

OPERATING AN EMPLOYER REPUTATION SYSTEM: LESSONS FROM TURKOPTICON, 2008-2015

M. Six Silberman[†] and Lilly Irani^{††}

I. INTRODUCTION

In November 2005, Amazon launched Mechanical Turk (“AMT”), a website where “requesters” can post tasks, called “Human Intelligence Tasks” or “HITs,” for workers to complete for pay.¹ Workers are required to agree that they are independent contractors, not employees, and that they are therefore not entitled to minimum wage or other employment benefits.² Requesters post tasks to the platform, workers choose and do tasks, and requesters then review and “approve” or “reject” the submitted work. Workers are paid for approved work; they are not paid for rejected work. Requesters can reject (i.e., decline to pay for) work for any reason.

AMT’s “application programming interface” (“API”) allows requesters to post and review (i.e., approve or reject) tasks automatically by writing software.³ Requesters can write software to manage complex workflows through the API. For example, a photo tagging task might be posted twice for two workers to complete. If the two workers produce the same answers, the requester’s software can pay both workers. If they produce different answers, the software can post the task a third time, perhaps at a higher price for a worker with a previously earned qualification. In this workflow, the workers in the “majority” are paid; the “dissenter” is assumed to be incorrect and is not paid. A vast variety of variations on this workflow are possible. Researchers in fields such as human-computer interaction and science and technology studies (including us) have called this management by software

[†] Project Secretary at IG Metall.

^{††} Assistant Professor in the Department of Communication at University of California, San Diego. This Article represents the views of the authors, not their organizations.

1. *Mechanical Turk Is a Marketplace for Work*, AMAZON MECHANICAL TURK, <http://mturk.com> (last visited May 5, 2016).

2. *Mechanical Turk Participation Agreement*, AMAZON MECHANICAL TURK, § 3, <http://mturk.com/mturk/conditionsofuse> (last modified Dec. 2, 2014).

3. *Amazon Mechanical Turk API Reference, API Version 2014-08-15*, AMAZON, <http://docs.aws.amazon.com/AWSMechTurk/latest/AWSMturkAPI/Welcome.html> (last visited May 4, 2016).

variously “automatic management,”⁴ “algorithmic management,”⁵ “algorithmic authority,”⁶ and “algorithmic governance.”⁷

In 2011, Amazon reported that AMT hosted “more than 500,000 workers from 190 countries,”⁸ but researchers and requesters estimate that AMT hosts about 50,000 “active workers,”⁹ 1,000 to 10,000 “full-time equivalents,”¹⁰ or about 7,300 “reachable” workers.¹¹ Workers can conceivably work from anywhere in the world, as long they understand the task language (usually English) and have a reliable internet connection. In practice, however, most workers are in the United States or India¹²—presumably at least partly because workers can receive payment only in dollars, rupees, or Amazon gift card points.¹³ As of 2015, about three-fourths of the active workers appear to be based in the United States.¹⁴ Some of these workers have limited access to “traditional” jobs; researchers report that for these workers, AMT acts as a sort of “safety net.”¹⁵

In 2008, in response to reports from workers describing conditions of low pay, slow pay, poor communication, and arbitrary rejections (i.e., nonpayment),¹⁶ we designed Turkopticon,¹⁷ a website and browser extension that workers can use to review requesters, mainly along criteria of pay, pay speed, fairness of evaluation, and communication. As of January 2016, 56,000 users have created accounts on the Turkopticon website and about

4. See, e.g., Lilly Irani, *The Cultural Work of Microwork*, NEW MEDIA & SOC'Y (Nov. 21, 2013).

5. See, e.g., Min Kyung Lee et al., *Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers*, PROC. 33rd ASSOCIATION FOR COMPUTING MACHINERY CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1612 (2015).

6. See, e.g., Caitlin Lustig & Bonnie Nardi, *Algorithmic Authority: The Case of Bitcoin*, PROC. 48TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES 743 (2015).

7. See, e.g., Claudia Müller-Birn, Leonhard Dobusch & James D. Herbsleb, *Work-to-Rule: The Emergence of Algorithmic Governance in Wikipedia*, PROC. 3RD CONFERENCE ON COMMUNITIES AND TECHNOLOGIES 80 (2013); Francesca Musiani, *Governance by Algorithms*, INTERNET POL'Y REV. 2 (2013); GOVERNING ALGORITHMS: A CONFERENCE ON COMPUTATION, AUTOMATION, AND CONTROL (New York University, May 16-17, 2013).

8. *Amazon Mechanical Turk Service Summary*, AMAZON MECHANICAL TURK, <https://requester.mturk.com/tour> (last visited May 5, 2016).

9. *50,000 Worldwide Mechanical Turk Workers*, TECHLIST (2015), <http://techlist.com/mturk/global-mturk-worker-map.php>.

10. Panos Ipeirotis, *Answer to How Many Active Workers Are There in Mechanical Turk?*, QUORA (May 2, 2013), <https://www.quora.com/How-many-active-workers-are-there-in-Mechanical-Turk>.

11. Gabriele Paolacci, *How Many People Can Your Lab Reach on Mechanical Turk?*, EXPERIMENTAL TURK (July 30, 2015), <https://experimentalturk.wordpress.com/2015/07/30/how-many-people-can-your-lab-reach-on-mturk>.

12. See, e.g., Panos Ipeirotis, *Demographics of Mechanical Turk: Now Live! (April 2015 edition)*, COMPUTER SCIENTIST IN A BUSINESS SCHOOL (Apr. 6, 2015), <http://www.behind-the-enemy-lines.com/2015/04/demographics-of-mechanical-turk-now.html>.

13. *Amazon Mechanical Turk Worker Web Site FAQs*, AMAZON MECHANICAL TURK, <https://www.mturk.com/mturk/help?helpPage=worker> (last visited May 4, 2016).

14. See Ipeirotis, *supra* note 12.

15. David Martin et al., *Being a Turker*, 224 (Proc. 17th Association for Computing Machinery Conference on Computer Supported Cooperative Work, 2014).

16. TURK WORK (Lilly Irani ed., 2008), <http://turkwork.differenceengines.com>.

17. TURKOPTICON, <http://turkopticon.ucsd.edu> (last visited May 4 2016).

35,000 use one of the two browser extensions.¹⁸ Since early 2009, these users have posted 290,000 reviews of 42,000 requesters. As on AMT itself, only a small fraction of registered users is “active”: in an average month, about 1,000 workers post about 5,000 reviews; in the period between December 16, 2015 and January 17, 2016, for example, 1,205 reviewers posted 5,205 reviews. Even within these “active” users, participation is very unequally distributed: within any given month, most users who have posted any reviews have posted exactly one, but a dozen or so post more than 50, and one or two post more than 100.¹⁹

To our knowledge, most “professional” AMT workers use Turkopticon. And in 2014, in an ethically fraught but instructive experiment,²⁰ a group of economists found that effective wages among requesters with “good” reputations on Turkopticon were about 40% higher than effective wages among requesters with “neutral” or “bad” reputations—and that requesters with good reputations attracted workers to their tasks at nearly twice the rate as requesters with bad reputations.²¹

Despite these apparent successes, and generally favorable portrayals in the media,²² Turkopticon has serious problems that threaten its long-term usefulness to workers. Workers often disagree about how to review “properly”; these disagreements can become heated and even vicious, destroying trust and goodwill and draining participants emotionally and mentally. Turkopticon is also occasionally a site of the harassment, insults, sexism, racism, profanity, baseless accusations, and occasional threats that trouble other online communities (and indeed offline worker communities and organizations). We have not developed robust processes for mediating these disagreements or moderating the harassments and incivility; indeed, our attempts to do so have generally produced further complications.

This Article describes AMT and Turkopticon, and attempts to draw lessons from our experiences with Turkopticon for broader efforts to develop broad-based worker power in the “on-demand economy.” Eight lessons are

18. Site registration figures are our own data. Browser extension user figures are from the CHROME WEB STORE, <https://chrome.google.com/webstore/detail/turkopticon/dgefbojfgdddniqnhmfimnengiloojpe> (last visited May 4, 2016), and *Add-Ons*, MOZILLA, <https://addons.mozilla.org/en-us/firefox/addon/turkopticon> (last visited May 4, 2016).

19. Figures in the previous three sentences are our own data.

20. For details on the experiment and its consequences from the perspective of Turkopticon users, see M. SIX SILBERMAN, HUMAN-CENTERED COMPUTING AND THE FUTURE OF WORK: LESSONS FROM MECHANICAL TURK AND TURKOPTICON, 2008-2015, at 92-94 (2015).

21. Alan Benson, Aaron Sojourner & Akhmed Umyarov, *Can Reputation Discipline the Gig Economy?* (Forschungsinstitut zur Zukunft der Arbeit Discussion Paper No. 9501, 2015), available at <http://ftp.iza.org/dp9501.pdf>.

22. We refrain from citing particular coverage here. For extended discussion of media coverage of Turkopticon, see Lilly Irani & M. Six Silberman, *Stories We Tell About Labor: Turkopticon and the Trouble with “Design,”* (Proc. 34th Association for Computing Machinery Conference on Human Factors in Computing Systems, 2016), available at http://wtf.tw/text/turkopticon_stories.pdf.

presented. The first four concern reputation systems specifically; the last four concern the development of worker power in the on-demand economy generally. First, reputation is important in online labor platforms; participants do avoid other participants with bad reputations, and the processes by which reputation is calculated and displayed affect market outcomes. Second, reputation in online labor platforms is not simple; “star” rating systems, even on multiple criteria, are likely to be inadequate. Third, successful reputation systems will evolve as participants and administrators learn what is important in the labor process, and as what is important itself changes. Fourth, for designers or regulators embarking on the task of adding reputation to an existing platform, the choice between designing an independent reputation system or adding a reputation system to the platform itself presents trade-offs; it is not at all clear that one or the other is always preferable. Fifth, alternative models for organizing labor platforms are imaginable that could offer strategies for overcoming some of the difficulties posed by both independent and integrated reputation systems. Our experiences suggest that these alternatives are worth exploring. Sixth, current platforms, including Turkopticon, concentrate power in the hands of a small group of operators. More democratic designs and governance strategies are imaginable, and likely worth exploring in the context of supporting the development of broad-based worker power. Seventh, the growth of the “on-demand economy” raises thus far understudied cultural questions about the casualization of service labor. Eighth and finally, robust interdisciplinary and cross-sectoral collaboration will be needed to develop broad-based worker power in the “on-demand economy.”

The Article proceeds as follows. Part II describes AMT in detail. It describes the kinds of tasks available, the kinds of requesters who post them, the process by which work is completed, common problems that arise in the course of work, and so on. It also describes the broader “ecosystems” of tools, practices, and discourses “around” it developed by both workers and requesters. Part III describes Turkopticon in detail. It describes its design and operation; outcomes and current problems; and ongoing discussions about possible future developments.²³ Finally, Part IV steps back from the technical and social details of AMT and Turkopticon to elaborate on the eight lessons summarized above.

23. Parts of the text in Parts I and II are taken from Silberman’s Ph.D. dissertation, *HUMAN-CENTERED COMPUTING AND THE FUTURE OF WORK: LESSONS FROM MECHANICAL TURK AND TURKOPTICON*, 2008-2015 (2015).

II. AMAZON MECHANICAL TURK

A. *Origin Story*

While AMT's origins were specific to Amazon, they were symptomatic of the problems of organizing large volumes of products, images, and other culturally meaningful but computationally challenging objects. Amazon representatives describe AMT as a system originally built to help "clean" data coming into Amazon's huge product catalog. Amazon was a clearinghouse for products from many different vendors, and could handle everything from payments to inventory and shipping logistics.²⁴ These varied vendors would sometimes upload entries for identical products. As a result, customers searching for products would see several search results for products that appeared identical. Amazon designers wanted to hide the duplicate entries, but the task of doing so computationally proved "insurmountable" for Amazon engineers.²⁵ Amazon also did not want to burden vendors with the task of marking other Amazon-listed products identical to that sold by the vendor. The system designers decided to displace this labor to a "crowdsourced" workforce. Amazon engineers built a site through which Amazon employees, in their spare work time, could contribute to the process of identifying and hiding the duplicate entries. This was successful, and it was eventually opened to workers and requesters outside Amazon. It was extended to support tasks other than duplicate product identification. A mechanism for paying workers was added. And the cheeky but truthful tagline "artificial artificial intelligence" was coined to describe the new service.²⁶

With these additions, AMT became a prototype for how to expand computer scientists' agencies over expanded pools and kinds of labor. It became simultaneously the next step in both artificial intelligence and cloud computing. In describing the system in a 2006 lecture at MIT (see Figure 1 below), Amazon CEO Jeff Bezos said, "You've heard of software-as-a-service. Well, this is human-as-a-service."²⁷ Inspired by AMT, competitors have sprung up to offer crowdsourced search engine optimization, sales lead

24. Birgitta Bergvall-Kåreborn & Debra Howcroft, *Crowdsourcing and Open Innovation: A Study of Amazon Mechanical Turk and Apple iOS* (Proc. 6th International Society for Professional Innovation Management Innovation Symposium, 2013).

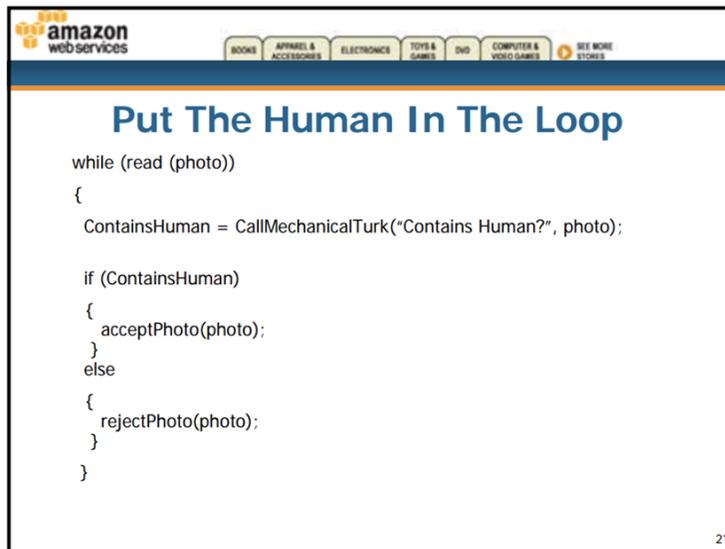
25. U.S. Patent No. 7,197,459 (Issued Mar. 27, 2007).

26. See also Jason Pontin, *Artificial Intelligence, With Help from the Humans*, N.Y. TIMES, Mar. 25, 2007, at BU5.

27. Jeff Bezos, Opening Keynote and Interview, Massachusetts Institute of Technology Emerging Technologies Conference (Sept. 27, 2006). For extended discussions of the term "human-as-a-service," see Lilly Irani & M. Six Silberman, *Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk*, at 612-14 (Proc. 31st Association for Computing Machinery Conference on Human Factors in Computing Systems, 2013), and Irani, *supra* note 4, at 5-9.

generation, and backoffice data processing for small-scale contracts. AMT prototyped an early iteration of the information processing sector of the “gig economy.”

Figure 1. A slide from Amazon CEO Jeff Bezos's 2006 keynote at MIT's Emerging Technologies conference. The slide shows how programmers who are “end users” of AMT can write code that calls on AMT to recruit humans to do tasks that could not otherwise be performed programmatically.



AMT offers a view into how speculative technological production, as well as “big data” industries, generate value through new labor processes. The work of processing cultural data, filling the gaps left by “real” artificial intelligence, is central to web industries that organize, store, and surveil large volumes of user generated text, images, and sounds, usually while searching for a profit. AMT and similar systems are central to calibrating search algorithms, offering companies a view into public social media sentiment about brands, and making sound and video searchable.²⁸ AMT also enables rapid innovation by offering firms and entrepreneurs rapid, flexible, and cheap access to a diverse and educated workforce for rapid marketing surveys and product testing;²⁹ among the most active AMT workers, nearly 58% already have a bachelor’s degree or higher.³⁰ Inexpensive and quick workforces lower the financial and time costs of experimentation and failure

28. Irani, *supra* note 4.

29. DAREN C. BRABHAM, CROWDSOURCING (2013).

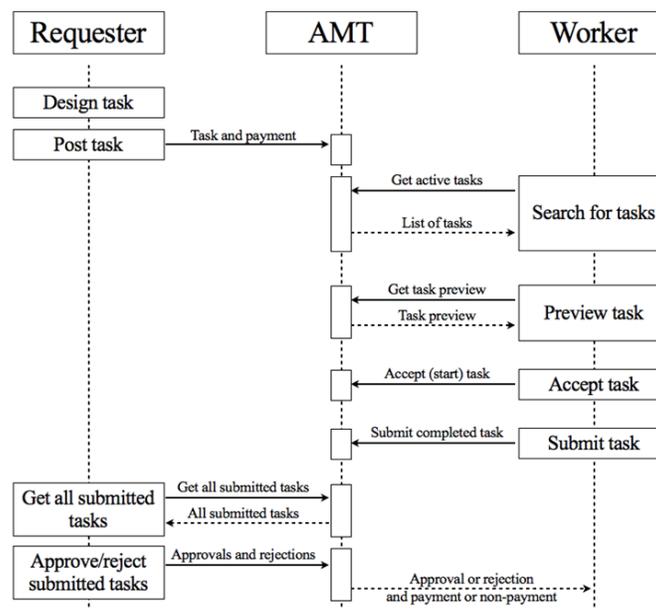
30. *Demographics of Mechanical Turk*, MTURK GRIND, <http://www.mturkgrind.com/threads/demographics-of-mechanical-turk.26341/#post-275423> (last modified Jan. 18, 2015).

for employers.³¹ This is one example of the labor conditions that enable rapid change and product speculation in innovative industries.³²

B. The Labor Process

The basic process of posting tasks to AMT and completing them has three main parts (see Figure 2): requesters post tasks to the site; workers choose tasks to do and do them; and the requester reviews—and “approves” or “rejects”—the work.

Figure 2. The Basic AMT Work Process.



A task may include visiting an assigned webpage and copying data from the page into a structured form provided by the requester. It may include transcribing small bits of audio. It may include classifying an image in accordance to an acceptable content policy. It may include clicking edges on pictures of a room to guide a robot’s computer vision algorithm. Requesters may post just a few of such tasks, or they may post hundreds of thousands. New or small-scale requesters can post these tasks using a web-based interface, but experienced requesters can automate the procurement of work and integrate it into existing code by using the AMT Application Programming Interface (“API”). Requesters include professors and graduate

31. Irani, *supra* note 4, at 730.

32. Bergvall-Kåreborn & Howcroft, *supra* note 24.

students in computer science research labs and quantitative social science departments, employees of big data startups and major corporations such as LinkedIn and Google, and intermediary companies that organize and broker crowdsourcing projects such as CastingWords and CrowdFlower.

An illustrative sample of HITs appears in Table 1; Figures 3 and 4 illustrate the AMT interface. Pay ranges from one cent to tens of dollars; most HITs pay less than \$0.50.

Table 1. Some tasks from Amazon Mechanical Turk

Title	HITs available	Reward (USD)
Find images of these real estate agents	136,725	0.04
Find the mobile app link	12,748	0.03
Type the text from the images, carefully. Productivity and bonuses guaranteed.	9,498	0.01
Judge the appropriateness of a product for a question	5,294	0.03

Figure 3. Details of a HIT from the AMT HIT List

Find Images of these Real Estate Agents [View a HIT in this group](#)

Requester: Clayton Gladieux HIT Expiration Date: Dec 31, 2014 (22 hours 50 minutes) Reward: \$0.04

Time Allotted: 7 minutes HITs Available: 133575

Description: Find the website for a given real estate agent, then find an image of that real estate agent

Keywords: real, estate, agents, images

Qualifications Required: Masters has been granted

Figure 4. Preview of a HIT

Find Images of these Real Estate Agents

Requester: Clayton Gladieux Reward: \$0.04 per HIT HITs Available: 133576 Duration: 7 minutes

Qualifications Required: Masters has been granted

Find the largest possible image of the given real estate agent, then copy the IMAGE URL and paste it into the provided field.

Please read the instructions before beginning work on this HIT:

- Do a Google search to find the real estate agent's website by searching for their name, location, and office. [Click here to search](#)
- Find the section of the website that lists the real estate agents that work at that office.
- Find the **largest possible image** for the given real estate agent next to "Real Estate Agent:" below the instructions.
- Right click the image and copy the **IMAGE URL**. Depending on the browser you are using, the right click menu item might say *copy image url*, *copy image location*, or *copy image address*. If you are using Internet Explorer, you will need to select *Properties* from the right click menu, highlight the address appearing under *Address (URL)*, and copy that URL.
- Note:** The IMAGE URL usually ends in .png, .jpeg, .jpg, .gif, or another image file extension. However, not every IMAGE URL ends this way. If the link opens the image in its own browser window, it will be approved.
- Enter the IMAGE URL in the provided text field.
- If you cannot find an image for the given real estate agent, answer question 2 to provide the reason why.

[Click here to search](#)

Office Name/Location: Intero Real Estate Svc. in Discovery Bay, CA

Want to work on this HIT? Want to see other HITs?

Requesters assign tasks with a reward for successful completion – the piece rate. Amazon charges the requester a fraction of the posted reward in addition to the reward price. Workers choose what tasks to do, then do them.

After the worker submits work for a task, the requester has full discretion to either “approve” or “reject” the task in accordance with their standards or convenience. Workers whose work is approved are paid. Workers whose work is rejected are not paid, although requesters may choose to keep and use the work in any case. Finally, requesters may choose to give some workers a bonus of any size. Amazon charges the requester a fee between 20% and 45% of the payment price, depending on the structure of the task and the worker pool the requester wants to access.³³

1. Seeing and Choosing Tasks

After a task is posted, it appears in the HIT listing, which by default lists ten HIT groups per page (see Figure 5).³⁴ Workers can sort the listings by how old the HITs are, how soon they will expire, how much they pay, how many tasks are available, how much time they allow, or (perhaps occasionally usefully) alphabetically.

Figure 5. Part of the AMT HIT List

Find Images of these Real Estate Agents		View a HIT in this group	
Requester: Clayton Gladieux	HIT Expiration Date: Dec 31, 2014 (22 hours 53 minutes)	Reward: \$0.04	
	Time Allotted: 7 minutes	HITs Available: 133580	
Inv_B_2		View a HIT in this group	
Requester: rohzit0d	HIT Expiration Date: Jan 19, 2015 (2 weeks 5 days)	Reward: \$0.00	
	Time Allotted: 48 minutes	HITs Available: 23666	
Product Categorization		View a HIT in this group	
Requester: PriceGrabber.com	HIT Expiration Date: Jan 29, 2015 (4 weeks 1 day)	Reward: \$0.03	
	Time Allotted: 20 minutes	HITs Available: 12189	
Linguistics Research: A likely (?) conclusion given a sentence		View a HIT in this group	
Requester: Textual Choreography	HIT Expiration Date: Apr 7, 2015 (13 weeks 6 days)	Reward: \$0.25	
	Time Allotted: 60 minutes	HITs Available: 7245	
Type the text from the images, carefully. Productivity and bonuses guaranteed.		View a HIT in this group	
Requester: CopyText Inc.	HIT Expiration Date: Jan 6, 2015 (6 days 20 hours)	Reward: \$0.01	
	Time Allotted: 10 minutes	HITs Available: 7026	
Geo Result Relevance-Tue Dec 23 20:16:50 PST 2014		View a HIT in this group	
Requester: Amazon Requester Inc.	HIT Expiration Date: Jan 24, 2015 (3 weeks 3 days)	Reward: \$0.00	
	Time Allotted: 60 minutes	HITs Available: 5083	

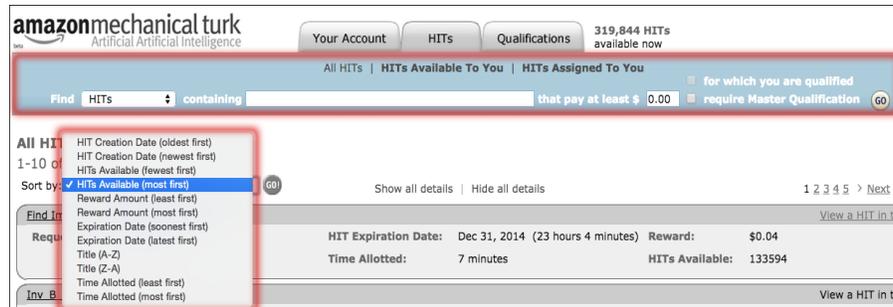
They can also search for HITs matching specific keywords, HITs that pay at least a certain amount, HITs that require the Masters qualification

33. Amazon Mechanical Turk Pricing, AMAZON MECHANICAL TURK, <https://requester.mturk.com/pricing> (last visited May 4, 2016).

34. The distinction between “HIT” and “HIT group” is subtle. Technically, a HIT is a single task; for example, manually transforming an image of a receipt into structured data by entering different values on the receipt into different fields in a form. A HIT group is a group of HITs that use an identical form and process but include different data elements; for example, a thousand receipt entry HITs with identical forms but different images (i.e., different receipts). Workers sometimes refer to large HIT groups as “batches” or “batch HITs,” but also use sentences like “I did a hundred of these [HITs]” to describe having completed many HITs in a single HIT group.

(explained below under “Discipline and Quality Control”), or HITs that are available to them (see Figure 6). (To produce results for the latter kind of search, the site checks the qualifications required by available HITs against the worker's qualifications.) When looking at a page in the listings or search results, a worker can click on the title of a HIT to see more information (specifically, the description, keywords, and required qualifications).

Figure 6. Workers Can Use Different Criteria to Sort the HIT List



They can also click a link to see the first page of the HIT itself. This page may be the entirety of the HIT, it may be a brief information page, or it may be entirely uninformative. After viewing this “preview,” the worker may choose to accept the HIT. The worker then has the allotted time to complete the HIT. While completing the HIT, the worker may choose to “return” the HIT. (Workers often do this if a HIT turns out to have technical problems, or to take longer than they expected.) Or the worker may run out of time, in which case AMT classifies the HIT as “abandoned.”

2. Discipline and Quality Control

Once the worker submits the HIT, the requester can approve or reject it. Workers are paid for approved HITs and not for rejected HITs. If the requester does neither, AMT eventually approves the HIT and pays the worker. The time after which this occurs is called by workers the “auto approve” or “AA” time of the HIT group. The default, and maximum allowed, auto approve time is 30 days; some prolific requesters set it to seven days, or even 24 or 48 hours. Workers often share this information once they know it, and attentive requesters know that workers value quick evaluation of their work. Requesters can also give workers bonuses when evaluating work (or even after they have approved or rejected work).

The option to reject work is broadly assumed to be intended to prevent workers from quickly submitting useless work in the hope of being paid anyway—and it appears to meet this aim relatively well. When rejecting work, requesters must offer some explanation for the rejection. But this is

enforced simply by disallowing requesters from leaving the explanation text field entirely blank, so sometimes they offer unhelpful “explanations,” such as “1,” “X,” or “.” The AMT participation agreement (typically referred to by workers and requesters as the “terms of service” or “TOS”) says (Sec. 3a): “Upon completion of Services [i.e., work] to Requesters’ reasonable satisfaction, Requesters must pay Providers [i.e., workers] for their Services.”³⁵

But the agreement does not define “reasonable,” nor would it be possible to do so in a manner that could be operationalized entirely algorithmically. In practice, requesters can reject work for any or no reason (beyond, for example, “X”). If a worker finds the rejection unreasonable, they can file a complaint with AMT. But AMT does not mediate in disputes between workers and requesters. In fact, the TOS explicitly disclaims any such responsibility (Sec. 3f, original emphasis):

Disputes between Requesters and Providers. Your use of the Site is at your own risk. Because Amazon Mechanical Turk is not involved in the actual transaction between Providers and Requesters, Amazon Mechanical Turk will not be involved in resolving any disputes between participants related to or arising out of the Services or any transaction.³⁶

Typically, workers filing complaints of unfair rejection with Amazon receive form responses to the same effect.

Several years after the launch of AMT, Amazon introduced an alternative quality control mechanism for requesters. Amazon declares certain workers—through mechanisms not made public—as “Masters” workers. Amazon describes these workers as an “elite” pool of workers who consistently surpass the bar set by Amazon’s “statistical monitoring.”³⁷ Panos Ipeirotis reported in late 2012 that “a current search reveal[ed] 20,744 workers” with the Masters qualification.³⁸ For requesters using the web interface, tasks are restricted by default to workers with the Masters qualification. To change this, the requester must know that this is the case and know where the option is to change the setting. The process by which a worker is given the Masters qualification is vigorously speculated about by workers—including relatively inexperienced workers who have, to their surprise, received the qualification, and experienced workers who, to their annoyance, have not.

35. *Amazon Mechanical Turk Participation Agreement*, AMAZON MECHANICAL TURK, www.mturk.com/mturk/conditionsofuse (last updated Dec. 2, 2014).

36. *Id.*

37. “What is a Mechanical Turk Master?,” *Amazon Mechanical Turk Worker Web Site FAQs*, AMAZON MECHANICAL TURK, https://www.mturk.com/mturk/help?helpPage=worker#what_is_master_worker (last visited May 4, 2016).

38. See Panos Ipeirotis, *Mechanical Turk Changing the Defaults: The Game Has Changed*, A COMPUTER SCIENTIST IN A BUSINESS SCHOOL (Dec. 5, 2012).

From the perspective of workers, both the decision to “approve” or “reject” work and (especially) the assignment of the Masters qualification are unaccountable and can seem arbitrary.

C. Requesters

AMT requesters are diverse. The most prolific historically have been intermediaries for others and have offered a variety of task types. (The two largest, CrowdFlower and CrowdSource, eventually built their own platforms and left AMT.) Many prolific requesters, however, even those who are intermediaries, specialize: Speechpad and CastingWords, for example, focus on audio transcription; p9r offers image transcription tasks; Tagasauris specializes in image metadata (“tagging”); and VidAngel, which appears to have removed AMT from its work process, used to pay Turkers to edit movies to make them family-friendly. Some requesters are special-purpose accounts for particular projects run by particular individuals or organizations. For example, social media giant LinkedIn posted business card transcription tasks through an intermediary using the requester name “Oscar Smith”; a Google speech recognition project posted tasks under the name “Project Endor”; and popular “microblogging” platform and social network Twitter used AMT to classify trending topics in real time.

While there are tens of thousands of requesters on AMT, a market power is concentrated in the small fraction of requesters who post most of the work; a 2015 study, for example, found that 10% of requesters post 98% of tasks.³⁹ As a result of this concentration of market power, the history of workers’ experiences in AMT is partly the history of specific requesters and their practices.

Workers often express deep ambivalence about their relationships with the most prolific requesters. Prolific requesters often start out beloved—for the volume of work they make available and the regularity with which they post it—but their relationships with workers often become strained as requesters grow, streamline their processes, and cut costs. As this slow process unfolds, workers may find themselves feeling trapped: a worker may want to stop doing tasks for a particular requester whose practices they feel have begun to adversely impact workers. But they may need the money and stability offered by a prolific requester. Even if the requester has lowered pay, workers familiar with the requester’s practices may continue to work for them instead of taking the time to look for other requesters who might treat them better or pay more. The “search costs” and “switching costs” associated

39. Sara C. Kingsley, Mary L. Gray & Siddharth Suri, *Accounting for Market Frictions and Power Asymmetries in Online Labor Markets*, POL'Y & INTERNET 7, 383 (2015); see Panos Ipeirotis, *Analyzing the Mechanical Turk Marketplace*, 17 XRDS 16 (2010).

with finding other reliable requesters may effectively constrain workers to persist in situations they no longer find satisfactory—a situation that mirrors some workers’ relationship with AMT generally.

The story of one particular requester, CrowdFlower, illustrates well the complexity of worker-requester relationships and the difficulty of assessing the overall impact of a particular requester on workers’ livelihoods and well-being. CrowdFlower, a San Francisco-based startup originally called Dolores Labs, no longer posts tasks to AMT but is well remembered in Turker memory as one of the most prolific, and controversial, requesters ever to post HITs.

CrowdFlower acted as an intermediary for other organizations and individuals who wanted to post large task batches to AMT but lacked the technical expertise or time to do it themselves. CrowdFlower built a platform that workers and other requesters could interact with that provided many features not offered by AMT itself. CrowdFlower’s platform managed workers and allowed requesters to manage posted and submitted work. In 2009, CrowdFlower cofounder Lukas Biewald was instrumental in collecting the initial corpus of Turkopticon reviews.⁴⁰ He was enthusiastic about the prospect of making information about requesters widely available to workers. He reasoned that such information would both pressure neglectful requesters to improve their practices and reward well-behaved ones—and in his view, at the time at least, CrowdFlower was among the better requesters.

By 2011, CrowdFlower was known among workers mainly for the diversity of its tasks (posted for a variety of paying clients) and their uniformly low pay. And in 2012, the firm became the defendant in a class action lawsuit filed by Oregon-based worker Christopher Otey. The suit alleged that although Otey and other crowd workers had been required to agree that they were independent contractors—not employees—before completing work for CrowdFlower, the degree of control the firm exerted over the work through its platform made them employees in practice, and in the eyes of the law. Because they were paid less than minimum wage but were, they alleged, employees in practice, CrowdFlower was in violation of the Fair Labor Standards Act, the U.S. federal law that regulates working time and minimum wage. The plaintiffs’ complaint alleged that Biewald had stated in public that they paid many of their workers \$2-3/hour; the complaint offered links to YouTube videos documenting these statements. (The videos have since been removed.) CrowdFlower stopped posting tasks to AMT in late 2013, focusing on other labor providers in its place. Sadly, despite the firm’s helpful involvement in Turkopticon’s early days (and despite the fact that the most recent review for CrowdFlower posted to Turkopticon was

40. *See infra* Part III.

posted in December 2013), CrowdFlower has more low ratings from Turkopticon users than any other requester. The lawsuit was settled in July 2015.⁴¹

D. Complications

In practice, the Turking process is complicated—for workers and requesters—by a wide range of unexpected outcomes, mistakes, miscommunications, and even abuses—on the part of both requesters and workers. These complications arise in part because doing remote work well through a complex computer information system is hard, even for well-intentioned participants, and in part because some market participants are not necessarily well-intentioned.

Perhaps the most well-known complication is a consequence of the rejection feature. Requesters may reject work for any reason, and workers have no technical or legal recourse within AMT against requesters who they suspect may have erroneously rejected their work—or done so maliciously, with the intent to use it anyway. Thus while illegal wage theft is common in other low-wage industries, AMT's rejection feature has effectively legalized wage theft in crowd work, as there is no way to distinguish between wage theft and legitimate and normal use of an intentionally designed platform feature. This feature gives requesters unique power over workers, and creates uncertainty among workers that some researchers have argued leads, at least in some parts of the market, to a “vicious circle” of low-quality work and low wages. For example, Panos Ipeirotis wrote in 2010:

Effectively, what Amazon Mechanical Turk is today [i.e., was in 2010] is a market for lemons A market for lemons is a market where the sellers cannot evaluate beforehand the quality of the goods that they are buying. So, if you have two types of products (say good workers and low quality workers) and cannot tell who is whom, the price that the buyer is willing to pay will be proportional to the average quality of the worker. So the offered price will be between the price of a good worker and a low quality worker. What [would] a good worker do? Given that good workers will not get enough payment for their true quality, they leave the market. This leads the buyer to lower the price even more towards the price for low quality workers. At the end, we only have low quality workers in the market (or workers willing to work for similar wages) and the offered price reflects that. This is exactly what is happening on Mechanical Turk today [i.e., was happening in 2010]. Requesters pay everyone as if they are low quality workers, assuming that extra quality assurance techniques will be required on top of Mechanical Turk.⁴²

41. *Otey v. CrowdFlower, Inc.*, 2014 WL 1477630 (N.D. Cal. 2014).

42. See Panos Ipeirotis, *Mechanical Turk, Low Wages, and the Market for Lemons*, A COMPUTER SCIENTIST IN A BUSINESS SCHOOL (July 27, 2010); Benjamin B. Bederson & Alexander J. Quinn, *Web*

While Ipeirotis later wrote that he believed subsequent changes to AMT had improved the situation,⁴³ the extent to which this is the case *for workers* is hard to discern. And while changes to AMT may have improved requesters' ability to secure quality work, Ipeirotis's 2010 observation that "there is a symmetric market for lemons on [the requester] side"—i.e., that workers struggle to identify good requesters—still appears relevant.⁴⁴ The difficulty for workers of distinguishing well-intentioned requesters from scammers—and the difficulty for requesters of distinguishing well-intentioned workers from cheaters—contributes to the need for complex quality control schemes on both sides of the market, adversarial worker-requester relations, and a climate of distrust, anxiety, and even hostility among workers, some of whom are quick to accuse others of "shilling" for requesters (or of *being* a requester) when others' accounts of their experiences are markedly different from their own.

This dynamic is augmented by other properties of the market, especially the scale of market interactions, the algorithmic management of work, the disparate expectations requesters and workers bring to their interactions, and the effect of rejection statistics on workers' ability to get work. One requester may receive work from thousands of workers in a single HIT group. Requesters may post and reviews tasks programmatically, through the API, without human oversight. Because programmatic review processes may be complex—and because they are handled by software written by humans—they are error-prone. Errors in workflows managed by software often lead to worker confusion, accusations of intentional requester malfeasance, and stress and wasted time for all parties. And when workers contact requesters to seek explanations for unexpected rejections, they are often dismayed to receive, in return, no response, a canned and irrelevant response, a slow response, or—perhaps worst of all—a dismissive response indicating that the requester is not inclined to spend time to figure out what happened to cause the worker to be rejected for a task worth a few cents. Requesters primed to expect quick and "frictionless" interactions with AMT often write software to post and review tasks, and ignore worker communications or offer cursory responses. One requester told us: "You cannot spend time exchanging email

Workers Unite! Addressing Challenges of Online Laborers 97 (Proc. 29th Association for Computing Machinery Conference on Human Factors in Computing Systems Extended Abstracts, 2011).

43. Panos Ipeirotis, Reply, *Is Mechanical Turk Really Broken?*, QUORA (Sept. 2, 2013), <http://qr.ae/B7gZn>.

44. In 2010, Ipeirotis wrote: "Scam requesters post HITs, behave badly, and cause good workers to avoid any newcomer. New requesters then get only low quality workers, get disappointed with the quality of the results[,] and . . . leave the market." Ipeirotis, *supra* note 42. The consistency in the responses to Lilly Irani's 2008 (TURK WORK, <http://turkwork.differenceengines.com> (last visited May 4, 2016)) and 2013 (TURKERS' BILL OF RIGHTS 2013 EDITION, <http://turkwork2013.differenceengines.com> (last visited May 4, 2016)) "Turkers' Bill of Rights" surveys suggests that not much had changed for workers between fall 2008 and summer 2013.

[with workers]. The time you spent looking at the email costs more than what you paid them. This has to function on autopilot as an algorithmic system . . . and integrated with your business processes.”⁴⁵

Workers, in contrast, expect—or at least wish—to be treated as “human beings, not algorithms,”⁴⁶ and to receive what they see as due consideration for their concerns from requesters. Responsiveness and communicativity are especially significant concerns for workers who rely on Turking income to meet basic needs; if a major problem occurs and a workers find themselves short payment for work they spent a significant amount of time on, can they rely on a requester to read, reply to, and act on their communications about the problem? If not, working for such a requester may pose a significant risk. These conflicting expectations, desires, and constraints lead, predictably, to frustration.⁴⁷

E. Ancillary Tools, Networks, and Practices

1. Worker Tools, Networks, and Practices

AMT is surrounded by a rich and complex “ecosystem” of technologies, communities, relationships, and practices “Turkers” use to get work done, share information, and support one another. According to Turkers, the most important parts of this ecosystem are the forums.⁴⁸ There are several worker forums, including, in no particular order, Turker Nation,⁴⁹ MTurk Crowd,⁵⁰ MTurk Grind,⁵¹ the MTurk and HITsWorthTurkingFor “subreddits”⁵² (i.e., sections on the content aggregator reddit.com), and MTurk Forum.⁵³ Each forum has different policies, politics, histories, and personalities—and differences between them sometimes erupt into inter-forum “drama.” But many experienced Turkers acknowledge the centrality of forums to earning money on AMT, especially for those “Turking for a living.” Discovering a forum appears to be a crucial turning point in the “careers” of financially

45. Irani & Silberman, *supra* note 27, at 614.

46. See, e.g., Mark Harris, *Amazon's Mechanical Turk Workers Protest: "I am a Human Being, Not an Algorithm,"* GUARDIAN, Dec. 3, 2014, <https://www.theguardian.com/technology/2014/dec/03/amazon-mechanical-turk-workers-protest-jeff-bezos>.

47. See, e.g., M. Six Silberman et al., *Sellers' Problems in Human Computation Markets* (Proc. 2nd Workshop on Human Computation, 2010); M. Six Silberman, Lilly Irani & Joel Ross, *Ethics and Tactics in Professional Crowdwork*, 17 XRDS 39 (2010); Bederson & Quinn, *supra* note 42; Martin et al., *supra* note 15.

48. Kristy Milland et al., *MTurk Workers: We're Here and We're Human* (Panel at Digital Labor: Sweatshops, Picket Lines, Barricades, Nov. 14-16, 2014).

49. TURKERNATION, <http://turkernation.com> (last visited May 4, 2016).

50. MTURK CROWD, <http://mturkcrowd.com> (last visited May 4, 2016).

51. MTURK GRIND, <http://mturkgrind.com> (last visited May 4, 2016).

52. *Turkkit*, REDDIT, <https://www.reddit.com/r/mturk> (last visited May 4, 2016); *HITs Worth Turking for*, REDDIT, <https://www.reddit.com/r/hitsworthturkingfor> (last visited May 4, 2016).

53. MTURK FORUM, <http://mturkforum.com> (last visited May 4, 2016).

successful Turkers; it is typically only by connecting to a community of more experienced workers that one can navigate AMT well enough to earn significant income.⁵⁴ Forum discussions are relatively open and unstructured. They include both specific information-sharing about tasks, requesters, and processes and more open-ended discussions in which newer workers “learn the ropes” and connect informally with others to build trust and community.⁵⁵ For example, the MTurk Grind forum hosts a thread about “Mechanical Turk software beyond [user] scripts,”⁵⁶ and another thread with resources and support for Turkers struggling with anxiety and depression.⁵⁷ Turker Nation hosts a famous “Requesters Hall of Fame/Shame” subforum⁵⁸—the original dedicated resource, predating Turkopticon, for explicit discussion of workers’ experiences with requesters. All four major forums host public “daily HIT threads” in which workers post information about live HITs. These threads are very active, with some accumulating over a thousand posts in a 24-hour period. And many of the forums have associated IRC (Internet Relay Chat) channels (i.e., chat rooms) for real-time communication.

Requesters also communicate with workers on forums. Requesters can build trust with workers by communicating publicly on worker forums, and can share information and get feedback about their tasks and processes. Crucially, communicating quickly with workers on forums can save requesters from losing previously earned goodwill in the event of technical problems that result in accidental rejections or wasted time for workers. Turker Nation, MTurk Grind, and MTurk Forum all have dedicated subforums where requesters are invited to introduce themselves and recruit workers to their HITs.

Forum discussions are complemented by highly structured information sharing practices supported by specialized software. Most special-purpose Turking software used heavily by Turkers is built and maintained by Turkers. (I know of only two exceptions to this generalization: Turkopticon and Dynamo, discussed below.) Most of this software takes the form of user scripts, relatively small pieces of code that users can download and run within their web browser. User scripts typically alter the appearance and/or functionality of particular websites. For example, one popular script, “HIT Scrapper,” scrapes as many pages of the AMT HIT list as the user specifies,

54. Milland et al., *supra* note 48.

55. For a qualitative study of Turker Nation, the oldest Turker forum, see Martin et al., *supra* note 15.

56. *Mechanical Turk Software Beyond Scripts*, MTURK GRIND (July 18, 2014), <http://www.mturkgrind.com/threads/mechanical-turk-software-beyond-scripts.25492>.

57. Eetha, *Sad Pandas Support Group*, MTURK GRIND (Aug. 2, 2014), <http://www.mturkgrind.com/threads/sad-pandas-support-group.26375>.

58. The subforum is private to registered users.

applies filters specified by the user, and displays the information concisely (see Figure 7).⁵⁹

Figure 7. Results from the HIT Scrapper Script, Built by Turkers

Requester	Title	Reward	HITS Available	TO pay	Accept HIT	
Sara Spencer [R T]	Find the Contact Info for the Owner of Some Websites [v]	\$0.05	3	4.09	Accept	R T
gbr [R T]	Transcribe data [v]	\$0.03	72	2.85	Accept	R T
gbr [R T]	Transcribe data [v]	\$0.01	5	2.85	Accept	R T
Mechanical Turk [R T]	Transcribe the characters in the images (50 images) [v]	\$0.01	2	1.00	Accept	R T
AE [R T]	Find the link and image that a forum thread is discussing [v]	\$0.05	5	3.40	Accept	R T
Jimmy Dregon [R T]	Business Card Transcription [v]	\$0.02	3	2.67	Accept	R T
gbr [R T]	Transcribe data [v]	\$0.04	4	2.85	Accept	R T
gbr [R T]	Transcribe data [v]	\$0.09	17	2.85	Accept	R T
AM [R T]	To be eligible, you are 18 years or older, and currently working / employed. [v]	\$0.35	1	2.73	Accept	R T
Steve Murch [R T]	Review transcription quality [v]	\$0.02	1	3.34	Accept	R T
Speechlink [R T]	MTurk payment authorization [v]	\$0.00	3	2.91	Accept	R T
Steve Murch [R T]	Review transcription quality [v]	\$0.03	3	3.34	Accept	R T
Vedrona Group LLC [R T]	Quick survey [v]	\$0.05	1	2.00	Accept	R T
GUAS [R T]	3 Minute Survey For Mpa (non Residents) [v]	\$0.75	1	1.95	Accept	R T
Terry2094014 [R T]	Command to an autonomous car [v]	\$0.04	6	3.00	Accept	R T
Terry2094014 [R T]	Describe POJ [v]	\$0.05	7	3.00	Accept	R T
CastingWords [R T]	Audio Transcription: George King George #193921 (avg rwd+bns: \$0.88) (05:13 mms) [v]	\$0.44	1	2.89	Accept	R T
CastingWords [R T]	Audio Transcription: George King George #193921 (avg rwd+bns: \$2.26) (13:15 mms) [v]	\$1.13	2	2.89	Accept	R T
CastingWords [R T]	Audio Transcription: George King George #193921 (avg rwd+bns: \$1.41) (08:15 mms) [v]	\$0.70	1	2.89	Accept	R T
CastingWords [R T]	Audio Transcription: George King George #193921 (avg rwd+bns: \$0.91) (05:15 mms) [v]	\$0.45	2	2.89	Accept	R T
CastingWords [R T]	Audio Transcription: George King George #193921 (avg rwd+bns: \$3.62) (21:15 mms) [v]	\$1.81	1	2.89	Accept	R T
Harlan Beverly [R T]	Transcript recipe [v]	\$0.20	10	1.83	Accept	R T
CastingWords [R T]	Audio Transcription: George X-ray #193903 (avg rwd+bns: \$3.98) (21:02 mms) [v]	\$1.79	1	2.89	Accept	R T
CastingWords [R T]	Audio Transcription: George X-ray #193895 (125% premium) (avg rwd+bns: \$1.12) (05:15 mms) [v]	\$0.56	1	2.89	Accept	R T
CastingWords [R T]	Audio Transcription: George Boston #193916 (avg rwd+bns: \$2.26) (13:15 mms) [v]	\$1.13	1	2.89	Accept	R T
Steve Murch [R T]	Transcribe a recipe [v]	\$0.30	7	3.34	Accept	R T

The script lets the user omit HITs posted by requesters the user specifies (a feature perhaps inspired by another script, “Block Requester,” which hides HITs from requesters the worker doesn’t want to work for from the HIT list⁶⁰), or show only HITs posted by specific requesters. The script can also programmatically generate a piece of code for vBulletin—the software that powers the major worker forums—that can be pasted into a forum post. The code lists the HIT title, requester name and ID (including a link to Turkopticon reviews), Turkopticon aggregate ratings, number of Turkopticon reviews, link to submit a Turkopticon review, HIT description, allotted time for the HIT, number of HITs available, reward, and qualifications required—all in a visually clear and concise layout (Figure 8).

59. Tjololo, *HIT Scrapper with Export*, GREASY FORK, <https://greasyfork.org/en/scripts/2002-hit-scraper-with-export> (last visited May 4, 2016).

60. Aphit (archived by clickhappier), *Block Requesters (Fixed)*, GREASY FORK, <https://greasyfork.org/en/scripts/6132-block-requesters-fixed> (last visited May 4, 2016).

Figure 8. A post to the MTurk Crowd “daily thread” produced by a specialized user script built by Turkers. The text in the black box is gathered together automatically by the script from different sources—the AMT HIT list and Turkopticon—and presented clearly and concisely. The fact that the posting user’s avatar is not merely a photograph of a cat but rather an animated dancing cat is, regrettably, impossible to convey in this static screenshot.



Hundreds of posts including programmatically generated information about HITs, often supplemented with the posting worker's comments, populate the daily threads on Turker forums. Other scripts allow workers to see the total value of their HITs pending approval (“Today’s Projected Earnings”), see more information about HITs, including the auto approval time (“Enhanced HIT Information Capsule”), and keep track of HITs they have done (“Mturk database”). Several browser extensions (“Turk Assist,” “Tools for Amazon’s Mechanical Turk”) attempt to integrate many of the most crucial features of multiple scripts into a single package.

2. Requester Tools, Networks, and Practices

An ecosystem of requester tools, networks, and practices parallels that made and used by Turkers. The requester world can be thought of having two parts: one inhabited mainly by academic requesters and one inhabited mainly by industry requesters. The two parts are tightly interwoven, but academic and industry requesters have different responsibilities and work within different institutional cultures. Thus the practices of academic and industry requesters, and the resources developed to support and explain them, differ somewhat. And researchers in different academic disciplines post different kinds of tasks to AMT; social scientists often post surveys and run experiments, while computer scientists often post tasks that incorporate workers into an algorithmic process, for example in computer vision or machine learning applications. These differences lead to different requirements: social scientists, for example, are often keen to ensure that they

are able to prevent workers from participating in a particular survey or experiment more than once, while the information processing work posted by computer scientists can often be done, at least in theory, by “any human.” To the extent that this is the case, computer science researchers posting “human computation” tasks to AMT may sometimes “look” more like industry requesters than like social science researchers.

Like Turkers, requesters teach one another, share information, and develop and circulate specialized software. But requester discourse circulates in a broader variety of media: forums, blogs, meetups, workshops, conferences, and peer-reviewed papers. Most software developed by requesters is for requesters, not Turkers, and as many requesters work in the information technology industry, many requesters are programmers or have relatively easy access to programming expertise. As a result, there is a great deal of specialized software available for requesters.

The differences between academic and industry institutional cultures appears most markedly in the distribution of information about management techniques—especially approaches to quality control—and in the distribution of software itself. Crowd work researchers and academic requesters openly discuss quality control techniques in blogs, academic conferences, and journals, the contents of which are sometimes free online. Some industry requesters participate in these discussions, or discuss their approaches and techniques in their own blogs, but it appears that many do not. Similarly, most software written for academic requesters is free or open source, but there appears to be few free software packages distributed for industry requesters. Many prolific requesters in both sectors write their own specialized software, but it appears that few industry requesters share this work with others. Most free or open source software for requesters aims to aid with workflow management and data analysis; for example, tools exist to prevent workers from retaking surveys (e.g., “TurkGate”⁶¹), integrate AMT more closely with other online survey tools such as Qualtrics (“QualTurk”⁶²), link statistical packages such as R to the AMT API (e.g., “MTurkR”⁶³), post and manage iterative tasks (“TurKit”⁶⁴), and manage the complex process of running behavioral experiments on AMT (e.g., “PsiTurk”⁶⁵). The most well-known free or open source tool for industry requesters may be Clockwork Raven, a toolkit for posting human judgment tasks such as sentiment analysis made by developers at Twitter.⁶⁶

61. Gideon Goldin & Adam Darlow, *TurkGate*, EXPERIMENTAL TURK (Apr. 22, 2013).

62. QUALFORK, <http://www.qualturk.com> (last visited May 4, 2016).

63. GITHUB, <https://github.com/leeper/MTurkR> (last visited May 4, 2016).

64. TURKIT, <http://groups.csail.mit.edu/uid/turkit> (last visited May 4, 2016).

65. PSITURK, <https://psiturk.org> (last visited May 4, 2016).

66. CLOCKWORK RAVEN, <http://twitter.github.io/clockworkraven> (last visited May 4, 2016).

At least as important as the auxiliary software they produce for other researchers, researchers in a wide range of disciplines also publish methodological papers and run workshops explaining how to use AMT for research.⁶⁷ This both spreads practical knowledge about how to use AMT to do research and socially legitimates the practice. In the overlapping human computation and human-centered computing literatures, a large and diverse body of work exists on quality control for crowd work generally, with some of this work focusing specifically on AMT.⁶⁸ Another body of work considers workers' motivations for working, mainly for the purpose of designing incentives that maximize quality and speed and minimize cost.⁶⁹

67. For example, in human-computer interaction, see Aniket Kittur, Ed H. Chi & Bongwon Suh, *Crowdsourcing User Studies with Mechanical Turk*, 453 (Proc. 26th Association for Computing Machinery Conference on Human Factors in Computing Systems, 2008), and Jeffrey Heer & Michael Bostock, *Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design*, 203 (Proc. 28th Association for Computing Machinery Conference on Human Factors in Computing Systems, 2010); in behavioral research, see Gabriele Paolacci, Jesse Chandler & Panagiotis G. Ipeirotis, *Running Experiments on Amazon Mechanical Turk*, JUDGMENT & DECISION MAKING 5 (2010), 411; Gabriele Paolacci & Jesse Chandler, *Inside the Turk: Understanding Mechanical Turk as a Participant Pool*, CURRENT DIRECTIONS IN PSYCHOLOGICAL SCIENCE 23, 184 (2014); Winter Mason & Siddharth Suri, *Conducting Behavioral Research on Amazon's Mechanical Turk*, BEHAVIORAL RESEARCH (June 30, 2011); Michael Buhrmeister, Tracy Kwang & Sam D. Gosling, *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?*, 3 PERSPECTIVES ON PSYCHOL. SCI. 6 (2011).

68. See, e.g., Rion Snow et al., *Cheap and Fast – But Is It Good? Evaluating Non-Expert Annotations for Natural Language Tasks*, 254 (Proc. 2008 Conference on Empirical Methods in Natural Language Processing); Julie S. Downs et al., *Are Your Participants Gaming the System? Screening Mechanical Turk Workers*, 2399 (Proc. 28th Association for Computing Machinery Conference on Human Factors in Computing Systems Extended Abstracts, 2010); Panagiotis G. Ipeirotis, Foster Provost & Jing Wang, *Quality Management on Amazon Mechanical Turk*, 64 (Proc. 2nd Workshop on Human Computation, 2010); Shailesh Kochhar, Stefano Mazzocchi & Praveen Paritosh, *The Anatomy of a Large-Scale Human Computation Engine*, 10 (Proc. 2nd Workshop on Human Computation, 2010); David Alan Grier, *Error Identification and Correction in Human Computation: Lessons from the WPA*, PROC. 3RD WORKSHOP ON HUMAN COMPUTATION (2011), 32; Steven Dow et al., *Shepherding the Crowd Yields Better Work*, 1013 (Proc. 15th Association for Computing Machinery Conference on Computer Supported Cooperative Work, 2012); Hyun Joon Jung & Matthew Lease, *Improving Quality of Crowdsourced Labels via Probabilistic Matrix Factorization*, 101 (Proc. 4th Workshop on Human Computation, 2012); Andrew Mao, Ariel D. Procaccia & Yiling Chen, *Better Computation Through Principled Voting* (orig. *Social Choice for Human Computation*), 136 (Proc. 4th Workshop on Human Computation, 2012); David Oleson et al., *Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing*, 43 (Proc. 3rd Workshop on Human Computation, 2011); Huaming Rao, Shih-Wen Huang & Wai-Tat Fu, *What Will Others Choose? How a Majority Vote Reward Scheme Can Improve Human Computation in a Spatial Location Identification Task*, 130 (Proc. 1st Association for the Advancement of Artificial Intelligence Conference on Human Computation and Crowdsourcing, 2013).

69. Shaili Jain & David C. Parkes, *The Role of Game Theory in Human Computation Systems*, 58 (Proc. 1st Workshop on Human Computation, 2009); Dana Chandler & John Horton, *Labor Allocation in Paid Crowdsourcing: Experimental Evidence on Positioning, Nudges and Prices*, 14 (Proc. 3rd Workshop on Human Computation, 2011); Siamak Faridani, Björn Hartmann & Panagiotis G. Ipeirotis, *What's the Right Price? Pricing Tasks for Finishing on Time*, 26 (Proc. 3rd Workshop on Human Computation, 2011); Yaron Singer & Manas Mittal, *Pricing Tasks in Online Labor Markets*, 55 (Proc. 3rd Workshop on Human Computation, 2011); Adish Singla & Andreas Krause, *Incentives for Privacy Tradeoff in Community Sensing*, 165 (Proc. 1st Association for the Advancement of Artificial Intelligence Conference on Human Computation and Crowdsourcing, 2013); Jens Witkowski et al., *Dwelling on the Negative: Incentivizing Effort in Peer Prediction*, 190 (Proc. 1st Association for the Advancement of Artificial Intelligence Conference on Human Computation and Crowdsourcing, 2013); Andrew Mao et al., *Volunteering Versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing*, 94 (Proc. 1st Association for the Advancement of Artificial Intelligence Conference on Human Computation and Crowdsourcing,

III. TURKOPTICON

We designed Turkopticon to intervene in the interface-mediated labor politics described in Part II. Over the years, Turkers have woven it into their everyday work practices. It has two main parts: a web database application and a browser extension. The web application lets workers review requesters. The reviews include qualitative and quantitative elements. The browser extension aggregates the quantitative elements of all reviews of a particular requester and adds them to the HIT listing next to HITs posted by that requester. This allows a worker to see what other workers have said about a requester before accepting work from them.

We built the prototype system in October 2008, launched it in January 2009, and have maintained it since then. As of January 2016, 56,000 users have created accounts on the Turkopticon website and about 35,000 use one of the two browser extensions. Since early 2009, these users have posted 290,000 reviews of 42,000 requesters. In an average month, about 1,000 workers post about 5,000 reviews; in the period between December 16, 2015 and January 17, 2016, for example, 1,205 reviewers posted 5,205 reviews. Participation is very unequally distributed: within any given month, most users who have posted any reviews have posted exactly one, but a dozen or so post more than 50, and one or two post more than 100. Most professional Turkers appear to use Turkopticon. Requesters report that the prospect of negative Turkopticon reviews influences their decision making while posting HITs and reviewing submitted work. One study of Turkopticon, conducted in the summer of 2014 by a group of economists at the University of Minnesota, found that Turkopticon reviews both accurately reflect requesters' propensities to reject work and affect requesters' ability to have work completed. Silberman is the system's de facto lead programmer and database administrator; we both work with Turkopticon's most active users, including the volunteer moderators, to manage community issues and support users who run into problems.⁷⁰

Turkopticon is named after the panopticon, a prison surveillance design devised by British philosopher Jeremy Bentham and most famously analyzed by French philosopher Michel Foucault. The panoptic prison is round with a guard tower in the center. The tower does not reveal whether the guard is present, so prisoners must assume they could be monitored at any moment. The possibility of surveillance, the theory goes, induces prisoners to discipline themselves. Turkopticon's name cheekily references the

2013); Chien-Ju Ho et al., *Towards Social Norm Design for Crowdsourcing Markets*, 94 (Proc. 4th Workshop on Human Computation, 2012).

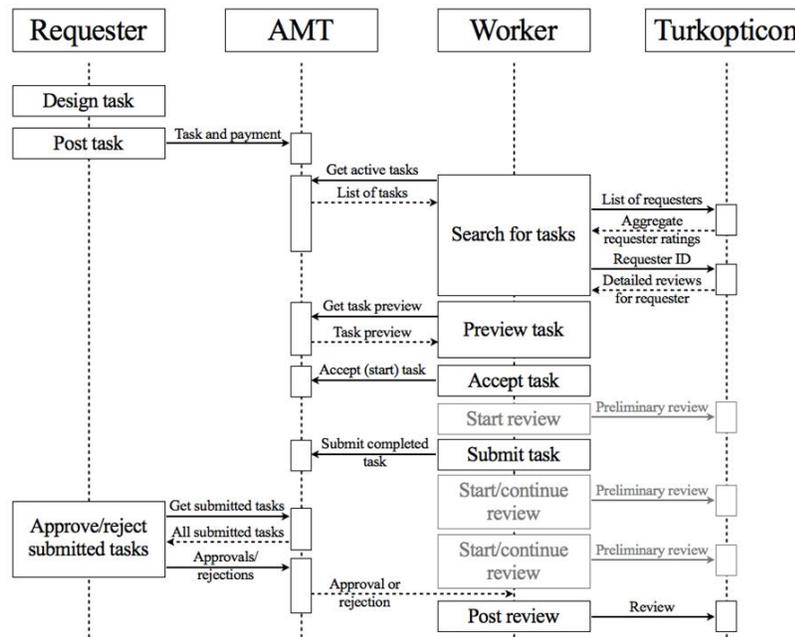
70. For data sources for the foregoing paragraph, see *supra* notes 18-21.

panopticon, pointing to our hope that the site could not only hold employers accountable, but induce better behavior.

A. Turking with Turkopticon

Using Turkopticon complicates the Turking process for workers (see Figure 9).

Figure 9. Turking with Turkopticon

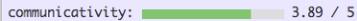
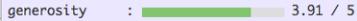
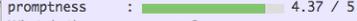


Turkers can interact with Turkopticon in a variety of ways. The simplest and most common is to view aggregated review data while selecting tasks from the AMT task list. Turkopticon adds new information to this interface (see Figures 10 and 11), and workers using Turkopticon look at and consider this information when choosing tasks.

Figure 10. The AMT HIT List with Turkopticon Installed

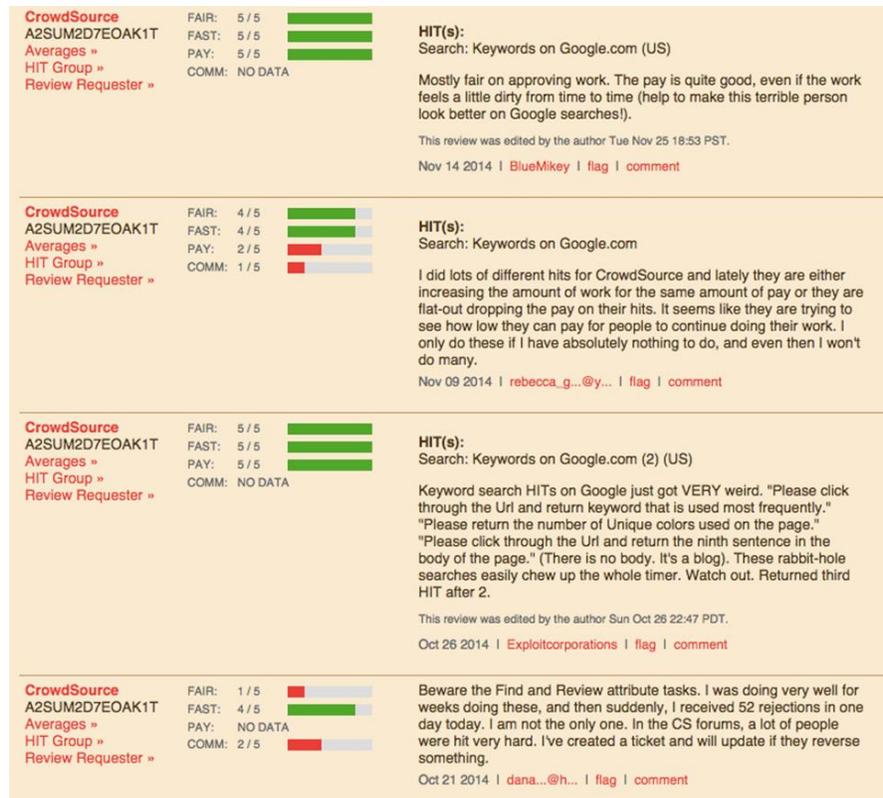
Read a Brief Body of Text and Determine Company Relations II (North America +) Not Qualified to work on this HIT (Why?) View a HIT in this group			
Requester: ▼ OCOMP	HIT Expiration Date: Apr 12, 2015 (14 weeks 4 days)	Reward: \$0.28	
	Time Allotted: 15 minutes	HITs Available: 20	
Search for company information Request Qualification (Why?) View a HIT in this group			
Requester: ▼ Houdini, Inc.	HIT Expiration Date: Mar 19, 2015 (11 weeks 1 day)	Reward: \$0.02	
	Time Allotted: 10 minutes	HITs Available: 1	
Business Phone Validation Request Qualification (Why?) View a HIT in this group			
Requester: ▼ Kevin Dodds	HIT Expiration Date: Mar 19, 2015 (11 weeks 1 day)	Reward: \$0.08	
	Time Allotted: 15 minutes	HITs Available: 2	
Colorado parent opinions regarding banning indoor UV tanning for those under age 18 Request Qualification (Why?) View a HIT in this group			
Requester: ▼ Robert Dellavalle	HIT Expiration Date: Apr 26, 2015 (16 weeks 4 days)	Reward: \$0.05	
	Time Allotted: 60 minutes	HITs Available: 1	

Figure 11. A User Can Mouse over the Arrows Added by Turkopticon to the AMT HIT List to See Aggregate Turkopticon Review Data about the Relevant Requester

REVIEW: Company's Current Job Listings (US) View a HIT in this group			
Requester: ▼ CrowdSource	HIT Expiration Date: Jan 13, 2015 (1 week 6 days)	Reward: \$0.03	
communicativity: 	3.89 / 5		HITs Available: 915
generosity: 	3.91 / 5		
fairness: 	4.32 / 5		
promptness: 	4.37 / 5		
Search: Keyword	View a HIT in this group		
Requester: What do these scores mean?	2015 (1 week 6 days)	Reward: \$0.04	
Scores based on 580 reviews		HITs Available: 48	
Terms of Service violation flags: 2			
Search: Keyword Report your experience with this requester »	View a HIT in this group		

If the aggregated quantitative information is not decisive in helping the worker make a decision, the worker may click a link in the Turkopticon interface element to view the individual reviews, including comments (some of which are lengthy), posted by other workers about the requester (see Figure 12).

Figure 12. Reviews on the Turkopticon Site



the reviews here before working for this requester” is a common statement in unfavorable Turkopticon reviews.

A worker may post a review of a requester after, or even before, they have completed a task posted by that requester. Workers sometimes post “preliminary reviews” that leave some parts of the review form blank in order to warn others about technical problems or an unusual or confusing task design (see Figure 13). Workers often post reviews after completing a task; at this point, they have a good deal of information about the task, but do not yet know how the requester will review it. Workers therefore often edit reviews after their work is approved or rejected by the requester—which may happen up to 30 days after the worker completes the task.

Figure 13. A Turkopticon Review without Ratings

<p>MTurk Booth A1WHARISL4ALTM Averages » HIT Group » Review Requester »</p>	<p>FAIR: NO DATA FAST: NO DATA PAY: NO DATA COMM: NO DATA</p>	<p>HIT(s): "Pilot testing some reading materials"</p> <p>\$0.25</p> <p>My experience was the same as what several other people have posted here.</p> <p>In the early parts of the survey it said there would be a base pay of \$0.25 for starting the study, plus additional performance-based bonus for each section that you read.</p> <p>There was a test reading at the beginning; I did that and then answered some questions, and then suddenly there was nothing more to read, just a completion code. No bonus for me, I guess. I went ahead and submitted the HIT.</p> <p>I have messaged the requester asking them to check to see if the survey was working correctly.</p> <p><small>This review was edited by the author Tue Dec 30 19:18 PST.</small></p> <p><small>Dec 30 2014 Cargo Plane flag comment</small></p>
--	--	---

B. Outcomes and Complications

Turkopticon appears to have changed the decision-making process in approving or rejecting work—at least among requesters who know about it. One requester told us that having a bad Turkopticon reputation made it effectively impossible to attract workers who would submit quality work. But it may be too easy to get a bad reputation: even a technical problem with a task can lead to a string of bad reviews if it is not addressed quickly. And Turkopticon has other problems. From an administrative perspective, many of the problems fall into two major categories: problems arising from disagreements about how to use the review form and problems arising from the absence of a strong connection between a worker and their Turkopticon account.

1. Disagreements: The Review Form

The review form, designed in 2008 after coding the responses to Irani's initial "Turkers' Bill of Rights" survey, includes 5-point Likert scale entries for four requester attributes—generosity of pay, fairness (ostensibly of approval, and rejection decisions, but used by workers for a variety of purposes), speed of pay, and communicativity. The site itself does not give guidance about what counts as, for example, a 5/5 for generosity of pay, or a 2/5 for communicativity; on the contrary, the first instruction on the review form is "Give the ratings you feel best describe your experience." This ambiguity was a strength in Turkopticon's early years, as it created space for workers to discuss their experiences, compare notes, and develop a collective sense of the range of requester behavior—and within that, what ought to be considered good and what bad. But in recent years, with veteran Turkers

having to some extent established a shared understanding of how to use the rating system, newcomers who lack this understanding—or users who disagree with the common usage—can cause tension.

In late 2013, for example, a worker posted a thread to the MTurk subreddit with the title “Incompetent users breaking TurkOpticon, what now?” The 500-word post began:

TurkOpticon is filling up with incompetent users. They don't know what good pay is, they don't know what is prompt, they don't know what is fair, and/or they rate every category based on ONE thing instead of on the actual categories. They've rendered the rating-at-a-glance feature to be completely unreliable; if you want to know whether a HIT is worth your time, you have to open up the Reviews page and look for someone who actually left a comment, and then decide if that reviewer is one of the people who has the same standards as you or if you have to disregard their rating because they're one of the ones that think \$1/1 hour deserves a 5/5 score. This takes time and causes aggravation, the very things TurkOpticon was meant to save us.⁷¹

In view of these concerns, a series of discussions on the Turkopticon mailing list in spring 2014 raised the prospect of adding several more “objective” fields to the review form. “How long did the requester take to pay you?” was proposed to replace the speed of pay rating. “Were you approved?” and “If not, do you think the rejection was fair?” were proposed to replace the fairness rating. And a combination of the reward and the time it took to do the task was proposed to replace the generosity of pay rating. In fall 2015, a new review form was drafted. The new form has not yet been implemented; it amounts to an almost complete redesign of the review system. The proposed design is however instructive in that it shows the detailed information workers have said they would like to be able to use to choose among tasks. Boxes 1 and 2 list the questions in the proposed form. Workers are asked to review specific HITs, not requesters (who may post many different HITs at once).

Figure 14 shows how an updated version of the browser extension might integrate this information into the AMT HIT list. (For the AMT HIT list without Turkopticon, see Figure 5; for the current browser extension, see Figures 10 and 11 above.

71. *Incompetent Users Breaking TurkOpticon, What Now?*, RMMTURK (Dec. 16, 2013), www.reddit.com/r/mturk/comments/1t17w3/incompetent_users_breaking_turkopticon_what_now.

*Figure 14. A Possible Integration of Information Collected by the Proposed New Review Form into the AMT HIT List. This Would Be Accomplished by an Updated Version of the Turkopticon Browser Extensions*⁷²

Extract Company Year Founded from Page	BROKEN (4)	DECEPTIVE (7)	TOS VIOLATION (3)	Request Qualification (Why?)	View a HIT in this group
Requester: OCMPS0				HIT Expiration Date: Dec 14, 2015 (6 weeks 1 day)	Reward: \$0.06 (avg. \$0.07 over 23 revs)
All HITS				Time Allotted: 20 minutes	HITS Available: 40
Average Wage: \$2.51 / hr (162)				Time Reported: 1m 31s (23)	Reviewed in: 4 days (23)
Rejected: 31/160 (19%)				Average Wage: \$2.37 / hr	Satisfactory Communication: 5/6 (83%)
Reviewed in: 8 unfair (5%)				Rejected: 4/21 (19%)	All 25 reviews >
Reviewed in: 6.5 days (162)				Rejected: 3 unfair	Review HIT >
Satisfactory Communication: 55/71 (77%)					
All 181 reviews >					

A version of the browser extension updated to represent the information collected in the proposed new review form would add over a dozen new pieces of information to the HIT listing. Because the proposed new review form collects both the reward and the estimated completion time, an average wage can be calculated from the review data. The rejection rate, at least from among those who have reviewed the HIT, can be calculated and displayed. The number of reviewers claiming that their rejections were unfair can also be displayed. Because requesters sometimes change the reward for a HIT, the average reward may be different from the current reward; differences can alert workers to increases or decreases in the reward. The average review time (and therefore payment delay for approved work) can be displayed, as can reviewers' satisfaction with the requester's communication about the HIT. If reviewers flag a HIT as broken, deceptive, or in violation of AMT TOS, this information can be displayed. Finally, average information for some of these statistics for all of the requester's HITs can be computed; in the design shown here they appear at the far left, under the requester's name.

72. The data are fictional. Full-size image at SHS1RDA, <http://i.imgur.com/sHS1rda.png> (last visited May 4, 2016).

Box 1. First part of the proposed new review form. A worker is asked these questions whether they are reviewing a single HIT or a "batch."

HIT ID [automatically filled, can be edited by reviewer] (required)
HIT Name [automatically filled, can be edited by reviewer] (required)
Requester ID [automatically filled, can be edited by reviewer] (required)
Requester Name [automatically filled, can be edited by reviewer] (required)

This HIT is broken.

If the HIT is broken, please explain how:
(required if "This HIT is broken" is checked)

This HIT is deceptive.

If the HIT is deceptive, please explain how:
(required if "This HIT is deceptive" is checked)

This HIT violates the Mechanical Turk Terms of Service (TOS).

If the HIT violates TOS, please explain how:
(required if "This HIT violates TOS" is checked):

Did you do more than one of these HITs? (required)

Yes No

Box 2. Second part of the proposed new review form. Similar questions with slightly different wording are shown if the user checks "Yes" in response to "Did you do more than one of these HITs?" in the first section.

This section is displayed if the user checks "No" in response to "Did you do more than one of these HITs?" in the first section.

How much did the HIT pay? US \$ ____ (required)

About how long did it take you to do the HIT? (required)
 ____ hours ____ minutes ____ seconds or I don't know

Has your work been approved or rejected yet? (required)

Yes No

How long did the requester take to approve or reject your work?

This question will only appear if the reviewer clicks "Yes" in response to "Has your work been approved or rejected yet?"

Less than one day or ____ days

Was your work rejected? (required)

This question will only appear if the reviewer clicks "Yes" in response to "Has your work been approved or rejected yet?"

Yes No

Do you think the rejection was fair?

This question will only appear if the reviewer clicks "Yes" in response to "Was your work rejected?"

Yes Maybe No

Please feel free to explain what happened here.

Did you try to communicate with the requester about the HIT?

Yes No

Did the requester respond to your satisfaction?

This question will only appear if the reviewer clicks "Yes" in response to "Did you try to communicate with the requester about the HIT?"

Yes No

If you have not already, please feel free to explain what happened here.

2. Identity: Harassment and Requester Self-Reviews

Some workers use Turkopticon to encourage others to harass requesters. Sometimes the impetus for this incitement arises from a misunderstanding. For example, new academic requesters who do not want multiple responses from one worker for a survey often use an AMT feature called a "block" to prevent workers from completing their HITs more than once. (There appears to be multiple kinds of block, although workers argue about this.) If a

requester blocks a worker, the worker may receive an automated email from Amazon informing them of the block and warning them that if they receive multiple blocks, their account may be suspended. This creates stress, especially for workers who rely on Turking income to meet basic needs. As a result, new requesters who do not realize they are doing anything “wrong” by blocking workers may be on the receiving end of significant worker anger. In October 2012, for example, one worker posted a review rating a requester 1/5 for fairness (and N/A for the other attributes), with the following explanation: “DO NOT COMPLETE THESE HITS, THIS PERSON WILL BLOCK YOU AND GET YOUR ACCOUNT SUSPENDED. BEWARE.”

In May 2013, another worker posted a comment underneath the review (“XXX” indicates redacted content):

Report this scammer. His website is registered through enom.com. The person it’s registered to is XXXXXXXXXX. Send complaints to enom. Then file complaints to their affiliate program. Their clixsense affiliate Id & account number is XXXXXXXX. Their rewarding ways affiliate Id is XXXXXXXXXX. Their inbox dollars & Jill’s Click corner affiliate ID is refXXXXXXX. Their Reality-Networkers Id is XXXXXXXX. And they are using paypal on their site. Paypal has this as their email address, XXXXXXXX@XXXXXXX.edu. Complain about their scam to paypal. If you search for the email address you will find [the requester’s name]. An [name of university redacted] student. He also has a facebook page also.

The requester, a graduate student, replied in a comment:

This post is defaming and insulting my name. I have done nothing this person has reported. I just came across this today. It is one thing to leave a review that is true, but it is another thing to name someone out, the reason he is mad is because it is a low paying hit. Please delete this comment. Thank you. I will work on increasing the pay on my hits.

The requester emailed us several days later. After an email exchange spanning a month in which the requester explained that he was receiving threatening emails and messages on social networking sites—sometimes dozens daily—and that he was working with his university to have his email address changed as a result, we censored the comment with the requester’s information, at that time an unprecedented exercise of administrative power.

We have also received reports of workers attempting to blackmail requesters into paying for work both parties know is bad by threatening to leave bad Turkopticon reviews. We have received reports that sometimes these efforts are successful.

Sometimes requesters review themselves. This is possible because we have no mechanism for verifying that a Turkopticon user posting a review has worked for the requester they are reviewing, or that they are even a worker at all. Niloufar Salehi, Ali Alkhatib, and Eva Ogbe developed a mechanism for linking a user’s account on a separate service to a worker’s

AMT account was developed in 2014 for the Dynamo platform.⁷³ Turkopticon users have expressed interest in adopting this model, but it has not yet been implemented. (And in November 2015, Alkhatib told us that Amazon had disabled the technique they were using.) For now, Turkopticon's volunteer moderators—who are expert Turkers and members of multiple Turker communities—rely on information from the Turkopticon database, contextual clues such as a reviewer's email address and review history, and direct communication with reviewers to determine if a review was posted inappropriately by a requester masquerading as a worker.

Finally, Turkopticon is afflicted by the harassment, insults, sexism, racism, profanity, baseless accusations, and occasional threats that trouble other online communities (and indeed offline worker communities and organizations). At present, Turkopticon uses a combination of simple automated filters and a relatively unsophisticated flagging system that lets users bring uncivil reviews to the attention of volunteer moderators. The ambiguities of this arrangement sometimes upsets the people who are “moderated” (or, in extreme cases, silenced). As a result, we occasionally receive angry or threatening emails from both workers and requesters, and the Institutional Review Board at the University of California, San Diego, where the server hosting Turkopticon is physically located, has received a few formal complaints against us over the years. Although more sophisticated approaches to online community moderation exist,⁷⁴ exploring and refining them is beyond Turkopticon's current organizational capacity.

IV. LESSONS FOR DEVELOPING WORKER POWER IN THE ON-DEMAND ECONOMY

What have we learned in the six years of Turkopticon's operation? Are these lessons relevant for efforts to support the development of worker power in the “on-demand economy”? In this concluding Part, we offer eight lessons that we hope will be useful for stakeholders in those ongoing efforts. The first four pertain to reputation systems for labor platforms specifically; the last four address the project of developing “on-demand” worker power generally.

73. Niloufar Salehi et al., *We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers*, 1621 (Proc. 33rd Association for Computing Machinery Conference on Human Factors in Computing Systems, 2015).

74. See, e.g., F. RANDALL FARMER & BRYCE GLASS, *BUILDING WEB REPUTATION SYSTEMS* (2010); Yubo Kou & Bonnie Nardi, *Regulating Anti-Social Behavior on the Internet: The Example of League of Legends* (Proc. Conference 2013); Yubo Kou & Bonnie Nardi, *Governance in League of Legends: A Hybrid System* (Proc. Foundations of Digital Games 2014); Laura Hudson, *Curbing Online Abuse Isn't Impossible. Here's Where We Start*, WIRED, May 15, 2014.

A. *Reputation Is Important*

First, unsurprisingly, reputation is important. Just as in other markets,⁷⁵ computerized reputation systems can significantly influence outcomes in labor markets. When they have it, people use reputation information in their decision making. The processes by which that information is created, filtered, aggregated, revised, and displayed influences those decisions and, ultimately, “macro” outcomes in the market such as the distribution of benefits and bargaining power between stakeholder groups. In a labor market, a reputation system serves functionally as a governance mechanism, “reifying” commonly accepted ideas about accepted (and unacceptable) practice and sanctioning participants who violate market norms.

B. *Reputation Is Not Simple*

Second, reputation is complex. Workers in different circumstances look for different things in a task or client. This means that a single 5-point Likert scale, or even several, as in the current Turkopticon review form, may be inadequate or even misleading. The significant increase in complexity between the current review form and the proposed new form shows that workers are interested in obtaining information about client behavior at many points in the workflow enabled by the market. In other discussions, workers have raised the possibility of deleting old reputation data when they know that management of a requester account has changed hands, or if a novice requester made a mistake that severely harmed their reputation but was quickly corrected. Technical and organizational mechanisms for *accountably* incorporating such knowledge into the aggregate reputation information seen by a worker browsing or searching for work will be complex, but will improve the fairness of the reputation system. Finally, despite the ongoing change of relevant criteria and worker expectations, some norms appear to have stabilized. For example, the current Turkopticon review form allows a reviewer to note that a requester’s HITs violate the AMT Participation Agreement (informally referred to by workers as the “TOS,” or “Terms of Service”). However, sometimes reviewers leaving an unfavorable review after a rejection they perceive as unfair or mistaken will indicate, falsely, that the requester’s HITs violate the TOS. They are typically asked by Turkopticon’s volunteer moderators to correct their review to avoid needlessly deterring other workers from doing the HIT, but a more efficient approach might be to simply separate the “TOS violation” “flag” from individual reviews that reviewers feel are “theirs,” and devise a collective

75. See, e.g., Michael Luca, *Reviews, Reputation, and Revenue: The Case of Yelp.com* (Harv. Busn. Sch. Working Paper No. 12-016, Sept. 2011).

workflow that gives experienced workers more influence in determining the status of that flag for a given HIT.

C. *Successful Reputation Systems Will Evolve*

Third, successful reputation systems will evolve. As in other domains, successful governance is an ongoing process: “problems” are rarely decisively solved; rather, the technical or organizational “solution” of one “problem” often creates another problem. Relevant review criteria and their interpretations may change over time, presenting system administrators with both technical and social challenges. And to know about these changes in the first place, administrators must take the time to remain “in touch” with stakeholders. In such a dynamic context, the notion—beloved of engineers—that a technical system can offer a solution to a social problem is likely to be misleading.

D. *The Choice between Independent and Integrated Reputation Systems Presents Trade-offs*

Fourth, the choice between *independent* and *integrated* reputation systems—for example, between, on one hand, establishing a reputation system independent from the labor platform it regulates, and, on the other, requiring or calling on platform operators to include sophisticated reputation mechanisms within their platforms—presents trade-offs. This choice was not available to us as operators of Turkopticon, but it may be available to regulators. The major concern in requiring platform operators to integrate nuanced, two-sided reputation mechanisms into their platforms is conflict of interest: “on-demand” platforms thrive economically on convenience *for the client*, and establishing and operating complex governance mechanisms by which workers can effectively police client behavior may not be perceived as aligned with platforms’ business interests. On the other hand, platform operators are likely to be better-positioned technically and organizationally to operate reputation systems that regulate their own platforms, as they already employ staff with the relevant expertise (community management, user research, software design and development) and have unique access to relevant data that they could easily publicize. Operators of independent reputation systems, in contrast, must—for now—work as volunteers (as in the case of Turkopticon), secure external funding, or convince workers to provide financial support. As a result, unlike well-funded labor platforms, independent reputation systems face an ongoing challenge of funding their work and establishing organizational sustainability. Additionally, independent reputation systems do not have direct access to relevant data (for example, in the case of AMT, requesters’ rejection rates). This means they

must use proxies (such as Turkopticon's "fairness" criterion), ask workers to enter the data manually (as Turkopticon's proposed new review form would do), or develop complex and perhaps brittle technological means to acquire it (Crowd-Workers.com, for example, a Turkopticon alternative, is a browser extension that scrapes the relevant data from the workers' account "dashboards").

A change to the AMT HIT list proposed by workers illustrates the dilemma. One use of Turkopticon is that it gives workers some information, however imperfect, about how likely a given requester is to reject (i.e., not pay for) their work. AMT staff, however, has access to complete data on this topic, and they could make it visible to workers. At least one worker responded to a call for design suggestions from AMT staff with a request that this data be made available to workers, but it has not been done.

E. Labor Platforms Can Be Organized Differently

Fifth, alternative organizational models are thinkable, and our experiences, which make visible the limitations of both independent and integrated systems, suggest that such alternatives may be worth exploring. For example, cooperatively owned and managed platforms may resolve the conflict of interest in integrated reputation systems by prioritizing financial sustainability rather than profit maximization and shareholder value as an organizational constraint rather than orienting platform design and operation—and a lively discourse is developing around the possibility of developing such platform.⁷⁶ Alternatively, government regulation or an independent certification organization could require or call on platform operators to include integrated reputation systems within their platforms; to meet certain technical, procedural, and organizational requirements; and to open the technical and organizational operations of those systems to independent audit. Yet another alternative is that platform operators could be pressured or legally required to fund reputation systems that are maintained by a formally independent organization but functionally integrated with the software used by workers and clients, in a manner reminiscent of the former separation between the editorial and advertising departments of newspapers. Relevant data typically accessible only to platform operators could be made accessible to the reputation system. Finally, but perhaps most quickly acted upon, existing independent systems operated by workers and other third parties, such as the forums and software tools described in Part II, can be financially, technically, and organizationally supported by government agencies, labor organizations, or foundations.

76. See, e.g., TREBOR SCHOLZ, PLATFORM COOPERATIVISM: CHALLENGING THE CORPORATE SHARING ECONOMY (2016).

F. Current Platforms Concentrate Power

Sixth, current platforms, including Turkopticon, concentrate power in the hands of a small group of operators. Platform operators make design and policy decisions that shape the general availability of information among participants as well as administrative decisions in specific cases. As Ajay Agrawal and colleagues write:

The position of the platform vis-a-vis the marketplace is . . . like that of a government that sets policies to encourage efficient market outcomes without dictating trades. The platform decides how often and in what context participants are exposed to each other, what information is collected by parties, and how this information is displayed. Platforms also set policies about what trades are permissible, how entry is gained, what contracts and prices are allowed, and so on. The platform may also make recommendations and set defaults.⁷⁷

Like governments who rely on tax revenues from a particularly lucrative market, however, platform operators' incentives make it implausible to expect them to be perfectly neutral; rather, their interests are typically more aligned with clients' than workers'. Platform operators set the "rules of the game," but they are also themselves "players." Further, most platforms—again, Turkopticon included—function to some extent as *autocratic* governments, with design, development, policy, and administration in the hands of a small group of technocrats. As the aggrieved email from a worker sent to Silberman's personal email address makes clear, the technical power of site operators vests them with significant social power—whether they want it or not. This may be beneficial for labor platforms aiming to maximize profits but it may be counterproductive for platforms aiming to support the development of broad-based worker power. It takes significant care and effort to distribute this power accountably, however, especially in a not-for-profit setting. It is possible to imagine more democratically accountable platforms that distribute power more broadly among stakeholders. The extent to which such platforms are however democratic in practice depends not only on the design of the technology and the social practices that it supports and hinders, but also the organizations and social practices enabling and surrounding the development, operation, maintenance, and evolution of the platforms.

77. Ajay Agrawal et al., *Digitization and the Contract Labor Market: A Research Agenda* (Nat'l Bureau of Econ. Res. Working Paper No. 19525, Oct. 2013).

G. The “On-Demand Economy” Raises Deep but Understudied Cultural Questions

Seventh, the growth of the “on-demand economy” raises deep, but thus far under-studied, questions about *cultural* trends expanding the casualization of service labor relationships of all kinds that may be beyond the scope of regulation or technological intervention to mitigate. Irani found that AMT allows programmers to maintain the ideology of non-hierarchical collaboration by rendering low valued labor invisible.⁷⁸ Similarly, freelance filmmaker Andrew Callaway noted of Postmates, an on-demand courier service, that the platform makes service labor more culturally palatable in a culture that disavows servitude. He writes:

Postmates couriers are told that it is strictly against the rules to shake a customer’s hand. Like all rules, this didn’t come from nowhere. The truth is that using sharing economy services can breed contempt for the workers. One creepy Uber driver can nurture disdain for all the lowly drivers. You never even have to see the person who is cleaning your house or your clothes. Plenty of people requested that I drop off their food at the door. Customers grow to love apps that make the worker anonymous. That way, you don’t have to feel guilty about having servants.⁷⁹

Even more than low cost, immediacy and convenience are the major marketing points of many on-demand services: ride-“sharing” service Lyft advertises “rides in minutes,”⁸⁰ delivery service Postmates offers “the best of your city delivered in minutes,”⁸¹ and general labor platform TaskRabbit assures potential clients that “help around the home is just a few clicks away.”⁸² The German language site of Ohlala, a Berlin-based platform for “paid dates”⁸³—described in the American tech news as “an Uber for escorts,”⁸⁴ although the site denies that it is an “escort app”—promises clients “paid dates” that are “immediate, anonymous, [and] uncomplicated.”⁸⁵ The apparently broad appeal of the ability to enter into a labor relation—whether the labor on offer be “data cleaning,” taxi services, house cleaning, food delivery, or sex—in a way that is artfully technologically mediated so that clients need not grapple with or even confront the “complications” of hiring a human being deserves careful examination.⁸⁶

78. Irani, *supra*, note 4.

79. Andrew Callaway, *Exploitation in a City of Instaserfs: How The “Sharing Economy” Has Turned San Francisco into a Dystopia for the Working Class*, CAN. CTR. POL’Y ALTERNATIVES (Jan. 1, 2016), <https://www.policyalternatives.ca/publications/monitor/exploitation-city-instaserfs>.

80. LYFT, <https://lyft.com> (last visited May 4, 2016).

81. POSTMATES, <https://postmates.com> (last visited May 4, 2016).

82. TASKRABBIT, <https://taskrabbit.com> (last visited May 4, 2016).

83. OHLALA, <https://ohlala.com> (last visited May 4, 2016).

84. Mike Butcher, *Ohlala, An Uber for Escorts, Launches in Berlin, Plans Global Rollout*, TECHCRUNCH (Aug. 10, 2015).

85. OHLALA, *supra* note 83.

86. For further discussion, see Lilly Irani, *Justice for Data Janitors*, PUBLIC BOOKS (Jan 15, 2015).

H. Interdisciplinary and Cross-Sectoral Collaboration Is Needed

Finally, robust interdisciplinary and cross-sectoral collaborations will be needed to improve the fairness of existing platforms, develop sustainable worker power, and develop and maintain new labor platforms that orient toward fairness and accountability—between workers, social scientists, designers, programmers, social entrepreneurs, regulators, labor organizers, and other stakeholders invested in developing technologies, organizations, business practices, and economic institutions that attend more conscientiously to the needs of workers—not just employers and customers.