataset containing 6-years worth of discussions produced by crowdworker communities in six online fora – `mTurkCrowd`, `mTurkForum`, `TurkerNation`, `mTurkGrind`, `Reddit HWTF`, and `Turkopticon`. We contribute a taxonomy to categorize discussions according to topic and function (e.g. Comment, Experience, Social), and trained a machine learning classifier to automatically categorize messages, thus enabling large-scale analysis of workers' discussions in fora.

To study the mutual influence of crowdworker communities and the mTurk market, we collected data for more than 2.6M HIT groups, including data about their publication and completion over time. We linked these HITs with related messages in fora, to create a dataset that is unique both in scale and diversity.[1]

We show that fora are differently popular, and the activities of their members encode different norms, preferences, and task consumption behaviors. We also show how requesters with higher communicativity and generosity are more likely to be mentioned in all fora. By applying time series analysis techniques, we find significant synchronicity between the temporal evolution of HITs availability in the market and discussions among crowdworker communities (with an average positive lag of 4 hours, and shortest positive lag of 45 minutes), with significant differences across fora and discussion categories. Finally, we present quantitative evidence of the positive effect that HIT groups' mentions in fora have on HIT consumption throughput (on average, a 59% improvement in the first hour). A targeted analysis of temporal synchronicity shows that the temporal progression of mentions in fora can help in the prediction of throughput, with an average positive lag of 30 minutes, and an average 340% boosting effect. We conclude the paper by discussing the results, their implications for stakeholders of microwork markets, and threats to their validity.

Related Work

Previous work addressed the complex mechanisms that regulate microwork markets from the perspectives of the composition of crowdworkers, properties of tasks and task-types, and incentive mechanisms. Ipeirotis (2010) indicate that mTurk crowdworkers are mainly from US and India. Difallah, Filatova, and Ipeirotis (2018) reveal that there are more than 100K workers available in mTurk and more than 2K are active at any time. The roles of task properties such complexity and clarity on task performance have been investigated in Yang et al. (2016) and Gadiraju, Yang, and Bozzon (2017), where it has been found that both properties can significantly affect crowdworkers' selection of tasks and influence task completion speed. Incentive mechanism have been studied by Gao and Parameswaran (2014) and Difallah et al. (2014), where optimal pricing schemes are investigated to reduce the cost without sacrifice for task completion speed and worker retention rate.

Most notably, Difallah et al. (2015) conducted a long-term analysis of the mTurk market, showing that the size and recency of HIT groups are two key features for the prediction of the completion time (throughput) of a HIT group.

The analysis of the impact of crowdworker communities on microwork markets received less attention. In the context of mTurk, crowdworkers organize in communities around a number of online discussion services (Gray et al. 2016; Irani and Silberman 2013; Yin et al. 2016) such as `Turkopticon`, `mTurkCrowd`, and `TurkerNation`. Members of crowd communities turn to these services to serve three fundamental needs for their members (Gray et al. 2016; Laplante and Silberman 2016; Wang et al. 2017): functional needs (e.g. skill improvement, information sharing), social needs (e.g. building community and trust between workers, providing collective protection), and psychological needs (e.g. providing moral support and encouragement to each other).

Despite being "invisible" to microwork platforms, crowdworker communities share a great amount of relevant information about HITs, requesters, and their own work (Silberman, Irani, and Ross 2010). Yin et al. (2016) highlight the presence of rich network topology around crowdworkers, which is built around online fora. Authors enabled their analysis by injecting a HIT in mTurk, to provide crowdworkers incentive to self-report their connections. Martin et al. (2014) conducted an ethnomethodological study on the crowdworker community active on `TurkerNation` to observe crowdworker discussions on the forum for a period of seven months and analyze crowdworkers' motivation and their attitude towards the different actors in the market (especially requesters). Their work clearly shows that crowdworkers regard their activities on mTurk as paid work, and often, as the main source of income. Therefore, crowdworkers strive for efficiency in work execution and fairness, and transparency in the way the work is evaluated and rewarded. Wang et al. (2017) further show that crowdworker participate in communities to find good HITs and their participation is often associated with higher income.

These studies inspired and informed our work. We validate — and expand upon — their results by means of scaled-up analysis of multiple fora, considering their whole history of existence. The resulting dataset combines information from the mTurk market. Our quantitative analysis has a broader spectrum; in particular, it includes the study of the relationships between activities in crowdworker fora and the properties, availability, and consumption rate of HIT groups, as well as the properties of requesters, in the mTurk market.

Dataset

We consider four general-purpose fora related to mTurk – `mTurkCrowd`, `mTurkForum`, `TurkerNation` and `mTurkGrind` – and two popular specialised fora – `Turkopticon` and `Reddit HWTF`. `Turkopticon` (Irani and Silberman 2013) is a system designed to focus on the evaluation of *requesters* (and their HIT groups), as performed by mTurk crowdworkers, who are also allowed to comment on ratings. `Reddit HWTF` (HITsWorthTurkingFor) is a subreddit devoted to the advertisement of HIT groups that community members deem worthy of attention. These fora were selected due to their popularity among workers (Yin et al. 2016). We included multiple fora in the analysis to account for the limited overlap of workers across fora (around 30%), and the homophily effect found in previous studies (Yin et al. 2016). We assume all the fora members to be also workers in mTurk, including administrators and moderators. We are aware that some of the registered members are requesters in mTurk, or scientists interested in studying crowdworker activities. We believe their number to be limited, and their impact on the forum activity to be marginal for the purpose of this study.

[1]The dataset is available for download at https://github.com/yangjiera/HCOMP2018_Worker-Communities.

Forum		Overall Statistics			Linked HIT Groups					Linked Requesters				
	Start	#Members	#Threads	#Messages	#HM	AvgHMW	% HMF	#HITs	% MH	#RM	AvgRMW	% RMF	#REQ	% MR
`mTurkForum`	07/12	4,926	1,889	1,427,856	233,294	47.36	16.30%	104,893	15.81%	217,489	44.15	15.19%	22,737	43.08%
`mTurkGrind`	10/13	3,217	943	948,775	229,504	71.34	24.08%	100,310	23.07%	233,063	72.45	24.46%	22,078	60.44%
`TurkerNation`	11/14	572	563	173,942	73,637	128.74	41.87%	41,659	5.44%	73,767	128.96	41.95%	11,610	45.69%
`mTurkCrowd`	01/16	616	131	177,669	40,771	66.19	21.88%	19,278	11.76%	41,522	67.41	22.29%	6,896	85.66%
`Turkopticon`	01/09	18,640	310,129	371,750	NA	NA	NA	NA	NA	371,797	19.95	94.50%	45,701	54.10%
`Reddit HWTF`	03/16	930	3,490	17,843	1,937	2.08	10.01%	1,649	2.58%	635	0.68	3.28%	415	9.21%
Overall		28,901	317,145	3,117,835	579,143			184,390		938,273			50,912	

Table 1: Descriptive statistics of the six targeted fora and links to mTurk HIT groups and requesters. Legend: **Start** – earliest crawled message; **#HM** – number of messages with Links to HIT groups; **AvgHMW** – average number of HM per user in forum; **% HMF** – percentage of HM in forum messages; **#HITs** – number of unique HIT groups mentioned in forum messages; **%MH** – percentage HIT groups in the market mentioned in messages; **#RM** – number of messages with Links to Requesters; **AvgRMW** – average number of RM per user in forum; **% RMF** – percentage of RM in Forum messages; **#REQ** – number of unique Requesters mentioned in forum messages; **%MR** percentage of requesters in the market mentioned in messages. *Ratio of mentioned HIT groups and Requesters are calculated within the timespan of existence of each forum.*

Dataset Creation

Fora. We focus on content that is publicly visible on fora, or available to registered users. We retrieved the whole history (until May 20th, 2016) of discussions and messages of all fora (except `Reddit HWTF`) using custom Web crawlers. `TurkerNation` is the new instantiation of an older forum (http://turkers.proboards.com) that migrated technological platform in 2014. We were not able to retrieve earlier data. `Reddit HWTF` content was retrieved using the official reddit API, which, unfortunately, sets limitations in the amount (and age) of accessible content. Therefore, our collection is limited to the period Mar. 27, 2016 to May 20, 2016. For each thread and message, we retrieved title, content, timestamp, and creator. Table 1 reports descriptive statistics of the resulting dataset. Overall, we collected more than 3.1M messages, produced by 28.9K members.

mTurk Market. The dataset includes Amazon mTurk activities spanning 6 years. We started with the dataset studied in (Difallah et al. 2015), which contains more than 2.56M distinct HIT groups, and 130M HITs produced from 2009 to 2014. To analyze the activities of more recent fora, we enriched the dataset with 46K HIT groups and 1.9M HITs collected between Apr. 11 and May 20 2016. All HIT groups are described by metadata, including their size at publication, title, description, reward, and allotted time. To study HIT groups consumption over time, we adopted the notion of HIT group *throughput* (the number of HITs in the group completed in a given time interval) proposed in (Difallah et al. 2015). Throughput information is obtained by crawling every 5 minutes the mTurk system, to retrieve, for each active HIT group, the amount of available HITs.

Linkage to mTurk

To enable our study, HIT groups and requesters must be identified in fora messages. We focus on explicit *mentions*, i.e., unambiguous references to HIT groups and requesters.

We parsed the text in threads and messages to extract and process `http` links towards mTurk pages of HIT groups and requesters.[2] This technique allowed us to achieve maximum linking precision. 22.84% of the messages in the dataset link to at least one HIT group, while 33.62% link to a requester page in the mTurk market. We retrieved a total of 184K distinct HIT groups (up to 20% of the total amount of groups available in the market during the considered fora lifetime) from 579K messages, and 51K distinct requesters (up to 85% of active requesters) from 938K messages. Table 1 summarizes the distribution of links to mTurk HIT groups (%MH) and requesters (%MR) across fora. The considered fora cover a partial yet numerically significant share of the mTurk market. There are also significant differences in terms of market coverage. For instance, the HIT groups mentioned in `TurkerNation` account for only 5% of the market. `mTurkForum` and `mTurkGrind` feature a better coverage (respectively 15% and 23%). Such differences can be partially explained by forum-specific "culture" and norms. For instance, the lower coverage of `Reddit HWTF` (less than 3% of the HITs in the market) could be explained by its mission statement[3], where members are asked to only report HIT groups with fair hourly retribution.

Message Categorization

We categorize messages in the dataset according to the type of discussion they include. Given the size of the dataset, we resorted to supervised learning for automatic classification. A manual annotation process was instrumented to create a training set of suitable size.

Annotation of Training Dataset. To minimize sampling bias, we randomly selected 10% of all threads from each forum, except `Turkopticon`. From each selected thread, we picked a random sample of at most 50 messages, which were empirically found to be representative of messages in a thread. In `Turkopticon`, given the amount (and topical homogeneity), we sampled 500 threads. The resulting $13,017$ messages were manually inspected to label messages.

We employed card sorting (Spencer 2009), a technique widely used in the design of information architecture to cre-

[2] HIT group links: https://www.mturk.com/mturk/preview?groupId=⟨tID⟩, where ⟨tID⟩ is the HIT group identifier); requester links: https://www.mturk.com/mturk/searchbar?selectedSearchType=hitgroups&requesterId=⟨rID⟩, where ⟨rID⟩ is the requester identifier.

[3] https://www.reddit.com/r/HITsWorthTurkingFor/wiki/index

Type	Accuracy	F-Score	Type	Accuracy	F-Score
Ask or Answer	0.86	0.27	Comment	0.98	0.60
Experience	0.80	0.46	Judgment	0.86	0.46
Rating	0.93	0.85	Social	0.75	0.74

Table 2: Performance of message type classification.

ate mental models and derive taxonomies from input data. From recent work on crowdworker communities, we elicited a number of message types (e.g. "problems, suggestions, tips" and "community communication and interests" from (Martin et al. 2014)). Then, using *open* card sorting, we synthesized and defined six types of messages, described as follows. 1) *Ask or Answer*: messages with questions about tasks, general purpose issues, or answers to previous questions. Example: "Anyone able to withdraw?" 2) *Comment*: messages with general comments about a HIT group, such as its availability, requirements, or presence of bugs (e.g. lack of completion code). Example: "Can't be on mobile." 3) *Experience*: messages that report the experience of a crowdworker in HIT execution, e.g. the amount of time spent on a task, or the amount of rewarded bonus. Example: "Projected Earnings for Today $70.00." 4) *Judgment*: messages where crowdworkers explicitly express compliment or criticisms about a HIT or a requester. Example: "$0.60 cent one is good, 0.36 hit sucks." 5) *Rating*: messages that include a reference to `Turkopticon` rating, or rating in other fora. Rating messages often serve as recommendation from crowdworkers to the community, as only HIT groups worthy of discussion are mentioned. Example: "This requester has actually joined Opticon just to flag negative reviews and accuse them of blackmail." 6) *Social*: messages where crowdworkers address the community with general-purpose social topics, e.g. greetings and jokes. Example: "Turtles for days Happy new year!"

We then applied closed card sort to categorize all messages in the training set. The first author of this paper created, in a digital form, a card for each message. By means of an online collaboration tool, other researchers (including the second and last author) were involved in assigning cards to message types. To reduce bias and strengthen the validity of results, all researchers reviewed and agreed upon the categorization of messages. Messages could belong to multiple types. For instance, it is common for workers to rate a task while sharing information about their experience, or expressing a judgment about the task.

Automatic Classification. We fed a multi-label Random Forest classifier with textual features of the annotated messages (bag of words, TF-IDF weighted) and trained it to predict the message's type. To account for the relative sparsity of some message types in the training dataset, we assess the performance of the classifier both in terms of accuracy and F-score in a 5-fold cross-validation setting. The classification performance is reported in Table 2. *Rating*, *Social*, and *Comment* messages are those identified more accurately by the classifier. The classification of *Experience* and *Judgment* messages is also accurate and with acceptable F-score. *Ask or Answer* messages are the most difficult to classify (i.e. low F-Score): without loss of generality, we exclude this category of messages from subsequent analysis, and leave the improvement of classification performance to future work.

54.95% of messages were classified as *Social*; 25.35% as *Rating*, 18.35% as *Experience*, 8.67% as *Judgment*, and 4.90% as *Comment*. Such distribution is consistent with the result of manual annotation. Results reveal that the amount of crowdworkers' social-related activities is comparable to their work-related activities (Rating, Experience, etc.). This result quantitatively supports the outcome of previous work (Laplante and Silberman 2016) and highlights the dual nature of online crowdworker communities, where both social and technical needs are addressed. Notably, the *Judgment* message type has a relatively low frequency compared to *Rating*, suggesting the presence of norms (i.e. `Turkopticon` ratings) related to the expression of opinions about requesters and HIT groups. An analysis of the linguistic properties (e.g. sentiment) of such judgment is an interesting topic for future work.

Influence of the Market on Fora Discussions

This section addresses **RQ1**, and investigates how the content and the dynamics of mTurk influence discussions in fora. We hypothesize that crowdworker discussions are influenced by **1)** properties of published HIT groups; **2)** temporal variations in the mTurk market demand; and **3)** properties of HIT groups' requesters.The investigation resulted in a number of insights, highlighted below with "ⓧ".

Which Properties of HIT Groups are Relevant for Crowdworkers' Fora Discussions?

We analyze five properties of a HIT group: 1) *Group Size*, i.e. the amount of HITs available at publication time; 2) *Reward*, i.e. the amount of monetary compensation associated with a successful execution of a task; 3) *Time Allotted* for task execution, as specified by the requesters; 4) *Requirement*, a Boolean variable[4] that encodes the specification of as approval rate threshold for the worker to be allowed to execute the HIT; and 5) *Task Type*, defined according to the taxonomy proposed in (Gadiraju, Kawase, and Dietze 2014). The task types of considered HIT groups is inferred with the classifier developed in (Difallah et al. 2015).

Table 3 reports descriptive statistics for HIT groups that are mentioned (respectively, unmentioned) by community members of different fora. The analysis provides a number of non-trivial insights, showing both the influence of

[4]A value of 1 encodes "Qualification or approval rate greater than x needed".

		Group Size	Reward (cents)	Time (minutes)	Requirement	Task Type %SU	%CC	%CA	%IA	%VV	%IF	%OT
`mTurkForum`	M	450.31 ± 3895.49, 1	90.62 ± 275.54, 50	70.05 ± 153.93, 60	0.55 ± 0.50, 1	68.77	15.08	0.12	7.85	1.08	6.62	0.50
	UM	34.79 ± 419.63, 1	378.31 ± 792.26, 52	131.22 ± 206.12, 60	0.14 ± 0.35, 0	19.66	55.37	1.24	4.69	15.68	3.27	0.08
`mTurkGrind`	M	493.39 ± 3422.52, 1	66.65 ± 344.92, 20	72.95 ± 177.07, 45	0.55 ± 0.50, 1	47.54	18.48	0.51	17.97	4.31	10.37	0.82
	UM	30.43 ± 370.04, 1	381.19 ± 793.75, 55	131.64 ± 205.98, 60	0.14 ± 0.34, 0	19.78	55.61	1.24	4.42	15.72	3.16	0.07
`TurkerNation`	M	541.55 ± 3563.56, 1	91.88 ± 285.69, 50	77.26 ± 174.98, 60	0.61 ± 0.49, 1	63.87	15.14	0.11	11.27	1.22	7.96	0.44
	UM	30.14 ± 366.03, 1	380.09 ± 794.05, 52	131.44 ± 206.02, 60	0.13 ± 0.34, 0	19.46	55.62	1.25	4.59	15.77	3.22	0.08
`mTurkCrowd`	M	367.00 ± 2599.24, 1	74.95 ± 210.28, 40	91.97 ± 212.11, 60	0.59 ± 0.50, 1	62.98	17.09	0.79	9.6	1.31	7.76	0.47
	UM	25.60 ± 320.21, 1	388.21 ± 802.23, 56	132.10 ± 205.04, 60	0.12 ± 0.33, 0	18.38	56.55	1.24	4.51	16.13	3.11	0.07
`Reddit HWTF`	M	123.13 ± 645.40, 1	46.51 ± 65.62, 32	55.75 ± 83.50, 60	0.65 ± 0.48, 1	83.30	5.03	0.24	5.72	0.46	5.03	0.22
	UM	40.43 ± 643.03, 1	377.28 ± 791.02, 53	131.05 ± 206.31, 60	0.14 ± 0.35, 0	19.76	55.27	1.23	4.73	15.61	3.31	0.09

Table 3: Descriptive statistics – mean (μ) ± standard deviation (σ), and median (m) – of metadata, and task type distribution for *mentioned* (**M**) and *unmentioned* (**UM**) HIT groups. Note that group size, reward, and time follow long-tail distribution, thus showing large standard deviation. To account for such a phenomenon, we use non-parametric hypothesis testing methods in our analysis. **Task Type**s: SU – Survey; CC – Content Creation; CA – Content Access; IA – Interpretation and Analysis; VV – Verification and Validation; IF – Information Finding; OT – Other types. Differences within fora are statistically significant (Mann-Whitney test, p-value < .001) for all the analyzed properties.

the market on fora, and the heterogeneous nature of fora communities. ① *Task type popularity in mentioned HITs significantly differs from the distribution of task availability in the market*. For instance, tasks of type Survey (SU) are the most mentioned in all fora, while previous work (Difallah et al. 2015) reports that Content Creation (CC) is the most available task type. ② *Task type popularity differs across fora*. For instance, Interpretation and Analysis (IA) and Verification and Validation (VV) task types are more popular in `mTurkGrind`, while `Reddit HWTF` emerges as the most polarized towards *Survey* tasks. ③ *Properties of Mentioned HIT groups differ across fora*. Differences in the distributions of Group Size properties are statistically significant across all fora (Mann-Whitney test, p-value < .001); differences of Reward values are significant between `mTurkGrind` and all fora (Mann-Whitney test, p-value < .001), `mTurkForum` and `Reddit HWFT` (Mann-Whitney test, p-value < .001), and between `mTurkCrowd` and `Reddit HWFT` (Mann-Whitney test, p-value < .001). ④ *Unmentioned HIT groups are similar across fora*. The properties of unmentioned tasks show no statistically significant difference across fora (Mann-Whitney test, p-value > .001). The result gives more value to the differences that emerge when analyzing mentioned HIT groups. ⑤ *HIT groups are more likely to be mentioned when having a large size, short time allotted, requirements for execution, and lower reward*. The result suggests that crowdworkers are more likely to discuss HIT groups if there is an opportunity for large amounts of work to be performed, or if there are limitations in their ability to execute tasks. In contrast, issues about rewards appear not to be relevant.

How Are Discussions in Fora Influenced by Task Availability in the Market?

In this section, we investigate the relationship between the dynamic properties of the mTurk market, and the discussion by crowdworker communities. We compare the temporal distribution of the amount of HIT groups *available* in the market, with the temporal distribution of the amount of HIT group mentions in fora.

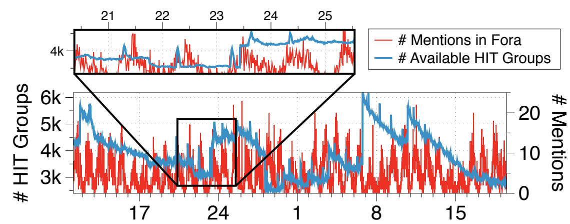


Figure 1: Time series of #available HIT groups in mTurk and #mention by crowdworker communities in Apr.-May 2016.

Analysis of Aggregated Fora Activities. We hypothesize the presence of a quantifiable relationship between the dynamics of work demand in the market and work-related discussions in crowdworker communities.

To test the hypothesis, we first analyze the temporal distribution of mentions across all fora, to study the relationship between the availability of HIT groups in mTurk and the whole set of online crowdworker communities. The analysis includes market data related to time intervals where all fora were active – i.e. three months, March 2016 – May 2016. Figure 1 shows an example of the two time series. As in previous work (Difallah et al. 2015), we observe a weekly periodicity for the market work demand (HIT groups availability). A similar periodicity is observed in the temporal distribution of mentions, with higher volumes of messages posted during weekdays. We find evidence of daily periodicity in both HIT groups availability and number of mentions, with a peak in the early morning (PST time). Compared with the curve of available HIT groups, the temporal evolution of discussions (mentions in fora) shows smaller variations across days. This finding suggests ⑥ the presence of an *upper bound* in the amount of HIT groups that can be discussed (or that are worthy of discussion) that is only partially dependent on the current market demand.

Co-locating the two time series on the time axis, we observe that the peak time of the #mentions distribution closely

Forum	F-Statistics	*p*-value	**Opt. Lag** (Minutes)
`Social`	3.0664	0.0001	225
`Rating`	2.8737	0.0002	225
`Experience`	2.7180	0.0003	240
`Judgment`	1.5173	*0.0510*	360
`Comment`	3.6913	0.0006	135

Table 4: Synchronization between the HITs availability in the market and HITs mentions across message types.

Forum	F-Statistics	*p*-value	**Opt. Lag** (Minuts)
`mTurkForum`	5.8429	0.0006	45
`mTurkGrind`	1.5599	*0.0663*	255
`TurkerNation`	1.8398	0.0166	270
`mTurkCrowd`	2.7972	0.0002	225
`Reddit HWTF`	2.3129	0.0080	225

Table 5: Synchronization between the HITs availability in the market and HITs mentions across different fora.

relates with the amount of HIT groups available in the market, with a delay in the range of 1 to 6 hours. We therefore test for Granger Causality (Granger 1969) between the two time series. Granger Causality is a technique for determining whether one time series is significant in forecasting another (Eichler 2012); specifically, it measures a statistical dependence between the past of a process and the present of another. It is a statistical test widely used in fields such as econometrics, data mining (Arnold, Liu, and Abe 2007) and machine learning (Peters, Janzing, and Schölkopf 2013). We stress that the Granger Causality test proves a temporal synchronicity between the two series, *but does not fully prove causality*. This is obvious, as we could not control, in our collected data, for other temporal and contextual factors possibly influencing the HIT mentions in crowdworker communities. In Granger Causality, a "lag" parameter captures the temporal delay between the two series for which better prediction is achieved. Optimal lag is usually selected by searching for the one with the lowest AIC/BIC (Akaike or Bayesian information criteria (Box et al. 2015)) within a predefined range.

Since the time series of available HIT groups shows large variations across days, we first de-trend its temporal distribution on a daily basis by applying Z-score normalization (Bollen, Mao, and Zeng 2011), such that similar variance can be obtained in different days.

The analysis provides two relevant insights. (7) *The temporal distribution of the #mentioned HITs in the fora is correlated with the de-trended distribution of the number of available HIT groups*. We find a significant statistical correlation (F-Statistics: 2.8681, p-value = .002), thus confirming the presence of a *quantifiable* relationship between work demand in mTurk and work-related discussions. (8) *The lag between task availability and discussion in fora is (on average, of 4 hours*, The highest correlation is achieved with a lag value of 4 hours. The result indicates that variations in market demand have a visible effect on the activity of crowd communities after **4 hours** (on average). Considering the deadlines of HIT groups in mTurk, typically in the order of days, the small lag indicates prompt discussions by crowdworker communities in reaction to HIT publication.

Analysis of Message Categories. We then investigate the presence of temporal correlation between market demand and fora messages that mention HIT groups with a specific message type. Results in Table 4 indicate that market demand is a significant predictor for the temporal distributions of all message types (with the exception of *Judgment*). *Comment* messages are the ones for which stronger prediction power (F-Statistics) and shorter delay (**2 hours** delay) can be observed. As *Comment* messages include information about task requirements and work-related issues, this result provides an indication of the minimum (averaged) "reaction time" that crowdworker communities can have to variations in market demand. *Social*, *Rating*, and *Experience* messages show an additional delay of 90-120 minutes. This result provides an additional insight suggesting that (9) *discussions about work execution temporally (but not quantitatively*[5]) *precede communication for other purposes*.

Analysis per Forum. Finally, we address differences in temporal correlation across the considered fora. (10) *Discussions in all fora* (except `mTurkGrind`) *are significantly correlated with market demand*. Results in Table 5 show a correlation between the amount of forum members and the temporal lag w.r.t. the market demand curve (One-tailed Mann-Whitney U Test. p-value < .01). The task availability time series has stronger prediction power on `mTurkForum`, while the lower is with `TurkerNation`. The lack of significant synchronicity for `mTurkGrind` is of interest, given the age and popularity of the forum. We hypothesize that this result is due to the difference in the distribution of preferred task types (CC, IA, and IF) HIT groups have an higher popularity than in other fora. Further investigations are left to future work.

Which Properties of Requesters Are Relevant for Crowdworkers' Fora Discussions?

We address the relationship that exists between requesters in the mTurk market, and discussions in fora. We consider as dependent variables the reputation scores assigned to requesters on `Turkopticon`, which include *Communicativity*, *Generosity*, *Fairness*, and *Promptness*.

Table 6 reports descriptive statistics for requesters that are mentioned (respectively, unmentioned) on different fora. As reputation scores are obtained from `Turkopticon`, we included in the "unmentioned" group those requesters that are part of the `Turkopticon` database, but not mentioned by the analyzed forum. The analysis provides two relevant insights. (11) *Requesters that are more communicative and generous are consistently preferred in all fora*. The difference in terms of *Communicativity* and *Generosity* between mentioned and unmentioned requesters is similar across fora (respectively ~ 0.3 and ~ 0.4 on a 5 point scale). (12) *Fairness and promptness are not differentiator properties of requesters*. *Fairness*

[5]The type distribution of messages linked to HIT groups is heavily skewed: *Rating*: 97.14%; *Social*: 63.64%; *Experience*: 57.15%; *Judgment*: 45.63%; *Comment*: 3.34%.

		Communicativity	Generosity	Fairness	Promptness
mTurkForum	**M**	3.35±1.41,3.50*	3.35±1.00,3.38*	4.42±0.87,4.85	4.40±0.80,4.71
	UM	3.19±1.50,3.20*	3.03±1.34,3.00*	4.24±1.20,5.00	4.21±1.14,4.81
mTurkGrind	**M**	3.37±1.37,3.50*	3.45±0.96,3.46*	4.45±0.80,4.85	4.41±0.76,4.69
	UM	3.19±1.51,3.20*	2.99±1.32,3.00*	4.23±1.20,5.00	4.22±1.12,4.80
TurkerNation	**M**	3.39±1.39,3.57*	3.44±0.96,3.46*	4.48±0.80,4.89	4.41±0.79,4.71
	UM	3.17±1.50,3.15*	3.01±1.31,3.00*	4.21±1.19,5.00	4.22±1.10,4.77
mTurkCrowd	**M**	3.40±1.40,3.62*	3.42±1.05,3.50*	4.47±0.87,5.00*	4.43±0.83,4.76*
	UM	3.12±1.49,3.00*	2.95±1.28,3.00*	4.18±1.18,4.86*	4.17±1.10,4.67*
Reddit HWTF	**M**	3.58±1.36,3.86*	3.81±0.80,3.86	4.64±0.64,4.91	4.60±0.59,4.81*
	UM	3.25±1.45,3.33*	3.16±1.20,3.17	4.32±1.05,4.95	4.29±0.99,4.75*
Turkopticon		3.27±1.45,3.40	3.21±1.17,3.22	4.32±1.05,4.93	4.30±0.98,4.75

Table 6: Mean (μ) ± std, deviation (σ), and median (m) – of reputation scores for *mentioned* (**M**) and *unmentioned* (**UM**) requesters. * marks properties with significant difference within forum (Mann-Whitney test, p-value < .001).

values are generally high for both mentioned and unmentioned requesters. This result strides with our previous findings, where we observed workers favoring tasks with lower reward but comparable execution time. Considering that previous results highlight the importance of intrinsic task properties such as complexity and clarity (Yang et al. 2016; Gadiraju, Yang, and Bozzon 2017), we hypothesize that the value of *Fairness* may be affected by such properties, that are not easily observable from HIT groups metadata. The investigation of this hypothesis is left to future work.

Influence of Fora Discussions on the Market

This section addresses **RQ2**, and investigates the presence and effect of a quantifiable relationship between discussions about HIT groups and their *consumption speed* (throughput)[6] in mTurk. We study differences in average throughput for HIT groups mentioned and not mentioned in fora. Then, we seek evidence of temporal correlation by using the same time series analysis technique introduced in the previous section. We consider the progressions of throughput at individual HIT group level, and compare them with the respective temporal distribution of their *mentions* in fora.

Analysis of Throughput Differences for Mentioned HIT Groups. To perform the analysis, we need to identify HIT groups featuring enough data points describing both the fora *mention* and the task *consumption speed* time series. The dataset developed in (Difallah et al. 2015) contains consumption data for 149K HIT groups; the addition of data from the 46K HIT groups collected between May and March 2016 yields a total of 195, 332 HIT candidate groups. From this set, 26, 204 groups are linked to *mentions* in fora. The task type distribution of these 26, 204 HIT groups is however different from the set of HIT groups (184K) having at least one mention. To improve the generalizability of results, we applied stratified random sampling (with strata corresponding to the distribution of task types in the original dataset), to obtain an analysis dataset having comparable distribution. The result is a set of 19, 122 HIT groups.[7]

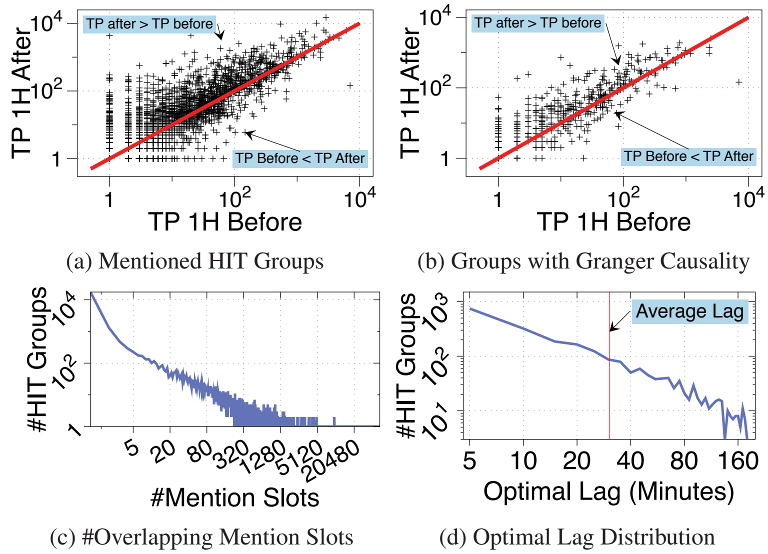
The analysis yields the following insight. (13) *Mentioned and unmentioned HIT groups have different throughput*. We found a statistically significant difference (Mann-Whitney test, p-value < .001) between the average hourly throughput of mentioned HIT groups (μ = 27.35, σ = 286.14, m = 0.09) and the average hourly throughput of unmentioned groups (μ = 19.53, σ = 137.95, m = 1). The result suggests the presence of an acceleration effect due, at least in part, to mentions in fora. To better characterize this acceleration effect, we compare the HIT group consumption speed *one hour before* and *one hour after* their earliest mention in fora. Figure 2a shows that the consumption of the majority of tasks is boosted (on average) by 59.26%.

Analysis of Temporal Correlation. We apply Granger Causality analysis to investigate the presence of a temporal relation between HIT group mentions and throughput variations. We discretize the *mention* time series into 5 minutes slots, to align it with the sampling rate of HIT groups consumption. Figure 2c shows the heavy-tailed distribution of the number of time slots that feature an overlap between the *consumption* and *mention* time series of a HIT group. We consider tasks with ≥ 5 overlapping slots. The resulting 4, 539 HIT groups feature an average hourly throughput of 22.10 (σ = 269.80, m = 0.06) and a comparable task type distribution (54.2% SU, 19.2% CC, 0.5% CA, 15% IA, 2.1% VV, 8.6% IF). The analysis provides two insights.

(14) *For low-reward and low-throughput HIT groups, the temporal distribution of their mentions in fora is related to the temporal distribution of their execution*. 1, 541 HIT groups (33.95% of the considered set) show significant Granger Causality (p-value < .05) between the temporal progression of their mentions in fora, and their consumption. When compared to the overall population of HIT groups mentioned in fora, these 1, 5K groups have lower averaged hourly throughput (μ = 12.6, σ = 115.23, m = 0.10), a

[6]The amount of HITs in a group that gets completed between two successive observations (typically, every 1 hour).

[7]Task types distribution: 59% SU, 16% CC, 0.6% CA, 11% IA, 2% VV, 8% IF, 0.3% Other.



(a) Mentioned HIT Groups (b) Groups with Granger Causality

(c) #Overlapping Mention Slots (d) Optimal Lag Distribution

Figure 2: Upper figures are hourly throughput (TP) – 1 hour before and 1 hour after the first mention in fora – of (a) HIT groups and (b) HIT groups with Granger Causality; lower figures are (c) distribution of #overlapping mention slots across HIT groups and (d) distribution of optimal lags in Granger-Causal HIT groups.

slightly different distribution in terms of task types (51.2% SU, 20.7% CC, 0.4% CA, 16.1% IA, 2.4% VV, 8.9% IF), lower reward (μ = 44.65, σ = 130.50, m = 15), higher allotted time (μ = 225.07, σ = 4591.08, m = 45), and comparable group size. (15) *Discussions in fora can have a quick and significant effect on the market*. The strongest causality is found when the average lag is set to 30.40 minutes (σ = 36.62 minutes). Figure 2d shows the log-log distribution of the optimal lag across HIT groups: the majority of groups achieve higher causality for lags *lower than **15** minutes*. The result shows that crowdworker discussions can have a quick effect on the market. To quantify such effect, we compare the HIT groups consumption *one hour before* and *one hour after* their earliest mention in fora. Figure 2b shows that the consumption of the majority of tasks is boosted, on average, by **3.4** factor (340%) after fora discussions.

Discussion

This section discusses the insights reported in the previous sections, their implications, and threats to their validity.

RQ1. There is a discrepancy between market demand (in terms of task types) and the mentions of HIT groups in fora. *Survey* tasks are in general the most mentioned, but there are forum-specific preferences for other task types (e.g. IA and VV tasks in `mTurkGrind`).

Crowdworkers are more likely to discuss HIT groups if there is an opportunity for large amounts of work to be performed, regardless of unfair hourly reward. We argue these results to be a clear indication of the dominant role that the market has on the task selection strategy of communities, where the need for guaranteed income prevails over issues of fair payment – alas an accepted (yet unpleasant) norm. Attitude towards (un)fair payment seems to greatly vary across fora. For instance, tasks mentioned by members of `mTurkGrind` are up to 38% less rewarding than the tasks preferred by other communities. Using time series analysis techniques, we found significant synchronicity between the temporal evolution of available HIT groups in mTurk, and mentions in fora. Peaks of market demand correspond to peaks in crowdworkers' discussion activity, with an average delay of 4 hours. There are significant differences in terms of prediction strength and average lag across fora and discussion categories. `mTurkForum` emerges as the forum where activities can be predicted by market variation with a minimum time delay (45 minutes); *Comment* messages (i.e. messages with comments about HIT groups) feature the smaller time lag. In terms of requesters' properties, we found *Communicativity* and *Generosity* to be the only properties commonly valued across fora. Crowdworkers in `mTurkCrowd` also favour *Fairness* and *Promptness*.

RQ2. We provide a quantification of the effect that activities in fora can have in terms of HIT groups consumption. We measured a statistically significant positive (μ = 42%) difference in the throughout of HIT groups mentioned in fora, and an average 59.26% increase of throughput in the first hour after the first mention. Time series analysis revealed that, for HIT groups featuring temporal synchronicity with mentions in fora, the average lag is of 30 minutes, and a 340% average throughput increase in the first hour after the first mention in fora. It is worth noting that these HITs are characterized by lower reward and higher allotted time: as such, they may not be appealing to workers in the market at a first sight. By being referenced in fora, the popularity of these HITs greatly increases, as our analysis proves. This also suggests that crowd fora, as a whole, can count on more resources to scout and select HITs worth completing.

Implications of the Results. We provide quantitative evidence of the impact that crowdworkers community operating in fora can have on the performance of tasks (and requesters) in microwork markets. We hope that such awareness could help a shift toward more transparent, possibly self-governed, marketplaces (Gaikwad et al. 2015), where the rights and duties of all involved stakeholders are explicit and accounted for. In this respect, microwork markets should consider the possibility of embedding social interaction capabilities in their platform, to support workers and requesters interaction, but also to engage worker fora with suggestions for HIT groups needing attention and to highlight requesters with worker preferences in terms of requester properties and task types. Our findings are also of importance for researchers and practitioners in the field addressing issues such as task assignment and optimization techniques for microwork campaigns: although task-related factors such as batch size and rewards are dominant for crowdworkers, there is evidence that confounding factor (e.g. task complexity and task clarity) can have an effect on the performance of microwork campaigns in terms of execution speed and workers composition; this work provides an additional evidence, by highlighting the impact that worker communities might provide.

Threats to Validity. Members of the selected fora might not be representative of the wider population of mTurk crowdworkers. This risk is mitigated by the popularity of fora among crowdworkers. Previous work shows that a relevant amount of workers (up to 60% in sampled population) are active in at least one fora (Yin et al. 2016). However, we consider three additional biases due to: 1) the omission from the study of other mTurk fora;[8] 2) the homogeneity of crowdworkers' country of origin;[9] and 3) the overrepresentation of crowdworkers sharing specific social and economic needs (e.g. the need for a guaranteed income). While we can't exclude the presence of these biases, we must acknowledge the importance of the investigated communities, as they represent a considerable share of the mTurk workforce.

The task and requester linkage procedure also represents a validity threat. To maximize precision, we only rely on explicit links in messages, thus failing to consider indirect references (e.g. members referring to requesters only by name). During the training set annotation activity more than 13K messages were analyzed, but only a minority of messages (approximately 5%) referring to HIT groups or requesters did not include a link. We therefore believe this limitation to have a negligible impact on the validity of our results.

Finally, in the study of RQ2, we believe the percentage of matching HIT groups, and the percentage of Granger-causal groups, to be explained by crawling problems in the 2.6M HIT groups dataset studied in (Difallah et al. 2015), and by the relatively low crawling frequency (5 minutes). We acknowledge such limitation, but we believe that it does not affect the importance of the results herein presented, but leaves space for future work.

[8]Authors are aware of less popular fora (e.g. "CloudMeBaby"), but decided to scope the analysis to the most popular ones.

[9]Crowdworkers in fora are mainly from US (Martin et al. 2014).

Conclusion

Crowdworker fora are a relevant part of the microwork ecosystem. In this work, we hypothesized that the activities in mTurk fora can influence – and can be influenced by – properties of the mTurk market and its actors. Based on a rich dataset linking 3.1M messages in online fora with mTurk 2.6M HITs groups, we found quantitative evidence of relevant relationships in both directions. Our findings are meaningful for a variety of aspect related to microtask crowdwork, e.g. the design of tasks, incentive and task allocation schemes, and novel microwork systems.

Future research might address intrinsic properties of HITs (e.g. clarity and complexity (Yang et al. 2016)) to provide a deeper understanding of the factors that drive task selection in fora. Similarly, the properties of both requesters (e.g. role in the market, type of submitted HITs) and crowdworkers (e.g. reputation) could be studied to better characterize their role in both fora and markets. Further research could also compare the properties of crowdworkers communities with other online communities that use the Web as a medium for collaboration (e.g. software engineering).

Acknowledgments

We would like to thank the crowdworkers who participate in fora discussions that facilitate our study. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. J.Y. was supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement 683253/GraphInt).

References

Arnold, A.; Liu, Y.; and Abe, N. 2007. Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 66–75. ACM.

Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1 – 8.

Box, G. E.; Jenkins, G. M.; Reinsel, G. C.; and Ljung, G. M. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Bozzon, A.; Brambilla, M.; Ceri, S.; Silvestri, M.; and Vesci, G. 2013. Choosing the right crowd: expert finding in social networks. In *Proceedings of the 16th International Conference on Extending Database Technology*, 637–648. ACM.

Difallah, D. E.; Catasta, M.; Demartini, G.; and Cudré-Mauroux, P. 2014. Scaling-up the crowd: Micro-task pricing schemes for worker retention and latency improvement. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

Difallah, D. E.; Catasta, M.; Demartini, G.; Ipeirotis, P. G.; and Cudré-Mauroux, P. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th International Conference on World Wide Web*, 238–247. International World Wide Web Conferences Steering Committee.

Difallah, D. E.; Demartini, G.; and Cudré-Mauroux, P. 2013. Pick-a-crowd: tell me what you like, and i'll tell you what to do. In *Proceedings of the 22nd International Conference on World Wide Web*, 367–374. International World Wide Web Conferences Steering Committee.

Difallah, D.; Filatova, E.; and Ipeirotis, P. 2018. Demographics and dynamics of mechanical turk workers. In *Proceedings of the 18th ACM International Conference on Web Search and Data Mining*, 135–143. ACM.

Eichler, M. 2012. Causal inference in time series analysis. *Causality: Statistical Perspectives and Applications* 327–354.

Gadiraju, U.; Kawase, R.; and Dietze, S. 2014. A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 218–223. ACM.

Gadiraju, U.; Yang, J.; and Bozzon, A. 2017. Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, 5–14. ACM.

Gaikwad, S. N.; Morina, D.; Nistala, R.; Agarwal, M.; Cossette, A.; Bhanu, R.; Savage, S.; Narwal, V.; Rajpal, K.; Regino, J.; et al. 2015. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 101–102. ACM.

Gao, Y., and Parameswaran, A. 2014. Finish them!: Pricing algorithms for human computation. *Proceedings of the VLDB Endowment* 7(14):1965–1976.

Granger, C. W. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* 424–438.

Gray, M. L.; Suri, S.; Ali, S. S.; and Kulkarni, D. 2016. The crowd is a collaborative network. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, 34–147. ACM.

Ipeirotis, P. G. 2010. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students* 17(2):16–21.

Irani, L. C., and Silberman, M. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–620. ACM.

Laplante, R., and Silberman, M. S. 2016. Building trust in crowd worker forums: Worker ownership, governance, and work outcomes. In *Weaving Relations of Trust in Crowd Work: Transparency and Reputation across Platforms. Workshop at ACM WebSci*.

Martin, D.; Hanrahan, B. V.; O'Neill, J.; and Gupta, N. 2014. Being a turker. In *Proceedings of the 17th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, 224–235. ACM.

McInnis, B.; Cosley, D.; Nam, C.; and Leshed, G. 2016. Taking a hit: Designing around rejection, mistrust, risk, and workers' experiences in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2271–2282. ACM.

Peters, J.; Janzing, D.; and Schölkopf, B. 2013. Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems*, 154–162.

Ross, J.; Irani, L.; Silberman, M. S.; Zaldivar, A.; and Tomlinson, B. 2010. Who are the crowdworkers?: Shifting demographics in mechanical turk. In *SIGCHI Extended Abstracts on Human Factors in Computing Systems*, 2863–2872. ACM.

Silberman, M.; Irani, L.; and Ross, J. 2010. Ethics and tactics of professional crowdwork. 17(2):39–43.

Spencer, D. 2009. *Card sorting: Designing usable categories*. Rosenfeld Media.

Wang, X.; Zhu, H.; Li, Y.; Cui, Y.; and Konstan, J. 2017. A community rather than a union: Understanding self-organization phenomenon on mturk and how it impacts turkers and requesters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2210–2216. ACM.

Yang, J., and Bozzon, A. 2016. On the improvement of quality and reliability of trust cues in micro-task crowdsourcing. In *Weaving Relations of Trust in Crowd Work: Transparency and Reputation across Platforms. Workshop at ACM WebSci*.

Yang, J.; Redi, J.; Demartini, G.; and Bozzon, A. 2016. Modeling task complexity in crowdsourcing. In *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing*, 249–258. AAAI.

Yin, M.; Gray, M. L.; Suri, S.; and Vaughan, J. W. 2016. The communication network within the crowd. In *Proceedings of the 25th International Conference on World Wide Web*, 1293–1303. International World Wide Web Conferences Steering Committee.