“It was a shady HIT”: Navigating Work-Related Privacy Concerns on MTurk

Shruti Sannon
Cornell University
Ithaca, NY 14850, USA
ss3464@cornell.edu

Dan Cosley
Cornell University
Ithaca, NY 14850, USA
drc44@cornell.edu

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188511

Abstract
As online digital labor platforms grow in popularity, research is needed to understand how workers navigate the unique privacy concerns that emerge during their work. We surveyed 82 Amazon Mechanical Turk (MTurk) workers about how they make decisions about revealing personal data while doing tasks. We find that many comply with privacy-invasive information requests for a range of reasons, including benefits outweighing costs, fears of losing access to work, and contributing to scientific research. Several workers also engage in privacy-protective behaviors, motivated by perceptions that the information request is unnecessary or violates policies, as well as concerns about data use and identification. We discuss how our findings can inform policy and design to protect workers’ privacy.

Author Keywords
Privacy; privacy-protective behaviors; digital labor.

ACM Classification Keywords
K.4.1. Computers and Society: Public Policy Issues: Privacy.

Introduction
Digital work is becoming increasingly widespread, with almost one in every ten Americans (8%) earning money via an online work platform in the past year

[10]. This percentage may increase as the gig economy becomes increasingly integrated into society.

Growing participation in digital labor raises new concerns about online privacy. On Amazon Mechanical Turk (MTurk), a popular crowdsourced labor site, workers (known as MTurkers) have reported a range of privacy concerns and violations that have arisen during their work on the platform [11]. These include concerns about data collection and profiling, unauthorized use, invasive practices such as being stalked or spammed, and deceptive practices such as phishing and scams.

Although these issues are not unique to MTurk, they may be magnified in work contexts where information and power asymmetries might lead workers to discount legitimate privacy concerns in order to obtain work. Further, MTurk's nominal anonymity may provide a false sense of security, since MTurkers can be de-anonymized while working on the site [6]. On the other hand, MTurkers appear to be more privacy-conscious than average Internet users [4], suggesting they may be relatively sophisticated in both their privacy decision-making and privacy-protective behaviors. Together, these factors make MTurk—and more generally, online and gig economy labor markets—a compelling context for studying both privacy concerns and related privacy-protective behaviors.

This paper makes two main contributions about how MTurkers make decisions about disclosing personal data when faced with concerning information requests during Human Intelligence Tasks (HITs). First, it identifies a range of reasons why MTurkers provide truthful personal information during HITs despite privacy concerns. Second, it identifies whether and how MTurkers engage in privacy-protective behaviors while working on the platform.

Understanding the privacy considerations and decision-making calculus of digital workers such as MTurkers is vital to ensuring that the designs and policies of digital labor platforms protect workers. Knowing how MTurkers make decisions around providing, withholding, and fabricating their data can also help researchers ethically collect accurate data in the face of power imbalances between requesters and workers.

Method

We draw from survey data collected on MTurk in June 2017 as part of a larger study (N=356) on online privacy [9]. Participants were asked to describe a time they provided personal information despite concerns for their privacy online, as well as examples of times they had lied to protect their privacy.

Of those examples, 91 (from 82 unique participants) concerned privacy considerations that arose during their MTurk work. These 82 participants were evenly divided by gender (45% male, 55% female), most often identified as Caucasian (82%), and averaged 34 years old. 10% were high school graduates, 43% had some college or an Associate's degree, and 47% had a Bachelor's degree or higher. Most were employed (61%) or self-employed (23%) and had worked on MTurk for over a year (76%). They averaged about 8 hours a day using the Internet. Participation was restricted to MTurkers in the U.S.

Results

Using an inductive approach, the lead author assigned open codes to these 91 examples, and then reiterated

coding of the dataset using focused codes to capture recurrent themes. Our results are organized into two main groups: reasons for providing accurate information despite discomfort, and reasons for engaging in privacy-protective behavior.

Providing Accurate Information Despite Discomfort
Over two thirds of participants' examples (67) described times they had provided personal data during HITs despite privacy concerns. Their reasons for complying fell into four main categories.

1. Direct Benefits of Compliance
MTurkers often complied with invasive requests when the benefits of providing the information outweighed the risks. A well-paying task or large bonus could entice MTurkers: "Usually I wouldn't do this but it was for $2 and I thought why not", as could hardship: "I was homeless at the time though so I really needed money and went and did it anyways." Monetary incentives could override other factors such as concerns about the requester: "A survey…that was very shady but I had fun with the hit and submitted all of my information in order to get a good bonus, I did it willingly."

Requests for email addresses were particularly frustrating, as this is identifying data and asking for it is explicitly against MTurk's policy. Here too the benefits sometimes outweighed the costs, as when requesters asked for email addresses they could use to offer follow-up studies. Workers sometimes complied with these requests, though several reported setting up a legitimate alternate email address for MTurk so they did not have to provide their personal email.

2. Consequences of Non-Compliance
Some MTurkers worried that lying or not complying with a request for personal information could harm their ability to get future work, either from individual requesters: "I'm afraid that if I don't I may be blocked from doing future work with the survey providers" or from Amazon itself: "I do however still put my real information so that I don't get stricken on MTurk."

Tasks that made invasive requests partway through, after a worker had already invested time in the HIT, required workers to make difficult decisions about abandoning the effort they had put in versus complying to get paid: "halfway thru the study they asked to link to my facebook and twitter. I was upset by this and felt it was wrong however I needed to give the information in order to complete the study and be paid so I decided to provide it."

3. Contributions to Research
Several participants complied with an invasive data request because they perceived academic research to be important: "I don't know who all is going to see that information necessarily, but ultimately generally give it since it's typically in the interest of some educational goal." These participants did not want to negatively impact research: "I gave the correct info in case it mattered to the survey and would affect it" and "I want to be honest for the integrity of the data." The esteem these participants had for the scientific enterprise also extended to their perception of academic practices. Some felt that academic researchers might be more mindful of their privacy than other requesters: "I usually try not to lie on surveys used for scientific research. Even if I think it's rather personal I know the information is anonymized (thanks to ethics boards

prohibiting lying on consent forms) and generally feel it's somewhat safer."

4. Perceived Protections

Requesters could engender a sense of safety when they included a statement about data collection: "[they] said they would not share my information with other parties", a practice some MTurkers came to expect: "most reputable requesters make an effort to protect privacy (or at least they make a statement about it)."

One MTurker looked to notions of reputation and recourse in making disclosure decisions: "I usually only provide sensitive information when the requester is rated on Turkopticon or, preferably when they list information such as their address and phone number—which many academic requesters do." Turkopticon is a site where MTurkers can comment on requesters and HITs [2]; it was only mentioned once in our data but is popular enough among seasoned MTurkers that we suspect requester ratings and reputation play a key role in experienced MTurkers' disclosure decisions.

Other Rationalizations

Besides those four categories, MTurkers gave several other rationalizations for complying with invasive requests. One was that if negative outcomes were manageable, compliance was acceptable, as with a request for an email address: "I figured I could always use my junk folder to filter spam that came from it." Another involved minimizing perceptions of potential risk: "But really the internet is so vast and I am just a single person that I dont really think it matters", or discounting concerns: "I was worried but I figured it was innocuous". Some MTurkers also concluded that disclosing personal information was a norm among MTurkers: "I asked other turkers if it was normal, and they thought I was silly to even ask. I gave that information out anyway although it felt weird at first."

Engaging in Privacy-Protective Behavior

Although MTurkers chose to disclose personal information despite privacy concerns for a number of reasons, just over one-fourth of participants' examples (24) described times when they had lied to protect their privacy during HITs. Their reasons for telling these lies could be categorized into three main themes.

1. Requested Data is Unneeded

The most common reason to lie was that the personal information being requested—often the participant's name or date of birth—was irrelevant for the task at hand: "They don't need this information for conducting research". Some participants guessed why the data was being requested in order to judge its relevance: "I lie about my name any time a hit on MTurk wants my name for the purpose of trying to humanize the task I guess. It doesn't matter so I make a name up." Another strategy was to reason about how much and how accurate the data needed to be, and to partially obscure information, for instance, by listing a neighboring town when asked for their location, or an accurate birth year but false date and month: "the most important part is just my age."

2. Lack of Policies and Policy Violations

Although some participants were willing to disclose identifying data even though asking for it violates AMT policy, others were not when a request "was against MTurk rules", sometimes lying in these cases: "I [lied] because it is a MTURK violation anyway." And, on the flip side of the sense of safety that statements about

data protection provided, their absence could make MTurkers wary: "the people creating [the survey] didn't have any sort of privacy policy or data protection policy. So I gave them junk data."

3. Concerns about Data Use and Identification
Participants also lied because they were concerned about how their data would be used. Sometimes they wanted to provide less information "so the people running the survey didn't know too much about me", worrying about it being "easily accessible". Some HITs also triggered wariness about requesters' intentions: "I felt it was a shady hit", for example, when a HIT's requests "were too similar to types of security questions". Finally, some participants lied about sensitive information such as household income or history of sexual assault.

Discussion and Implications for Design

Our findings suggest that MTurkers navigate many privacy concerns that arise from power and information asymmetries on the platform. Access to valuable work, appropriateness of requests for personal data, and perceptions of requesters' data practices each play a key role for at least some MTurkers in deciding whether to provide accurate responses. They use their limited information about the requester and the purpose of the task to make these judgments; if they view the requester as trustworthy and the data requested as needed, they are more likely to provide accurate responses.

These findings align well with prior work around MTurkers' motivations. For instance, MTurkers will work harder if they are provided with information about the requester [7]; such information likely increases trust, which some of our participants mentioned as a key driver. Previous work also suggests that MTurkers feel strongly about fairness, accepting responsibility for a rejection if it is clearly their fault, and expressing outrage if the error lies with the requester [8]. In our case, asking for needed information is likely seen as more fair than asking for unneeded information.

Extending this focus on trust and fairness to the context of data collection, this suggests that most MTurkers aren't fabricating data to scam requesters. Rather, our findings suggest they carefully weigh the decision to provide inaccurate data and do so to protect themselves or, as with one participant who lied because "they didn't pay me enough for the risk [of providing accurate data]", to harm unfair requesters.

Because MTurkers often lied when they perceived a requested data point to be irrelevant for the HIT, a straightforward design implication for requesters is to clarify why any specific data point is being requested to increase user trust and compliance. This is surprisingly rare in other online settings that involve data collection, such as web forms and account sign-ups, and may be helpful in these contexts as well. That said, Kittur et al. point out that a balance needs to be struck between providing workers with more context and other concerns such as efficiency and confidentiality [5].

MTurkers also tend to give academic requesters the benefit of the doubt because they have practices such as providing a consent form prior to a HIT that provides contact information and describes data collection, use, and storage. Non-academic requesters could be encouraged (or mandated through policy) to mirror such practices to reduce both perceived and actual

privacy risks. Requesters could also draw on differential privacy techniques for collecting crowdsourced data that obscure individual responses [3]; these techniques could also reduce the burden on MTurkers to engage in privacy-protective behaviors. Research on open collaboration systems suggests that, despite commitment and incentives, privacy concerns can adversely impact participation [1]; similarly, mitigating MTurkers' privacy concerns may increase productivity and quality of work.

Longer-term, since some MTurkers felt that they had to comply with invasive requests to avoid the negative consequences of non-compliance, more ways to reduce the asymmetry in power between requesters and workers are needed. A potential solution could be to make it easier for MTurkers to flag privacy-invasive requesters on the platform, or to report them for making requests that violate Amazon's policies.

Acknowledgements

Data collection was part of a larger study on online privacy supported by the National Science Foundation (Award #1405634). This study was conducted while Dan Cosley was serving at the NSF and does not necessarily reflect the views of the NSF.

References

1. Andrea Forte, Nazanin Andalibi, and Rachel Greenstadt. 2017. Privacy, anonymity, and perceived risk in open collaboration: A study of Tor users and Wikipedians. In *Proceedings of CSCW '17*, 1800–1811.
2. Lilly C. Irani and M. Six Silberman. 2013. Turkopticon: interrupting worker invisibility in Amazon Mechanical Turk. In *Proceedings of CHI '13*, 611–620.
3. Thivya Kandappu, Vijay Sivaraman, Arik Friedman, and Roksana Boreli. 2014. Loki: A privacy-conscious platform for crowdsourced surveys. In *Proceedings of COMSNETS*, 1–8.
4. Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy attitudes of Mechanical Turk workers and the US public. In *Proceedings of SOUPS*, 37–49.
5. Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of CSCW '13*, 1301–1318.
6. Matthew Lease, Jessica Hullman, Jeffrey P. Bigham, Michael S. Bernstein, Juho Kim, Walter Lasecki, Saeideh Bakhshi, Tanushree Mitra, and Robert C. Miller. 2013. Mechanical Turk is not anonymous. *SSRN*.
7. Jennifer Marlow and Laura A. Dabbish. 2014. Who's the boss?: requester transparency and motivation in a microtask marketplace. In *Proceedings of CHI '14*, 2533–2538.
8. David Martin, Benjamin V. Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of CSCW '14*, 224–235.
9. Shruti Sannon, Natalya N. Bazarova, and Dan Cosley. (2018). Privacy lies: Understanding how, when, and why people lie to protect their privacy in multiple online contexts. To appear in *Proceedings of CHI '18*.
10. Aaron Smith. 2016. Gig Work, Online Selling and Home Sharing. *Pew Research Center.* Retrieved from http://www.pewinternet.org
11. Huichuan Xia, Yang Wang, Yun Huang, and Anuj Shah. 2017. "Our privacy needs to be protected at all costs": Crowd workers' privacy experiences on Amazon Mechanical Turk. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 113.