Mechanical Turk is Not Anonymous

[Version 1.0, March 6, 2013]

Matthew Lease[⋆], *Jessica Hullman*[⋈], *and Jeffrey P. Bigham*[⋒], *Michael Bernstein*[♯], *Juho Kim*[†], *Walter S. Lasecki*[⋒], *Saeideh Bakhshi*[⋄], *Tanushree Mitra*[⋄], *and Robert c. Miller*[‡]

[⋆]School of Information
University of Texas at Austin
ml@ischool.utexas.edu

[⋈]School of Information
University of Michigan
jhullman@umich.edu

[⋒]Department of Computer Science
University of Rochester
{jbigham,wlasecki}@cs.rochester.edu

[♯]Department of Computer Science
Stanford University
msb@cs.stanford.edu

[† ‡]MIT CSAIL
[†] juhokim@mit.edu
[‡] rcm@csail.mit.edu

[⋄]School of Computer Science
Georgia Institute of Technology
{saeideh,tmitra3}@gatech.edu

ABSTRACT

While Amazon's Mechanical Turk (AMT) online workforce has been characterized by many people as being anonymous, we expose an aspect of AMT's system design that can be exploited to reveal a surprising amount of information about many AMT Workers, which may include personally identifying information (PII). This risk of PII exposure may surprise many Workers and Requesters today, as well as impact current institutional review board (IRB) oversight of human subjects research involving AMT Workers as participants.

We assess the potential multi-faceted impact of such PII exposure for each stakeholder group: Workers, Requesters, and AMT itself. We discuss potential remedies each group may explore, as well as the responsibility of each group with regard to privacy protection. This discussion leads us to further situate issues of crowd worker privacy amidst broader ethical, economic, and regulatory issues, and we conclude by offering a set of recommendations to each stakeholder group.

Keywords

crowdsourcing, human computation, privacy, regulation

1. INTRODUCTION

While Amazon's Mechanical Turk (AMT) online workforce has been characterized by many researchers, journalists, and bloggers as being anonymous (§2), we expose in §3 an aspect of AMT's system design which can be exploited to reveal a surprising amount of information about many AMT Workers, which may include personally identifying information (PII). This risk of PII exposure may surprise many Workers and Requesters today, as well as impact current institutional review board (IRB) oversight of human subjects research including AMT Workers as research subjects.

To assess the potential impact of such PII exposure, we investigate current Worker expectations with regard to their privacy (§4.1), and we enumerate several risks such unanticipated exposure may pose to each stakeholder group: Workers (§4.2), Requesters (§4.3), and AMT itself. In regard to the latter, we compare and contrast this case with past incidents of surprising PII exposure involving other companies, particularly AOL and Netflix, and discuss Federal Trade Commission (FTC) actions in those specific cases (§4.4).

Should this privacy vulnerability be allowed to persist, we speculate upon how platform use by Workers and Requesters might evolve over time (§4.5). We then consider potential measures each stakeholder group might pursue to either correct or at least mitigate risk of exposing PII or other private information (§5.1). Following this, we assess whether we really ought to be surprised about lack of AMT Worker anonymity, given a variety of previously known (but relatively less discussed) aspects of AMT (§5.2).

As academic researchers, we have particular insight and duty to reflect upon our own community's responsibility (and past mistakes) in safeguarding crowd worker privacy. We consider this in some detail (§5.3). Such reflection further leads us to consider these issues of crowd worker privacy amidst broader ethical, economic, and regulatory issues surrounding the crowd work industry at large, and particularly our own impact on it and the workers involved (§6). We conclude by offering a set of recommendations, to both academic Requesters and to AMT, for steps each party may take to further strengthen what we suggest is our joint stewardship of worker privacy (§7.1) and overall industry health.

Posted online to the Social Science Research Network (SSRN) at:
http://papers.ssrn.com/sol3/__papers.cfm?abstract_id=2228728.

2. BACKGROUND

2.1 Amazon's Mechanical Turk

Launched in 2005, Amazon's Mechanical Turk (AMT)[1] [13] provides a massive globally-distributed, online marketplace for work in which *Requesters* post tasks to be completed by *Workers* (aka, *Providers*). By 2007, AMT had grown to encompass 100,000 Workers in more than 100 countries [53]. Today, AMT boasts over 500,000 Workers from 190 countries[2]. While the scope of tasks one can post on AMT is largely unrestricted, AMT has predominantly attracted entry-level *micro-tasks* in which human workers engaged in data-processing continue to produce higher quality results

[1] https://www.mturk.com
[2] https://requester.mturk.com/tour

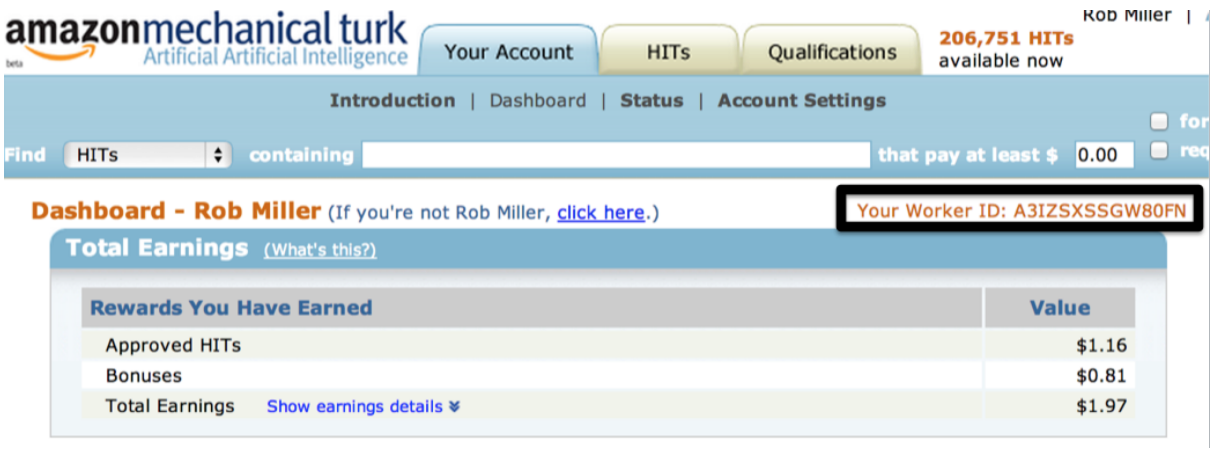

Figure 1: Example AMT WorkerID, shown from Worker's Dashboard.

than even state-of-the-art, automated methods. As one of the most prominent examples of market-based *crowdsourcing* [29, 14], *human computation* [54, 38, 67], and crowd work [35], AMT tasks are posted via an open call which allows any qualified crowd worker to choose to respond.

In addition to industrial application, AMT has also been used to power *citizen science*, a movement within scientific fields that seeks to engage non-domain experts in scientific work [33], as well as many applications based on collective intelligence or wisdom of crowds [61, 68, 60]. Support for such varied usage, along with being first-to-market and helping to define the industry, has helped AMT capture both the attention and imagination of the information technology community since its inception.

A particularly noteworthy aspect of AMT is the degree to which it enables crowd labor interactions to be automated in comparison to traditional labor practices. Requesters can post tasks programmatically to the marketplace in large volume, as well as express formal qualification requirements that Workers must meet in order to *accept* a given task. Once a Worker completes a task, the Requester may programmatically collect the Worker's output and provide payment in response. As Amazon CEO Jeff Bezos has described AMT, "You've heard of software-as-a-service. Now this is human-as-a-service" [12]. By streamlining online work so thoroughly, AMT has enabled data processing tasks to be accomplished with an incredible level of efficiency.

2.2 Safeguarding Worker Privacy

As a further stride towards optimizing efficiency of work production, AMT's design incorporates a novel method of how Requesters and Workers are identified to one another: via an uniquely assigned 14-character alphanumeric string (the *RequesterID* and *WorkerID*, respectively). This enables AMT interactions to take place without either party knowing the real-world identity of the other. The Requester sees only the WorkerID of Workers accepting tasks, and while Workers can see Requesters' account names, Requesters are not required to use their real names (and often do not).

In describing AMT, Amazon's VP in charge of the platform, Sharon Chiarella, explained that this "minimal expressiveness of Worker representations" was designed not only to promote efficiency of online work, but also to simultaneously protect Workers against any discrimination on the basis of gender or race [31]. The official AMT Blog[3] provided a brief synopsis of the 2010 NetWork conference in which Chiarella participated in a panel [43]:

> Anonymity of Workers was discussed... Sharon agreed that it is not an individual's identity that's important but the quality of their work.

In addition to the use of alphanumeric identifiers and statements by AMT personnel, Worker privacy on AMT is further protected by AMT's Terms of Service (ToS): "[Requesters] may not use Amazon Mechanical Turk for ... collecting personal identifiable information" from Workers[5]. The policy is also further elaborated:

> What are some specific examples of HITs that violate Amazon Mechanical Turk policies? HITs requiring disclosure of the Worker's identity or e-mail address, either directly or indirectly.

AMT's separate Participation Agreement [4] also forbids any Requester use of AMT for "invasion of privacy".

As adoption and use of AMT has become increasingly widespread, such privacy safeguards have garnered particular attention from both academic researchers and popular media alike in their characterizations of AMT:

- "Both workers and requesters are anonymous..." [51]
- "...it is the norm for workers to remain anonymous on Mechanical Turk. Worker IDs are anonymized strings and do not contain personally identifiable information." [41]
- "...although identifiable via worker IDs, workers are anonymous to requesters, which protects respondent anonymity..." [57]
- "Mechanical Turk is purposefully an anonymous labor market..." [44]
- "...Mechanical Turk's labor force is decentralized, anonymous, and invisible..." [17]

In particular regard to human subjects research at universities, the anonymity of Workers has been highly valued in facilitating AMT's use, since ensuring the privacy of subjects'

[3] `http://mechanicalturk.typepad.com/blog`

identifying information is often otherwise a complicated aspect of experimental and data design.

Given all the background above, as well as AMT's growing scientific usage for human subjects research, it was a remarkable surprise to our group of researchers to discover that AMT's workforce is actually not quite so anonymous as many in our community have come to believe.

3. THE VULNERABILITY

At the recent CrowdCamp Workshop[4] less than two weeks ago (February 23-24, 2013), co-located with the ACM Conference on Computer Supported Cooperative Work and Social Computing, our team of faculty and student researchers accidentally discovered that the same 14-character alphanumeric string to uniquely identify a worker for AMT is also used to uniquely identify Amazon customers across all Amazon properties. In particular, anyone who writes an Amazon product review, rates a product, etc., and publicly shares personally identifying information (PII) in their associated Amazon Profile, can be directly linked to any work they perform on the AMT platform. And this linkage is exceedingly easy to exploit: just type any AMT WorkerID into a standard Web URL for locating Amazon profiles[5].

Our team stumbled upon this finding while assembling a collection of AMT datasets for the purpose of studying longitudinal aspects of AMT Worker behavior over time. To enlarge our existing data collection, we decided to try searching the Web for additional AMT datasets which included outputs from others Workers whose WorkerIDs were already present in our existing collection. Imagine our surprise when, after entering a given WorkerID as a search query, the returned search results contained not only the Worker's name but even their picture! When we announced our unexpected discovery to fellow researchers at our CrowdCamp workshop, the initial reaction was one of audible gasps and disbelief, followed by stunned silence. Of all the thirty-some highly-educated researchers in the room, many of whom had used AMT regularly for years, no one had known. Our impression is that this aspect of AMT's design has been there since its inception eight years ago, and it is quite remarkable it has not been more widely noticed before now.

As we reflected upon the implications of this simple attack for revealing AMT Worker information, we began to wonder how many workers could have (or already had) disclosed PII or other private information as an Amazon customer, without realizing it could also be linked to their AMT Worker account? Of course not all Workers who have Amazon profiles have made them public, some may have shared falsified names or pictures, and the amount of other personal information shared will naturally vary per person. Because this form of identifiable human subjects data is sensitive and requires regulation by IRBs, we restrict ourselves to speculation for now. We thus *hypothesize* that perhaps 50% of Workers have profiles, maybe 50% of those include real names, and perhaps 50% of those also include pictures. If confirmed, such numbers would certainly suggest cause for concern that real crowd worker PII is being exposed. Of course, a hypothesis is not established fact. We are now seeking IRB approval to collect specific statistics to answer these questions, which would provide data-driven estimates

[4] `crowdresearch.org/crowdcamp`

[5] `www.amazon.com/gp/pdp/profile/<WorkerID>`

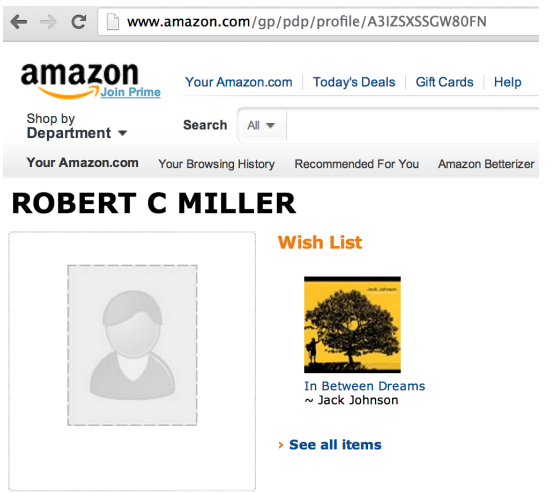

Figure 2: Amazon customer profile corresponding to the AMT WorkerID shown in Figure 1.

of the extent and severity of Worker PII or other private information which can be linked from Amazon profiles to specific AMT Workers.

It is important to restate that all of the Worker PII in question is only that which the Workers themselves have chosen to voluntarily disclose in their Amazon Profiles. Moreover, AMT does not provide a public listing of AMT WorkerIDs that can be exploited in this fashion; one must identify WorkerIDs directly using AMT as a Requester, or discover them via some other method, such as online listings of WorkerIDs provided by others (§4.1.3). The situation would seem to exemplify wider concerns about data mining at large in our information age. While neither the Amazon Profile or the WorkerID are individually a problem, it is still unsettling whenever such linkage occurs in ways that seem surprising to users and their understandings and expectations of system behavior and privacy protections.

4. IMPLICATIONS OF EXPOSING PII

4.1 Worker Perceptions of their Anonymity

To appreciate the potential ramifications of potential PII exposure by AMT Workers, it is important to understand current AMT Worker expectations and perceptions of their privacy. To investigate this, we pursue three approaches: 1) How does AMT itself shape Worker perceptions (e.g., system design aspects such as identifying workers by alphanumeric IDs, statements made by AMT personnel, and stated privacy protections and policies); 2) How do AMT Workers describe their own perceptions when surveyed; and 3) How have AMT Workers described their beliefs about their privacy in online forums?

4.1.1 How AMT Shapes Privacy Perceptions

To begin, AMT's Worker FAQ assures Workers that "Amazon Mechanical Turk treats your information with care. We protect the security of your information..." [3]. Moreover, the AMT Privacy Notice [2] states, for example, that, "Amazon Mechanical Turk knows that you care how information about you is used and shared, and we appreciate your trust

Current Perceptions	**Can the Requesters you work for on Mechanical Turk find out your...**	**Yes**	**No**	**Not Sure**
	first and last name?	196	475	329
	current location (e.g., city or region where you are working)?	484	271	245
	home address?	74	726	200
	email address? (Assume you have not emailed the Requester).	345	440	215
Values	**How much do you value the anonymity of your...**	**Median**	**Mean**	**Std. Dev.**
	first and last name on MTurk?	6	5.44	1.62
	current location information on MTurk?	5	4.96	1.72
	home address on MTurk?	7	6.31	1.22
	email address on MTurk?	5	5.04	1.69
Potential Reactions	**Do you expect you would continue working if current policy was changed so that your full name was available to Requesters on MTurk?...**	**Count**		
	Yes, I would continue to work the same amount	319		
	Yes, I would continue to work, but not as much.	91		
	I'm not sure if I would continue to work or not.	217		
	No, I would not continue to work on MTurk.	43		

Table 1: Worker Survey Results.

that we will do so carefully and sensibly. . . .". The document's section entitled "How Secure Is Information About Me?" provides assurance of AMT's use of appropriate technological standards to safeguard Worker PII: "We work to protect the security of your information during transmission by using Secure Socket Layer (SSL) software, which encrypts information you input." In the AMT official blog, the Worker who moderates the Turker Nation discussion forum[6], *SpamGirl*, cites AMT's Requester polices [5] in expressing a concern that a posted task was in violation of stated policies by requiring Workers to register for another site. The official AMT response to SpamGirl indicates that because "[the Requesters] provide the credentials for registration, there's no violation of the registration – the intent of that rule is to prevent exposure of PII..." [7]. Overall, AMT documentation appears extremely consistent in providing Workers with strong assurances of their privacy in using the platform.

On the other hand, Amazon has never used the term "anonymous" in platform documentation or policies (this point will be further discussed in §5.2). Moreover, a Worker uses the same login credentials to access both their regular Amazon account and the AMT platform, and proceeds of their AMT work can be used to make Amazon purchases from this same account. The crux of the question would seem to be whether or not Workers realized that their WorkerID was used in the URL of their Amazon profile in a way that could be directly accessed or found by Web searches?

4.1.2 Surveying Perceptions of AMT Workers

How do AMT Workers understand their privacy, and what concerns do they have? To begin answering these questions, we surveyed AMT Workers in February 2013 regarding their knowledge and attitudes with respect to the privacy of their personal information on the platform. Workers with a 95% or above approval rating were eligible to complete the 13 question survey for a reward of $0.25. Upon consenting to participate, Workers first selected "Yes", "No" or "Unsure" to four questions querying their knowledge of current privacy policies on the platform for four types of information: "Can the Requesters you work for on Mechanical Turk find out your first and last name; current location; home address; email address, assuming you have not contacted the Requester?" A second set of four questions asked Workers to rate the degree to which they valued the anonymity on Mechanical Turk of each of the same four pieces of personal information using a 7 point likert scale where a choice of "1" equated to "Don't Value at All" and a choice of "7" equated to "Value a Great Deal". Workers then answered optional demographic questions about the country from which they worked, their gender, the total duration they have been a Worker on Mechanical Turk, and their education level. A final question inquired how the Worker expected to react if she learned that Requesters could identify her by name on the system, with choices ranging from continuing to work the same amount to no longer working on the system.

[6] `www.turkernation.com`

What country are you working from?	**Response**	**Prevalence**
	USA	714
	India	226
	Other	60
What is your gender?	**Response**	**Prevalence**
	Male	659
	Female	326
How long have your been working on MTurk?	**Response**	**Prevalence**
	Less than 1 week	79
	Several weeks	222
	Several months	207
	6 months to 1 year	214
	Over 1 year	274
What is your highest achieved education level?	**Response**	**Prevalence**
	High school	98
	Some college	355
	Bachelor's degree or equivalent	434
	6 months to 1 year	214
	Graduate school	105

Table 2: Survey Demographics.

Results: The results along with the survey questions are shown in full in Table 1, with demographic information shown in Table 2. 1000 Workers participated in the survey. Consistency was checked where possible, for example by ensuring that Workers who had selected "Don't Value at All" for their valuation of the privacy of their name, or Workers who had responded "Yes" when asked if their name was currently available to Requesters, did not also select the option indicating they would quit working on the system if this information was made publicly available. No inconsistent responses were found. The results indicate that Workers are most likely to believe that Requesters can access their general location information (Yes: 484, No: 271, Unsure: 245), yet are much less likely to believe that Requesters could find out their home address (Yes: 74, No: 726, Unsure: 200), followed by their first and last names (Yes: 196, No: 475, Unsure: 329) and email addresses (Yes: 345, No: 440, Unsure:

215). Workers placed relatively high value on the security of their first and last name (mean: 5.44, 95% CI: 5.34 to 5.54), as well as their city or region location (mean: 4.96. 95% CI: 4.85 to 5.07), their email address (mean: 5.04, 95% CI: 4.93 to 5.14), and in particular their address (mean: 6.31, 95% CI:6.23 to 6.38). When asked how they would react if AMT Requesters could access their full name, the majority of Workers who answered the question indicated that they would continue to work the same amount (319), followed by those who were unsure of whether they would continue to work on the platform (217), those would work less (91), and those would discontinue working on Mechanical Turk (43).

Limitations. A larger response size would enable even narrower estimates for confidence intervals characterizing beliefs of the overall Worker population. Repeating the survey at different days and times would likely increase sample diversity and representativeness. Selection bias is likely due to Workers who like or do not like to take surveys vs. perform other available marketplace tasks. There may also be specific Worker populations that are may be more or less concerned about the issue, more or less likely to inadvertently disclose PII or other private information, and more or less at risk from such a disclosure should it occur. Survey questions may not have been fully understood and could potentially be rephrased, or revised to assess different beliefs.

4.1.3 Forum Discussions Regarding Privacy

We also searched online discussion forums popular with Workers to learn how Workers have expressed their beliefs regarding their privacy as AMT Workers. To the best of our knowledge, there are three primary forums where such Worker discussion takes place: Turker Nation (mentioned earlier), mTurk Forum[7], and Turkkit-Reddit[8]. In searching the forums since finding the defect, we have learned that one year ago, AMT Workers report finding exactly the same vulnerability which we have only now "discovered" [63]. Their discussion thread tells a strikingly similar tale of discovery to our own. The Turker Nation moderator, SpamGirl, reported having found an AMT dataset online containing WorkerIDs and suggested that, "... everyone should google their worker ID to see what's found." One of the Workers' responses succinctly summarizes the entire problem:

> Googling my ID has 3 results. One shows my Amazon product reviews (had no idea they were linked to my Turker ID), 2nd shows my Amazon profile (includes my first initial and last name, date of birth, location, and all the products on my Amazon wishlist)... This makes me highly uncomfortable and I have no idea what to do about it... I was able to make a few adjustments regarding my ID linking to my public Amazon profile and what information shows: removed my name, changed my DOB to private, removed my location, and changed my recent purchases and my wish list to private. Now my ID only links to my product reviews. It still shows my name in the search result if I google my ID but I'm too late to do anything about that. I should have been more careful with the information before now. I don't think I had ever visited my Amazon profile until today to see what information was public.

Posts in the discussion thread by numerous other works further elaborate upon the issue:

- "...I guess my reviewer profile is linked to my Mturk number! I had no idea that would be the case..."
- "It's starting to look like Amazon is the culprit. They're using the same id's for AMT, and other Amazon sites."
- "...I'm kind of irritated that my worker ID is associated with my Amazon product reviews..."
- "... Amazon needs to separate the Mturk numbers from seller numbers to protect our privacy.."
- "I think this is outrageous though. Makes me concerned about trusting privacy agreements."
- "Mine pulled up my Amazon wish list which revealed my identity. It seems to me that so called "anonymous" tasks on mTurk (like surveys) are not anonymous after all."
- "This is NOT at all GOOD news . . . for academic surveyers. And how many people are going to do their lovely surveys from now on - when the data they collect is so easily mine-able..."
- "Not a big deal to me. If they start releasing demographic information with the ID then I would be upset. What can someone do with my user ID?"

One conclusion we draw from these comments is that at least some Workers have been aware of this issue for at least a year. It is not clear how widely their findings have been shared with the rest of the AMT Worker population. Nonetheless, it would seem AMT Workers anticipated much of what we academic researchers are only now recognizing. The Workers not only identified the same vulnerability via Profile linkage, but considered many possible implications of this exposure, their own uncertainty regarding possible ramifications, potential corrective measures (e.g., changing privacy settings, removing information from one's Profile, or opt-ing out of using AMT), and practical challenges involved in enacting such corrective measures.

It is premature to assume that comments found in a single discussion forum thread are accurately representative of Worker beliefs at large, but such comments at least suggest some cause for concern. Our later reflection on Requester responsibilities (§5.3) further discusses both these comments and others we have found in online AMT Worker forums.

4.2 Potential Risks to Workers

What are the ramifications of this vulnerability of AMT Workers? To the extent this linkage is indeed surprising to Workers, there would seem to be some unexpected loss of privacy, seemingly at odds with the aforementioned messaging by AMT regarding protection of Worker privacy (§4.1.1).

To consider a simple example of how this might could change the future landscape of AMT work, recall Andy Baio's famous "Faces of Mechanical Turk" [9]. Baio wondered about the Workers, "... who are these people? ...what do these people look like, and how much does it cost for someone to

[7] `mturkforum.com`
[8] `www.reddit.com/r/mturk`

reveal their face?" Baio posted an AMT task asking Workers to submit a picture of themselves holding up a note stating why they worked on AMT. Thirty Workers answered the call, and Baio published a mosaic of their pictures, putting a collective face on AMT's workforce. While Baio's task was clearly contrary to AMT's prohibition of collecting PII from Workers, it still required voluntary participation: there was no way for Baio to have obtained these Workers' pictures without their express consent. Fast forward to today: instead of 30 Workers voluntarily providing pictures (without names), a Requester could now automatically generate such a mosaic with potentially many thousands Workers, with each picture linking to that Worker's profile. By posting their pictures on their Amazon profiles, did Workers intend to provide similar consent for a Requester to collect their PII in this fashion? Again, we suspect many Workers would be surprised to have their faces show up in such a photo collage, identifying them as AMT Workers. Perhaps they intended and believed their work on AMT was private, or even that others' knowledge of this could jeopardize the Worker's other employment?

One of the very laudable strengths of AMT has been its clear intent to support blind hiring practices online, where withholding of demographic information precludes the possibility of discrimination based on race, gender, etc., and where the value of an individual's contribution is assessed entirely upon the quality of his or her work products. This aspect of AMT has provided technophiles with a triumphant story of the Internet enabling us to gain traction on an age-old, otherwise entrenched problem of social injustice. With crowd worker identities no longer safeguarded, what other safeguards might be used instead to guard against this risk of discrimination practices in online work?

Crowd workers already face various risks in performing online work that some Requesters may not be aware of. Risks include task-directed installation of malware on their computers, performing tasks which violate the ToS of other sites (e.g., writing disingenuous reviews), being engaged in tasks which involve illegal or unethical behaviors for which they may lack the sufficient context to recognize, or disclosing PII [62]. Exposure of PII in ways Workers have not anticipated could give rise to other, new forms of exploitation by Requesters or other malicious parties which is exactly what Amazon has tried to protect workers from by prohibiting Requesters from collecting PII [5]. A Requester unhappy with a Worker's performance could use the worker's PII to damage their reputation in other online venues or even attempt to track down the Worker's physical location.

4.3 Potential Risks to Requesters

Most visible to academic researchers, exposing PII of crowd workers has clear consequences for inclusion of AMT Workers in human subjects experimentation for university studies. In the USA, human subjects research at universities is overseen by Institutional Review Boards (IRBs)[9] that ensure research activities expose human subjects to minimal risk. Across the USA, many academic researchers have communicated to their IRBs that AMT Worker PII is unavailable, thus minimizing risk to crowd workers participating in research studies. For example, consider the following excerpts from what we believe to be a fairly typical IRB protocol for research conducted on AMT today:

- "While we will include an optional demographic form with our online task, Mechanical Turk provides only worker IDs, so there is no way for Requesters to gather information on a worker's identity."
- "[email, address, and/or telephone lists]... will not be obtained by the researchers. The identification of workers on Mechanical Turk is kept anonymous . . . "
- "Will the study team access any data that is linked to a subject's identity by name or other identifier or code?" "No"
- "Explain how the subjects' privacy will be protected." "Mechanical Turk is designed to prevent work Requesters from obtaining any personal information."

Such claims would now appear to be on rockier ground, meaning that existing IRB protocols for studies proposed, or already approved protocols for studies underway, would need to be reviewed and likely revised by researchers and their IRBs to reassess risk to the crowd workers involved in the studies. We further discuss this issue in §5.1 below.

Looking ahead, how this AMT vulnerability affects regulation of human subjects experiments using AMT and similar systems will naturally depend upon the subsequent actions taken by AMT. For example, were AMT to break the linkage between Profiles and AMT Worker accounts, it is possible that existing IRB protocols could continue to remain in their current form without revision after all. Nevertheless, this lapse in privacy protection, whether intentional or not, could still serve to reform future regulation of human subjects research using AMT or other crowdsourcing platforms. For example, the breach of privacy information may compel IRBs to approach AMT studies using the same regulations currently applied to undergraduate research pools of identifiable participants. New studies may have to be more conservatively designed. Progress of human subjects research may be diminished, and new discoveries may be impeded by PII exposure neither intended nor desired.

It should be noted that not all research involving AMT is human subjects research and therefore subject to IRB governance. Researchers should consult their own university's IRB documentation and personnel to make such a determination about whether proposed research does in fact constitute human subjects research. That said, the greater risk for harm to AMT Workers, via potential PII disclosure, is likely to make universities more conservative in their governance and determinations. As we reflect below on the responsibilities and mistakes of academic Requesters to date with regard to safeguarding crowd worker privacy (§5.3), we note there is cause for concern even with regard to research which likely was not formerly considered to be human subjects research.

4.4 Potential Risks to AMT

The aforementioned risks to Workers or Requesters of course may impact their use of AMT. Moreover, upsetting the balance of one value associated with the design and/or use of a technological system (e.g., privacy) can alter the realization of other interrelated values: security, credibility, trust, and community, to name a few. Perhaps we face a decrease in the workforce as crowd workers lose trust in AMT and stop working for the platform or online altogether. This might compound with other fears that AMT may no longer

[9] en.wikipedia.org/wiki/Institutional_review_board

be accepting new international crowd workers, which AMT personnel have denied [8]. Since the success of crowdsourcing ultimately rests entirely upon its human capital [37], what might this mean for the ongoing viability of the platform to meet Requesters' needs? Could there be a ripple effect outward from AMT to the larger market, with crowd workers on other crowdsourcing platforms becoming concerned by their the risk of PII disclosure on other platforms as well?

AMT is far from the first case of a company inadvertently disclosing its users' data. As discussed in an earlier study [69], the Federal Trade Commission (FTC) has recently begun to aggressively protect consumers from data breaches by commercial entities, including scrutinizing the release of supposedly "anonymous" data (also see [50]). In 2005, the FTC found that BJ's Wholesale Club violated the "unfair or deceptive practices" standard by failing to adequately protect its customer records from thieves. Shortly thereafter, the FTC filed a similar complaint against DSW when hackers broke into the company's database. The agency found that DSW failed to protect its customer's private data and thus violated the deceptive acts prohibition. Then, in 2006, the FTC extended its reach even further in a complaint against CardSystems Solutions (CSS). CSS provided businesses with products that authorized credit card transactions. The FTC found that CSS violated privacy regulations by failing to protect the personal information it collected by storing data in an insecure format, failing to assess the vulnerability of its system, and not implementing strong protections against hackers [56].

What is more, the FTC also regulates how businesses treat supposedly anonymous user data. Recently, some companies have found that they can source innovative business ideas by sharing user information to the crowd. In 2006 AOL released data from 650,000 users and 20 million search queries to support research. While the company attempted to anonymize the data, a New York Times article revealed that one could still find the identities of individual users [10]. In response, the Electronic Frontier Foundation (EFF) filed a complaint with the FTC, requesting it act against AOL [23]. AOL ultimately fired the individual responsible and effectively shut down its research division [49].

Later in the same year, Netflix released one hundred million anonymized user records as part of its "Netflix Prize" Contest. In this contest, the company offered one million dollars to the first team to significantly improve Netflix's movie recommendation algorithm [11]. The initial contest was so successful the company decided to hold another one. However, two researchers discovered it was "surprisingly easy" for a malicious party to use Netflix's data, combined with a little other information, to find the identities of the users in the dataset [47]. Soon thereafter, a class action suit was filed against the company and the FTC entered the picture. Fearing legal troubles and agency pressure, Netflix cancelled its formerly scheduled second contest [49].

Maureen Ohlhausen writes that the FTC's views on data security have evolved from a "notice and choice" approach, where an online business would remain safe by adhering to its stated privacy promises, through a harms-based model, to today's hybrid approach [48]. In 2010, the agency proposed the new hybrid framework for protecting consumer privacy, broadening its scope even further. Now it applies to all commercial entities that collect information from consumers, online or offline, whether they interact directly or indirectly with consumers. The approach includes "any data that can reasonably be linked to a specific consumer, computer, or other device" (page 44). Instead of focusing on privacy promises, this model looks at company actions likely to cause physical or economic harm or intrude into the lives of their customers.

The case here of AMT appears to be most similar to that of Netflix, in that once more it would appear to be "surprisingly easy" for a malicious party to exploit this vulnerability in order to obtain PII of AMT Workers which they had not realized was linked from their Amazon profiles. Nevertheless, Amazon appears to have had the best of intentions in protecting crowd worker privacy, and its Privacy Notice [2] informs Workers, "What Choices and Access Do I Have? ...you can always choose not to provide information..." Nevertheless, did AMT Workers have a reasonable expectation of privacy in their roles as Workers based on their understanding of AMT's stated documentation and policies, as well as Workers' own use of both AMT and Amazon's other properties? The question remains as to how AMT and the FTC might act in response to this seeming surprise of PII exposure.

4.5 A Silver Lining to PII Exposure?

While unexpected exposure of AMT Worker PII would seem to be problematic, what if the element of surprise were removed and the current system were to persist as is? Let us imagine AMT simply updates its documentation and policies to further clarify the issue, notifies Requesters and Workers of the changes, and allows Workers time to revisit their choices about what personal information they choose to share in their profiles. Might there be any potential advantages to such a new version of AMT vs. the version to which we have been traditionally accustomed?

Certainly there are many market or research studies for which demographics of participants are important to know. For example, uTest (`utest.com`) requires its workforce to provide demographic information during registration. uTest customers specify test requirements such as demographics, OS, browser, etc., and uTest proceeds to identify and invite qualified testers from its online community. With a seemingly anonymous workforce on AMT, controlling for demographics has been far more difficult, requiring Requesters to solicit demographic information directly from Workers [55, 30] for tasks like remote usability testing [40]. To some extent, Amazon Profiles may offer the opportunity to obtain some measure of demographic information directly, or to corroborate demographic information reported by participants via their Amazon Profiles. Of course, it is unclear how many AMT Workers might continue to share demographic information after the change was announced, which may compromise the magnitude of this potential benefit.

There have also been many studies on best practices for quality assurance with AMT, and concerns of multiple-identity Sybil attacks [39] or use of automated routines (robots) to perform work instead of individuals performing the work themselves (akin to software robots playing online poker). Various studies have sought to correlate work quality with demographics, or to detect, prevent, and mitigate cases of worker fraud, such as those cited above, by verifying each AMT WorkerID in fact corresponds to single, real person [28]. Associating an Amazon Profile with each Worker account

would provide further corroborative evidence.

Of course, as popularized in the recent film *Catfish*[10], an online profile does not a real person make. Again, however, the potential benefit may provide only short term rather an enduring protection against online fraud or poor quality work, as spammers modify their Profiles once notified of the change. Moreover, attempts to assess or predict work quality on the basis of demographics would seem to be ideologically at odds with AMT's long-standing goal of preventing discrimination online [31].

Another intriguing possibility is that anonymity itself may fundamentally contribute toward poor quality work, leading to ephemeral working relationships in place of more enduring ones. For example, some research has reported that establishing long-term relationships with Workers made it unnecessary to apply many of the statistical procedures commonly used on AMT for quality assurance [37]. oDesk (`odesk.com`) workers are similarly known to those who hire them for often longer-term relationships.

The idea of such alternative crowd work models inspired a past pilot project [36] which explored inviting AMT Workers to voluntarily de-anonymize to establish trust relationships, advertise skills and expertise, and gain access to higher paying work and economic mobility, etc. At the time, this was achieved by letting AMT Workers link their WorkerID to an external identity source (e.g., OpenID, Facebook, Google+, etc.). Now, however, Amazon has already provided every Worker a built-in opportunity to voluntarily de-anonymize by creating an Amazon Profile page, where such skills and expertise could be listed. Amazon Profile pages could become the LinkedIn equivalent for crowd workers engaged in micro-work. Of course, such a move would mark a significant pivot in AMT's existing platform strategy: besides exposing PII, this could compromise the programmatic efficiency of online micro-work under their formerly anonymous workforce. Of course, such a decision need not be all or nothing: Workers without Profiles could continue to perform entry-level data processing, and those with skills might more easily gain access to work requiring greater expertise. A diverse ecology of work and skills might come to flourish, in time.

5. THE SEARCH FOR A SOLUTION

5.1 Potential Next Steps

Should AMT wish to prevent this vulnerability from being exploited, the most obvious solution would be to issue unique IDs to AMT Workers which are not linked to other Amazon properties or online profiles. While such a fix would seem to be consistent with AMT's communicated intent to safeguard Worker PII, it is unclear how difficult or expensive it would be in practice for Amazon to enact such a change, how many of their other systems would need to be updated in response, and how long this would all take to accomplish. Work history and *Approval Rating* would need to propagated to each Worker's new WorkerID. A host of other issues to which we, outside the company, are not aware of could preclude Amazon from pursuing this. Our impression from communication with AMT regarding this issue is that such a change is unlikely to occur [15].

Given this, it would seem to be a shared responsibility to ensure Worker and Requester communities are informed of this vulnerability and take appropriate actions. For Workers, Amazon's profile system allows individuals at any time to delete their PII or to mark their entire profile as private rather than public. We therefore advocate immediately informing and educating AMT Workers about this issue such that they can make informed decisions as to how to best protect their own PII. Of course, it is important to recognize this may be more easily said than done. It may be difficult to effectively communicate the issues involved such that Workers feel sufficiently comfortable to fully appreciate the issues and potential ramifications at stake so that they can make informed decisions. One intriguing idea would be to develop a site like *PleaseRobMe*[11], which promoted awareness of how Twitter users discussing their travel could be linked to their home addresses and exploited for home robbery. Similar design activism here would seek to inform by provocation, exploiting this vulnerability in an obvious way online so Workers could more fully appreciate nature of the vulnerability.

A related issue is also ensuring that Amazon's interface for making Profile changes is sufficiently simple to use that it is easy to protect one's PII, should they choose to do so (just think of all the reported problems of people trying to control their privacy settings on Facebook). The aforementioned example described one Worker doing this with no reported problems (§4.1.3).

This solution would also require Workers to withdraw from benign social interactions in using Amazon's main site in order safeguard their roles as anonymous crowd workers, and some Workers may not be willing or recognize the need to make this social sacrifice in order to protect themselves. Another option, therefore, would be for Workers to create a new, secondary account, using one account for their Amazon social interactions and shopping, and another for their role as an AMT Worker [15]. As noted above, however, it would not be possible to port one's Amazon or AMT history from an old account to a new one without AMT support. The Worker would need to remember login credentials for both accounts, and be careful not to accidentally use the wrong account for the wrong role. It is not clear if funds earned on AMT under one account could be easily transferred to the other, of if the Worker would need to make purchases with AMT funds using his AMT account, but switch to his other account to write and share his product reviews for such purchases?

There also remains an immediate and ongoing risk to any currently exposed Worker PII until each Worker acts to protect his or her own PII. Moreover, Workers are not the only ones affected by this. On one hand, this likely impacts many currently approved IRB protocols for human subjects research using AMT. On the other, we note that Requester PII could be exposed by the same vulnerability, i.e., if a Requester did not realize his RequesterID was also linked to his Amazon Profile. Given that Requester account names are already publicly available, however, we consider this last concern to be relatively minor.

Regarding IRB, individual researchers and institutions can one-by-one revisit and revise their existing protocols to ensure that crowd workers involved in human subjects research are sufficiently safeguarded. Levels of risk to subjects vary depending on the nature of the data. For example, data

[10] `http://en.wikipedia.org/wiki/Catfish_(film)`. 2010.

[11] `pleaserobme.com`

on visual judgments (e.g., [27]) is likely to be regarded as less sensitive than study records containing data on people's most private moments in the home (e.g., [16]). Some researchers may wish to take precautionary actions without waiting for an AMT response, such as by mapping WorkerIDs to a new set of identifiers and removing any records of the original IDs and the mapping.

5.2 How Surprising Is This, Really?

5.2.1 Why Should We Be Surprised?

There seem to be several good reasons for being surprised. We began this article with excerpts from a variety of published articles that explicitly described AMT's workforce as being anonymous. Moreover, we reported the collective surprise at our CrowdCamp scientific workshop when we announced the vulnerability in AMT's design we had found to the other researchers in attendance. We also cited discussion of anonymity in the official AMT Blog, as well as a variety of other language more suggestive in nature. As one more example of this, consider a thread from AMT's official discussion forum[12] entitled *Worker Privacy/Anonymity* [1] in which a Requester states, "I don't want to know the real names of workers, nor do I want to have the ability to find out. I want providers to know that their responses are anonymous...". This Requester's question desire seems to mirror many academic Requesters using AMT for human subjects research. The response from AMT personnel in this case was that, "Personal information about a worker is not available to a requester. From your point of view, the identifiable information for a worker is going to simply be a unique id." While such PII is indeed not available on AMT, we have seen it might be obtained elsewhere if the Worker has chosen to share it on his Amazon profile.

5.2.2 Why We Should Not Be Surprised

There are a variety of very good reasons why perhaps Requesters should not be so surprised after all. For example, we mentioned earlier that Workers identified the same defect a year ago, suggesting our community has not paid sufficient attention to AMT Worker forums. Moreover, we mentioned that unintended exposures of private information have become increasingly common in today's environment of big data collection and networked systems (§4.4).

Another point worth considering is that to the best of the authors' knowledge, AMT's Website and policies have never used the term "anonymous" to characterize its workforce. Even in the example above, the term "anonymous" appears multiple times in the Requester's question but never in the AMT official response. Searching the AMT site for the term "anonymous" returns only a single unrelated result, and historical searches performed by the authors using Internet Archive[13] site snapshots similarly fail to produce any evidence of the workforce being described as anonymous. While AMT may not have gone out of their way to issue any press releases attempting to correct public misperception about such anonymity, it would seem that misapplication of the term "anonymous" began with users rather than with AMT itself. And it is well beyond the purvue of AMT to try to track and correct every inaccurate characterization of its platform by others not in its employ.

[12] `https://forums.aws.amazon.com/forum.jspa?forumID=11`

[13] `archive.org/web/web.php`

Another reason for lack of surprise is AMT's support for email exchange between Workers and Requesters without any anonymizing proxy. While Requesters cannot directly view workers' names or email addresses, a "NotifyWorkers" platform function takes a list of WorkerIDs as input and sends a regular email to each Worker from the Requester's email address. As the developer documentation describes it [6], "The NotifyWorkers operation lets you... send a Worker a message without having to know his or her name or e-mail address." The problem, however, is that while the Requester need not know this information beforehand, responses from notified Workers typically occur outside of AMT's purview via regular email channels, thereby disclosing the Worker's email address and typically their name (i.e., PII). Moreover, even if a Requester never attempts to contact workers in AMT, Workers can and often do send unsolicited email to Requesters via AMT, in order to clarify an aspect of submitted work, or to ask for more information about a task or details of payment, among other reasons. We should note that AMT is reported to provide a worker-equivalent form of privacy-preserving communication from Workers to Requesters, though we have not yet received such a communication in practice. Ideally, Amazon should provide an easy way for Workers to reply by email or an embedded link in a Requester message via such a privacy preserving proxy.

5.2.3 PII Disclosure for Tax Reporting

Another strong argument against AMT's workforce being construed as anonymous is its well-documented and oft-discussed compliance with US law enforcement and tax reporting requirements. While the former does not concern PII disclosure to Requesters, the latter does. First, let us momentarily put aside recent 2011 changes in Section 6050W of the Internal Revenue Code [19] (with corresponding changes in AMT policies). Instead, we first concentrate on how AMT tax policy prior to this may have contributed toward popular perceptions. First, recall that AMT's Participation Agreement [4] defines "Provider" as a synonym for being an AMT Worker: " "Provider" means you, if you use the Site to perform Services for a Requester." It further states:

> . . . In addition to the disclosures described in our Privacy Notice, we may disclose to Requesters your name, address, data on HITs you have completed, and Provider Tax Information... such as a Social Security Number or Employer Identification Number. You hereby consent to our use and disclosure of Provider Tax Information...

Similarly, AMT's Privacy Notice [2] states:

> We release account and other personal information when we believe release is appropriate to comply with the law... if you are a provider of a service on the Amazon Mechanical Turk site, we will release your name and address only to requesters for whom you provide services so that those requesters can comply with tax and other legal obligations they might have.

Thus, Requesters must have access to Worker PII for compliance with IRS reporting requirements. How does AMT communicate Worker PII to Requesters in practice?

While AMT could provide Requesters with this PII for every Worker who ever works for them, it seems this PII is only shared once Workers cross an IRS-specified threshold requiring tax reporting [3]: "... information is only provided to Requesters who have provided an EIN and for whom you have exceeded the IRS tax reporting threshold." With regard to Worker anonymity, this would mean that so long as a Requester's payments to a Worker stayed below this threshold, AMT would not disclose the Worker's PII. However, once this threshold was exceeded, the Requester would have a legal obligation to obtain this PII in order to file the appropriate paperwork. Perversely, this also means a Requester could, without violating AMT policies, obtain any Worker's PII from AMT by simply paying the Worker $600 (though as noted earlier, some workers appear willing to disclose their PII directly for far less compensation). In short, while Worker PII clearly was obtainable via tax reporting requirements, it seems this too did little to dislodge the popular notion that AMT's workforce was anonymous.

As briefly mentioned above, AMT's tax reporting policy was recently changed:

> Dear Amazon Mechanical Turk Worker, The IRS published new regulations under section 6050W. As a result of this change, Amazon Payments, as the third party payment network provider for Mechanical Turk transactions, is responsible for reporting transactions that meet the IRS thresholds via Form 1099-K, this reporting includes Worker earnings. As a result of this change, Mechanical Turk will no longer provide the Mechanical Turk tax report to Requesters. [46]

While this would seem to suggest PII will no longer be disclosed to Requesters for tax reporting purposes, reality is not nearly so clear. First, neither AMT's Privacy Notice nor its Participation Agreement have been revised for this change in IRS tax reporting requirements. AMT personnel have communicated to us that while their interpretation of 6050W indeed means Requesters not having to report Workers' income, apparently some Requesters disagree and still value AMT's willingness to provide this tax reporting information to them. Thus it appears AMT will not be revising its PII disclosure policy in regard to 6050W changes until such time as definitive regulatory clarity is clearly established [15].

5.3 Assuming Responsibility as Requesters

Thus far, this article has predominantly focused on AMT's role in regard to safeguarding crowd worker privacy. We turn now instead to an equally important consideration: that academic Requesters assume responsibility for any of their own actions that may potentially compromise crowd worker PII. Mistakes can always occur despite good intentions; what matters is that we recognize and learn from any such mistakes. A key observation, though, is that while Requesters can only propose actions for AMT to consider, we have full control over our own choices and behavior. With this in mind, we use this experience as an opportunity to review our own understandings about what constitutes appropriate actions with regard to protecting Workers' personal information.

5.3.1 Sharing Data Online

At the end of the Worker Perceptions section above, we mention an AMT dataset containing WorkerIDs that Turker Nation moderator SpamGirl found online. In fact, this data was posted by one of the authors of this paper as part of a sourcecode repository. While the author deleted the data long ago to correct the mistake, there appears to be no way with the particular source control site being used to remove entire file histories (except perhaps by deleting the entire open source project and repository). Thus, as demonstrated by earlier Worker discussion forum comments (§4.1.3), information can be remarkably difficult to purge once released online. To make this particular situation worse, the data in question contained not only WorkerIDs, but also evaluatative information associated with a subset of those WorkerIDs, "RejectedTurkers: A list of turker IDs who are doing substandard work and might need to be blocked." While a Requester may certain utilize such a list internally as one component of quality assurance, disclosing such a list may lead to other Requesters blocking a particular Worker, potentially impacting that his ability to obtain future work.

Further back in time, Workers found in October 2011 that a different author of this paper had shared an AMT dataset that also included WorkerIDs (though without any evaluative characterizations) [64]. While the dataset in question was thought to be innocuous (a single A/B poll question with no correct answer, simply intended to demonstrate how AMT operates during a live presentation), crowd workers were still angered to find their WorkerIDs posted online. One wrote:

> The good news is, unlike the last time this crap happened, there's no "extra" identifying information beyond the worker ID–things like income levels, country of residence, etc. Still not cool at all, however. Have you emailed him and/or his department head explaining that a worker ID counts as identifying information and is not to be included in supposedly anonymous survey results?

Another Worker wrote, "It would be nice for him to tell others to not make the same mistake." (SpamGirl contacted the author at the time and requested that the dataset be taken down immediately, and it was).

Moreover, we have focused in this article on the central problem being the linkage to Amazon Profiles exposing Worker PII. However, even before Workers discovered such linkage they still expressed significant concerns about their WorkerIDs being disclosed in general. Even if a WorkerID does not identify the real life person, it does uniquely identify his crowd worker persona, as the last example highlighted: "...a worker ID counts as identifying information and is not to be included in supposedly anonymous survey results." This seems to be a widespread occurrence by many researchers, and those AMT Workers who have responded appear to be unanimous in objecting to it. In our review of the earlier post cited above [63], we recognize at least three other academic researchers (not authors of this paper) with whom the AMT Workers have identified similar incidents of online WorkerID disclosure. In addition to this, our own online searches during the CrowdCamp uncovered many more academic datasets online that include WorkerIDs. We recommend such practice discontinue.

Interestingly, one Worker remarks in this same post, "I'm

not sure how this is different than our hall of shame?" That is, Turker Nation posts flagged RequesterIDs, which might be construed as similarly raising ethical questions and/or violating AMT policies. In fact, perhaps considering that Requester names are publicly available, unlike that of Workers, such sharing of RequesterIDs might impact not only their ability to recruit crowd workers, but could also harm their personal reputations. As mentioned earlier, it is well known that many Requesters (including some of the authors) do not use their real names in their AMT accounts. Some Requesters even change their names to rejuvenate an online reputation that has become sullied [65]. One might argue that both parties should be governed by the same AMT policies, and Requesters should abide by the same principles of consideration and respect they so often decry the absence of in spammers. On the other hand, significant differences in power and knowledge held by the two parties also might suggest Workers deserve more preferential treatment at the negotiating table [59].

5.3.2 *Other Lessons*

Given the novelty of AMT, new Requesters regularly make the same mistakes that more experienced Requesters once made, while even experienced Requesters are continuing to learn themselves (as the examples above and this overall article demonstrates). While academic Requesters often discuss building new software packages or identifying best practices in order to improve time, cost, ease, or quality of performing crowd work, relatively less attention has been directed toward disseminating knowledge regarding what mistakes we Requesters have made with regard to crowd worker privacy, as well as what steps we might take to help keep such mistakes from recurring. As such, a recommendation of this paper is that we academic Requesters direct more attention to this issue.

In searching the academic literature for prior studies related to our finding, we became aware of another study entitled "Conducting Usable Privacy & Security Studies with Amazon's Mechanical Turk" [34]. Somewhat ironically, the study investigates feasibility of using AMT for privacy studies (utilizing demographic profiling of the Workers) without any consideration of the Workers' own privacy. We do not mean to single out this study for reprimand, but to the contrary, to suggest this study seems quite representative of many published AMT scientific studies. Because AMT has generated such excitement about what can be accomplished with it, it has sometimes been too easy to forget about the underlying, invisible workforce powering the research community's many innovative projects built atop AMT.

Consider briefly the simple practice of collecting crowd worker demographics. As mentioned earlier, many AMT Requesters have solicited demographic information directly from workers since AMT does not provide it [55, 30]. While AMT's Participation Agreement [4] explicitly forbids Requester use of AMT for "invasion of privacy", does collecting demographics information constitute such an invasion? Perhaps not, since those who choose to participate in such studies volunteer this information. One step further, recall that AMT more specifically prohibits Requesters from collecting any PII from Workers. While broad demographic information is typically not considered PII (describing traits shared by many people), increasingly specific demographic information can be used to uniquely identify people.

Recent AOL and Netflix incidents serve as a stern warning that in our Internet age, with all kinds of data captured about us and made available in ways that are extremely difficult for us to monitor or control, it is hard to anticipate how various data about us might be linked together to generated unexpected PII. In a similar vein, if a Worker discloses PII by emailing a Requester, does failure to purge such emails from one's email archive constitute collection of worker PII? Academic Requesters have ethical and legal obligations to comply with both AMT's ToS and IRB-governance (with human subjects research).

6. BROADER CONCERNS

While we regularly seek to engage in conscientious reflection in drafting our IRB protocols, we must also continually evaluate the potential impact the decisions in our daily lives have upon our fellow citizens. In considering the impact of this particular privacy vulnerability on larger practice, this sections seeks to situation this issue of crowd worker privacy amidst broader ethical, economic, and regulatory issues surrounding the crowd work industry at large, and how academic researchers might best engage with it.

6.1 A New Research Focus?

As seen with earlier outsourcing, global market forces are increasingly moving computer work to regions of the world where it can be completed more quickly and affordably. Crowd work savings arise from increase in labor supply, lower cost of living in other geographic regions, and the ability to decompose work into very fine-granularity units which can be efficiently and affordably distributed. While early demographic studies suggested that crowd work was typically performed for supplemental rather than primary income, subsequent studies have indicated crowd work is increasingly become a source of primary income, especially in developing economies [55, 30]. Relatively low wages, depersonalized work, and asymmetric power relationships have led some individuals to raise ethical concerns that we may be building a future of crowd-powered computing on the backs of exploited workers in digital sweatshops [45, 17]. At the same time, crowdsourcing is conversely being seen as "The New Sewing Machine" [52], creating new opportunities for income and social mobility in regions of the world where local economies are stagnant and local governmental structures may discourage traditional outsourcing firms. Just as consumers can choose to buy fair trade goods or invest in social choice funds, some crowdsourcing services now offer guarantees of worker protections and living wages (SamaSource[14], MobileWorks[15], CloudFactory[16]). Recently filed litigation [58] questions whether individuals engaged in certain forms of crowd work should be legally classified as employees rather than independent contractors under the Fair Labor Standards Act (FLSA), a direction of possible regulation only speculated upon earlier [69].

While optimizing worker behaviors [42] and investigating technological opportunities and challenges is clearly important and valuable work, how might we best wrestle also with questions about what is ethical, legal, and sustainable economic practice in crowd work [59, 22]? For example, is it

[14] `samasource.org`
[15] `www.mobileworks.com`
[16] `cloudfactory.com`

truly preferable to pay people nothing, i.e., having workers play online games which generate work products as an output (which the workers may not be aware of, and which may generate revenue which the workers do not benefit from) [38], than to pay them low wages and clearly communicate they are performing work [22, 21]?

6.2 The Danger of Abstraction

Computer Scientists love abstractions that allow us integrate diverse software modules while encapsulating pesky details of the modules' internal characteristics. AMT's novel design lets us write programmatic function calls that look like we are calling upon any other Artificial Intelligence (AI) subroutine, except that the results often outperform those of our traditional AI modules. Crowdsourcing poses a particular danger with regard to the tendency to abstract: we may forget there are real people behind the abstraction, we may impact their lives in ways that do not penetrate the abstraction, and we may not realize or fully appreciate those impacts we are having. Terminology such as "Human Computation", "Human Processing Units (HPUs)"[20], and "Remote Person Calls (RPCs)" offer conceptually useful (and amusing) ways for us to think about the computation of crowdsourcing, but the opacity and humor of this same terminology may also serve to perpetuate the invisibility of a global workforce that is by its very distributed nature difficult to put a face on. As has been succinctly noted elsewhere, "abstraction hides detail" [59]: some details may be worth keeping conspicuously present.

To help us understand crowd workers in a way that statistics on crowd demographics do not seem to make as visceral to us, Andy Baio created a collage of worker faces [9]. Leila Chirayath Janah, who founded the non-profit SamaSource platform for crowd work, regularly gives talks showing people living in African refugee camps performing online crowd work as one of their only opportunities to earn income and exert some measure of control in otherwise chaotic living environments. As a form of design activism, Lilly Irani and collaborators built Turkopticon, combatting the invisibility of crowd workers and raising collective awareness of worker concerns [59, 32]. Despite such efforts, we still know relatively little about the lives and conditions of the many crowd workers powering today's crowdsourcing applications. When we observe only the work products and remain ignorant of its source, we risk assuming a worker's economic or privacy decision is fully informed and freely chosen. We remain unaware of which worker decisions truly are free and informed.

6.3 An Opportunity to Raise Awareness

While exciting applications of crowd work exist with great potential to transform and advance our society, less ideal uses of crowd work also exist. Consider the case of outsourcing of dirty digital jobs, an online equivalent of the traditional practice of offloading of locally undesirable jobs to immigrant workers. Because it is not acceptable for a social network customer to see a pornographic or violent image posted via social media, we might instead hire crowd workers to sift through such images click after click so we are not exposed to it ourselves [26]. While such a content moderation task should ideally be automated — after all, computers cannot become emotionally disturbed from constant exposure to such imagery — our best state-of-the-art AI software is unfortunately not effective enough to flag all of the "muck" that gets posted online. So we utilize *human computation* instead via crowd work. Just as janitorial work cleans our local restrooms, crowd work cleans our social networks.

In another vein, human history shows us that when one group has power over another, with money or other incentives at stake, the lesser group may be compelled to perform work they would otherwise not choose [18]. The recent case of prisoners compelled to perform gold farming in online games is a modern example [66]. As the industry of crowd work grows, how might we monitor and mitigate the corresponding growth of such practices with at-risk populations?

Much of the general public likely remains unaware of the crowd work industry today, despite the massive amount of online information processing now being performed in this manner. Why should we be concerned by such worker invisibility? When AOL failed to properly anonymize its search logs before releasing them, there was an outcry; ubiquity of search engine use meant most people could personally relate to the plight of those whose privacy was lost. When the Netflix challenge compromised customer PII, there was anger from many Netflix subscribers who could imagine their own privacy being similarly compromised. What sort of public response should be expected if some crowd workers discover they are not as anonymous as they might have thought? After all, what is at stake might be construed as just the privacy of some remote group of Internet workers, engaged in a profession that is difficult to explain, often performing entry-level data processing jobs, and many of whom live beyond America's borders. Will anyone really care beyond those immediately impacted?

6.4 A Lesson in Value Sensitive Design

We suggest broader lessons can be learned from this experience as well. Value sensitive design stresses the role of stakeholder beliefs and values in shaping the behavior with a technological system, and the way in which the system itself can promote or undermine such values [24]. The importance of informed consent is one example that researchers have explored through study of how browser cookies are used and understood [25]. Above, we contribute another case study with relevance to such approaches by identifying various factors that have informed the AMT system and it's common usage. The results clearly show how the act of designing a system such as AMT does not end with those who create the software, but continues to unfold in system use. As discussed above, misunderstandings in Worker and Requester reception of communications by system designers, as well as a lack

of sufficient communication between different user groups (e.g., Requesters and Workers), appears to have contributed to a misperception of AMT Worker anonymity. By untangling the various threads that played a role in the dilemma now faced by Requesters and Workers alike, designers and researchers of crowdsourcing systems have an opportunity to learn from this experience about diverse factors that contribute to the actual realization of crowdsourcing platforms.

A goal in the practice of value sensitive design is often to avoid privileging particular groups of stakeholders over others. In §5.3, we note that various plausible views might be taken with regard to the question of whether one group of stakeholders (Workers, Requesters, system designers) should be privileged in the design of a crowdsourcing platform. While AMT appears to have been designed with the protection of its various stakeholders as a goal, the case illustrates how even subtle imbalances tacitly permitted by system design can significantly influence community dynamics [59]. As researchers devoted to the study of crowdsourcing and human-computer interaction, we possess both training and responsibility to contribute to the discussion of such issues faced by Requesters and Workers alike, and to use what we learn to improve the design of future crowdsourcing systems.

7. CONCLUSION

While Amazon's Mechanical Turk (AMT) online workforce has been often characterized as being anonymous, we have identified an aspect of AMT's system design which can be easily exploited to reveal a surprising amount of information about AMT Workers, often including personally identifying information (PII). We believe this PII exposure may come as a surprise many Workers and Requesters today, as well as impact current institutional review board (IRB) oversight of human subjects research at many universities currently involving AMT Workers as participants.

Our research helps shed light upon the potential multifaceted impact of such PII exposure for each stakeholder group: Workers, Requesters, and AMT itself. We discussed potential remedies each group may pursue, as well as the responsibility of each group with regard to privacy protection. This discussion led us to further situate issues of crowd worker privacy amidst broader ethical, economic, and regulatory issues. Our experience leads us now to propose a set of recommendations with regard to safeguarding worker privacy.

7.1 Recommendations

The ecology of AMT use today comprises a diverse intermixing of at least three stakeholder groups: AMT itself, Requesters, and Workers. We propose related but distinct recommendations for each stakeholder group. As academic researchers, our recommendations for Requesters focuses particularly on this subgroup; others may propose additional recommendations specific to industrial Requesters, or consider other vertical groups of Requesters or Workers.

AMT. Our review of AMT's documentation and policies, as well as statements made by its personnel, suggests AMT deeply values and respects Worker privacy. At the same time, it has never explicitly promised Worker anonymity, nor does it seems particularly interested in doing so now. We nevertheless believe the "attack" we have described represents a serious vulnerability to privacy in AMT's design, at least with respect to Worker and Requester expectations today. Therefore, we strongly encourage AMT to take steps to remedy this vulnerability. An ideal solution would directly address the design issue we have identified which enables this attack to occur. Short of this, AMT may explore avenues for better educating Workers, Requesters, and the public at-large about the AMT's lack of anonymity, as well as specific actions Workers can take to avoid unintentionally disclosing PII which can be difficult to undo afterward.

Requesters. Academic Requesters should also acknowledge their own responsibility to anticipate and manage the impact of their actions upon Worker privacy. Those conducting IRB-sanctioned human subjects research with AMT are further obligated to minimize risk to participants by educating themselves and their subjects about the capabilities and limitations of the experimental apparatus (i.e., AMT) to safeguard participant privacy under their experimental protocol. Even in cases when WorkerIDs are not personally-identifying, they still represent unique identifiers and must be appropriately handled as such (e.g., unrestricted sharing of WorkerIDs online should be particularly discouraged).

Workers. Ultimately, Workers are responsible for safeguarding themselves and protecting their own privacy. Each individual should educate himself to make appropriate, informed decisions regarding what is in his own best interests. In this particular case, Workers should learn and appreciate the capabilities and limitations of AMT in regard to protecting their privacy, and they should take appropriate action to protect themselves. The responsibility of AMT and Requesters is then to make a good faith, concerted effort to support Workers in this vein.

7.2 Opportunity vs. Responsibility

AMT has been truly transformative for both industrial and scientific practice, deeply impacting research directions, enabling new progress on many traditionally difficult problems, and helping us imagine new, innovative applications for better serving our society's needs. Of course, each new crowdsorucing opportunity may also impose an accompanying cost which may be challenging to fully discern in its complexity. An ongoing challenge we face as a community is how to develop a fair, transparent, and consistent practice for fully appraising and balancing what may often be competing interests. With the power to greatly impact the lives of crowd workers around the world, researchers have a moral responsibility to acknowledge our dual role as both stewards and beneficiaries of crowd work, and to understand and appreciate the multi-faceted impact our crowdsourcing activites have upon the lives of the many workers involved.

Conflict-of-Interest Disclosure

The first author has several ties to AMT and the crowdsourcing industry, having: 1) received Amazon Web Services research awards for use of Amazon's cloud computing; 2) having received Amazon sponsorship funds for the National Institute of Standards and Technology (NIST) Crowdsourcing Track; and 3) having spent a mini-sabbatical working in residence at CrowdFlower.

Acknowledgments

We thank the CrowdCamp 2013 organizers for hosting the event which brought the authors together to collaborate, and Sean Munson in particular for his comments on portions of

this work. We also acknowledge the earlier contributions of Stephen Wolfson regarding inadvertent data disclosures by other companies and resulting FTC enforcement. We thank Cathy Marshall for the pointer to Turkkit-Reddit, and Lilly Irani letting us preview her work in preparation [31].

We particularly thank Sharon Chiarella at AMT for making time to speak with us [15] regarding issues raised in this paper, as well as her clear interest in supporting academic researchers using AMT and helping them to identify ways to conduct responsible research given AMT capabilities. Last but not least, we thank the many crowd workers who are powering the innovation which enables much of crowdourcing practice and research today.

This work was supported in part by an NSF CAREER award, DARPA Young Faculty Award N66001-12-1-4256, and a Temple Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors alone and do not express the views of Amazon, CrowdFlower, or any of the funding agencies supporting the authors' research.

8. REFERENCES

[1] Amazon Mechanical Turk Forum. Discussion Thread: Worker Privacy/Anonymity. `https://forums.aws.amazon.com/thread.jspa?messageID=104498&#104498`. Visited March 5, 2013.

[2] Amazon.com. Amazon Mechanical Turk Privacy Notice. `https://www.mturk.com/mturk/privacynotice`. Last Updated: August 5, 2009. Visited March 5, 2013.

[3] Amazon.com. FAQ: Worker Web Site FAQs. `https://www.mturk.com/mturk/help?helpPage=worker`. Visited March 5, 2013.

[4] Amazon.com. Participation Agreement. `https://www.mturk.com/mturk/conditionsofuse`. Last updated: November 1, 2012. Visited March 5, 2013.

[5] Amazon.com. Requester FAQ: Policies. `https://requester.mturk.com/mturk/help?helpPage=policies`. Visited March 5, 2013.

[6] Amazon.com. Requesters and Workers. `docs.aws.amazon.com/AWSMechTurk/2008-02-14/AWSMechanicalTurkRequester/Concepts_RequestersAndWorkersArticle.html`. Visited March 5, 2013.

[7] Amazon.com. Some thoughts on invalid HITs. *The Mechanical Turk Blog*, Decemeber 17, 2010. `http://mechanicalturk.typepad.com/blog/2010/12/some-thoughts-on-invalid-hits-.html`. Visited March 5, 2013.

[8] Anonymous. The reasons why amazon mechanical turk no longer accepts international turkers. *Tips For Requesters On Mechanical Turk*, January 17, 2013. `http://turkrequesters.blogspot.com/2013/01/the-reasons-why-amazon-mechanical-turk.html`.

[9] A. Baio. The Faces of Mechanical Turk, 2008. November 20. `waxy.org/2008/11/the_faces_of_mechanical_turk`.

[10] M. Barbaro, T. Zeller, and S. Hansell. A face is exposed for aol searcher no. 4417749. *New York Times*, August 9, 2006.

[11] R. Bell and Y. Koren. Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79, 2007.

[12] J. Bezos. Opening Keynote and Keynote Interview. September 27, 2006. techtv.mit.edu/videos/16180-opening-keynote-and-keynote-interview-with-jeff-bezos. Visited March 5, 2013.

[13] J. J. Chen, N. J. Menezes, A. D. Bradley, and T. North. Opportunities for crowdsourcing research on amazon mechanical turk. In *CHI Workshop on Crowdsourcing and Human Computation*, 2011.

[14] E. H. Chi and M. S. Bernstein. Leveraging online populations for crowdsourcing: Guest editors' introduction to the special issue. *IEEE Internet Computing*, 16(5):10–12, 2012.

[15] S. Chiarella. Personal Communication. March 5, 2013.

[16] E. K. Choe, S. Consolvo, J. Jung, B. Harrison, and J. A. Kientz. Living in a glass house: a survey of private moments in the home. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pages 41–44, New York, NY, USA, 2011. ACM.

[17] E. Cushing. Amazon mechanical turk: The digital sweatshop. *UTNE Reader*, January/February 2013.

[18] danah boyd. What is the role of technology in human trafficking?, December 7, 2011. `http://www.zephoria.org/thoughts/archives/2011/12/07/tech-trafficking.html`.

[19] F. Data. Section 6050W of the Internal Revenue Code, November 2010. `http://www.firstdata.com/downloads/marketing-merchant/255-001_irs_merchantuw111810fff.pdf`. Visited March 5, 2013.

[20] J. Davis, J. Arderiu, H. Lin, Z. Nevins, S. Schuon, O. Gallo, and M. Yang. The HPU. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 9–16, 2010.

[21] A. Felstiner. Sweatshop or paper route?: Child labor laws and in-game work. In *Proceedings of the 1st Annual Conference on the Future of Distributed Work (CrowdConf)*, San Francisco, September 2010.

[22] K. Fort, G. Adda, and K. B. Cohen. Amazon mechanical turk: Gold mine or coal mine? *Computational Linguistics*, 37(2):413–420, 2011.

[23] E. F. Foundation. Eff demands ftc investigation and privacy reform after aol data release, August 14, 2006. `https://www.eff.org/press/archives/2006/08/14`.

[24] B. Friedman. Value-sensitive design. *interactions*, 3(6):16–23, Dec. 1996.

[25] B. Friedman, P. H. Kahn, and A. Borning. Value sensitive design and information systems. In *Human-Computer Interaction and Management Information Systems: Foundations. M.E. Sharpe*, pages 348–372, 2006.

[26] R. Harmanci. The googler who looked at the worst of the internet. *BuzzFeed*, 2012. August 21. `http://www.buzzfeed.com/reyhan/tech-confessional-the-googler-who-looks-at-the-wo`.

[27] J. Heer and M. Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 203–212, 2010.

[28] P. Heymann and H. Garcia-Molina. Turkalytics: analytics for human computation. In *Proceedings of the 20th international conference on World wide web*, pages 477–486. ACM, 2011.

[29] J. Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.

[30] P. Ipeirotis. Demographics of Mechanical Turk. Technical Report CeDER-10-01, New York University, 2010.

[31] L. Irani. Idealogical work. *In preparation*, 2013.

[32] L. Irani and M. Silberman. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceeding of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 2013.

[33] A. Irwin. Constructing the scientific citizen: Science and democracy in the biosciences. *Public Understanding of Science*, 10(1):1–18, Jan. 2001.

[34] P. G. Kelley. Conducting usable privacy & security studies with amazonŠs mechanical turk. In *Symposium on Usable Privacy and Security (SOUPS)(Redmond, WA*, 2010.

[35] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton. The Future of Crowd Work. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 1301–1318, 2013.

[36] J. Klinger and M. Lease. Enabling trust in crowd labor relations through identity sharing. In *Proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, pages 1–4, 2011.

[37] S. Kochhar, S. Mazzocchi, and P. Paritosh. The anatomy of a large-scale human computation engine. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 10–17. ACM, 2010.

[38] E. Law and L. v. Ahn. Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(3):1–121, 2011.

[39] B. N. Levine, C. Shields, and N. B. Margolin. A survey of solutions to the sybil attack. Technical report, University of Massachusetts Amherst, Amherst, MA, 2006.

[40] D. Liu, R. Bias, M. Lease, and R. Kuipers. Crowdsourcing for usability testing. In *Proceedings of the 75th Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, October 28–31 2012.

[41] W. Mason and S. Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1):1–23, 2012.

[42] W. Mason and D. J. Watts. Financial incentives and the Performance of Crowds. In *SIGKDD*, 2009.

[43] N. Menezes. Net:Work Conference Recap. *The Mechanical Turk Blog*, Decemeber 9, 2010. `http://mechanicalturk.typepad.com/blog/2010/12/network-recap.html`. Visited March 5, 2013.

[44] C. Mims. How mechanical turk is broken. *MIT Technology Review*, January 3, 2010.

[45] S. Mitchell. Inside the online sweatshops. *PC Pro Magazine*, 2010. August 6. `www.pcpro.co.uk/features/360127/inside-the-online-sweatshops`.

[46] mTurk Forum. Discussion Thread: Tax information i got from amazon! `http://mturkforum.com/showthread.php?5300-Tax-information-i-got-from-amazon`. Visited March 5, 2013.

[47] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 111–125. IEEE, 2008.

[48] M. K. Ohlhausen. The FTCŠs new privacy framework. *Antitrust*, 25:43–46, 2011.

[49] P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57:1701–1777, 2010.

[50] A. Osterhaug. Privacy without borders: The ins and outs of outsourcing. *Data Breach Examiner*, 3, August 2011.

[51] G. Paolacci, J. Chandler, and P. Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419, 2010.

[52] P. Paritosh, P. Ipeirotis, M. Cooper, and S. Suri. The computer is the new sewing machine: benefits and perils of crowdsourcing. In *Proceedings of the 20th international conference companion on World wide web*, pages 325–326. ACM, 2011.

[53] J. Pontin. Artificial intelligence, with help from the humans. *New York Times*, March 25, 2007.

[54] A. J. Quinn and B. B. Bederson. Human computation: a survey and taxonomy of a growing field. In *2011 Annual ACM SIGCHI conference on Human factors in computing systems*, pages 1403–1412, 2011.

[55] J. Ross, L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 2863–2872. ACM, 2010.

[56] M. D. Scott. The FTC, the Unfairness Doctrine, and Data Security Breach Litigation: Has the Commission Gone Too Far. *Admin. L. Rev.*, 60:127, 2008.

[57] D. N. Shapiro, J. Chandler, and P. A. Mueller. Using mechanical turk to study clinical populations. *Clinical Psychological Science*, 2013.

[58] A. Shaw. Some initial thoughts on the otey vs. crowdflower case, January 9, 2013. `http://fringethoughts.wordpress.com/2013/01/09/some-initial-thoughts-on-the-otey-vs-crowdflower-case/`.

[59] M. Silberman, L. Irani, and J. Ross. Ethics and tactics of professional crowdwork. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):39–43, 2010.

[60] C. Sunstein. *Infotopia:How Many Minds Produce Knowledge*. Oxford University Press, USA, 2006.

[61] J. Surowiecki. *The wisdom of crowds*. Anchor, 2005.

[62] Turker Nation Forum. Dicussion Thread: *** IMPORTANT READ: HITs You Should NOT Do! ***. `http://turkernation.com/showthread.php?35-***-IMPORTANT-READ-HITs-You-Should-NOT-Do!-***`. Visited March 5, 2013.

[63] Turker Nation Forum. Dicussion Thread: Privacy, shmivacy. `http://turkernation.com/archive/index.php/t-6065.html`. Visited March 5, 2013.

[64] Turker Nation Forum. Dicussion Thread: Requester posting Worker IDs. `http://turkernation.com/archive/index.php/t-1384.html`. Visited March 5, 2013.

[65] Turker Nation Forum. Dicussion Thread: Requesters Who Change Their Name. `http://turkernation.com/archive/index.php/t-2886.html`. Visited March 5, 2013.

[66] D. Vincent. China used prisoners in lucrative internet gaming work. *The Guardian*, May 25, 2011.

[67] L. von Ahn. *Human Computation*. PhD thesis, Carnegie Mellon University, 2005. Tech. Report CMU-CS-05-193.

[68] H. Wallach and J. W. Vaughan. Workshop on Computational Social Science and the Wisdom of Crowds. In *NIPS*, 2010.

[69] S. Wolfson and M. Lease. Look Before You Leap: Legal Pitfalls of Crowdsourcing. In *Proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 2011.