June 2020

A Study of Ethics in Crowd Work-Based Research

Huichuan Xia
Syracuse University

ABSTRACT

Crowd work as a form of a social-technical system has become a popular setting for conducting and distributing academic research. Crowd work platforms such as Amazon Mechanical Turk (MTurk) are widely used by academic researchers. Recent scholarship has highlighted the importance of ethical issues because they could affect the long-term development and application of crowd work in various fields such as the gig economy. However, little study or deliberation has been conducted on the ethical issues associated with academic research in this context. Current sources for ethical research practice, such as the Belmont Report, have not been examined thoroughly on how they should be applied to tackle the ethical issues in crowd work-based research such as those in data collection and usage. Hence, how crowd work-based research should be conducted to make it respectful, beneficent, and just is still an open question.

This dissertation research has pursued this open question by interviewing 15 academic researchers and 17 IRB directors and analysts in terms of their perceptions and reflections on ethics in research on MTurk; meanwhile, it has analyzed 15 research guidelines and consent templates for research on MTurk and 14 published papers from the interviewed scholars. Based on analyzing these different sources of data, this dissertation research has identified three dimensions of ethics in crowd work-based research, including ethical issues in payment, data, and human subjects. This dissertation research also uncovered the "original sin" of these ethical issues and discussed its impact in academia, as well as the limitations of the Belmont Report and AoIR Ethical Guidelines 3.0 for Internet Research. The findings and implications of this research can help researchers and IRBs be more conscious about ethics in crowd work-based research and also inspire academic associations such as AoIR to develop ethical guidelines that can address these ethical issues.

A Study of Ethics in Crowd Work-Based Research

by

Huichuan Xia

B.A. Central China Normal University, 2006

M.S. Peking University, 2010

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Information Science and Technology

Syracuse University

June 2020

ACKNOWLEDGMENTS

Acknowledgments may be the hardest part to write in my dissertation. As a Chinese proverb says, certain debts of gratitude cannot be expressed or repaid by words, and there is always a delicacy to acknowledge the right persons in the appropriate extents. Meanwhile, there is certainly a risk of forgetting somebody that deserves my thanks because, after years of hardship in pursuing a Ph.D. and moments of ecstasy after obtaining it, I may have developed amnesia. Nonetheless, I still feel obliged to write my acknowledgments here.

First and foremost, I thank my advisor, Dr. Jennifer Stromer-Galley, for her education of my knowledge and cultivation of my character throughout these years. She is the model researcher, teacher, and advisor for me. Her rigor in researching, passion in teaching, and dedication in advising are all lifetime examples and benchmarks for me. I will certainly carry them on and embody them to my students and children in the future. Here, I can use a "small" example to demonstrate Dr. Jennifer Stromer-Galley's education and character: Throughout my Ph.D. program years, she is the *only* professor I know that would carefully read and examine every sentence in detail in a manuscript to her (yes, including this dissertation document) from the broad research aspects to the tiny grammatical errors. I personally regard her as my "academic mother" who has nurtured me from a graduate pupil to Dr. Xia.

Then, I thank my committee members, and I am fortunate to have the chance to consult with them. Individually, Dr. Caroline Haythornthwaite is such a scholar and person that her erudition, grace, and status are self-evident, and you immediately realize it when you talk with her. It is like when you are basking: You know the sun is there, but you feel pleasant and illuminated rather than being dizzied or intimidated away. Dr. Ingrid Erickson is a radiant and passionate scholar and has what I call the "Stanfordian" spirit. She can stimulate your critical

thinking and challenge your ideas, but you will never feel cold or dull talking with her. Dr. Carsten Østerlund is a master at prioritizing and balancing. The more I know about him, the more I want to be a scholar like him. Thus, besides being a prominent scholar and after finishing up a paper, I can take a break and sail with my family and children into the sea.

Next, I thank my internal reader, Dr. Jeffery M. Stanton, for his inspiration and education. His quantitative teaching has a lasting influence on me, and he had also inspired my dissertation research as the chair of the IRB at Syracuse University when I consulted him about research ethics. In addition, I had attended his *concerts* several times, how cool is that! (I also miss my time sang with the iBand). Also, I thank my external reader, Dr. Michael Zimmer, and the chair of my dissertation defense, Dr. Lael Schooler, for their insights, feedback, and advice on my dissertation research. I will maintain active consultation with them in the future.

I thank Dr. Steve Sawyer and Jennifer Barclay for providing a great Ph.D. program. I thank my Ph.D. cohorts - they are all admirable and adorable. I thank Dr. Yang Wang and Dr. Yun Huang. I had my first and several publications with them after I came to the U.S., and they also showed me the efforts to become prolific scholars as well as their lessons in how to treat advisees. I thank Dr. Lu Xiao for her support during my difficult periods and the invitations to her home for the hot-pot and fun. I thank Dr. Brian McKernan, Dr. Kevin Crowston, Dr. Jason Dedrick, Dr. Ping Zhang, Dr. Bryan Semaan, Dr. Barbara Kwaśnik, and Dr. Daniel Olson-Bang for their advice on my research and teaching. I thank Dr. John Warren for his conducting and education in music. I feel purely happy to sing in with him and the Syracuse University Oratorio Society. Also, I thank the participants in this dissertation for their precious insights and time.

Last but not least, I sincerely thank my parents and grandparents. I will ensure that their love and cultivation for me will pass to my offspring. This dissertation is dedicated to them.

Table of Contents

CHAPTER 1: INTRODUCTION

Chapter 1 is the introduction of this dissertation work and provides an overview of its findings and contributions. I start this chapter by introducing the research background, motivation, and boundary of this dissertation work. Then, I define some of the key terms that will be used throughout this document. Afterward, I present the research questions and how they were developed. Finally, I reveal several key findings and contributions from this dissertation work.

Background

Ethical issues have long been a subject of discussion within the academic community. Many of these issues emerged soon after World War II when Nazi scientists abused concentration camp prisoners for experiments in the name of science. The Nuremberg Code (1947) and the Declaration of Helsinki (1964) were two products of a series of discussions on how to make clinical experiments with human subjects more ethical. In academia, and in the U.S. specifically, the Belmont Report (1979), by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, was the most influential attempt to address different ethical issues in academic research and it not only focuses on clinical experiment but also extends to social research with human subjects.

In 1978, the Department of Health and Human Services (DHHS) established the requirement for Institutional Review Boards (IRBs) in all U.S. institutions that receive federal funding for their research with the premise that "investigators should not have sole responsibility for determining whether research involving human subjects fulfills ethical standards" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, p. vii). The role of IRBs is to ensure that academic research with human subjects is ethical.

The ethical standards are later proposed and represented in the Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research in 1979. The Belmont Report (1979) proposed three ethical guidelines, i.e., respect for persons, beneficence, and justice. They have become the cornerstones for academic research and the benchmark for IRB directors and analysts to review a study with human subjects. After that, various academic associations have further interpreted and amended these principles in the context of their specific disciplines. For example, in 1992, the Association for Computing Machinery (ACM) designed the Code of Ethics and Professional Conduct ("the Code") for all computing professionals to guide and their conduct (ACM, C. M., 1992). This Code highlights the importance of protecting individual privacy and data integrity from "unauthorized access or accidental disclosure to inappropriate individuals" (ACM, C. M., 1992, p. 4). In 1992, the American Psychological Association (APA) defined the Ethical Principles of Psychologists and Code of Conduct ("the Ethics Code") and further revised it in 2002 to guide psychologists conduct research with a high ethical standard (American Psychological Association, 2002).

As more applications, platforms, and software on the internet are sprouting and developing, new ethical challenges arise for researchers. For example, on the internet, the boundary between public space and private space has blurred, such as whether online chatrooms should be considered public space for all the internet users that researchers could freely observe or private space that researchers should solicit chatroom participants' consent to record their conversation (Ess et al., 2002). There also exists a tension between anonymity and identification of internet users, that is what steps should be followed to protect an individual's privacy and prevent unauthorized data linkage and triangulation (Markham and Buchanan, 2012). To address

such issues, the Association of Internet Researchers (AoIR) has released three versions of Internet Research Ethics guidelines (IRE) for internet research and internet researchers.

AoIR's first version of IRE (AoIR IRE 1.0) in 2002 focused on providing ethical recommendations for researchers, ethicists, and students who study human interactions enabled by the internet, the organizations that fund and oversee internet research, and the academic societies and groups that promote and incorporate internet research (Ess, 2002). AoIR IRE 2.0 published in 2012 was updated to deal with the dramatically expanded scope and contexts of internet research; this version proposed a more comprehensive list of internet-specific ethical questions for research stakeholders to consider (Markham and Buchanan, 2012). Most recently, AoIR IRE 3.0, published in 2019, has incorporated ethical concerns in different research contexts, for example in Big Data online research, as well as offered further considerations of ethics for different research stages. AoIR IRE 3.0 highlights ethical challenges of obtaining informed consent particularly in Big Data research contexts (Franzke et al. 2020).

Another notable ethical guideline specifically pertaining to computer and information security research is the Menlo Report published by the U.S. Homeland Security in 2012. Unlike the Belmont Report, ACM and APA ethical guidelines, or the AoIR IREs, the Menlo Report is not only purposed for academic researchers but also for "corporate and independent researchers, professional societies, publication review committees, funding agencies" (Bailey et al., 2012, Section A.1). The Menlo Report focuses on ethical issues in information and communication technology research (ICTR) and has adapted the Belmont principles in this research context; meanwhile, the Menlo Report has added a fourth principle, "Respect for Law and Public Interest," which recommends ICT researchers to comply with the relevant ICT laws and policies

and to be transparent and accountable in their research methods, results, and behaviors (Bailey et al., 2012).

Alongside the evolution of research principles, new mechanisms for people to work via digital platforms has arisen. Commonly called crowd work, it is "the performance of tasks online by distributed crowd workers who are financially compensated by requesters (individuals, groups, or organizations)" (Kittur et al., 2013, p. 1302). Crowd work is supported by crowd work platforms like Amazon Mechanical Turk (MTurk) that allows requesters to recruit workers online to do their tasks. Since its emergence, however, crowd work evokes both promises and problems. On the one side, crowd work mobilizes a geographically distributed global workforce and provides remarkable flexibility and productivity for people who want to outsource their tasks efficiently (Kittur et al., 2013). On the other side, crowd work tasks are usually transient that only take a short time to complete (Kuek et al. 2015) and are scalable to a large number of unknown and geographically distributed crowd workers (Schulze et al. 2011), which make a malicious act, e.g., a phishing task, hard to detect and trace to its publisher (Xia and Mckernan, 2020).

In the context of crowd work, academic research published on a crowd work platform, or what I call *crowd work-based research* in this dissertation document, inherits several fundamental ethical issues pertaining to internet research, such as the tension between privacy and de-anonymization. Furthermore, crowd work-based research could potentially induce additional ethical challenges, for example, previous research has discussed the potential dehumanization of crowd workers (Silberman et al., 2010); the commodification of crowd workers' labor (Aloisi, 2015; Bergvall-Kåreborn and Howcroft, 2014); privacy risks and violations of crowd workers (Xia et al., 2017), and the crowd work platform's lack of regulation

and responsibility of requesters' conduct, which makes crowd workers vulnerable to various harms (Silberman and Irani, 2013; Xia and Mckernan, 2020).

Research Motivation

Within such a background, this dissertation is motivated by three research gaps in the existing scholarship around ethical issues in crowd work. The first gap is that there has been little examination on what I see as a paradox in crowd work-based research. The paradox is that, on the one side, for some crowd workers, doing tasks to earn money is an essential or even mandatory revenue to maintain their living; on the other side, ethical research participation is founded on a principle of voluntary consent, as I detail in the literature review. Hence, academic researchers may situate their research participants as wage-contractors instead of research volunteers. Since wage-contractors are employed and paid by their employers, they would have more obligation to obey their employers' instructions. If academic researchers assume such a wage-contractor relationship in crowd work-based research, they may be inattentive to the voluntariness of research participation and fail to fully respect the autonomy of crowd workers as research subjects. Also, a related issue is the effect of *immediate gratification* (Acquisti, 2004), which could adversely influence some crowd workers to take tasks without reviewing carefully whether the tasks are benevolent and whether task publishers are trustworthy (Xia and Mckernan, 2020). As a result, the voluntariness of research participation maybe not be as much based on informed consent as for the pursuit of immediate monetary incentive. Hence, more research is needed to understand this paradox.

The second research gap is a dearth of empirical investigation with academic scholars and IRBs' with regard to their perceptions of ethics and their practices of ensuring ethical research on crowd work platforms. For example, Belmont's principles of respect, beneficence,

and justice guide academic researchers' conduct in crowd work-based research, but how academic researchers and IRBs perceive and apply these principles is largely unknown in research on a crowd work platform. Furthermore, there has been little investigation on whether researchers' perceptions of ethics can be different from IRBs' interpretations, and there has been little comparison between researchers' and IRBs' ethical considerations. This lack of knowledge has not been fully discussed in the current scholarship, and I propose these to be essential to investigate so as to provide ethical insights and guidance for both researchers and IRBs to conduct and review crowd work-based research in the future.

The third research gap is that ethical issues with crowd work have been drawing growing discussion and critiques (e.g., Silberman et al., 2010; Irani and Silberman, 2013; Salehi et al., 2015; Sheehan, 2018), yet the Belmont principles that supervise and approve various research on crowd work platforms such as MTurk have not been thoroughly examined in terms of their application and feasibility to this emerging research context. Multiple academic associations have proposed ethical guidelines, but none of them has updated or specified to deal with the ethical issues in crowd work-based research. AoIR IREs, including the most recent AoIR IRE 3.0, which are widely referenced in internet studies, have also not focused on the emerging ethical challenges in the context of crowd work. Hence, I want to examine these ethical guidelines and explore their application and limitations in directing crowd work-based research.

To sum up, these research gaps call for this dissertation work as there is a need to investigate ethical issues with crowd work-based human subjects research.

Research Boundary

Apart from the motivations to this dissertation work, it is important to clarify and set a boundary for this dissertation work. In the first place, I focus on academic researchers' and

IRBs' ethical perceptions and practices in crowd work-based research for the following reasons. First, academic researchers and IRBs are most familiar with the ethical principles in academic research, and thus, in this pioneering work into ethical issues in crowd work-based research, I start from interviewing these two groups. Second, I focus on interviewing researchers who are well-known in their studies in crowd work and IRB directors and analysts who have reviewed research that uses crowd work platforms such as MTurk to collect data. Hence, my research participants are quite knowledgeable about crowd work-based research. Third, for this dissertation work, I excluded crowd workers to participate in my interviews because they are presumably less familiar with ethics in academia and since they do both academic and non-academic tasks, they are susceptible to confuse the different natures of these two types of tasks. I plan to investigate crowd workers' perceptions of academic tasks and requesters in my follow-up research after this dissertation work.

In the second place, I choose MTurk as the crowd work platform to focus my research. I do so because it is the first crowd work platform and the most popular and familiar among scholars and IRBs, which is evident in the number of publications that use MTurk for data collection (e.g., Callison-Burch, 2009; Gottlieb et al., 2012; Xia et al., 2017) and take it as the target for deliberating crowd work and crowdsourcing (e.g., Irani and Silberman, 2013; Kittur et al., 2013; Salehi et al., 2015; Silberman et al., 2018). In this regard, MTurk is the benchmark for crowd work platforms, and academic research on MTurk is the most popular and representative form of crowd work-based research. I note, however, that my dissertation is not merely a study of a specific crowd work platform, i.e., MTurk, I focus on the practices of crowd work-based research, such as its inherent payment and reputation mechanisms, which are not exclusive to MTurk. Further, I do not focus on the specific design aspects of the MTurk platform for

academic research, such as the platform's usability for academic scholars. Meanwhile, my research findings are generalizable to other crowd work platforms that are similar to MTurk, such as MicroWorkers, ShortTasK, ClickWorker, RapidWorkers (Nagrale, 2018) even though they have not been widely used for academic research purposes (Peer et al., 2017). Finally, I am aware that there are alternative platforms such as Prolific that scholars are increasingly using (Chandler et al., 2019; Palan and Schitter, 2018), but these platforms are much newer than MTurk, and not many researchers have migrated from MTurk to them (Chandler et al., 2019). Hence, to probe researchers' and IRBs' thoughts about ethics in crowd work-based research, MTurk is still the most suited platform to focus this dissertation research.

In the third place, I focus on *academic research* on crowd work platforms. While academic studies constituted a large portion of tasks, most tasks on crowd work platforms are still non-academic; for example, there are tasks from business, non-profit organizations, and others (Hitlin, 2016). My research lens primarily focuses on ethical issues with academic tasks that the researchers and IRBs are most familiar with. In this sense, industrial research projects, such as those published on MTurk by corporations such as US Foods, or by organizations such as Allen Institute for Artificial Intelligence ("Amazon Mechanical Turk," n.d.-c), are excluded from the analytic scope of this dissertation work. Finally, this dissertation work is situated in the broader context of the gig economy that is comprised of crowd work and "work-on-demand via app" where conventional work is outsourced and undertaken through mobile applications (De Stefano, 2015). As such, even though my research lens focuses on academic research on crowd work platforms, certain phenomena discussed in this document, such as the abdication of responsibilities among different parties, apply to other gig economy phenomena as well, such as a ridesharing company's treatment of drivers as independent contractors. MTurk is both the

paradigm of crowd work and the prelude of the gig economy. Thus, the ethical problems in MTurk may also find traces in various other contexts in the gig economy. I will study them further in my future work after this dissertation.

Key Terms

Ethics

It is difficult to give an accurate definition of ethics, but it is a significant concept that dates back to ancient civilizations. In Western culture, Socrates, Plato, and Aristotle were the first to reflect on how to be virtuous and pursue happiness in life; in the East, Chinese philosopher Confucius raised the concept of *Ren* to instruct a person to be benevolent with others. Since then, great philosophers in the modern era from Immanuel Kant, Jeremy Bentham, J. S. Mill, to John Rawls all interpreted ethics from different perspectives such as duty, utility, and justice.

Roughly speaking, Kant judged the ethical value of a person's action by examining its motive rather than its consequence. The action is ethical only if the motive is good regardless of whether the consequence turns out to be favorable. Kant proposed two principles to assess whether a motive is good and must be obeyed as a duty in order to act. First, one should consider whether an action can be universalized as a law so that every person can also act as such without generating any contradiction. Second, one should consider whether an action has treated themselves as well as the other persons involved not only as a means to certain ends but also as ends in themselves.

The utilitarianists such as Jeremy Bentham and J. S. Mill offer a different set of principles to consider ethics as opposed to Kant. They suggested considering the consequence of a person's action to assess its ethical value, specifically whether a person's action could

maximize utility, i.e., the balance of happiness or benefits over pain or costs. At the collective level, utilitarian ethics advocates actions that would result in the greatest welfare for the greatest number of people; at the individual level, utilitarian ethics encourages persons to pursue the most desired pleasures from undesired ones to maximizes their utility, i.e., their happiness.

Justice is a key concept of ethics since Aristotle's teleological philosophy. Justice is also an essential component of academic research ethics, as proposed in the Belmont Report (1979). There are three main approaches to justice from the stance of welfare, freedom, and virtue (Sandel, 2010). Utilitarianism holds that justice is achieved when welfare is maximized for the majority of people even at the expense of sacrificing minority or individuals' interests to various degrees. The freedom stance relates to libertarianism and egalitarianism. Libertarians believe that justice is maintained as long as the involved stakeholders voluntarily decide or agree to take certain action or sign some contract without any coercion; egalitarians believe that justice means equal opportunity to success and equal access to resources, and advantageous groups should help disadvantageous groups to make benefits and risks distribution more evenly (Sandel, 2010).

The virtue stance of ethics includes several alternative perspectives. Some philosophers believe justice means respecting an individual's right and dignity regardless of majority's benefits, and there needs to be an initial stage, or as John Rawls depicted, a "veil of ignorance" when all people are unaware of each other's social status and natural talent to decide a fair distribution of opportunities (Rawls, 1971). Some teleological philosophers such as Aristotle perceive that justice means the fair match and honor of a telos (i.e., purpose) of a practice with people who best fulfill this telos. For example, the best flutes should be given to the best flute player to attain its purpose to be played beautifully and the best rewards should be given to the best flute players for fulfilling the flute's purpose (Sandel, 2010).

Crowdsourcing

Howe (2006) gave the first definition of crowdsourcing in the *Wire* magazine, which is: "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call." Howe's original definition emphasizes that a crowdsourcing task originates from an open call with the flexibility on who can take the task. It also indicates that the open call is undertaken by a large and undefined network of people.

Since Howe (2006)'s definition, derivative conceptualizations of crowdsourcing with different specificity emerged. For instance, Brabham (2013) gave a more specific definition in his book *Crowdsourcing*: "[It is] an online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals" (p. xix). His definition stresses the *problem-solving* nature of crowdsourcing and an *online platform* to carry out the process of crowdsourcing. Another noteworthy definition is from Estellés-Arolas and González-Ladrón-de-Guevara (2012) who developed a synthesized and lengthy definition of crowdsourcing based on a review of forty relevant definitions. In particular, their definition highlights that the motivation for crowd workers to participate in a crowdsourcing task can be either monetary or non-monetary.

Crowd work

Kittur et al. (2013) defined crowd work as "the performance of tasks online by distributed crowd workers who are financially compensated by requesters (individuals, groups, or organizations)" (p. 1302). This definition emphasizes that crowd work has a monetary incentive, the crowd workers are distributed via an online platform, and the requesters could be of different identities. The term *crowd work* is sometimes used interchangeably with *crowdsourcing*, but

there are two distinctions between them. First, crowd work is financially compensated, and thus, it is different from volunteer-based crowdsourcing projects such as OpenStreetMap, which offer no monetary incentive. Second, crowd work is distributed and performed online and thus differs from *work-on-demand via apps* (De Stefano, 2015). In this sense, crowd work excludes mobile crowdsourcing applications such as Waze. This dissertation focuses on the concept of crowd work defined by Kittur et al. (2013)'s and avoids using the inflated concept of crowdsourcing.

Research

The Belmont Report (1979) defined research as "an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships)" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part A). Meanwhile, research is usually a formal protocol with a defined object and a set of procedures to achieve that object (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part A). This dissertation relies on the definition and the boundary of the term "research" from the Belmont Report.

Crowd Work-Based Research

Crowd work-based research is the term coined in this dissertation to study the research ethics on crowd-based internet platforms where work is conducted and paid for. Several scholars have used various terms to describe this method of collecting data for basic research. For instance, Sheehan (2018) used the term *crowdsourcing research* to denote the research data collection with MTurk. However, it is a murky notion because it neither differentiates crowdsourcing from crowd work nor clarifies the relation between *crowdsourcing* and *research*. Graber and Graber (2012) used the term *crowdsourcing model of research* but their conception

of *crowdsourcing* as a superset of the research protocols was exemplified by *Foldit*, which actually is more like a citizen-science platform. Williamson (2016) explored the ethical issues in *crowdsourced research* which also refers to the research on MTurk, but as I have clarified in the definitions above, crowdsourcing and crowd work should not be equated and the conceptualization of research on MTurk should not be inflated.

Hence, even though somewhat awkwardly named, I use the term *crowd work-based research* in this dissertation to set a clear boundary of the research type that I am interested to explore. Imitating Markham and Buchanan's (2012) definition of *internet research*, and referencing from Gadiraju et al.'s (2014) taxonomy of crowd work tasks, I give a working definition of crowd work-based research as encompassing the following academic inquiries on a crowd work platform to

(a) survey crowd workers (Buhrmester et al., 2011). For instance, my survey study on MTurk disclosed several privacy violations and concerns among the MTurk workers (Xia et al., 2017).

(b) conduct user-testing for academic tasks with crowd workers (Kittur et al., 2008). For instance, Liu et al. (2012) tested the usability of a graduate school's website with MTurk workers.

(c) leverage crowd workers' labor for content creation, such as media transcription, translation, and tagging (Callison-Burch, 2009). For instance, Marge et al. (2010) leveraged MTurk workers to transcribe audio recordings from native and non-native English, male, and female speakers.

(d) harness crowd workers for information retrieval and identification (Higgins et al., 2010). For instance, Gottlieb et al. (2012) harnessed MTurk workers to annotate the geo-location information from some random videos on the internet.

(e) use crowd workers to conduct psychological and user experience (UX) experiments. Exemplary academic research: (Casler et al., 2013; Gardner et al., 2012). For example, Casler et al. (2013) converted a behavioral, face-to-face psychological experiment to an online test on MTurk and found that MTurk workers provided quite valid results comparable to those in the lab environment.

Research Questions

Before revealing my research questions, I hope to reflect on my path to them. I have had a lasting interest in crowd work since I first used it as a means to collect research data. When I launched my first survey on MTurk in 2013, I was impressed by how cheap and fast it was to obtain responses. More incredibly, except for a few obvious cheaters, the data quality was satisfactory. My positive impression of using MTurk for research purposes motivated me to research crowd work and crowdsourcing, and I have become a believer, to this day, of the power of collective intelligence that amass the ideas and wisdom from a large population of people to address simple as well as complex research tasks. However, gradually, I also started to see several dark sides of harnessing crowd work for research data collection. For instance, when I conducted an international survey about privacy on MTurk, I noticed that many task requesters were intrusive to inquire much personal and sensitive information, e.g., sexual and religious orientations, from MTurk workers; some requesters even publish malicious tasks, e.g., a link to a phishing website, that hacked MTurk workers' bank accounts (Xia et al., 2017). Meanwhile, on the MTurk worker's side, I found that some MTurk workers' cheating was not a trivial or

isolated case but a consistent phenomenon; not the least, Amazon is hands-off to monitor or regulate bad behaviors of either workers or requesters.

On the other side, I am a person who like contemplating ethics. I was brought up in a Confucian culture. Confucianism is a deductive moral system that emphasizes intuition and extrapolation. For example, a moral relationship between the father and son in a family should be extrapolated to the relationship between an emperor and their servants. However, Confucianism does not emphasize logical arguments. For example, what the logical connection is between a moral relationship in a family and a moral relation in an imperial court, and how can the former be readily extrapolated to the latter was not thoroughly articulated or justified in Confucianism. Meanwhile, I am interested and impressed by Western philosophy's rationality and empirical inquiries that have descended from Aristotle to moral philosophers such as Bentham, J. S. Mill, Kant, and Rawls. I also had the concrete experience and comparison between my academic training in mainland China and in the U.S. where I got to know the role of IRB and the guidance of the Belmont principles in academic research here, which are lacking in mainland China.

Hence, combining my research interests in crowd work and ethics together, I am curious: What does ethics mean in crowd work? When academic researchers conduct studies on a crowd work platform such as MTurk like I often do, what ethical considerations do they have in mind? Since the IRB will review and monitor research on a crowd work platform, and the IRB is guided by the Belmont Report, then how does IRB think about ethics in crowd work, and how do the Belmont principles apply in this context? Finally, would the context of crowd work impose certain ethical challenges to researchers, IRBs, and the Belmont Principles? These curiosities together with the literature that I have reviewed, which I will present in the next two chapters, generate the following research questions:

RQ1: What do academic researchers perceive as the ethical issues with crowd work-based research?

The first research question aims to empirically probe academic researchers' perceptions of the ethical issues with crowd work-based research. For example, how do they understand research ethics in this context generally?; what are their research practices to respect and abide the ethical principles of the Belmont Report to meet IRB's examination and approval?; how do they deal with certain research dilemmas, for example, between ensuring data quality and rejecting crowd workers?

RQ2: How do IRB directors and analysts interpret and enforce the federal government's research mandates in the context of crowd work-based research?

The second research question probes the same themes as those in RQ1 but from the perspective of IRB directors and analysts. IRB directors and analysts are not like IRB board members who are also academic scholars. I prefer to interview IRB directors and analysts because they do the reviews but have not conducted research themselves. Hence, they have more distinctive roles from IRB board members. I interview IRB directors and analysts on how they perceive ethics in crowd work-based research, and how they judge the obedience and violation of the ethical principles in crowd work-based research. I also probe their experiences and perceptions of ethical violations, e.g., how do they perceive and handle complaints from crowd workers, if any. I also inquire into how they perceive research ethics in crowd work and how they deal with various research dilemmas that would also face the researchers.

RQ3: How do the existing guidelines specific for crowd work-based research, such as those drafted by IRBs and researchers for using crowd work for academic purposes, consider various ethical issues in crowd work-based research?

A few researchers and IRBs have drafted their own guidelines or templates for conducting research via crowd work. To cross reference the researchers and IRBs' thoughts in interviews, this research question aims to understand what ethical issues have been covered or considered in the existing guidelines specific for crowd work-based research. I have collected these guidelines and conduct a document analysis on them. Meanwhile, since the Belmont Report and AoIR IRE 3.0 are not directly addressing ethical problems in crowd work-based research but are likely to be referenced by researchers and IRBs, I will also discuss the application and limitations of these two documents.

Key findings and contributions

This dissertation research has made four findings and contributions. First, this dissertation work is a pioneering effort to focus on ethical issues in crowd work-based research. Research ethics has been an extensively deliberated topic since the Belmont Report, and contemporarily, a great deal of research effort has been spent on examining and considering ethics in research, such as AoIR IRE 3.0, and the PERVADE (Pervasive Data Ethics) project to study Big Data ethics ("PERVADE," 2017). In the context of crowd work, however, research ethics has not been extensively studied. Ethics in crowd work-based research embeds particular challenges and controversies because it entangles the for-profit and monetization purpose of a business crowdsourcing platform with the knowledge generalization and volunteer-oriented purpose of academic research. Hence, the respect of crowd workers as human subjects is not only about respecting autonomy but also about avoiding dehumanization; benefits and costs are not only related to the research outcome but also to the monetary payment and worker reputation protection; justice is not only about an equitable sampling of research subjects on a crowd work platform but also about the distribution of benefits to the populations not on this platform. To my

best knowledge, little empirical research so far has investigated these issues. This dissertation work sets out to uncover and discuss these ethical issues, and thus, contributes to fill this knowledge gap and shed light on the domain of research ethics broadly and guide academic research with crowd work specifically.

Second, this dissertation work has not only uncovered the scope and variety of ethical issues but also analyzes the origin of these ethical issues. Specifically, it deliberates the "original sin" of various ethical issues in crowd work-based research, which comprises Amazon's fundamental stance of "Human-as-a-service," which sets the tone of dehumanizing crowd workers subsequently; Amazon's confusion of terminology, which still baffles the users of MTurk and the other crowd workers; as well as Amazon's abdication of responsibilities between MTurk, crowd workers, and requesters, which has become a model for the subsequent crowd work platforms for framing these responsibilities loosely by positioning crowd workers as independent-contractors. More specifically, Amazon's stance of "Human-as-a-service" set the basic of crowd work and debases crowd workers as a singular, inhuman, group term of "service" and further commodifies their labor. It gives rises to the ethical issues such as dehumanization, fair payment, and justice. Amazon's confusion of terminology, such as its position of MTurk workers as independent contractors, muddles the boundary between wage-earners and research volunteers and thus creates ethical concerns regarding undue influence, minimum wage, and voluntary participation. Amazon's abdication of responsibilities renders MTurk, crowd workers, and requesters not accountable to each other's duty and behaviors, and it engenders ethical issues such as cheating and fraud. Thus, this dissertation work has made a theoretical contribution to the interpretation of the ethical issues and their origin in crowd work-based research.

Third, though researchers and IRBs are abiding with the Belmont principles in administering crowd work-based research, little study has investigated the application and limitations of the Belmont principles in this research context. This dissertation work has examined how IRB directors and analysts as well as researchers interpret and implement the Belmont principles in crowd work-based research. It also discovers its limitations. For example, certain ethical themes such as ethical issues in payment and data in crowd work-based research are not well covered by the Belmont principles; certain subthemes in each ethical principle, such as dehumanization and empowerment in respect for persons, are also not incorporated in the Belmont Report. More fundamentally, this dissertation work argues that the Belmont Report's stringent delineation of the boundary between research and practice is not suitable for crowd work-based research. Thus, this dissertation work contributes to the ongoing deliberation of the Belmont Report's application and limitation in contemporary research contexts (Vitak et al., 2016; Friesen et al., 2017) and can help IRBs and researchers be more cautious about the ethical issues beyond the Belmont principles in crowd work-based research.

Fourth, this dissertation can contribute to the future development of academic ethical guidelines for research, such as that of AoIR. Although AoIR IRE 3.0 has mentioned that research in a crowdsourcing platform can induce certain ethical concerns such as fair payment and data quality, it has not examined this type of research and the associated ethical risks thoroughly enough. This dissertation demonstrates that crowd work-based research is a form of internet research but embeds particular ethical challenges. For example, an informed consent is hard to obtain in both the initial stage as well as the dissemination stage, i.e., the stage of publicizing research findings and data, of crowd work-based research because a research session is transient and the same batch of crowd workers are hard to identify and contact (Xia and

McKernan, 2020). Meanwhile, it is also an open question how to instantiate AoIR IRE 3.0's call for protecting researchers in a crowd work setting. This dissertation has also discussed the limitations of AoIR IRE 3.0 for crowd work-based research, and thus can contribute to the future versions of it.

Finally, although this dissertation focuses on MTurk as a case study of ethics in crowd work-based research, its findings can apply to the other crowd work platforms that scholars leverage to collect research data. More broadly, the implications of this dissertation can be referenced by scholars interested in ethical issues in the gig economy. Crowd work is an essential constituent of the gig economy (De Stefano, 2015), and MTurk exemplifies the paradigm of crowd work. The ethical issues uncovered in this dissertation work, such as exploitation, undue influence, dehumanization, empowerment, and abdication of responsibilities, are presumably not uncommon in the other platforms in the gig economy such as Uber and Airbnb. Future work can build on this dissertation to explore the other ethical issues in the gig economy apart from the academic research ethics.

CHAPTER 2: LITERATURE REVIEW PART I: ETHICS

Chapter 2 is the literature review of ethics and consists of two parts. In the first part, I review academic research ethics, from the Belmont Report to AoIR IREs. These literatures of ethics are most pertinent to academic research. In the second part, I review a few philosophical perspectives of ethics, such as Kantian and utilitarian ethics and different stances of justice. These literatures are the foundation for academic research ethics and can still inspire ethical deliberation and reflection in crowd work-based research.

Evolution of Research Ethics from the Belmont Report to AoIR

This section traverses the evolution of academic research ethics from its genesis shortly after WWII till the most recent AoIR IRE 3.0. Such an evolution demonstrates the development and complexity of research ethics as well as researchers' efforts to address them and improve the earlier versions of ethical guidelines.

Genesis of Research Ethics

During WWII, many Ally captives in concentration camps were abused by Nazi German scientists in biomedical experiments. Such inhumane treatments on human subjects were revealed during the Nuremberg Trial and afterward raised substantial critique and reflection on the relation between science and ethics, the purpose of conducting experiment with human subjects, and the researcher's role (Edgar and Rothman, 1995). Hence, soon after WWII, how to protect human subjects in research started to draw more attention and deliberation.

The Nuremberg Code (Code, N., 1947) could be seen as a first attempt to tackle various ethical issues in research experiments with human subjects. It proposed ten principles to direct experimental conduct. For example, the first principle is "The voluntary consent of the human subject is absolutely essential," and the last principle is "During the course of the experiment, the

scientist in charge must be prepared to terminate the experiment at any stage" (Code, N., 1947, p. 1). The central principle is a human subject's voluntary and informed consent to participate, which has been regarded as a basic human right by the Nuremberg Code (Shuster, 1997). Even though the Nuremberg Code has never been mandated as law by any nation or formally adopted as ethical guidelines by any major medical association, it has expounded and extended the traditional Hippocratic ethics of physicians to scientific research and has also inspired more specific ethical guidelines for human subjects research later (Shuster, 1997).

Entering the 1960s, new ethical issues and challenges emerged from the human subjects research. For example, thalidomide, which was still labeled as an experimental drug back then, was promoted across the U.S. without sufficient evaluation of its effects that led to thousands of children born with birth defects; the psychedelic drug LSD was distributed to children by Harvard professors Alpert and Leary; cancer cells had been tested on debilitated patients at a Jewish Hospital (Edgar and Rothman, 1995). During this period, Beecher's (1966) monumental work revealed a set of unethical practices across the variety of clinical research projects and noted that informed consent was frequently not obtained, or not in sufficient detail to allow patients to be genuinely aware of their participation or the potential consequence that could befall them in an experiment. For instance, Beecher (1966) listed an example that some investigators conducted an experiment to test the effect antibacterial drug *sulfadiazine* on rheumatic fever, however, the experiment subjects "were not informed, did not consent and were not aware" (p. 69) that they had been treated and tested.

In this context, the Declaration of Helsinki (DoH) was proposed by the World Medical Association (WMA) to specifically target at the ethical issues in clinical research and to emphasize the physician-investigators' obligations to the human subjects (Shuster, 1997). Since

DoH's first publication in 1964, WMA has made seven revisions, and the most recent one was issued in 2008. DoH is often regarded as the cornerstone and most widely recognized ethical guide for medical research (Crawley, 2003; Macklin, 2003). However, DoH has two limitations. First, it focuses on the ethics of medical research, and as such, may not be readily adaptable to behavioral research in social science, which also invites ethical considerations. Second, DoH's succinctness may fall short to clarify the "increasing complexity of research issues" (Carlson et al., 2004, p. 705) and perplex researchers as to how to operationalize and apply these principles in practice. Hence, the emergence of the Belmont Report (1979) and the establishment of IRB accordingly (in the U.S.) seemed to be a natural development.

The Belmont Report

The Belmont Report (1979) established a new milestone in human subjects research ethics. It not only addressed clinical research and medical experiments but also extended to social science and behavioral research that involve human subjects. The objective of the Belmont Report is "to provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 3). Hence, besides the general ethical principles, it also discussed the boundaries between practice and research as well as the applications of ethical principles. Since the Belmont Report has a significant influence on research ethics, it deserves a more detailed review.

The Belmont Report (1979) clarified the boundaries between practice and research because the deviant practices from standard were often named as experimental when these terms were not carefully defined. The report thus differentiated the boundary between practice and research such that *practice* denotes the interventions that "have a reasonable expectation of

success" (p. 3) to enhance a patient or client's well-being, whereas *research* is an activity to test hypotheses, draw conclusions, and generalize knowledge, and usually requires a formal protocol with an objective and a set of procedures to achieve that objective. In this regard, certain practices might be experimental but are not research, thereby "radically new" practices should be constrained by formal research procedures to be reviewed to ensure their safety and effectiveness (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 3-4). Such a boundary between practice and research may become more porous in social science. Though for social scientists such as sociologists, political scientists, anthropologists, their professional practice is essentially research, for criminologists who are members of or collaborating with law enforcement, or for social scientists who are public policy makers/contributors, their research and professional practice may need to be carefully distinguished (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol 1., p. 1- 23).

Most essentially, the Belmont Report (1979) proposed three fundamental ethical principles to conduct human subjects research: (1) respect for persons, (2) beneficence, and (3) justice. *Respect for persons* requires researchers to treat human subjects as autonomous agents who can deliberate their goals and actions. In case individuals lack such autonomy, this principle then requires the researchers to protect them depending on the risk of harm and likelihood of benefit. Respect for persons further demands that the research participation is voluntary and adequate information must be given to the participant (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 4-5). This principle derives from the right that "each of us possesses to be treated as a person, and in the duty which all of us have, to have respect for persons, to treat a person as such, and not as an object" (National

Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 3- 74). Meanwhile, this principle serves as "the basis for our sense of moral responsibility and is considered apart from any interest we might have in respecting other persons (e.g., that such respect is useful or will tend to protect us)" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 8- 5). We could readily trace this principle back to Kant who believed and contended that every person deserves to be respected as an autonomous and rational human being and therefore must be treated as ends in themselves rather than merely as certain means or means to certain ends.

The principle of *beneficence* has two general rules: the Hippocratic oath of "do no harm" and the maximization/minimization of possible benefits/harms. The possible harms, notably, should consider both the probability of a harm's occurrence and the magnitude of its consequence (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 5). More specifically, the Hippocratic Oath for physicians is: "I will follow that system of regimen which, according to my ability and judgment, I consider for the benefit of my patients" and it was restated at Geneva in 1948 to include "the health and life of my patient will be my first consideration" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 1- 29). Though the Hippocratic Oath emphasizes the pursuit and protection of patients' benefits, it's not adequate alone to researcher and subject relationship in research, and thus, it has to be combined with subject's informed consent (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 2, p. 26- 13).

There are various types of benefits and harms involved in human subjects research. For example, it could benefit society as a whole by developing or perfecting diagnostic modalities

and improving our understanding of human biology and psychology; it could also benefit research subjects with direct health enhancement and economic compensation (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 2- 36 – 2- 42). On the other side, human subjects research could also pose physical, psychological, social, economic, and legal risks to research subjects and society alike (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 2- 6 – 2- 29). Given the variety and complexity of these benefits/risks, it would be appropriate to avoid calculating the balance of them by any mathematical modeling, and researchers should also consider both magnitude and probability (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 2-48). Thus, the benefits and harms in human subjects research are recommended to be adequately described instead of mathematically calculated (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 2- 50). Here, we could perceive a utilitarian assumption behind this ethical principle, which is to maximize the balance of benefits to harms.

Third, the principle of *justice* requires the researcher to deliberate and explain how human subjects should be treated equally such as who should take the risk of research and who should receive its benefits (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 5-6). Justice in human subjects research is twofold: "fairness in distribution" and "what is deserved" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 5). As such, the principle of justice requires researchers to consider at least the following approaches: (1) treat each subject with an equal share of benefits and risks; (2) treat each subject according to their

individual need; (3) treat each subject based on their effort; (4) treat each subject according to their contribution; and (5) treat each subject according to their merit (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 5). The proposal of the justice principle was a reaction to some improper research activities in which human subjects from a poor area were recruited to burden the potential experimental risks whereas rich people enjoyed the benefits from such research (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, p. 6). To the Belmont Report, "what is deserved" denotes directly to the meaning of justice and "fairness of distribution" points to the expression of "distributive justice" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, vol. 1, p. 6- 1).

To demonstrate these ethical principles, the Belmont Report (1979) further proposed three practical guidelines for academic research: (1) obtain informed consent from research subjects; (2) assess research risks and benefits on research subjects, and (3) select research subjects with justice criteria and process. More specifically, informed consent requires the researcher to disclose adequate information to the research subjects and make sure they understand such information. The research subjects can then decide voluntarily whether or not to participate. Meanwhile, the researcher should assess research risks not only in terms of their magnitude but also their likelihood. In parallel, the researcher should evaluate benefits for research subjects as well as for research outcome in the long term. Finally, the fair selection of research subjects indicates that the researcher should exhibit social justice to avoid distributing research benefits to their favored groups while burdening the risks to undesirable persons while also ensure individual justice to treat each human subject equally (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979).

The Institutional Review Board (IRB) is a product of the Belmont Report, and it is charged with ensuring that institutions that conduct research and that receive federal funding for research follow the ethical guidelines and applications in the Belmont report. First, IRB must obey what counts as *research*, which is officially presented in the Belmont Report Appendix Vol. 1: "research (involving humans) is any manipulation, observation, or other study of a human being – or of anything related to that human being that might subsequently result in manipulation of that human being – done with the intent of developing new knowledge and which differs in any way from customary medical (or other professional) practice" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978, p. 1- 6-7; 1- 23). By definition, IRB is "an institutional review board established in accord with and for the purpose expressed in the policy" of human subjects' protection and is regulated by the Department of Health and Human Services (US Department of Health and Human Services, 2009, p. 4). IRB's primary purpose is to monitor human subjects research in the U.S. universities that receive federal funding (US Department of Health and Human Services, 2009; Vitak et al., 2016; Martin, 2007).

IRB has many federal regulations for which it must ensure institutional compliance. For example, IRB must have at least five members with diverse backgrounds to provide a thorough and multiangled review of human subjects research (US Department of Health and Human Services, 2009, p. 6). The IRB membership cannot be only men or women to avoid any bias of gender. As well, at least one member should be unaffiliated with the institution; at least one member's primary concerns are in scientific areas, and at least another member's primary concerns are in nonscientific areas (US Department of Health and Human Services, 2009, p. 6). To approve research, IRB must assess the following requirements: (1) minimize risks to human

subjects; (2) reasonable risks comparing to anticipated benefits to subjects and to knowledge development; (3) equitable selection of research subjects; (4) obtain and document informed consent from each subject or the subject's legally authorized representative; and, (5) monitor and guarantee appropriately the research subjects' safety and their data privacy (US Department of Health and Human Services, 2009, p. 6-7). Moreover, IRB has specific requirements for the review procedures, the criteria for its approval of research, and the additional protections for vulnerable groups, such as pregnant women (US Department of Health and Human Services, 2009).

Since the emergence of the Belmont Report and IRB, there have been some disputes and critiques. To the three Belmont principles, they were challenged on their application in online research ethics. For example, how should researchers ensure and obtain informed consent if research subjects are unaware of being monitored or unable to choose whether their data can be collected (Vitak et al., 2016; Boyd and Crawford, 2012). Meanwhile, minimization of risks to subjects online is hard to achieve because deanonymization could be done beyond the scope of research context and researcher's control (Vitak et al., 2016; Markham and Buchanan, 2012). Moreover, "fairness in distribution" and "what is deserved" become hard to assess because online participation does not represent the U.S. or global demographics (Vitak et al., 2016; Hargittai, 2015), and hence the benefits from research findings with online populations may not be distributed to the populations who do not have an online accessibility and may only benefit those who are tech-savvy and enjoying the internet already. In crowd work-based research, it is further arguable that, the three Belmont principles could face more challenges to concretize and interpret because crowd work-based research has more specific motivations, characteristics, and controversies than those in online research in broad.

There are also critiques on IRB in terms of its review board constituency and overly conservative view of human subjects research. First, the IRB review board is a localized committee and its majority members are affiliated with the institution and dominate the IRB discussions (Edgar and Rothman, 1995). Hence, the members from the institution might intentionally avoid setting standards for the other researchers that could backfire their own research in future or setting standards that might impact the relationship with the applicants in the same institution (Edgar and Rothman, 1995). Second, the IRB's actual instantiation of reviews usually assumes that the researcher always has power advantages over human subjects, and it holds a postpositivist view of conducting research with objectivity and distance as the principal researcher stance (Boser, 2007). Such a limitation potentially impedes the approval and progress of qualitative participatory research where the distinction between the researcher and the researched is blurred. For example, in community-based research, the participants who are affected by the research also shape the research questions and procedures; in this regard, the power-*over* relationship between researchers and research subjects presumed by the IRB has become a power-*with* relationship in such type of research (Boser, 2007).

To sum up, the Belmont Report and IRB have enhanced the implementation of research ethics and the protection of human subjects. They serve as the cornerstone to subsequent research ethical guidelines beyond medical research and the benchmark for human subjects research in the other countries (e.g., the Ethical Conduct for Research Involving Humans and the Canadian Institute of Health Research). Meanwhile, however, the advancement of the internet and associated platforms and technologies have provided researchers with new research settings and techniques and brings new challenges to the Belmont principles and IRB oversight. In this

regard, it is worthwhile to review the more recent ethical guidelines for research that is conducted on the internet or is facilitated by internet such as in terms of data collection.

Internet research ethics

Markham and Buchanan (2012) defined the internet research to include the following types of inquiries: (1) utilize the internet to collect information, e.g., through online survey or web crawling; (2) investigate how people use and access the internet, e.g., through observing social network activity and usage; (3) utilize or engage in "data processing, analysis, or storage of datasets, databanks, and/or repositories available via the internet;" (4) research on software, code, and internet technologies, (5) examine the "design or structures of systems, interfaces, pages, and elements;" (6) employ "visual and textual analysis, semiotic analysis, content analysis, or other methods of analysis to study the web and/or internet-facilitated images, writings, and media forms;" and (7) research on governments, industries, corporations, and military forces' usage, regulation, and large-scale production of the internet (p. 3-4). Internet research can include and be affected by various emerging contexts and devices. Hence, internet research should be considered in "its broadest sense" (Markham and Buchanan, 2012, p. 4).

Internet research has numerous benefits. For example, compared to offline research, the internet affords researchers to conveniently access and recruit a geographically and culturally distributed pool of potential research subjects with diverse backgrounds (Frankel and Siang, 1999; Kraut et al., 2004); internet research was less likely to cause physical harms; some subjects may find it more convenient to participate in research on the internet than in the physical world if they have online access (Frankel and Siang, 1999); or they may find it more comfortable to participate in certain research online than in the physical world, such as the research about sex (Binik et al., 1999). Meanwhile, the Email-based interview could better clarify concepts and

engage and empower participants more than in face-to-face based interview; for example, the participants can change the sequence of the questions they prefer to answer in order and ask the researchers to provide more background and details about the research goals and methods. (Murray and Sixsmith, 1998). Internet communities such as mailing lists, chatrooms, and website discussion boards offer rich data for researchers that would be difficult to obtain in a physical context (Eysenbach and Till, 2001). Finally, the internet itself can be examined as a social phenomenon such as how computer-mediated communication on the internet differs from face-to-face communication offline (Kraut et al., 2004).

Internet research could also bring new ethical challenges and dilemmas. First, research participants' privacy may be compromised from unsecured data transmission, intentional linkage of different data sets to divulge identifiable information, and inadvertent identity disclosure during data collection and storage (Frankel and Siang, 1999; Ess et al., 2002; Kraut et al., 2004). Second, informed consent can be complicated in internet research. For example, the boundary between public and private is blurred on the internet so that when informed consent is required becomes a controversy such as whether researchers should obtain informed consent in collecting user-generated content in public records (Frankel and Siang, 1999; Walther, 2002). Hence, when deindividuation and anonymity are assumed in internet research subjects, the need to ensure the implementation of informed consent and participants' comprehension may be ignored or complicated (Postmes and Spears, 1998; Frankel and Siang, 1999; Kraut et al., 2004). Finally, justice in internet research could add new controversies. For instance, internet samples may be biased toward those who have access to the internet or who can afford to do so; further bias could also occur by research subjects' self-selection and drop-out of online surveys (Kraut et al.,

2004). Meanwhile, the risks and benefits of internet research are difficult to identify and calculate, which further challenges the distribution of justice (Frankel and Siang, 1999).

In this context, several initial efforts were made to provide guidance to internet research. In 1999, the National Institutes of Health (NIH) and the American Association for the Advancement of Science (AAAS) sponsored a workshop to discuss and address various emerging ethical issues regarding human subjects in internet research. The workshop produced a report called "Ethical and Legal Aspects of Human Subjects Research on the internet," which was cited by diverse research disciplines (Frankel and Siang, 1999; Walther, 2002). However, a central critique to it is that some problematic internet research methods mentioned in the report do not constitute as research with *human* subjects, such as the analysis of messages posted in an online chat room or on a bulletin board system (BBS) on the internet (Walther, 2002). There were also discipline-oriented guidelines for internet research such as Kraut et al.'s (2004) recommendations for conducting psychological research online and Keller and Lee (2003)'s application of the American Psychological Association (APA)'s ethical codes for internet research with human subjects. Most notably, the Association of internet Researchers (AoIR) proposed three versions of ethical guidelines for internet research in 2002, 2012, and 2019 respectively.

AoIR IRE in 2002 (AoIR IRE 1.0) defined internet research's audience, purpose, rationale, and approach (Ess et al., 2002). The guidelines stressed that the internet-based/oriented research has introduced new ways and venues (e.g., email, instant messaging, and web pages) of human behavior and interaction scholarship, and along with such, has raised risk and safety issues to human subjects online (Ess et al., 2002). These guidelines are more specific than the

Belmont's ethical principles, while more interdisciplinary than the ethical codes from various academic associations.

AoIR IRE in 2012 (AoIR IRE 2.0) dealt with the expanding scope and context of internet research and identified three emerging tensions of the conceptualization of human subjects, the boundary between private and public, as well as how to handle personal data online (Markham and Buchanan, 2012). The updated guidelines also posed a set of ethical questions specific to internet researchers to consider such as: "How is the context defined and conceptualized?" "How is the context (venue/participants/data) being accessed?" "Who is involved in the study?" "What is the primary object of study?" "How are data being managed, stored, and represented?" "How are texts/persons/data being studied?" and "What particular issues might arise around the issue of minors or vulnerable persons?" (Markham and Buchanan, 2012, p. 8-11).

AoIR IRE in 2019 (AoIR IRE 3.0) further highlighted the protection of ethics in different stages of research, the acquisition of informed consent in Big Data research, and the protection of researchers' physical and psychological health that may be risked by research subjects' threats and research contents with extreme violence (Franzke et al. 2020). The different research stages and the associated ethical considerations include: (1) the initial research design stage with the considerations of presumed ethical risks that could happen potentially; (2) the initial research processes stage with the ethical considerations of how to protect research subjects' data from being de-identified and how to store data securely; (3) the analyses stage with the consideration of preventing subjects' data from being re-identified; (4) the dissemination stage with the consideration of how to share data appropriately and whether it is necessary to obtain consent from research subjects for data sharing and publication; and, (5) close of the project. Particularly, AoIR IRE 3.0 recommended internet researchers to pseudonymize their data or try to acquire

informed consent in the dissemination stage of research since in Big Data projects, informed consent is hard to acquire from every research subject at the onset of the research projects. AoIR IRE 3.0 also proposed protecting the researchers who may be threatened both physically and mentally by research subjects such as by their threats or deliberate retaliation or research content that may contain extreme violence (Franzke et al. 2020).

The internet has been growing and changing over time, and internet research ethics is also developing and adapting. It is impossible to develop a guideline that can address or prevent all the ethical issues in internet research. Thus, an inductive and flexible approach to the growing and complicated ethical issues in internet research is preferred, and as AoIR IRE 2.0 suggested:

> …it becomes clearer that an adaptive, inductive approach can yield potentially more ethically legitimate outcomes than a simple adherence to a set of instantiated rules. (Markham and Buchanan, 2012, p. 5)

Philosophical Perspectives of Ethics

Though not a philosophy-oriented dissertation research, it is necessary to provide an overview of some philosophical perspectives of ethics which can be applied to examine and reflect crowd work-based research ethics. The philosophical perspectives of ethics covered in this section include Kantian and utilitarian ethics as well as three philosophical interpretations of justice.

Kantian ethics

Immanuel Kant is often regarded as the greatest modern philosopher. His philosophy of ethics is well represented in his monumental work *Groundwork of the Metaphysic of Morals*. I will use H. J. Paton's classic translation of this book (Kant, 1964) and reference to a few other sources, primarily M. J. Sandel's (2009) book *Justice* and W. A. Wick's introduction to Kantian

ethics (Kant, 1995). Kant believed that the only good without any qualification or condition is a good will regardless of any judgment of various consequences that such a good will may end up with. For example, in practice, a will with good motive may result in unexpected or even undesirable consequences, but regarding its moral value, such a will still represent the only good - in itself and unconditional (Kant, 1964, p. 61-62). By unconditional, Kant meant that besides good will, there are still many good things or results in various aspects, but they just represent conditioned good under certain circumstances, only a good will contains good in all conditions and as such, represents unconditional good (Kant, 1964, p. 17-18).

To overcome hesitation to carry out a good will such as the temptation to calculate the benefits and risks in doing so or from its consequence, people must abide the motive of duty instead of the motive of inclination (Sandel, 2010, p. 111-112). Duty is to be acted from not for the purpose of the action (i.e., with *a posteriori* motive) but for the reverence of *law* (i.e., with *a priori* motive) (Kant, 1964, p. 68). People can respect law and act from duty instead of from inclination (e.g., motivated to take this action in order to maximize the happiness of the majority of people) or from self-interest (e.g., maximize my own happiness at the expense of other people's pain) because human beings are not only sentient beings but also rational beings (Kant, 1964, p. 69; Sandel, 2010, p. 118).

What is such a law like for human beings to respect and act from? Kant regarded such a law as the *categorical imperative* (Kant, 1964, p. 69). *Imperative* means it is a directive for action in certain ways and forbids inconsistencies; *Categorical* means it is unqualified and unconditional as opposed to hypothetical or conditional (Kant, 1964, p. 70-71). Because this law is a categorical imperative, it is universal in the sense that every human being must abide it and respect it unconditionally. Therefore, since this law is universal and unconditional, there can only

be one of it serving as the principle law, and the other laws such as "You should not kill" or "You should not lie" are all derived from this universal, principle law (Kant, 1964, p. 30).

The next question is, how does a person know whether they have acted from duty and respected the universal law? Kant proposed two formulae. First, a person must act in no other way except that they can affirm "that I can also will that my maxim should become a universal law" (Kant, 1964, p. 70). For example, a person may be inclined to tell a white lie to appease her friend's anguish, which *seems to be* a good will for an action out of altruism. But when she asks herself "can my telling a lie for the others' benefit be a universal law?", then she would immediately realize that it certainly cannot because if everybody tells a lie like this, there is no conception of *lie* (which must be distinguished from the truth) anymore. Thus, telling a lie regardless of its purpose is not an action from duty and cannot be a categorical imperative, which in turn manifests that it is unethical (Kant, 1964, p. 31).

Kant's second formula requires a person to "act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end" (Kant, 1964, p. 96). Humans are ends in themselves because they are rational beings and thus have the dignity and deserve to be respected by themselves as well as by the other human beings (Sandel, 2010, p. 122). Seemingly different on the surface, the second principle is equivalent to the first principle in logic. To treat a person only as a means is to disrespect and neglect their end or their purpose. But if it is to be universalized, then everybody would treat the others as means and at the same time be treated by the others as means. As such, nobody's end or purpose would be fulfilled which becomes self-contradictory to treat others as means in order to fulfill their own ends. Therefore, because from duty we should

act on maxims that we will to be universal laws, we have to treat the other people as well as ourselves not only as a means but also as an end.

In this logic, because everyone is their own end (i.e., to fulfill their own purpose) and treats the others also as their ends, there should be no need of any external power to "lay down the law" to each and every individual without subjecting themselves to it (Kant, 1995, p. xx). Hence, every human, as a rational being, is autonomous and should be treated with each other as such. A human, as a rational being, is free when they act from duty for the good will because they are not subject to any external law, purpose, or temptation to carry out an action. For example, a wealthy person may *appear to be* free because he has money to buy anything or even do anything, but in Kant's view, he is not because he is subject to and constrained by the utility and purpose of money. On the contrary, a miserable man, having tasted all the bitterness of life and longs badly for death, still preserves his life not due to fear or hope, is a *free* man, because he acts from duty to treat himself as an end - as the ultimate purpose of a living person to be alive (Kant, 1964, p. 65-66).

Utilitarian ethics

Utilitarian ethics is a contrast to Kantian ethics, and such a contrast embodies in two fundamental ways: Kant is an outstanding rationalist, utilitarianists are hardcore empiricists; Kantian ethics emphasizes the good motive, utilitarian ethics applauds good consequence. In this subsection, I base my review on Jeremy Bentham's *Introduction to the Principles of Morals and Legislation* and John Stuart Mill's *Utilitarianism* in the book *John Stuart Mill and Jeremy Bentham Utilitarianism and Other Essays* (Bentham and Mill, 2004). I also reference from Alan Ryan's introduction in this book, and M. J. Sandel (2010)'s interpretation of the utilitarian ethics in his book *Justice*.

Jeremy Bentham is the pioneer of utilitarian ethics. He founded the utilitarian ethics with a famous dictum "Mankind governed by pain and pleasure" (Bentham and Mill, 2004, p. 65). This dictum induces his subsequent argument that for pain and pleasure alone, human beings decide what to do and what ought not to do and differentiate between the standard of right and wrong (Bentham and Mill, 2004, p. 65). Then, Bentham provided his definition of utility: "By utility it meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good or happiness or to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is involved" (Bentham and Mill, 2004, p. 66). Party could refer to an individual or a community, as Bentham elaborated: "if that party be the community in general, then the happiness of the community; if a particular individual, then the happiness of the individual" (Bentham and Mill, 2004, p. 66).

Based on it, Bentham proposed his principle of utility, which undergirds utilitarian ethics: every action must be judged by its tendency to promote the utility of the party whose interest is concerned or to diminish that utility (Bentham and Mill, 2004, p. 65). According to this principle, on an individual level, a person's action is ethical so long as it tends to maximize their happiness or minimize their pain; on a collective level, a community or government's action is ethical if it seeks for the most benefits for the majority of its members or people. Such a somewhat selfish and calculative principle for ethics inevitably provoked much criticism, but at least Bentham briefly remarked that an individual's interest is also linked to the interest of the community to which they belong. He claimed that a community is composed of individuals who constitute it, and these individuals are seen as the community members and as such, the interest of the community is the sum of each individual member's interest (Bentham and Mill, 2004, p. 66). Hence, the community's interest is closely related to each individual member's interest, and

it is meaningless to discuss a community's interest "without understanding what the interest is of the individual" (Bentham and Mill, 2004, p. 66). In Bentham's view, every moral judgment must draw, even implicitly, on the principle of maximizing human happiness, at least in the long term (Sandel, 2010, p. 34-35).

Bentham also proposed a set of qualitative measures for different types of pleasure and pain. For example, the value of either pleasure or pain of a single person can be measured by its "intensity," "duration," "certainty or uncertainty," and "propinquity or remoteness" (Bentham and Mill, 2004, p. 86-87). Meanwhile, pleasures or pains are either *simple* or *complex* (Bentham and Mill, 2004, p. 89). Nonetheless, there have been numerous critiques, and two central objections to Bentham's utilitarian ethics are about its degradation of individual's right and dignity as well as its reduction of moral value to a single scale of pleasure and pain (Sandel, 2010, p. 48).

As a pupil of Bentham and developer of utilitarianism, J. S. Mill tried to defend utilitarian ethics by reconciling it with individual's right and humanizing it with less calculation of utility in every single case but more reckoning for the long-term benefits. Responding to various criticism to utilitarianism, Mill clarified that utility is not about *usefulness* but equates to *pleasure* (Bentham and Mill, 2004, p. 277) and he interpreted Bentham's principle of utility as that "actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness" (Bentham and Mill, 2004, p. 278). Happiness is "intended pleasure, and the absence of pain" and unhappiness or pain is "the privation of pleasure" (Bentham and Mill, 2004, p. 278). Mill further distinguished the qualitative assessment of pleasure and the quantitative assessment of it. For example, a human's pleasure cannot be reduced to a swine's pleasure because a human has higher faculties to desire and enjoy higher-end pleasures beyond

the sensuous ones. Therefore, some pleasures are more valuable and desirable in quality, and the calculation of pleasure over pain should not be dependent on the quantity of pleasure only (Bentham and Mill, 2004, p. 278; Sandel, 2010, p. 53).

A follow-up question to Mill's utilitarian ethics is how do we know whether one pleasure is qualitatively higher than the other? Mill proposed a test: "Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure" (Bentham and Mill, 2004, p. 279). However, human beings do not necessarily always prefer "high quality" pleasures than "low quality" pleasures (Sandel, 2010, p. 54-55) and what people desire for their pleasure is not necessarily what they ought to desire (Russell, 1945, p. 778). Finally, in my perspective, relying judgment of the quality of different pleasure types on another person's experiences of them is deviant from self-experience and reflection and could potentially lead to a fallacious *halo effect* of bias or mislead those people who do not have such experiences to the pleasures that they *ought not* to desire.

Three interpretations of justice

Apart from Kantian ethics and utilitarian ethics, three interpretations of justice also worth reviewing because they can serve as research lens to examine the Belmont principle of justice. Broadly speaking, justice is concerned with political science and the whole society; narrowly speaking, justice can be related to interpersonal treatment and relationship (Sandel, 2010). Justice in research ethics, in my view, lies in the middle but still can be interpreted through different philosophical approaches.

The three representative approaches to justice are welfare, freedom, and virtue (Sandel, 2010, p. 19). The first approach emphasizes that justice is achieved through the maximization of

welfare. In a society, it means to increase economic prosperity so that most people's standard of living can be improved. Utilitarianism, as reviewed above, represents this approach of justice (Sandel, 2010, p. 19). According to this approach, it is just to sacrifice an individual or minority's interests for the benefit of the majority, and it is just to maximize personal happiness and liberty as long as it won't hurt other people because in the long term a person's success and happiness could benefit the whole society.

The second approach to justice maintains that individual freedom and rights must be respected. This approach can be further divided into a libertarian perspective of justice and an egalitarian perspective of justice (Sandel, 2010, p. 20). Libertarians believe that it is just when a contract is upheld by consenting people who voluntarily choose to make this contract. A typical example of this stance is that a person owns herself, and thus, organ selling and even assisted suicide is permittable and just, as long as the person has agreed to do so voluntarily without coercion (Sandel, 2010, p. 71). The egalitarian perspective of justice emphasizes the fairness of distribution of benefits and risks and equal opportunity for success and access to resources. Thus, policies should remedy social and economic inequality and help disadvantageous groups for equal opportunities with advantageous populations.

The third approach to justice is bound up with virtue and good. From a Kantian perspective, justice concerns with individual's right and dignity. Therefore, it is unjust to sacrifice an individual or minority's interests for the sake of the majority's benefits. Justice infers that every human being should be treated as the end in themselves rather than as a means to an end otherwise. John Rawls can be seen as a modern philosopher who inherited Kantian ethics and made it explicit to a theory of justice (e.g., Rawls, 1971, p. 118, 121). Rawls (1971) proposed two principles of justice. First, every person must have equal basic liberties such as to

vote and to hold public office, freedom of speech and assembly, property ownership, and freedom from arbitrary property seizure by the government. Second, social and economic inequalities are only allowed when such inequalities could benefit the least well-off people. For example, if taxing more on rich people can provide better health, education, and welfare system to the poorest than that in a more stringent equal society, then the wealth inequality between rich and poor and the taxing system are just (Sandel, 2010, p. 152).

Another stance to this third approach to justice is the Aristotelian ethics of virtue which includes two key ideas: (1) justice is teleological, which means that justice is judged by the telos (purpose, end, or nature) of a practice, and (2) justice is honorific, which indicates that justice is also judged by what virtues deserve honor and reward related to the telos (Sandel, 2010, p. 186). For example, suppose we want to distribute flutes and want to know when the distribution is just, then, according to Aristotle, it is just when we distribute the best flutes to best flute players because it fulfills the purpose of flute, which is to be played well. On the contrary, it is unjust to distribute flutes to people who offer the best price or to people who have the best looking because it is not honoring the telos of flute (Sandel, 2010, p. 187-188).

CHAPTER 3: LITERATURE REVIEW PART II: CROWD WORK

Chapter 3 is the literature review of crowd work and consists of two parts. In the first part, I review the nature, motivations, characteristics, and controversies of crowd workers. In the second part, I review the landscape and methodological issues of crowd work-based research.

The nature, motivations, characteristics, and controversies of crowd work

In this section, I first review the nature of crowd work, e.g., who the "crowd" refers to and what "work" means in the context of crowd work. Then, I review crowd work motivations and depict the complex influences of money. Afterward, I review the characteristics of crowd work. Finally, I describe and discuss some controversies of crowd work.

Crowd work's nature

Amazon's Mechanical Turk is the precursor and embodiment of crowd work, and to this day, still defines crowd work's nature. The name, Mechanical Turk, is originally from an 18th-century chess-playing device, which was presumed as automatic by the public but in fact was manipulated by a professional chess player hidden underneath the device. Amazon MTurk is "a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually" ("Amazon Mechanical Turk," n.d.-c). It is based on the idea that many jobs can still be done more effectively by human beings than by computers, such as transcribing audio recordings ("Amazon Mechanical Turk," n.d.-c).

Even though there has been some research on the demographics of MTurk workers (e.g., Hara et al., 2018; Difallah et al., 2018; Ipeirotis, 2010), the population size and constituents of MTurk workers are still unknown, and perhaps only Amazon has the accurate demographic data of MTurk workers. Amazon defines MTurk workers to be "virtually anyone…The only

requirement…is a computing device connected to the Internet and to be at least 18-years-old" ("Amazon Mechanical Turk," n.d.-c). Besides this description, Amazon puts two additional notable terms on MTurk workers. First, MTurk states that MTurk workers are "people who want to earn money in their spare time" ("Amazon Mechanical Turk," n.d.-c). It infers that MTurk workers' primary purpose is to earn money in a flexible way. Second, MTurk states that MTurk workers are "a global, on-demand, 24x7 workforce" ("Amazon Mechanical Turk," n.d.-c). It infers the diversity and global distribution of MTurk workers. Finally, Amazon clearly positions MTurk workers as "independent contractors" that "will not be entitled to any of the benefits that a Requester or Amazon Mechanical Turk may make available to its employees…[and] are not eligible to recover worker's compensation benefits in the event of injury" ("Amazon Mechanical Turk," n.d.-b). Basically, these descriptions and terms also define the meaning, motivation, and position of a "crowd worker" to a crowd work platform such as MTurk

Also, Amazon makes it quite clear that MTurk is primarily built for business. For example, it makes explicit statements such as "MTurk enables *companies* to harness the collective intelligence, skills, and insights from a global workforce" and "For *businesses and entrepreneurs* who want tasks completed, the MTurk service solves the problem of accessing a vast network of human intelligence with the efficiencies and cost-effectiveness of computers" ("Amazon Mechanical Turk," n.d.-c). In comparison, MTurk has only a very tiny mention of academic research in a Q&A about asking MTurk workers' demographic information: "Surveys should only be posted for legitimate *academic* or market research purposes" ("Amazon Mechanical Turk," n.d.-c). In terms of the nature of "work," MTurk refers to it as "Work for Hire." First, it means that MTurk's work is for the benefits of requesters, and "all ownership rights, including all intellectual property rights, will vest with that Requester" ("Amazon

Mechanical Turk," n.d.-b). Second, it means that MTurk workers will "waive all moral or other proprietary rights that you may have in that work product" ("Amazon Mechanical Turk," n.d.-b). Hence, it is evident that MTurk is leaning toward the requester's interests and advantages.

Meanwhile, Amazon does not define research or what can be done on MTurk or differentiate between academic research and business research. In general, Amazon states that MTurk can be used to "conducting simple data validation and research to more subjective tasks like survey participation, content moderation, and more" ("Amazon Mechanical Turk," n.d.-c). It exhibits two use cases on its website, one about Machine Learning and the other about Business process outsourcing, to demonstrate what MTurk can do to help business. It demonstrates that MTurk is not purposed for academic research, and as that can be expected, MTurk has no mention of voluntary participation or the protection of MTurk workers' autonomy.

Although MTurk is the most popular crowd work platform, there are a few other notable ones. For example, Prolific has been a new and strong competitor to MTurk and drawing a growing attention in academia (Palan and Schitter, 2018). The other crowd work platforms such as ClickWorker and CrowdFlower (now it has transformed itself into Figure Eight) also have been discussed in academia but have not been widely adopted by academic researchers (Vakharia and Lease, 2015). Despite of these new and old competitors, MTurk is still the benchmark crowd work platform for academic research and the most familiar one to academic researchers (Chandler et al., 2019; Hitlin, 2016).

Crowd work motivations

Crowd work appeals to both crowd workers and task requesters. As regards crowd workers, money is the primary motivation for them to participate and undertake crowd work tasks, which can be inferred in the literature review above about MTurk as well as from the

published works (e.g., Alkhatib et al., 2017; Durward et al., 2016; Lee et al., 2014; Crowston, 2012; Horton et al., 2011; Ipeirotis, 2010). This is true not just in the United States but in other countries such as India (Horton and Chilton, 2011; Ipeirotis, 2010a), and by young and older crowd workers alike (Hitlin, 2016; Brewer et al., 2016). However, even though money is a primary motivation, it is a complicated motivation as well. Prior research suggested that there were intrinsic reasons for crowd workers such as the enjoyment of the task, contribution to a community, personal and topical interest, and desire to learn new knowledge (Kaufmann et al., 2011; Alam and Campbell, 2012; Brabham, 2010, 2012). In addition, there are extrinsic motivations. For example, acknowledgement of contributions, desire for recognition from the requester and peer workers, and the opportunity to build an experience/expertise portfolio so as to advance one's career could all motivate people to participate and contribute to crowd work (Alam and Campbell, 2012; Brabham, 2012).

Although money is the primary motivation, payment is usually quite low in crowd work. For example, Ipeirotis (2010b) found that "25% of the HITs create on Mechanical Turk have a price tag of just one cent, 70% of the HITs have a reward of five cents or less, and 90% of the HITs come with a reward of less than 10 cents" (p. 19). Pew Research also found that the majority of HITs on MTurk are paid less than 10 cents (Hitlin, 2016). Such a phenomenon of low payment in crowd work is more recently corroborated by Hara et al. (2018), who found that the median hourly wage on MTurk was only about $2/hour and 96% of MTurk workers earned below a federal minimum wage. Despite such minuscule payment, to some people, crowd work earnings are their main income. For example, many Indian people depend on MTurk for their main source of earnings because comparatively speaking, their local work payment is even lower than that on MTurk on an hourly basis (Litman et al., 2015; Gupta et al., 2014; Crowston, 2012).

To MTurk workers in the U.S., most of them are doing tasks on MTurk as a supplementary earning, yet still, 12% of them reported that it was their primary source of income (Hitlin, 2016). As such, monetary payment on MTurk is a primary motivation and has a significant influence on many MTurk workers' living.

Furthermore, it is noteworthy that there has not been an established "minimum wage" mechanism in most crowd work platforms as there now are in online freelancing sites. The major differences between a crowd work platform and an online freelancing site are that freelancing sites often require freelancers to have certain expertise or professional skills to work for large projects whereas crowd work has a much lower entry bar; also, freelancing projects are in contract with some business clients, which need a period of time to accomplish whereas crowd work can be published by a variety of requesters and can be completed usually in minutes (Kuek et al., 2015). Without a "minimum wage" mechanism, requesters of crowd work platforms often decide their amount of payment based on a rule of thumb. For example, Horton and Chilton (2011) estimated that the minimum payment to motivate an MTurk worker to accept and perform a task was around $1.38/hour, and academic researchers advocated for paying MTurk workers at least with a federal or state minimum wage (Silberman et al., 2018).

Next, a following question is how monetary motivation affects contribution quality in crowd work. Given that crowd work payment is so little, it is natural to assume that higher payment would predict better contribution from crowd workers; however, the answer is not so simple. Prior research has revealed varied and even conflicting results to this assumption. Some prior studies suggested that increasing monetary payment was likely to increase contribution quantity but not necessarily quality (Buhrmester et al., 2011; Mason and Watts, 2009), but there were at least two critiques to such a conclusion. First, MTurk workers' location of residence had

not been differentiated to examine monetary incentive's effects. For example, Litman et al. (2015) found that though payment rates did not affect U.S.-based MTurk workers' data quality, they significantly affected India-based MTurk workers' contribution quality, yet Buhrmester et al. (2011)'s study has not examined the effect of payment on MTurk workers from U.S. and India separately. Second, prior studies that did not find a direct effect due to payment usually examined data quality of simple tasks for MTurk workers, which did not need to make much effort (Yin et al., 2014). When designing a more complex task with a *performance-based* compensation scheme for MTurk workers, the results changed. Performance-based compensation includes a baseline rate plus a bonus to reward an MTurk worker's outstanding contribution to the complex task. Yin et al. (2014) and Ho et al. (2015) both found that with such a combination of complex tasks and performance-based compensation, the amount of monetary incentive would predict contribution quality (Yin et al., 2014; Ho et al., 2015).

To sum up, money is the primary motivation for crowd workers but also a complicated one to predict their contribution quality. It may be due to the other extrinsic and intrinsic motivations besides money that could affect contribution quality. As Crowston (2012) contended, it would be hard to test various motivation theories in crowd work because it is difficult to tease out the monetary impact and test the effects from other motivations for crowd workers.

In comparison to the abundant research on crowd workers' motivations, fewer studies have looked into the motivations for requesters to outsource their tasks to crowd workers, even though the benefits of crowd work may be telling to explain requesters' motivations. For example, crowd work provides a source of inexpensive yet high-quality data (Buhrmester et al., 2011), a convenient site to conduct survey and collect responses, as well as a representative and

diverse sample of crowd workers (Buhrmester et al., 2011; Mason and Suri, 2012). Meanwhile, crowd work platform can harness "wisdom of the crowd" (Surowiecki, 2005) as well as data mining for commercial and political purposes, thus enterprises and politicians also could leverage crowd work for them (Brabham, 2010; The Guardian, 2015). Furthermore, to some requesters, there might be a surreptitious motivation because crowd work could render "dirty deeds done dirt cheap" (Harris, 2011). In fact, recent studies have revealed that some requesters would cheat, manipulate, and even stalk crowd workers to send targeted ads, hack their financial accounts, and divulge their sensitive information (Sannon and Cosley, 2018; Xia et al., 2017).

Crowd work characteristics

Crowd work also exhibits various characteristics. First, crowd work enables crowd workers to work with the flexibility of time, space, and accountability. Unlike in-office work, crowd workers are not confined with office hour or space. They may undertake crowd work tasks when they have leisure at home; want to earn supplementary income on vacation, or simply intend to kill time during travel or boredom. As recent surveys suggested, crowd workers do not need to be "centralized in the same physical location" and can spend their leisure time to earn extra money instead of leisurely pursuits (Kuek et al., 2015, p. 10; Ipeirotis, 2010a).

Second, crowd work offers workers the opportunity to earn money, spend time, have fun, contribute to a community, and pursue self-development and peer-acknowledgement. These characteristics relate to various motivations for crowd workers (Alkhatib et al., 2017; Durward et al., 2016; Ipeirotis, 2010a; Kaufmann et al., 2011). As prior research indicated, crowd work is a supplementary source of income to most crowd workers and gives them more choice and liberty to spend their leisure time with flexibility. As such, crowd work complements crowd workers' need or desire to earn extra money, their boredom or availability, as well as their internet

accessibility and computer skills (Von Ahn, 2008; Quinn and Bederson, 2011). When perceiving these characteristics in a concrete crowd work platform such as MTurk, they can be perceived from a social-material perspective (Faraj and Azad, 2012), that is, they are not only about MTurk's technological features, e.g., to post tasks, recruit, and pay, nor only about MTurk workers' perception of these features, e.g., low payment, convenient access, abundant tasks, but about the actions and interactions MTurk workers exert on these features, e.g., some may use MTurk honestly as a primary income, some may spam tasks on it to make easy money, some may work to promote research, some others may work for fun and pastime.

Third, crowd work could have a misconceived assumption of anonymity to crowd workers. It is a misconception because workers may assume the maintenance of their anonymity (Crowston, 2012), yet such an assumption is not true (Kandappu et al., 2015; Lease et al., 2013; Xia et al., 2017). In fact, crowd workers' individual identities can be de-anonymized with data triangulation, for example, a seemingly anonymous MTurk worker can be identified by cross-referencing her MTurk ID with her Amazon ID because Amazon applies the same user ID for shopping and MTurk (Lease et al., 2013; Xia et al., 2017). Although some crowd workers are aware that their anonymity is not assured, such a misconceived assumption of anonymity deserves more caution from both crowd workers and academic researchers. For example, in crowd work-based research, such a misconception could lead researchers to collect more personal information from crowd workers than necessary.

Crowd work also has appealing characteristics for requesters. First, crowd work provides requesters with the ability to distribute a task or the components of a task to a large and undefined group of crowd workers to undertake, who may not know each other and can be located geographically. This characteristic represents the essence of crowdsourcing, that is,

harnessing collective intelligence and action from a large, diverse, and distributed network of people (Howe, 2006; Malone et al., 2010). Such a characteristic can be further specified according to different types of crowd work. For example, tasks that involve knowledge discovery, broadcast search, a computation that is easy for humans but challenging for computers (e.g., audio transcription), creative idea and product development all can be significantly facilitated and improved via such a characteristic (Brabham, 2013; Von Ahn, 2008). Conventional contract-based outsourcing could only delegate tasks to one or a few organizations (Brabham, 2013; Schenk and Guittard, 2011), but crowd work can enable requesters to reach out to a more diverse and larger population of workers.

Second, crowd work enables requesters to access an on-demand, elastic, inexpensive, and globally distributed workforce. Such characteristics are appropriately expressed by MTurk's slogan: "Access a global, on-demand, 24x7 workforce" ("Amazon Mechanical Turk," n.d.-c). Unlike conventional organization or outsourcing, there are always crowd workers available to do tasks because crowd work platforms such as MTurk have gathered a workforce worldwide who are available. Requesters also have more elasticity with tasks and crowd workers because they do not need to worry about the limitation of scalability of in-house workforce. Such elasticity can also be embodied in requesters' selection of an international sample. For example, a requester could publish a survey and specify its respondents to be only in the U.S., only in India, and neither in the U.S. nor in India (Xia et al., 2017). In addition, crowd work offers requesters an inexpensive workforce that can reduce their cost because as aforementioned, there is no minimum wage in crowd work and crowd workers are usually paid very poorly.

The third characteristic of crowd work for requesters can be regarded as a *hidden characteristic*, which is requesters' capability of *manipulation*. Crowd work could enable

requesters to manipulate workers by cheating them to undertake tasks with malicious purposes or by crafting a series of related tasks to compromise crowd workers' privacy and security. For example, requesters could mislead and leverage crowd workers' intelligence to decipher CAPCHAs or even hack accounts (Lasecki et al., 2014; Harris, 2011); requesters could also design a sequence of surveys to profile crowd workers gradually and divulge their privacy (Xia et al., 2017; Kandappu et al., 2015). Such a hidden characteristic for requesters needs more awareness and regulation because no technique or policy so far has designed to tackle it.

Crowd work controversies

Despite numerous advantages and opportunities of crowd work, there are also controversies in it. The three primary ones, I propose, are requesters' exploitation, crowd work platform's abdication from assuming its responsibility, crowd workers' misplaced trust.

First, it has been a longtime controversy on whether crowd work is a type of exploitation and whether requesters are a new type of bourgeoise. Several scholars critiqued that crowd work gives requesters a superior power and privilege to exploit crowd workers by treating them as merely a service, an API call, an inexpensive data source, or even like a commodity (Silberman et al., 2010; Irani and Silberman, 2013; Aloisi, 2015; Bergvall-Kåreborn et al., 2014). To make crowd workers' situation worse, there's no employment involved or any legal protection on the crowd workers' interests. As such, protection strategies such as data minimization could hardly effectuate because presumably, the requesters would always want to pay less to get more from a *commodity* and to maximize the benefits from a streaming *API*. In this context, several third-party forums and tools such as TurkerNation ("TurkerNation," n.d.), Turkopticon (Irani and Silberman, 2013), and Dynamo (Salehi et al., 2015) have emerged to unify the voice for crowd workers to rate and review requesters, and thus, balance the power dynamics with requesters.

On the other hand, some researchers argued that crowd work could not be simply tagged as a type of exploitation. First, there are different participation modes in crowd work besides Amazon's MTurk, and some of them are not exploitative in nature because crowd workers and their labor are not alienated (Hansson et al., 2016). For example, in a relational crowd mode of participation, crowd workers are both producers and consumers and related to each other without a hierarchy such as in the case of Waze (Hansson et al., 2016). Second, crowd work can be regarded as a contemporary instantiation of piecework in history, and therefore similarly, whether it is exploitative is conditional, i.e., we need to ask under what conditions does crowd work directly or indirectly harm crowd workers (Alkhatib et al., 2017). Finally, many crowd workers do not feel being exploited by requesters even though they have been presented with various definitions and explanations of exploitation (Busarovs, 2013). Thus, in a pluralistic view, crowd work is not only embodied by MTurk, and whether crowd work is exploitative is still an open question.

The second controversy is around the crowd work platform's role. Current crowd work functions as a tasking and transaction platform between the requester and the crowd workers. On the one side, crowd work platforms and requesters are stakeholders of each other. Requesters need crowd work platforms such as MTurk to publish their tasks and recruit crowd workers while crowd work platforms earn commission from a portion of requesters' payment. On the other side, crowd work platforms treat crowd workers as independent contractors instead of employees, and therefore they have avoided the liability to protect crowd workers' interests as that normally attached to an employment relationship (Aloisi, 2015; De Stefano, 2015). As MTurk's policy declares, crowd workers will "not be entitled to any of the benefits that a Requester or Amazon Mechanical Turk may make available to its employees, such as vacation

pay, sick leave, and insurance programs, including group health insurance or retirement benefits." It also reminds requesters not to "engage a Worker in any way that may jeopardize that Worker's status as an independent contractor performing Tasks for you" ("Amazon Mechanical Turk," n.d.-b).

Meanwhile, the crowd work platform's role of mediation between crowd workers and requesters abdicates a crowd work platform from assuming any legal responsibility to protect both sides' interests. Hence, crowd workers have to be self-dependent to protect themselves and be alert to any malicious task on MTurk; requesters have to be self-regulated to avoid manipulating or mistreating crowd workers. But since requesters and crowd work platforms are stakeholders in interest and more closely interdependent with each other than with crowd workers, it is usually crowd workers who are in a more vulnerable position (Xia and Mckernan, 2020). For example, MTurk only has a general policy and a weak warning to requesters: "You may not use Amazon Mechanical Turk for...collecting personal identifiable information; fraud; disrupting or degrading the operation of any website or internet service; direct marketing; spamming, etc." ("Amazon Mechanical Turk," n.d.-a). However, there is no mandatory prohibition to ban these activities, and many requesters apparently have disrespected and disobeyed these policies (Irani and Silberman, 2013; Xia et al., 2017; Sannon and Cosley, 2018).

The third controversy is about crowd workers' misplaced trust and their trustworthiness. According to my empirical study on MTurk, many MTurk workers explained that they did not have any privacy concern on this platform because they trust the reputation of Amazon (Xia et al., 2017). But it is a somewhat misplaced trust because what matter more to crowd workers is a requester's reliability rather than a crowd work platform or its parent company's reputation (Xia and Mckernan, 2020). In fact, prior work has demonstrated the importance of developing a

reputation system to let the crowd workers rate the requesters so that some requesters' bad behaviors could be deterred (Irani and Silberman, 2013). On the other side, crowd workers' trustworthiness should also be carefully considered. Prior studies have indicated that there are crowd workers who would spam survey questions, game tasks, and cheat requesters, and have proposed strategies to detect and filter the spamming crowd workers (Gadiraju et a., 2015; Lee et al., 2014; Wang et al., 2012; Difallah et al., 2012). But other scholars demonstrated and believed that most crowd workers are accountable and trustworthy (Vaughan, 2017; Salehi et al., 2015; Irani and Silberman, 2013).

In certain situations, crowd workers' trustworthiness is even more essential. For example, there are crowd work applications such as VizWiz (Bigham et al., 2010) that help a requester with visual impairment to take a photo at an unknown object or location via the app and outsource the photo to the other visible users to identify the object or location. In these situations, requester with visual impairment has to trust that crowd workers would not steal her card number or other personal identifiable information (PII) if she accidentally reveals it in the photo. To sum up, how much trust can requesters put on crowd workers is also debatable.

The landscape and methodological issues of crowd work-based research

Academic research based on crowd work has been quite active in recent years and is still growing. According to a recent study by Pew Research, 36% of the unique requesters on MTurk were from academia, which represented the largest percentage of requesters on MTurk (Hitlin, 2016). Though it is impossible to make a comprehensive review of crowd work-based research, I identify the primary categories of it from Gadiraju et al. (2014)'s taxonomy of crowd work tasks. Below, I depict the landscape of crowd work-based research based on these primary categories and then reflect the potential ethical issues within each type of crowd work-based research.

Crowd work-based survey research

Crow work-based survey research is one of the most common types of academic research based on crowd work. Crowd work platform such as MTurk has been frequently utilized for conducting surveys since its nascent time. Buhrmester et al. (2011)'s highly cited work advocated MTurk as a new source of "inexpensive yet high-quality data." The authors found that MTurk participants were more diverse than either standard internet samples or American college samples; participants could be recruited rapidly with low cost; monetary incentive rates were not related to data quality; and data obtained from MTurk had comparable reliability to that of data obtained from traditional methods (Buhrmester et al., 2011). Some of my conducted research projects are examples in this category. Moreover, MTurk allows requesters to recruit crowd workers around the globe and specify the nation/region for the recruitment. For one of my survey studies as an example, in the first branch of it, I could specify that I would only recruit the MTurk workers currently in the U.S.; in the second branch, I could confine my sample to be from India only; in the third branch, I could specify that the sample should neither be in the U.S. nor India (Xia et al., 2017). Such an operation made my survey sample more diverse and comprehensive, and I was able to compare between samples from different areas.

Potential ethical issues in crowd work-based survey research can arise in the following aspects. First, survey questions can be sensitive and perceived as privacy intrusive. Though such an ethical issue could occur in internet research more broadly, some crowd work's particular characteristics as aforementioned, such as the convenience for requesters to access a constantly available workforce, the monetary incentive for certain crowd workers as their main revenue, the problem of dehumanization and presumed anonymity, and the lack of monitor and regulation of crowd work platforms, could introduce more challenges to researchers' self-regulation, crowd

workers' protection, and privacy intrusion detection in crowd work-based survey research. From my own survey study, many MTurk workers have complained that they have encountered too private questions in many survey studies on MTurk (Xia et al., 2017).

Second, crowd work-based survey research can be deliberately crafted to form a series of surveys to gradually triangulate and profile a crowd worker's preference and identity even each individual survey only asks plain questions (Kandappu et al., 2014, 2015; Xia et al., 2017). Since researchers can often target and recruit the same pool of crowd workers (such as through their MTurk ID) such deliberate data triangulation and gradual divulging are difficult to be aware by crowd workers. In comparison, with general internet research, such intentional and targeted data triangulation is harder to implement.

Crowd work-based content creation

Content creation refers to the crowd work tasks about media transcription, translation, and tagging (Gadiraju et al., 2014). It is a common type of task on MTurk, but it is more popular from small business firms than from academia (Hitlin, 2016). Also, the basic content creation task such as transcription is often regarded as a typical example of human computation in the sense that though it is simple to do by humans, it is notoriously difficult to accomplish by computer or automation (Von Ahn, 2008). Prior research found that some related work to content creation such as evaluating and revising machine-translated texts can be done efficiently and creatively with a low cost by crowd workers (Callison-Burch, 2009).

As an example, Marge et al. (2010) recruited MTurk workers to transcribe their audio materials. They found that MTurk workers' transcription was accurate and reliable. Particularly, if combining several workers' transcriptions together, the final result would be further improved (Marge et al., 2010). The authors also proposed that compared to traditional transcription

methods for audio data, using MTurk would be much less expensive but with comparable quality; meanwhile, paradoxically, MTurk workers' transcription accuracy was not significantly related to the amount of payment (Marge et al., 2010).

Potential ethical issues in crowd work-based content creation can arise when the content for creation involves sensitive information of other people, and by transcribing such information, these people's privacy may be compromised. For example, Lasecki et al. (2015a) conducted a study to ask MTurk crowd workers to code behavioral videos. They highlighted the possibility that crowd workers may learn and reveal other people's sensitive information through such video viewing and coding process.

Crowd work-based information retrieval and identification

Crowd work-based information retrieval and identification is related to what Brabham (2013) called "broadcast search" that aims to solve empirical problems with "the right answers" not known by the requesters in advance and require crowd workers' expertise or knowledge. It is also similar to some citizen science projects such as Zooniverse that recruit volunteers without payment to discover patterns, identify outliers, and search for new knowledge. For example, Higgins (2010) leveraged MTurk workers' intelligence to find Arabic nicknames to contribute to the existing Named Entity (NE) lexicons, and they demonstrated that MTurk was viable for this type of task. Gottlieb et al. (2012) recruited MTurk workers to annotate the geo-location of random videos on the web and argued that such a kind of information identification task was difficult for either humans or machines. The authors applied various effective techniques such as setting a high-quality baseline and used iterative internal testing before outsourcing the task to crowd workers (Gottlieb et al., 2012). They found that these techniques were important to

improve crowd workers' capability and quality to solve such complicated tasks for accurate information identification.

Potential ethical issues in crowd work-based information retrieval and identification can occur when crowd workers are maneuvered to extract or identify someone's credit card information or handwriting characters (Lasecki et al., 2014; Lasecki et al., 2015b). Though such an intentional manipulation of crowd workers is unlikely to be seen from academic researchers, it is still an alarming phenomenon on crowd work platforms. Thus, it deserves more specific ethical regulation in academia to avoid any grey area such as exploiting crowd workers to retrieve certain scientific data that should have been purchased or officially requested by researcher or institution.

Crowd work-based experimental study

Crowd work-based experimental study, such as user testing and system prototyping, is also popular on MTurk. For example, Liu et al. (2012) conducted a usability testing study on MTurk to evaluate a school website's usability for students. They found that the usability testing via MTurk was significantly faster and cheaper than that in the lab setting, and the participants were also more diverse. However, they also found that the contribution quality from crowd workers was lower than that from lab participants, and the MTurk workers were less focused on their work and less interactive with researchers than their lab counterparts. As another example, the TRACE (Trackable Reasoning and Analysis for Collaboration and Evaluation) project that I had participated, was also a crowd work-based experimental study, and we had conducted experimental studies with MTurk workers to test their reasoning processes and variances with different TRACE versions and functionalities.

Potential ethical issues in crowd work-based experimental study could arise when sampling is skewed, and justice is not distributed fairly (e.g., test with Indian crowd workers with lower pay and research benefit than with the U.S. crowd workers) or informed consent is not informative enough to make crowd workers comprehend the intention and potential consequence of experiment. Informed consent might be more challenging for longitudinal study as both researchers and crowd workers may ignore it over time.

The methodological issues of crowd work-based research

Though the landscape of crowd work-based research is broad and diverse, various methodological issues of it are in debate since the launch of MTurk and are likely to continue evolving with it. These methodological issues are primarily concerned with data quality and validity in research on MTurk.

To start with, numerous studies in their respective research fields have claimed that the data collected from MTurk workers were of comparable and even superior quality to that from the other venues (e.g., Callison-Burch, 2009; Paolacci et al., 2010; Buhrmester et al., 2016; Kees et al., 2017a); nonetheless, there are still on-going doubts on the data quality issues on MTurk. These doubts were based on the observations that some MTurk workers were deliberately cheating in their responses (Difallah et al., 2012) or using virtual private servers (VPS) to conceal their real IP address in order to take certain tasks (Dennis et al., 2019) and even using automatic scripts to answer surveys (Dreyfuss, 2018). Prior research has also investigated the factors that correlated with data quality on MTurk; for example, as I have reviewed above, the monetary incentive is found to have a complex relationship with data quality. More recently, whether the Mechanical Turk Masters (MTM) Qualification of MTurk workers is positively related to their data quality has also been controversial. Although Peer et al. (2014) found that

the worker's reputation is a sufficient condition for their data quality, Lovett et al.'s (2018) more recent work revealed that some MTM workers themselves had confessed that they did not always provide reliable data.

Meanwhile, the academic debate over the validity issues in research on MTurk is no less unsettling than the data quality issues. Such debate is well embodied in an "academic fight" between scholars. On the one side, Kees et al. (2017a) first published a paper to advocate for using MTurk as a valid alternative means to collect research data besides the professional panels and student subject pools. On the other side, Ford (2017) wrote a comment and critiqued Kees et al.'s (2017a) advocacy and pointed out that the "cheaters and speeders" on MTurk would severely defect the validity of research conducted on this platform. Then, Kees et al. (2017b) rebuked Ford's (2017) critique and argued that "cheaters and speeders" are not "unique" to MTurk and can engender a similar level of validity problem in the other pools. Additionally, scholars were concerned about the validity issues on MTurk due to MTurk workers' prior knowledge or familiarity with various research questions and experimental manipulations (i.e., their "non-naïvety") (Chandler et al., 2012) and MTurk workers' being not as attentive to questions as student participants and not representative to a specific population for research (Goodman et al., 2013; Stewart et al., 2017).

The methodological issues are closely related to the ethical issues in crowdsourced survey research on MTurk (Haug, 2018), which partly motivated my exploration of the ethical issues in this dissertation work. However, as I will uncover and discuss further, the ethical issues in crowd work-based research are not only related to a particular crowd work platform or the methodology per se but also are stem from the problematic origin of crowd work.

CHAPTER 4: RESEARCH METHODOLOGY

Chapter 4 describes the methodology of this dissertation research. First, I describe the research paradigm that has been applied to guide the qualitative inquiry. Then, I introduce the research design including the interview and document analysis, sampling strategies, and privacy and data confidentiality considerations. Afterward, I present the data collection process, challenges, and data validation strategies. Finally, I reflect on my positionality evolution as well as the potential biases that may be exhibited in this dissertation work.

Research Paradigm

To explore crowd work-based research ethics, I have contemplated which inquiry approach is most appropriate. My methodological choices were primarily informed by three considerations: (1) the research paradigm or philosophical assumption that would guide the research, (2) the research design, and (3) the specific methods of data collection, analysis, and interpretation (Creswell, 2014). Meanwhile, the selection of a research approach depends on the nature of the research problem and the researcher's personal experience as well as the audiences for the research (Creswell, 2014). The research paradigm or philosophical assumption is a basic set of beliefs about the world and the nature of research that guides a researcher's actions (Guba, 1990; Creswell, 2014).

In this dissertation work, I follow the constructivist paradigm, which holds that individuals understand the world based on their living and working experiences, and they develop subjective meanings of such experiences (Creswell, 2014; Lincoln and Guba, 1985). The constructivist paradigm is typically associated with and guides qualitative research, and researchers with this worldview looks for complex views from different perspectives and avoids narrowing the subjective and diverse meanings individuals assign to their experiences into a few

categories or ideas (Creswell, 2014). The individuals' subjective and varied meanings in a specific context are constructed through the interaction with the researcher, and the researcher interprets these meanings based on his background and stance.

Meanwhile, there are several common characteristics of qualitative research, which can help in determining whether an inquiry should be conducted with a qualitative approach (Creswell, 2014; Hatch, 2002; Marshall and Rossman, 1989). Specifically, qualitative research typically focuses on a natural setting, which means the researchers collect data in the field or at the site where the participants experience the issue under study; qualitative data collection could be in the form of interviews and observation over a period of time; qualitative research requires the investigators to collect data firsthand by interacting with the participants, observing their behaviors, or examining documents (Creswell, 2014). Also, typically, the researcher collects data from multiple sources, such as interviews, observations, documents, and from multimedia instead of relying on a single data source; in addition, qualitative research is emergent in the sense that the understanding develops as more data are collected and more reflections are made; it is also reflexive: the inquirers are expected to reflect on their role in the research as well as on their own background, experience, and culture that could influence the interpretation of the themes emerging from the data (Creswell, 2014).

The characteristics of the constructivist paradigm and qualitative research pertain to my research to explore scholars' and IRBs' understanding of ethics in crowd work-based research, who have abundant experiences in this context and to collect data from multiple sources including interviews and documents. Finally, the constructivist paradigm and qualitative research characteristics suggest that it is important to reflect and explicate my role as a researcher in this study in terms of the personal background, values, and bias that potentially could influence my

interpretations of the findings (Creswell, 2014). Hence, I have a separate section in this chapter to describe and reflect on my positionality and potential biases in this dissertation work.

Research Design

Interview and document analysis

I collected and analyzed my dissertation data from multiple data sources. The first data source includes interviews with academic researchers in different research disciplines as well as IRB directors and analysts in different institutions to understand their perceptions of ethics in crowd work-based research. The interview was semi-structured and aimed to understand how academic researchers in different disciplines and IRB directors or analysts in different institutions perceive ethics in crowd work-based research. Some questions came from the questions typical in an IRB application, such as the most common risks and benefits in crowd work-based research; some questions were from my prior research and pilot interviews (Xia et al., 2017), such as an inquiry about their comparison between academic research ethics in general and crowd work-based research ethics in particular; finally, some questions were developed through my interview process as new themes emerged, such as a question about minimum wage. I have included my interview protocols in the Appendix.

The second data source includes documents of two types. The first type of documents includes guidelines for crowd work-based research from IRBs, academic organizations, and researchers. The second type of documents includes published papers from my recruited researchers. These documents provide me with an alternative analytical lens to cross-reference with my interview participants' perceptions of ethics. In particular, searching and analyzing these guidelines also let me understand the status quo of documented regulations for crowd work-based research.

Furthermore, these two types of data sources enable me to compare within and between each. For example, I compare researchers' ethical views versus IRBs' ethical positions in crowd work-based research. I also compare the guidelines and publications in terms of how they frame ethics and itemize their ethical concerns. Finally, between the interview and document data, I cross-reference their consistencies and disparities. Overall, these two distinctive data sources and their internal varieties enrich my data analysis and improve my research validity.

Sampling strategies

Based on my experience and knowledge in crowd work-based research, I identified 50 scholars as the sampling pool for my interviews. These 50 scholars are from different research disciplines in both public and private universities, and their research in MTurk, crowd work, crowdsourcing, and the gig-economy has been published in top tier journals and conferences and have been cited extensively. Although the population of researchers who conduct crowd work-based research may be quite large, these 50 scholars are the pioneers and representatives of this population. Hence, their views of ethics in crowd work-based research are most valuable to my dissertation research.

As regards the IRB participants in my dissertation research, I targeted the whole population of IRB staff in both private and public universities in the U.S. The IRB staff include IRB directors, IRB analysts, IRB administrators, and IRB coordinators. My recruitment preference was also in this order. First, among these four types of IRB staffs, IRB directors are the most experienced and knowledgeable in dealing with ethical issues. Second, IRB analysts also have first-hand experiences in reviewing ethical issues in IRB protocols, especially exempt IRB protocols. IRB analysts are at a junior level, and their review of expedited applications or full board applications, if any, would have to be examined by the IRB associate or IRB director

again. Based on my prior research experience, MTurk research is often categorized into exempt from IRB review, so IRB analysts may have experiences in reviewing ethical issues in crowd work-based research that is exempt. Third, IRB administrators and IRB coordinators may have experience reviewing exempt protocols, particularly if the IRB office is small. It should be noted that I excluded IRB directors and analysts who are usually professors because their epistemology of ethics in crowd work-based research may be more similar to the professors that I recruited and interviewed. I hope to learn from IRB's perspectives on this topic distinctive enough from academic scholars' views.

Finally, in terms of collecting guidelines for crowd work-based research, I conducted an extensive information search on the internet using keywords such as 'crowd work,' 'crowdsourcing,' 'Amazon Mechanical Turk,' 'MTurk,' 'ethics,' 'guidance,' 'guidelines,' 'requesters' in conjunction with the keywords such as 'online,' 'IRB,' 'institution,' 'university,' 'organization.' In terms of collecting published papers, I did so after I interviewed with a researcher and after obtaining their consent.

Privacy and data confidentiality protection

Privacy and data confidentiality protection have been a central and prioritized consideration in my dissertation research. I applied the following strategies to protect my interview participants' privacy and the data confidentiality of my interview recordings, transcriptions, and collected documents. First, I keep my interview participants' anonymous by removing their personally identifiable information (PII), such as their names and institutions in my interview transcriptions as well as in this dissertation document. I assign each interview participant with a letter 'P' such as 'P1,' 'P2,' and 'P32.' Also, I protect my interviewed researchers' privacy by removing any PII or linkable information from their publications that I

have collected and analyzed. Furthermore, I conducted all the interviews in locked and private rooms, and thus, no privacy risk of shoulder-surfing had ever occurred.

Second, I transcribed all the interview recordings by myself and did not hire any professional transcriptionist. Thus, I have prevented the potential data breach risk in sharing my interview recordings with them. Moreover, after fully transcribing an interview recording, I have permanently deleted the recording files from my computers and cloud storage. During the analysis on my collected documents and papers, for the private guidelines that are not publicly accessible and for the collected papers, I have removed all the identifiable information such as their titles, authorships, and publishers. Also, I have assigned each document with a letter 'D' such as 'D1,' 'D2,' and 'D29.' Furthermore, all my computers and cloud storage are password-word protected, and only I have access to it.

My dissertation research design, sampling plan, privacy, and data confidentiality protection strategies were approved by Syracuse University (SU) IRB in April 2019. It took a long time for the approval because SU IRB perceived my dissertation research to be more than minimal risk. They were concerned about the reputational risks to my potential participants if their views about ethics were de-anonymized and divulged. Therefore, my dissertation protocols to SU IRB had to go through an expedited rather than an exempt review.

Data Collection

The interview questions

My interview questions to researchers and IRB staffs share some commonalities but also include specific questions to each camp. My interview questions can be viewed in the Appendix of this dissertation document. Below, I introduce my interview questions to researchers and IRB staffs respectively.

My interview with researchers started with a few general questions where I asked them to introduce themselves in terms of research domain, community, and methods. I also asked them when they started to cast their research attention to MTurk and what motived them to do so. Most of my recruited researchers are well-known in the research field of crowd work. Since I am also interested in research that collects data from MTurk workers or discusses MTurk's features and potentials, it was fast for me to build a rapport with my interviewed researchers. Through these initial ice-breaking questions, I gathered the background and research discipline information of these researchers.

My second part of the questions was about the ethical items related to the IRB review application for research on MTurk. For example, I would ask the researchers what their inclusion and exclusion criteria are to recruit MTurk workers and how they ensure informed consent with MTurk workers to participate in their research. These questions should be most familiar with a researcher if they often launch studies on MTurk. Besides, I would ask some probing questions such as how they design and evaluate their payment standard and whether there were any uncommon risks or benefits to MTurk workers that they had ever encountered or reflected. Also, I designed and asked these ethics-related questions to the IRB review application to cross-referencing the IRB respondents' opinions about them. My third part of the questions to the researchers was also related to ethics but beyond the IRB and MTurk. For example, I asked the researchers about the ethical challenges in crowd work-based research more broadly, and how they compared ethical issues in crowd work-based research specifically and in academic research in general.

In addition, I was curious to learn their philosophical deliberation about ethics beneath their practical ethical considerations in crowd work-based research, and thus, I inquired into their

fundamental ethical stances in conducting research on MTurk. Because the question was a bit abstract, I provided the interviewees with a self-reflection on the shared similarities between the Belmont principles and Kantian and utilitarian ethics: "I reflected that the Kantian view of ethics that each individual human being's dignity and autonomy must be respected is underneath the Belmont principle of respect for persons whereas the utilitarian ethics of maximizing the happiness for the majority of people is behind the Belmont principle of beneficence How do you think about it?" If an interviewee was not familiar with these schools of thoughts, I would stop the question.

In terms of my interview questions to the IRB respondents, they also included three parts. The first part of the questions served as an icebreaker. I asked the IRB respondents how long they have been in their position and how often they review an application to conduct research on MTurk. Also, I probed whether these applications were exempt, expedited, or full board review and from which research discipline or department they usually came from. In parallel to my second part of questions to researchers, the second part of the questions to the IRB respondents was also related to the IRB application review for research on MTurk. For instance, I asked them how they evaluated payment in a study on MTurk and how they assessed the ratio of research benefits and risks.

My third part of the questions to the IRB respondents was about their interpretations of the Belmont principles in the context of crowd work-based research. I asked these questions to address my second research question about how IRB directors and analysts interpret and enforce the federal governments' research mandates. Apart from these questions, to be consistent with my final questions to the researchers, I inquired into the IRB respondents' opinions about the

comparison between ethics in crowd work-based research and in academic research in general, as well as their philosophical stance of ethics.

Finally, it should be noted that my interview was semi-structured, and I did not stick to these questions rigidly when I interviewed my research respondents. Sometimes, I would use one respondent's idea as a probing question to ask another respondent. For instance, I used one IRB director (P7)'s comment about justice as a probing question to ask another IRB director (P4) about her opinions. In fact, my interviewed researchers and IRB respondents provided me with many thoughts and ideas for follow-up and probing questions that were not included in my initial interview protocols, such as whether the monetary payment on MTurk can be regarded as a sort of research benefit and how they perceived using a minimum wage as a benchmark of payment in research on MTurk.

Recruitment of interview participants

My recruitment started from May 2019 until January 2020. It turned out to be a challenging and time-consuming process. My recruitment encountered three challenges. The first challenge was that it was hard to get my initial contacts to respond to me and to agree to be part of the study. Part of the challenge may have been due to my email recruitment script. SU IRB required me to use their approved scripts to invite potential participants. These email scripts include an invite academic scholar, a follow-up invitation to an academic scholar if they did not respond in two weeks, an invite to IRB staff, and follow-up invitation to an IRB staff if they did not respond in two weeks. These invitation scripts contained the rationale of my dissertation work, my basic information as the researcher, my advisor's basic information as the PI, the interview procedures, and the potential risks.

These email invitation scripts are formal and comprehensive, but they are also lengthy and somewhat impersonal. I had tried my best to make them as concise as possible when I wrote my dissertation research protocols, but SU IRB still required to include the essential information as mentioned above. Also, because I must comply with SU IRB's approved scripts, I could not use personal tones or any personal connection to invite the researchers or IRB directors that I knew previously. Such restrictions in my invitation messages, though understandable, challenged my data collection process.

A second challenge was partly related to the first reason that the potential participants were not responsive to my invitations, especially in the first few months of the recruitment. I started to send out invitations in May 2019 but did not get any response until early July 2019. In all the cases of recruiting academic researchers, I had to send a follow-up invitation and wait for another week or a few weeks to get a response. In many cases, I had to send two rounds of invitations. It means that if I received no response from a scholar after sending her a follow-up invitation in a few weeks, I would try to invite this scholar again with the original invitation email. If an invitee still did not respond after sending a second round of follow-up invitation, I would drop them from my sampling pool because I did not want to harass a potential participant.

The recruitment of IRB directors and staff turned out to be even more difficult. Originally, I planned to recruit IRB directors and staff from the same institutions as the scholars so that their views may be compared in the same institutional contexts. However, such a plan did not progress very well. I often met with the situation that I tried to send invitations to multiple staffs in an IRB, but none of them replied or they just rejected my request. I even tried to call them by phone for recruiting, but such a phone-calling strategy was not so effective either.

Hence, after consulting with my advisor, I decided to open my sampling pool of IRB potential participants to include all the IRB staffs in the top 100 universities in the U.S.

A third challenge was that the participants' time was hard to schedule, and in several cases, I had to wait weeks or even months to schedule an interview. Partly, it was because these participants, especially several famous academic researchers, were quite busy and occupied by their heavily-loaded duties and work; partly, it was also due to my busy schedule when I was in an internship during summer 2019 and had to teach a course in fall 2019. Thus, several interviews were postponed for weeks or even months due to various time conflicts in my participants as well as my own schedules.

To tackle these recruitment challenges, I used the following strategies. First, I was persistent in recruiting a potential participant. For example, if a scholar or IRB staff responded and showed interest to participate, I would send back a personal message (as SU IRB did not mandate it to be formalized) and told them my appreciation of their participation and the importance of their views on my dissertation. Second, as aforementioned, I would send another round of invitations if a potential participant did not respond to my first round of follow-up invitation. It turned out that a few participants responded and accepted my second round of invitations even though they did not reply in my first round. Third, I expanded my sampling pool of IRB staffs and extended the recruitment period from my original plan so as to accommodate late respondents and participants who did not have earlier availabilities for an interview.

Finally, it should be noted that even though the recruitment was challenging, and many potential participants were not responsive, I did not end my recruitment process merely based on the number of participants that I have interviewed. My recruitment continued until the saturation of themes emerged in both scholars' responses and the IRB staffs' responses.

Demographics of the participants

In the end, I contacted 225 IRB directors and analysts in every university in the top 100 universities in the U.S. and was able to recruit 17 of them in total. Meanwhile, I contacted 50 researchers that I know are active and prolific researchers in crowd work and was able to recruit 15 of them in total. Among the 17 IRB staffs, eight are IRB directors; three are IRB associated directors; four are IRB assistant directors; two are IRB analysts. 10 IRB participants are from private universities, and seven are from public universities. Among the 15 researchers, five are full professors; six are associate professors; three are assistant professors; one is an adjunct professor with a doctorate degree and has published extensively cited papers on MTurk. These fifteen scholars can be categorized into seven different research disciplines based on their educational backgrounds, research interests, and positions: Information Science/Computer Science (4), Communication (2), Political Science (2), Engineering (2), Business (2), and Psychology (3).

The demographics of these 32 interview participants are listed in Table 1 below:

Table 1. Demographics of my Interviewed Participants

NO.	ROLE	PARTICIPANT ORIGIN
P1	IRB Associate Director	Private university
P2	IRB Director of Research Protection	Private university
P3	IRB Assistant Director	Private university
P4	IRB Analyst	Private university
P5	IRB Director of Clinical Research	Private university
P6	IRB Assistant Director	Private university
P7	IRB Director (upcoming)	Private university
P8	IRB Assistant Director	Public university
P9	IRB Associate Director	Public university
P10	IRB Associate Director	Public university
P11	IRB Director	Public university
P12	IRB Director	Public university
P13	IRB Executive Director	Private university

P14	IRB Assistant Director	Public university
P15	IRB Director	Private university
P16	IRB Analyst	Public university
P17	IRB Director	Private university
P18	Full Professor	Information/Computer Science
P19	Assistant Professor	Engineering
P20	Full Professor	Political Science
P21	Associate Professor	Information/Computer Science
P22	Associate Professor	Engineering
P23	Full Professor	Political Science
P24	Adjunct Professor (with a Ph.D.)	Psychology
P25	Associate Professor	Information/Computer Science
P26	Associate Professor	Communication
P27	Assistant Professor	Communication
P28	Full Professor	Business
P29	Assistant Professor	Information/Computer Science
P30	Full Professor	Psychology
P31	Associate Professor	Psychology
P32	Associate Professor	Business

Collection of documents

The collection of documents for analysis in this dissertation work consists of two parts: the collection of guidelines for crowd work-based research, and the collection of published papers from my interviewed scholars. In the end, I collected fifteen guidelines in total. Thirteen of them are from eleven IRBs, one is from an academic organization, and the other one is from an academic community. These are all the existing crowd work-based research guidelines that I can find so far from IRBs and the other academic groups or organizations. Two guidelines are collected from two scholars. One guideline is from a scholar who also participated in my interview. He created this guideline for his graduate students to reference when they launch a research project on MTurk. I obtained his consent to share it with me and analyze it for my dissertation research purpose. The other guideline is from Dr. Michael Buhrmester who published it on his personal website for public view and access. I have tried several times to

reach out to invite him to my dissertation interview but never received a response. These fifteen guidelines are varied by length and format. Some guidelines are as short as one page, and some are as long as twenty pages. Some guidelines are published online as slides; some are embedded in more general guidance for human subjects research.

Besides these fifteen guidelines for crowd work-based research, I have also collected and analyzed fourteen published papers from the fifteen scholars that I have interviewed. During my recruitment, once I scheduled an interview with a scholar, I would send them a consent form in which I solicited their consent to audio record the interview as well as to let me collect and analyze their publications to complement the interview data. I usually chose this scholar's most cited paper related to crowd work. A particular case was that one highly cited paper in the research field of crowd work was co-authored by two scholars in two different institutions, who both participated in my interview. Hence, the final number of my collected papers is fourteen instead of fifteen. All my interviewed scholars consented me to collect and analyze their publications.

These fourteen published papers are varied in their publication venues, research nature (human subjects research vs. analytical review), and platforms (MTurk vs. non-MTurk crowd work platforms). More specifically, nine papers were published in journals; five were published in conference proceedings; six were human subjects research projects, which means they had recruited and involved crowd workers in their research to do surveys, experiments, or online behavioral tracking; eight were analytical reviews, which means they did not involve any human subjects in their research but used literature review or analysis on secondary data; thirteen papers are related to MTurk, and one paper is about other non-MTurk crowd work platforms.

Properties of the documents

The properties of these 29 documents are listed in Table 2 below:

Table 2. Properties of my Collected Documents

NO.	TYPE	SOURCE / REFERENCE
D1	Research Guideline	American Psychological Association / (Palmer and Strickland, 2016)
D2	Research Guideline	Stanford University / (Bailey, 2017)
D3	Research Guideline	Duke University / ("The model MTurk consent form," 2019)
D4	Research Guideline	Private guideline from an academic scholar
D5	Research Guideline	Colorado State University / ("Consent & Recruitment for mTurk population," 2015)
D6	Research Guideline	Furman University / ("IRB Consent Example_Amazon MTurk study," n.d.)
D7	Research Guideline	University of Texas / ("IRB Guidelines and Suggestions for Using Mechanical Turk (MTurk) for Social/Behavioral Research Projects," 2015)
D8	Research Guideline	University of California, Berkeley / (Committee for Protection of Human Subjects, 2020)
D9	Research Guideline	University of Massachusetts Amherst / ("MTurk Guidance," 2019)
D10	Research Guideline	M-Turk Guide by Dr. Michael Buhrmester / (Buhrmester, 2018)
D11	Research Guideline	University of Michigan / ("Amazon Mechanical Turk 'Workers' are not anonymous," 2013)
D12	Research Guideline	University of Illinois at Urbana-Champaign / ("Application for Review of Research Involving Human Subjects," 2013)
D13	Research Guideline	University of Indianapolis / ("Guidelines: Human Research Protections and Amazon mTurk," n.d.)
D14	Research Guideline	Iowa State University / ("Use of Amazon Mechanical Turk," 2020)
D15	Research Guideline	Academic Community ("We Are Dynamo Wiki") / ("Guidelines for Academic Requesters," n.d.)
D16	Research Paper	Journal/Analytical Review
D17	Research Paper	Journal/Human Subjects Research
D18	Research Paper	Conference/Human Subjects Research
D19	Research Paper	Conference/Analytical Review
D20	Research Paper	Journal/Analytical Review
D21	Research Paper	Journal/Human Subjects Research
D22	Research Paper	Journal/Analytical Review
D23	Research Paper	Conference/Human Subjects Research
D24	Research Paper	Journal/Analytical Review
D25	Research Paper	Journal/Human Subjects Research
D26	Research Paper	Conference/Analytical Review

D27	Research Paper	Journal/Analytical Review
D28	Research Paper	Conference/Human Subjects Research
D29	Research Paper	Journal/Analytical Review

Data Analysis

Simultaneous data collection and analysis is a recommended practice in qualitative research (Merriam, 1988; Marshall and Rossman, 1989). As described earlier in this chapter, I met with several recruitment challenges, and my data collection period spanned for more than half a year. Hence, during this extended period, I tried to keep my initial round of data analysis simultaneous with my data collection. For example, after each interview, I would take notes, and if the interviewee were an academic scholar, I would read their most cited paper that discussed MTurk's characteristics as a platform for academic research or used it as a platform for data collection. Also, I transcribed all the interviews by myself instead of a professional transcriptionist. As such, I could refresh my memory of an interview as I transcribed it and developed an intimate recollection and reflection on the particular responses and common themes to the questions. I also perceived self-transcribing to be more efficient because my interviews were so sporadic and may not have fit well into the professional transcriptionist's workload or time flow.

My second round of data analysis started after I finished collecting all my data. My strategy was to print out all my interview transcriptions and collected documents, and then read them through while highlighting key verbatim and writing down my notes and comments by pen. Such a strategy was essentially an inductive analysis and coding process. I would write down any remarks and themes that came up in my mind and made comments on any potential relationship between them. For example, I noticed that some emerging themes of scholars' opinions on payment seemed to be contrary to some IRB directors' perspectives. I also noticed that

guidelines on crowd work-based research varied significantly in terms of detail levels, particular emphases, and contemporariness. After the second round of data analysis, I had formed a basic sketch of the potential codes and themes that could emerge from the data.

My third round of data analysis was more rigorous. I imported all my transcriptions and collected documents to Atlas.ti so that I could make my coding more systematic and visualize the relations and patterns between the codes across multiple documents. Meanwhile, I followed Braun and Clarke (2006)'s proposed steps in conducting a thematic analysis. First, based on the initial two rounds of preliminary data analysis, I had familiarized myself with the interview transcriptions, guidelines for crowd work-based research, and my interviewed scholars' published papers. Second, I read through all my interview transcriptions and collected documents again on Atlas.ti and used it to code them again. Then, I compared my codes on Atlas.ti with the codes that I had written down during my second round of data analysis on papers. Third, I updated my codes on Atlas.ti with reference to my previous codes and notes on papers. Then, I searched for the themes that emerged from collating these initial codes on Atlas.ti. Fourth, I reviewed the themes for their fit for relation with the codes as well as with the entire data sets, including my interview transcriptions and collected documents. Fifth, I defined and described each theme with refined specifics. Finally, I generated a set of deductive codes grounded in the Belmont Report and compared them with the themes that I inductively coded and generated from my dataset. Based on it, I made some final adjustments to my themes and generated a narrative to present my findings.

Data Validation

As the strategies for validating findings, although research validity and reliability are traditionally associated with quantitative design, they have been applied in the qualitative inquiry

to evaluate the rigor and credibility of the study (Creswell and Miller, 2000). Qualitative validity requires the researcher to examine the accuracy of the findings by employing certain procedures such as triangulating different data sources, and the researcher's approach is consistent and applicable to the other researchers and projects (Gibbs, 2007).

I have applied the following strategies to improve the validity of my data. First, as a part of my research design, I have gathered data from a diversity of types and sources. I have interviewed scholars from different research disciplines and IRB directors and analysts in different institutions. I have also collected guidelines for crowd work-based research from various sources and published papers in different disciplines and publication venues. Such a diverse dataset allows me to compare and cross-reference these data. Second, I have described and reflected on the evolution of my positionality as well as the potential biases in this dissertation project. I have also looked for negative or discrepant information countering the themes actively during my data analysis phase. Third, I maintained regular consulting with my advisor about my research procedures and findings and made revisions based on her advice. Finally, I will do member-checking and invite some interviewees in my research to examine my findings and give me feedback. These practices are beneficial and important to improve validity in a qualitative research project (Creswell, 2014).

Positionality Evolution

Qualitative research requires the inquirer's interpretation and reflection. A qualitative inquirer's positionality, such as their background, bias, prior experience, could influence how they conduct the research, communicate with the subjects, and interpret the findings (Creswell, 2014). As such, I will explain my positionality in this dissertation work and reflect on how it has evolved and transformed.

My positionality in crowd work-based research evolved through four stages. In the first stage, I regarded myself as a researcher who heavily relied on a crowd work platform, namely MTurk, for data collection and research implementation. I conducted my first human subjects research project on MTurk back in 2013 when I administered a large-scale survey to explore internet users' privacy concerns about targeted ads. I also wrote my first IRB application for that survey project. I remember that I was delighted and surprised at how fast my IRB application got approved, and how quickly my survey was responded by MTurk workers. Meanwhile, I found my survey data from MTurk in good quality except for a few obvious cheaters. Thenceforth, MTurk became the easy and only recruitment platform for my survey studies. Moreover, it introduced me to the research field of crowdsourcing, and I started to believe in the power and efficiency of collective intelligence and crowd work. I published my first human subjects research paper based on the data collected from MTurk. At this stage, my positionality in crowd work-based research was overwhelmingly positive and optimistic.

In the second stage, my optimism on crowd work-based research had gradually and partly transformed into a combination of curiosity and skepticism. What concerns, if any, did these crowd workers have, given they were paid so low and being asked to provide data by so many requesters? This marked the second stage of my positionality in crowd work-based research when I became more familiar with MTurk than before and had read many papers about it. In particular, I wanted to explore one topic with MTurk workers, which was their privacy concerns and experiences. I chose this topic because my research projects at that time were about privacy in different contexts, and I was curious about privacy issues on MTurk. Back then, there wasn't much publication on privacy in crowd work, and my work would be exploratory and pioneering. Moreover, I had a dispute with my research supervisor because he did not believe my exploration

of privacy issues on MTurk would amount to any significant findings. After all, most surveys seemed generic and benign as far as we had seen.

However, I made up my mind to continue exploring this topic and used my own money to administer a survey to inquire about MTurk workers' privacy concerns and experiences around the globe. As could be imagined, without any funding support, I was very thrifty in using my own money to pay MTurk workers even though I had solicited their responses to a survey with many questions, most being open-ended. As a result, my survey still received many detailed responses sharing their privacy concerns and depicting their privacy compromise experiences. Among these responses, I perceived a sense of voluntary disclosure of many MTurk workers' privacy experiences because they wanted their voice of privacy concerns to be heard and various privacy-intrusive practices on MTurk to be revealed and addressed. Meanwhile, however, some MTurk workers also complained to me that I had been too stingy in payment, given that they had to input so much to complete my survey. My positionality at this stage, therefore, was a mixture of sympathy and guilt. I was sympathetic to MTurk workers and advocated with my pioneering paper not to neglect privacy vulnerabilities in the context of crowd work. Meanwhile, I also felt a bit guilty because those MTurk workers had provided me with so much information, whereas I had paid them with so little (even though I provided bonuses to a few MTurk workers).

In the third stage, my mixed feeling of sympathy and guilt pushed me to side with MTurk workers and a few researchers who advocate for MTurk workers' rights and empowerment, such as Dr. Lilly Irani from UCSD. Also, I started to hold a critical view of many MTurk requesters, including myself, because we paid crowd workers too little while requesting too much. I reflected that my previous delight and optimism on the recruitment efficiency and data quality of crowd work-based research were based on the jeopardy, disrespect, and even exploitation of

crowd workers. I also posited that even if the purpose of a crowd work-based research project would benefit crowd workers eventually, it could not justify the exploitation of a sample of crowd workers in that project. For example, I still believe that my pioneering research on MTurk workers' privacy concerns and experiences is beneficial to the whole crowd workers community because my paper had disclosed their privacy vulnerabilities and could potentially hold requesters more knowledgeable and accountable in respecting crowd workers' privacy. Despite it, I reflected that the valid intention and potential benefit of my research could not justify my nominal payment to those crowd workers in my sample. Hence, I decided to focus my dissertation topic on ethics in crowd work-based research, which I perceived as a higher and broader advocacy topic than exploring crowd workers' privacy concerns and experiences. I positioned myself as if I were a Bourgeoisie-turned Bolshevik that purported to voice and fight for crowd workers as if they were Proletariats that need to be liberated.

In the current stage, however, my positionality in crowd work-based research has evolved again through my dissertation research. Such an evolution occurred during and after I have interviewed 15 researchers from different disciplines, 17 IRB directors and analysts from different institutions; analyzed 14 MTurk guidelines by researchers and IRBs, and after I have critically reflected on the nature and characteristics of crowd work-based research. More specifically, my positionality has changed in the following aspects.

First, my attitude toward crowd workers has changed from sympathy to empathy. I still understand their concerns and expectations of various ethical issues such as privacy and exploitation, but I dropped my compassion to crowd workers in the context of crowd work-based research, at least to a certain degree. In this context, crowd workers do have the autonomy to choose between different academic HITs and quit a task without a penalty or even with full

payment. More often than not, it might be crowd workers' lack of differentiation or knowledge in academic HITs from non-academic HITs that render them to perceive academic HITs as exploitative as compared with non-academic HITs. Meanwhile, there is a portion of crowd workers who would cheat in HITs, such as by using a bot or script to automate their labor, which leads to poor data quality and research validity. Moreover, ironically, some crowd workers would abuse their power given by some academic scholars to negotiate with IRBs or some other academic scholars. For example, as I will describe in Chapter 6, several researchers and IRB directors in my interviews complained that some MTurk workers would threaten them by posting a bad review and rating on Turkopticon.com if they were not duly paid, even though these researchers and IRB directors had evidence that these MTurk workers did not pay sufficient attention to their tasks at all. To sum up, therefore, my positionality with crowd works has evolved from unconditional sympathy to eclectic empathy.

Second, my attitude toward requesters (including both academic and non-academic requesters) on MTurk is still critical, but the rationale has changed. As aforementioned, I held a critical view of requesters previously because I perceived their payment to be too little. Through this dissertation work, however, I realized that whether a nominal payment equates to exploitation is debatable. For example, some IRB directors discourage a relatively high payment rate, such as that in the standard of the federal minimum wage, because they worried that it might be coercive to some crowd workers. Some researchers also pointed out that whether a payment was nominal depended on where a crowd worker was living. Fifteen cents for a 10 mins survey is minuscule to crowd workers in the U.S., but it becomes a comparatively high payment rate for crowd workers in India or South Africa. Therefore, in the current stage, my positionality toward requesters' exploitation in payment has been less critical than before but more open-

ended and neutral. On the other hand, however, I become more critical on requesters' insatiable pursuit for fast, cheap, and convenient sampling and data. I will elaborate on this perception in more detail in the Discussion Chapter.

Third, my positionality toward crowd work in the current stage has become more dubious and critical. Previously, I held a positive and optimistic attitude toward crowd work because it could harness collective intelligence to solve problems more efficiently and effectively than individual human beings, teamwork, or even computers in certain types of tasks such as audio transcription. Through this dissertation work, however, I started to reflect more on its dark sides. Take MTurk as an example, it holds little responsibility in protecting either crowd workers' or requesters' interests. It has no mechanism to screen workers' data quality or requesters' accountability. On the other hand, MTurk has no regulation or policy on payment, which gives rises to various contingent issues related to payment. For instance, requesters have to make arbitrary payment rates and crowd workers would complain about low renumeration that is often due to arbitrary rates. In addition, MTurk is not proactive in protecting privacy and data confidentiality. As my previous research indicated, there were many privacy intrusive tasks and security risks on MTurk. However, MTurk still have no policy regulation or technical features to monitor or resolve these risks.

Potential Biases

Related to the evolution of my positionality, I also want to reflect and state the potential biases that could influence my interpretation of the research findings. First, I assume that the interview respondents in my dissertation research already have their awareness and understanding of ethics in crowd work-based research. This assumption is based on the fact that the researchers that I have interviewed have been active in conducting or evaluating crowd work-

based research; the IRB directors and analysts that I have interviewed showed their willingness to talk about ethical issues in crowd work-based research when they responded to my invitation. As such, my dissertation work excludes researchers who may have strong ethical opinions on crowd work-based research but never publish any work about crowd work; it also excludes researchers as well as IRB directors or staff who may not care so much about ethics in crowd work-based research. Such exclusions may generate a bias on the significance level of ethics in crowd work-based research from my dissertation findings because many researchers and IRB directors who did not participate in my research may not take crowd work-based research ethics as seriously or knowingly as my interview respondents.

Second, my philosophical preference of ethics in crowd work-based research is leaning toward deontological ethics that emphasizes respecting individual human being's dignity and autonomy rather than the utilitarian ethics that proposes maximizing the majority's benefit and happiness even at the cost of a few people's risk and pain. In the context of crowd work-based research, therefore, I am particularly mindful of my interview respondents' viewpoints on how they understand and implement the Belmont Principle of respect for persons, and how they assess the ratio of benefits over risks to implement the Belmont principle of *beneficence*. I am also sensitive to my respondents' opinions on autonomy, voluntary participation, benefit, and minimum risk because these concepts are closely related to deontological ethics and utilitarian ethics. My preference toward deontological ethics may generate a bias in my interpretation of the research findings because I may be more critical to some respondents' viewpoints in favor of emphasizing the minimum risk on crowd workers while neglecting the potential risk or consequence of crowd workers being merely treated as a means to an end but not an end in themselves.

Third, due to my prior research experiences with privacy and MTurk, I may be more sensitive to privacy issues in crowd work-based research than the other researchers if they conduct a similar study on ethics in this context. Privacy is an essential ethical issue but has not been discussed extensively in either empirical or theoretical aspect of crowd work. My large-scale survey (Xia et al., 2017) and literature review (Xia and McKernan, 2020) are two pioneering efforts in these two aspects. As such, I am particularly keen on my interview respondents' consciousness and perception of privacy issues in crowd work-based research, and whether they differentiate privacy and data confidentiality protection. Such prior research experience and sensitivity in privacy may also generate a bias in my dissertation work, as I could potentially overemphasize privacy in proportion to the other ethical issues in crowd work-based research and be more critical to the respondents who may not take privacy as seriously as one of their focal ethical considerations.

CHAPTER 5 – ETHICS IN PAYMENT ISSUES

Chapter 5 introduces various ethical concerns pertaining to payment issues in crowd work-based research. Payment is an essential component in a crowd work platform, and thus, it is also an integral part of crowd work-based research. There have been numerous empirical and theoretical papers discussing payment issues on MTurk, which primarily focused on probing the average payment amount (Horton and Chilton, 2011; Hara et al., 2018) or discussing fairness (Irani and Silberman, 2013; Salehi et al., 2015). However, few studies had investigated ethical issues associated with payment empirically in the context of crowd work-based research. From my interviews and document analysis in this dissertation work, I found that ethical considerations in payment are a salient major theme in both researchers' and IRBs' concerns about ethics in crowd work-based research. These ethical concerns are around undue influence, fair payment, and whether compensation should be regarded as a direct benefit. These concerns are also reflected in part in the guidelines for crowd work-based research and publications that I have collected.

After identifying these themes, I analyze them with multiple theoretical lenses. For example, I refer to the Belmont Report (1979)'s definition of undue influence to interpret interview respondents' concerns about the payment to crowd workers to participate in research. I apply Karl Marx's (1867) notion of exploitation to analyze some interview respondents' concerns about the payment being too low and unfair to crowd workers for their time and effort. I refer to the Belmont Report (1979) again to discuss respondents' struggles of whether the payment could be perceived as a benefit in crowd work-based research. Apart from these analyses, I compare the IRB respondents' and researchers' opinions about payment issues, as well as between my interview findings and document analysis findings.

Ethical considerations in undue influence

Undue influence occurs when an "excessive, unwarranted, inappropriate, or improper reward or other overture" is offered to research subjects to obtain their compliance, and it could also occur when certain ordinary inducement is presented to vulnerable subjects (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part C., Section 1). By definition, undue influence is different from coercion that occurs when "an overt threat of harm" is intentionally presented to research subjects to obtain their compliance (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part C., Section 1). Prior research has also found that IRB respondents sometimes misconceived coercion with undue influence by mixing up the physical harm with the distortion of risk assessment (Largent et al., 2013). Such a misconception is also reflected in my interviewed IRB respondents and researchers, where several of them used the term "coercion" or "coercive" to denote the potential undue influence imposed by payment in crowd work-based research.

In comparison, IRB respondents were more likely to raise ethical concerns about the undue influence of payment on MTurk for academic research than the researchers in my interviews. Some of these IRB respondents' concerns were preemptive, meaning that they were aware that the payment amount on MTurk is usually minimal but were still concerned about undue influence as their immediate reflection on payment. When I asked the reason for such a preemptive concern, some IRB respondents would say it was just what they were looking at on MTurk. An IRB director explained:

> What we really look at, like many other studies, is to see whether the payment is unduly influential on somebody's interest in participating the research. That really is what we will take a look at. (P14)

She was aware that the research payment on MTurk was quite low, but she did not take it as a reason to mitigate her concern about the undue influence or differentiate it from that in the other research venues with payment for participation. She still worried that the payment could distort MTurk workers' decision to participate in academic research.

Some researchers also raised ethical concerns about the undue influence of payment on MTurk, but with a different rationale. They grounded their concern on the observation that the payment amount in academic research on MTurk, which was usually in a federal or state minimum wage rate, was high relative to the normal payment amount on MTurk. Thus, they were concerned that such a comparative value of the payment in academic research on MTurk could exert undue influence. For example, a researcher pointed out:

> [I]f you are legitimately offering a payment rate to a survey that is tied with the minimum wage, I think you are probably offering one of the higher wage rates on MTurk…then people don't want to get rejected from your study; they don't want to be banned, and therefore, the monetary incentive is a lot more significant. That's where IRB is worried about, like, oh, you [a MTurk worker] keep on doing this study because you need this money. (P22)

The federal minimum wage is $7.25 per hour in the U.S., with the lowest state minimum wage in Wyoming and Georgia, which is $5.15 per hour, and the highest minimum wage in Seattle, which is $16 per hour ("Workers in Majority of U.S. States to See an Increase in Minimum Wage in 2020," 2019). By way of contrast, the median hourly wage on MTurk is only about $2

per hour, and only 4% of the tasks on MTurk paid above the federal minimum wage (Hara et al., 2018). In this regard, if an academic task pays the federal minimum wage, it will be higher than 96% of all the tasks on MTurk. Thus, this amount of payment can have a significant influence on MTurk workers to choose which tasks on MTurk to take. Even though P22 did not contend that such a relatively high payment could be unduly influential on MTurk workers, he noted that MTurk workers might be more worried about being rejected or banned than they might otherwise be in taking tasks on MTurk. He also speculated a concern of the IRB that such a payment might influence MTurk workers' voluntary participation and influence their decision regarding withdrawal from the study. I will discuss such a concern about voluntary participation in more detail later.

The payment rate in an academic study on MTurk can also be significantly higher in a U.S. standard than that in many places in the world. This international disparity of payment standard is a concern that some respondents had when they considered the influence of payment in international research on MTurk. For example, a researcher said:

> In the U.S. case, I never had this issue of paying so much as a type of undue influence, but it could happen in cases where people are doing research in other places like South Africa. Like 15 cents in the U.S. is not much but is equivalent to a day's wage [there]. Then if people say that you should pay $3, then it becomes a week's wage. Then that would be an undue influence. (P20)

This researcher noted that a nominal payment to the subjects in the U.S. could be a large amount to the subjects in South Africa. Thus, the payment could be an undue influence if using the same standard in the U.S. to recruit research subjects in countries with substantially lower minimum wages. It could be a legitimate concern and potential problem if a researcher uses a U.S.

minimum payment standard to pay MTurk workers in India as well. India has the second-largest number of MTurk workers, which was estimated to represent 16% of the MTurk population (Difallah et al., 2018) while the Indian minimum wage is only $3 for an eight-hour work in a day (37.5 cents per hour) (Thomson Reuters Foundation, 2019). Thus, paying a U.S. minimum wage on MTurk will be a much higher standard than the minimum wage in India.

However, it is an open question about how often an academic project on MTurk in the U.S. would recruit international MTurk workers such as those in India and pay in the same standard. In this dissertation work, some researchers, such as P18 and P24, told me that they would only conduct MTurk research based on the U.S. population; some other researchers, such as P21 and P29, said that they would apply the same payment standard to both the U.S. and Indian MTurk workers because they wanted to be fair and did not want to treat the latter as "secondary citizens." I will present and discuss my interviewees' opinions about fair payment in more depth later in a separate section.

Finally, a theme of the ethical concerns about the undue influence of payment is neither related to the amount of payment nor its comparative value but instead is about the nature of crowd work-based research. This is a conceptual concern about an ethical paradox between the monetary incentive nature of crowd work and the voluntary participation nature of research. For the respondents who voiced this concern, payment was considered an undue influence if it induces or corresponds with the subjects' purpose to earn money rather than their *ought-to* purpose to participate in and contribute to research voluntarily. However, this is also an ethical consideration that is far from a consensus, and as one IRB respondent said, it is a "perpetual question."

To start with this theme, a few IRB respondents perceived that participating in research is ultimately a voluntary process, which should not be influenced by payment. However, paradoxically, most people take tasks on MTurk because of payment instead of research. P4 explained this paradox:

> It is a paradox because people on the MTurk platform are specifically there to make money for the most part…you are absolutely within your right to say that I am not volunteering to participate [in research], but it's ultimately a voluntary process, and it's important to remain voluntary rather than having that be, you know, unduly influential and saying yes to participate in a study because it's so much money that you can't say no. (P4)

This IRB director argued that most often, people on MTurk were motivated by earning money, and they had the right to participate in academic research with such motivation. However, ultimately, research participation ought to be a voluntary process, and a subject in crowd work-based research on MTurk must remain voluntary to choose whether to participate in a research project without being affected by the amount of payment. If this research subject said yes to participate only because of the payment, which they would not otherwise do, then this payment becomes an undue influence. Hence, P4 saw a delicate line between MTurk workers being motivated by money while not unduly influenced by it when they decide to participate in an academic study on MTurk.

Another IRB director agreed with P4's viewpoint about a delicate line between monetary motivation and the undue influence of payment. However, she posited that payment could influence MTurk workers to participate in research as long as such an influence was not *undue:*

> What we really look at is that the compensation should not be UNDULY influential [emphasized by the respondent], meaning that it can influence people. A lot of people realize that their time is worth money, and that is not unethical for people to participate in research in order to obtain compensation as long as that compensation is not unduly influential. (P14)

Given most MTurk workers are motivated by money on MTurk, how does the IRB know whether a payment would have an undue influence on them? A few IRB directors proposed that if it was possible to identify and distinguish between the MTurk workers who earn money as their primary or supplementary income and the MTurk workers who just take tasks for some fun or loose money, then we might mitigate the potential undue influence of payment by only recruiting the latter group of MTurk workers. This seems to be a good idea, and prior research has indicated that there is a decent and growing portion of MTurk workers who earn money on MTurk as their primary income (Silberman et al., 2018; Hitlin, 2016). However, we still do not know who are these MTurk workers that are not motivated by money or what is their motivation for a specific academic project on MTurk between money and research. Because of such lack of knowledge, one IRB director speculated that it was the reason that the payment is normally low on MTurk to prevent any undue influence:

> That's potentially why the incentive or compensation is so low…you want people to participate in [sic] their own free will; you want them to know what they are getting into; you don't want them to be induced to be doing it because of the compensation. (P15)

On the other hand, not all the respondents agreed with the notion of undue influence. In particular, one IRB director, P17, doubted and critiqued its existence on MTurk and in academic research more broadly. He argued that before we probed into the question of undue influence of

payment, we had to understand and respect MTurk workers as cognizant adults that could make rational choices. The IRB's duty, he contended, was to inform them of research risks and benefits to let them make a choice rather than focusing on whether the payment was too much:

> I say this a lot in our IRB meetings that I am not a big believer in the "undue influence of payment argument." If someone is participating in a study because they want the money, that is a rational choice that they are making. And I don't think that just because you are poor, your judgment is overwhelmed by money. Our job as the IRB is to make sure that people are well-informed about the risks that they might face. And once they are fine about the risks, if they are motivated to take those risks for money, that seems fine. (P17)

The Belmont Report (1979), which was largely developed to address medical and clinical research, situated its notion of undue influence on research subjects being swayed to take dangerous drugs or medical operations that they would not necessarily do if without a large compensation. However, P17 believed that monetary incentive would not be unduly influential regardless of whether it was conventional clinical research or social research on MTurk, so long as the subjects understood the risks.

There are two potential problems in P17's viewpoint here. First, we cannot be sure that the MTurk workers who participated in our research are over 18 years old to be able to make a rational decision. In fact, another IRB respondent, P14, raised it as her ethical concern in conducting academic research on MTurk. Second, I posit that there is a nuance between voluntary decision-making and volunteering. The former is related to the autonomy and free will of a person; the latter is related to the motive of a person's choice. Hence, if academic research seeks for voluntary participation, P17's denial of undue influence on a rational subject is reasonable. But if academic research emphasizes volunteering for research, then I argue that the

purpose of a subject must be aligned with the purpose of research, which is for knowledge rather than money. The Belmont Report (1979) also mixed voluntariness and volunteering together, and it could be an origin of the dispute in the undue influence of payment in research on MTurk now.

Ethical considerations in fair payment

Fair payment has not been raised as an ethical consideration in the Belmont Report nor in AoIR IREs for internet Research. The Belmont Report (1979) had no mention of fair compensation to research subjects but rather focuses on the risk of an "inappropriate or improper reward" on the voluntariness of research. The two volumes of the Belmont Report Appendix (1979) also focused their discussion about compensation as regards the undue influence rather than fairness. Similarly, AoIR IRE 3.0 had no guidance for what would constitute a fair compensation for internet research.

In contrast, however, fair payment has been a major ethical concern in crowd work-based research based on my findings. Some researchers and IRB respondents were concerned about the payment being too low and exploitative of crowd workers' time and effort; some researchers proposed using a federal or state minimum wage as the benchmark of a fair payment; some other respondents argued that the fairness of payment would be determined by crowd workers' autonomous choice instead of by researchers or IRBs; still, some respondents proposed that the fairness of payment could be gauged and determined by the market mechanism of a crowd work platform. In general, the researchers were advocating for fair payment and implementing it with some minimum wage standard, whereas some IRB respondents would cast doubt on any arbitrary standard or definition for the fairness of compensation.

To start with, exploitation is a complicated issue and has often been deliberated on regarding the labor relationship between the proletariat and bourgeoisie. According to Karl Marx

(1867), exploitation occurs when capitalists pursue the maximization of the surplus value created by workers, which is the extra value produced by workers deducting the labor costs in the form of workers' wage. Such a pursuit of surplus value is enacted by mass production in a capitalist's private business, whereas the workers rely on the job for their survival, and a nation or state protects a capitalist's private business. However, Karl Marx's classic theory of exploitation has been controversial and one critique is that Marx's term of exploitation applies between different classes in an economy (i.e., between the proletariat and bourgeoisie) but not between individuals in a specific marketplace (e.g., between an employer and workers) (Wolff, 1999).

In the context of crowd work, especially in the case of MTurk, there has been an ongoing controversy on whether crowd workers are being exploited. Some scholars, such as Irani and Silberman (2013) and Pittman and Sheehan (2016), posit that MTurk was like a digital sweatshop, and MTurk workers were being exploited in terms of time and labor because their payment was minuscule. Some other scholars, such as Deng and Joshi (2013), posit that crowd work could potentially both exploit and empower crowd workers depending on the extent of a crowd work platform that could afford crowd workers to revise the meaning and social environment of their work. Still, some researchers, such as Busarovs (2013), surveyed and found that many MTurk workers did not perceive themselves as being exploited and were ready to take more crowd work tasks.

The controversy of whether exploitation exists in crowd work-based research is also manifest in my interview respondents' opinions. For example, one researcher told me that he used to pay MTurk workers very little in his research projects, and his colleague termed MTurk a digital sweatshop. A MTurk worker even compared him to a capitalist:

> [The payment] it's really shockingly low. So, every time I would talk about that, then people would say, well, it [MTurk] is just a digital sweatshop. And then, on Turkopticon, a worker wrote like a super scathing review of me. She was incredibly articulate, and she's comparing me to a capitalist. (P21)

Turkopticon is a website and online forum where MTurk workers could rate and review requesters. Turkopticon is a 3rd party site not owned by Amazon but maintained by Dr. Lilly Irani, Dr. Silberman, and their colleagues and volunteers. Its purpose is to balance the power dynamics between MTurk workers and requesters and give the former a channel to voice and review requesters. Turkopticon has been a popular forum among MTurk workers, and many scholars who conduct research on MTurk take workers' reviews on Turkopticon seriously. Later in the interview, the researcher told me that since the MTurk worker's review of him, he started to do research on helping MTurk workers to find tasks with high payment rates and also increased his payment to the level of a state or federal minimum wage. A few IRB respondents also agreed with the existence of exploitation on MTurk because they observed that the payment was normally extremely low.

However, does an extremely low payment necessarily indicate that it is exploitative and unfair? Many IRB respondents did not think so. Their immediate and primary concern about payment, as I introduced earlier, was whether it would be an undue influence. Exploitation or payment being too low was simply not their first thought about research compensation on MTurk. Some IRB respondents also argued that payment should not be the motivation for participating in academic research at all. Hence, exploitation does not apply in this context. For example, one IRB director said:

> Payment should not be the motivation for participating in research. And so, that's not a concern that our IRB has discussed in depth. (P8)

Some IRB respondents did not have a concern about too little payment leading to exploitation but rather leading to a low response rate in recruitment:

> I can't think of a circumstance where we've said, "that's not enough money to give somebody to ask that person what you want them to complete." We might say, "Wow, you are probably not likely to get a lot of people to participate in your study because we can't imagine somebody would give up the amount of time for that amount of compensation." (P15)

Her concern about a low payment was not its exploitative effect on MTurk workers but rather its consequence on research such as that the recruitment might be more difficult, and the sample size might be small. Similarly, a few IRB respondents evaluated payment not from MTurk workers' position but from researchers' position. They denied the existence or potential of exploitation because the research on MTurk that they have reviewed was always of minimum risk and short. Thus, they perceived that a low payment is commensurate with the risk level and task duration of crowd work-based research on MTurk. For example, when being asked about any concern about exploitation, two IRB directors responded as follows:

> No, we ask them how long the survey would take, so if it is something to be very long, then we would worry about it, but most often, it's like 5-10 mins, so it hasn't been an issue [of exploitation]. (P2)

> That [exploitation] is something we don't normally consider especially with these very minimal risk studies. (P8)

Furthermore, a camp of IRB respondents posited that MTurk workers have the autonomy to choose tasks based on their assessment and acceptance of payment and other factors. MTurk workers are not forced to accept a low payment rate if they decide not to. Therefore, it is a fair game on the table, and no exploitation is involved. An IRB director explained:

> [I]t's still their [MTurk workers'] decision whether or not to do it [a research study]. No one is telling them that they have to answer this survey, but they are choosing to answer this survey. That goes back to autonomy. If they think that [payment] is too low, then they shouldn't answer that survey. It's an offer on the table, and if you want to take this offer to answer this survey, this is the money you get, and this is your choice. I don't see how that could be considered as exploitation. (P13)

In her perspective, MTurk workers can decide whether or not to do a study and choose whether or not to accept its payment rate. If they decide to participate in a study, it means that they also choose to accept the amount of research compensation. As such, MTurk workers' autonomous decision-making and choice-making leave no room for exploitation. Another IRB director held a similar view on exploitation with P13 above, but she argued it from the perspective that MTurk is not an employment-based workplace. Thus, MTurk workers are not contracted or obliged to accept a payment rate if they feel that the rate is not worth their time:

> MTurk is not a platform of employment; it is ultimately not an employment contract, so [sighed], it is voluntary and on that a participant who feels that the level of compensation is an exploitation of their time, they are certainly within their rights to choose not to participate in that task. (P4)

Finally, one IRB director questioned the normalization of exploitation on MTurk and academic researchers' responsibility. She argued that if an extremely low payment had been

normalized on MTurk, where most tasks were paying very little, and MTurk workers were still taking these tasks, whether and when should academic researchers take the responsibility to deviate from such a norm? She argued:

> Because the payment on MTurk is extremely low, I don't think that it is necessarily germane to research itself but that's the part of landscape of MTurk. So, if MTurk itself is exploitative, do we have to hold research to a higher standard and when do we hold research to a higher standard than, you know, people use MTurk for a lot of other commercial purposes as well? If that [MTurk] is an exploitative situation but it's normalized in a certain sense, where is the responsibility of researchers lie in interacting in that exploitative situation. (P14)

She noted that the extremely low payment on MTurk had been quite common, but such a norm was not related to or determined by academic research but by the MTurk platform. Since MTurk includes both academic and commercial tasks on it, she questioned whether and when academic tasks should have a higher standard than commercial tasks. On the one side, P14's question is legitimate because MTurk is not purposed for academic research, and academic researchers cannot decide for MTurk what a normative payment means; on the other side, P14's question is also problematic because academic research is supposed to have a higher standard than commercial research in how it treats its subjects, and that is exactly the reason why the Belmont Report was published, and the IRB was established several decades ago.

Furthermore, following P14's argument, we can ask, what is an appropriate standard of payment on MTurk for researchers to abide by, and by what criteria should a standard be deemed fair? While most IRB respondents told me that they did not use any benchmark to evaluate research compensation on MTurk, many researchers in my interviews had chosen to use federal

or a state minimum wage to determine their payment. However, such a choice, when I inquired into it with the IRB respondents and some other researchers, received many critiques. These critiques include: (1) there is a large variance between the minimum wage in a state, a federal, and an international level; (2) MTurk workers are independent contractors, and therefore a minimum wage does not apply to them, and (3) it is hard to estimate a fair task completion time and scale up a minimum wage accordingly. For example, an IRB director commented on the variance of a minimum wage:

> I think it is an interesting idea, but the minimum hourly wage varies significantly, and MTurk people are pulled from lots of different places. Would it be the federal wage or the wage of Seattle? Seattle's minimum wage is significantly higher than the federal minimum wage, so that's where the question mark of whose minimum wage [you were talking about]. (P14)

She pointed out that the MTurk population is diverse, and a researcher may not know where their research subjects come from. Hence, if a researcher chooses a particular state's minimum wage or the federal minimum wage as her benchmark of payment, then it may be fair to MTurk workers in certain states but not necessarily in the others. As she said, Seattle has a significantly higher minimum wage than the federal standard, and if using the federal minimum wage as the benchmark of payment to research subjects from Seattle, it would be a rather low standard. A researcher added that the minimum wage also varied in different countries and areas, and thus, it is difficult to ascertain a minimum wage standard that is fair internationally:

> If you wanted to settle upon a minimum wage, what wage would you pick? It's a super interesting and difficult to answer question. An easy answer would be a federal or state minimum wage, but there's this weird factor that comes into play on crowdsourcing

platforms where you're drawing workers from all around the world. A fair wage in Mumbai is going to be different than in Minneapolis. (P21)

Additionally, a researcher critiqued the validity of using a minimum wage as the benchmark of payment on MTurk because MTurk workers are not employees but contractors. A minimum wage only applies to employees, and therefore, it does not apply to MTurk workers. He contended:

> Minimum wage applies to employees, not contractors, and things are being done on MTurk on a task basis, so hourly wage does not apply…like you delivering packages, and you are contractors of Amazon, and Amazon will pay by the packages that you delivered or by how much time you spend. So, it's up to you, if you deliver more packages, you can be paid more. (P28)

Finally, some respondents argued that it is hard to estimate how long a research project would take MTurk workers to complete, and thus, it is hard to scale the payment up accurately to an *hourly* wage. Hence, for them, it is better to let the MTurk market determine a fair payment. For example, one IRB director explained:

> MTurk is market-driven, and the amount that people get paid is kind of the amount that is driven by the MTurk worker population…I feel that the MTurk market is a better way than minimum wage because doing minimum wage you have to have a very accurate estimate on how long it will take people to do your work because it's hourly. (P17)

It is noteworthy that P17's argument may be a bit flawed. Several researchers told me that they were able to and would make an accurate estimation of how long their task on MTurk would take people to complete. They would launch a pilot study with their task on MTurk and observe how long it would take MTurk workers to complete. Based on such pilot data, they would then

calibrate the task time and payment amount. Still, some IRB respondents concretized P17's idea of how could the MTurk market determine a fair payment. For example, an IRB director argued that MTurk workers could determine what a fair payment is for them:

> [W]e let them [MTurk workers] decide what the fair payment is, and if they want to do the work for the payment, then that's their choice. But we don't have a standard that it has to be a certain amount. (P4)

P4's argument relates back to the notion of autonomy and voluntary participation: if MTurk workers choose to take a task and accept its payment, then the payment is fair to them.

Ethical deliberation on compensation as a benefit

Besides the ethical concerns about undue influence and fair payment, there is also an ethical deliberation that emerged from my interview responses on whether compensation in crowd work-based research can be regarded as a benefit to crowd workers. In the Belmont Report, the research benefit refers to "something of positive value related to health or welfare" in a research context (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part C. Section 2). However, the meaning of "welfare" is not specified in the Belmont Report. On the other hand, AoIR IRE 3.0 stated that "Research participants may benefit in some way from collaborating or participating (personally, professional or otherwise)" (Franzke et al., 2020, p. 71). The meaning of "personal, professional, or otherwise" is also not elaborated further. Meanwhile, there is not an explicit warning or statement in these canonical guidelines of research ethics that a research benefit *must not* be monetary.

Hence, it becomes a gray area for both the IRB respondents and researchers in my interviews; they interpret the research benefit using different approaches. Some of them insisted

that the benefit cannot be monetary, and as such, compensation in crowd work-based research is by no means a benefit to subjects. One IRB director argued:

> We specifically don't consider, and I say 'we' in terms of my IRB, and I would say it's across the IRBs in all my experiences, that compensation is never considered a direct benefit of participation. (P4)

P4's view is common among the IRB respondents, but why can compensation never be considered as a direct benefit of research participation? Some IRB respondents explained that because research benefits aim to justify research risks, and they should be commensurate with each other. However, compensation, no matter how much money it involves, cannot justify research risks, and thus, cannot be treated as a type of benefit. One IRB director explained:

> One of the biggest tasks of IRB when we review a study is to make sure that the risks of the study are commensurate with the benefits of the study. So, that's one of the reasons that we don't consider compensation to be a benefit because if you have a really high-risk study, paying people a ton of money doesn't justify exposing people to higher risk. Say if an extremely high-risk study, there is not much benefit to the society; there are no extra benefits to the subjects, but I pay them a ton of money, I don't want that money I pay them to offset the risk. (P14)

P14's argument may be applied to academic research in general, but given most IRB respondents also agreed that academic research on MTurk is usually of minimal risks and the payment is also nominal, can compensation be considered as a benefit to MTurk workers who aim to earn money on MTurk? Some IRB respondents still replied no because the only thing that the IRB considers, and the IRB wants MTurk workers to consider is research risks and benefits without weighing them against compensation. One IRB director explained:

> We make sure that benefits are separate from monetary compensation…the only thing that IRB considers and the only thing that we hope the Turker considers is the risks and the benefits of doing the project and that they don't consider the compensation--that they don't try to weigh their risks against the compensation. (P12)

On the other hand, however, some researchers posited that academic studies on MTurk usually have little benefit to MTurk workers other than the monetary incentive, and the monetary incentive is the motivation for most MTurk workers to participate in academic studies on MTurk. One researcher commented:

> For academic studies, I don't see many benefits to the worker other than the money...I think a lot of the IRB discussion about that the benefit should not just be monetary is kind of inherently false because that's why people are doing the things. We are not helping them in some way by having them fill out 50 questions on self-efficacy, like "I wish your life was changed healthily by that." But to me, that's not the real world. The money is an important part, especially for MTurk [workers] doing it. (P22)

P22 argued that there is hardly any concrete benefit to MTurk workers in an academic study. Some IRB or scholars may claim that the benefit from their survey is to help the participants to improve their self-efficacy, but in reality, to improve self-efficacy is not most MTurk workers' aim to participate. In P22's view, a researcher's purpose and most MTurk workers' purpose is not aligned. Thus, if the IRB assumes that MTurk workers are motivated by any non-monetary research benefit, it is inherently wrong. In fact, a few IRB respondents also acknowledged it even they knew that in policy, research benefits could not be monetary. For instance, one IRB director said, "The benefits are mostly, from our perspectives, the incentives and payments that they can

get out of it. I don't think anyone is benefiting from going through the types of survey questions that are there." (P7)

Finally, an IRB director attempted to reconcile the two opposing views on whether the payment could be perceived as a benefit to MTurk workers. She acknowledged that payment could be a sort of benefit to MTurk workers but only to the extent of rewarding their time and effort. The primary benefit, she argued, should be identified as an added value to the research subjects beyond a monetary reward:

> We think benefit and compensation as two separate things when we approach it. Though yes, similarly speaking, the fact that MTurk workers are getting paid is a sort of benefit to them. But that amount of payment should be in relation to the amount of time and effort being put it. When we look at benefit, [what] we are looking at is any added value to them of doing this beyond that…so if you are studying a specific population and collecting their information in your survey, could the finding of that research benefit that population, like help their community. (P1)

To sum up, it seems to be a gray area in the current policies and regulations for academic research on MTurk on whether payment can be categorized as a research benefit.

Document analysis of ethical issues in payment

To cross-reference the findings from my interview data, I also analyzed the crowd work-based research guidelines and my interviewed researchers' publications in terms of their ethical considerations in payment-related issues.

The ethical concern about "undue influence" on MTurk is not prevalent in these documents. From an IRB's perspective, a reason could be that a general IRB application template for human subjects research already requires a description of how to minimize undue

influence or coercion. Hence, none of the research guidelines from IRBs highlights it specifically for MTurk. On the other hand, none of the guidelines from researchers and academia, as well as the publications that I have collected, has mentioned undue influence or coercion. However, noticeably, one public research guideline "M-Turk Guide" proposes to "simply pay more:"

> Simply pay more. When you enter in how long you expect the study to take and how much you will pay, MTurk calculates the hourly wage you're paying. Our work and others have shown that workers are sensitive to how much they are getting paid – the more you pay, the quicker the data rolls in. Everybody wins. (D10)

This guideline is not an official document from any IRB but a researcher, Dr. Michael Buhrmester (he did not participate in my interview). He suggested that paying more is better because prior research had indicated that a higher rate would induce faster responses from MTurk workers, and MTurk workers would also be compensated more. Hence, it is a win-win situation. However, given many IRB directors in my study were concerned about undue influence on MTurk workers, IRBs might not agree with this "simply pay more" suggestion in this guideline and recommend it to researchers. I will discuss about the implication and consequence of "simply pay more" in more detail in my Discussion Chapter.

On the other hand, the ethical concern about "fair payment" on MTurk is more salient, particularly in the research guidelines and publications by scholars that I collected. For example, one guideline, "Guidelines for Academic Requesters 2.0," explicitly emphasizes a fair payment to MTurk workers:

> Pay Turkers fairly. They are a workforce, not a volunteer study population
>
> Crowdsourcing workers are a labor force. Many depend on income from crowdsourcing as critical income. Crowdsourcing workers are legally considered contractors and

> therefore are not protected by any minimum wage laws. When requesters pay a fair wage and treat workers like people, both sides receive positive results. (D15)

This guideline directly addresses some IRB respondents' and researchers' concerns above. It emphasizes that MTurk workers are a workforce but also neither volunteers for academic studies nor employees protected by minimum wage laws. The guideline proposes to pay fairly based on MTurk workers' identity as such, but is MTurk workers' identity of being contractors a sufficient condition for a fair payment? Plus, is any minimum wage a necessary condition for a fair payment? Some respondents would say no to either question. First, a few IRB respondents would still treat MTurk workers as volunteers instead of contractors when participating in academic research. Hence, any standard of payment seems not applicable to MTurk workers. Second, some IRB respondents would argue that the fairness of payment is not related to the payment amount per se, but whether it is commensurate with research time and effort. Also, it is notable that, like D10 above, which seeks for an "everybody wins" consequence, D15 also proposes to achieve a consequence that if pay fairly, both academic requesters and MTurk workers would receive "positive results." Hence, both D10 and D15 implicate a consequentialist stance of ethics.

This same guideline further recommends academic requesters to pay at least according to the norms of the minimum wage in the MTurk community:

> Pay (at least) community norms of minimum Turking wage. Underpayment of crowd workers is anything less than the current federal minimum wage in the United States. Since Turkers work independently, they are responsible for their own computers, electricity, taxes, health care, etc. Different workers consider fair pay anywhere from $6 an hour to $22 an hour. (D15)

It stresses a standard of "underpayment of crowd workers" and also suggests the fair payment criteria from MTurk workers' perspective (the authors of this guideline have communicated with many MTurk workers to understand their evaluation of fair payment).

As regards IRB guidelines for research payment on MTurk, they often do not have any item about it or concern about exploitation. An only exception is an IRB consent form template for MTurk that gives a payment rate to researchers:

> The survey should take [**7**] minutes to complete. After completing the survey, you will be paid [**.50**] for your participation. (D15)

This consent template recommends a payment rate of 50 cents for a 7 mins survey on MTurk. If converting to an hourly wage, it would be approximately $4.30 per hour. It is obviously below the federal minimum wage, and below any state-level minimum wage in the U.S. The other IRB guidelines for crowd work-based research that I have gathered do not have such a specified rate for MTurk studies.

In comparison, a researcher's private guideline for MTurk studies gives a much more detailed and generous policy about payment:

> Price. Estimate the time it will take to do the task. Set the reward to $0.15 per expected minute. Round up, if needed. When you test the task, you will check your estimate. Note that "fair pay" is hard to define for a distributed marketplace. For our purposes, we will define "fair pay" as $9-10 per hour, averaged over all workers who do the task properly. (D4)

In this guidance, this researcher has detailed instructions on how to estimate a payment to MTurk workers at a rate per minute. Also, he notes that fair payment is hard to define but provides a

rule-of-thumb rate of $9 - 10 per hour. This rate is much higher than D15's $4.30 hourly rate and is also higher than the federal minimum wage ($7.25 per hour).

In research publications, "exploitation" is more frequently mentioned than "fair payment" and as my literature review indicates, exploitation is a recurring theme in publications about MTurk. One paper introduces its aim is "in part to draw attention to commodification and exploitation in large-scale crowdsourcing markets" (D23). Another paper elaborated the potential exploitation by academic requesters on MTurk workers:

> [Prior] study found at least two thirds of [MTurk] workers considered themselves at least somewhat exploited. Wages were one reason, and the problem is particularly relevant for academics—those in the humanities and social sciences have especially limited funds for data collection. At the same time, journals require appropriate levels of statistical power (derived in part from a robust number of participants) for a study's results to be considered valid. As a result, researchers are put in a precarious position of balancing subject needs with research integrity. (D24)

This paper illustrates a tension between the limited research funding in humanities and social science and the academic journals' requirement of a robust sample for statistical power, which could lead to a limited compensation to each research subject on MTurk. In turn, such limited compensation could trigger some MTurk workers' perception of being exploited by academic researchers. In fact, one researcher in my interview raised a resonance: "you know, not everybody has the funds to be doing a lot of research." (P25)

Finally, as regards benefit, several IRB guidelines for research on MTurk explicitly state that there is no direct benefit to research subjects, and compensation is separated from benefit. For example, one IRB guideline says:

> The research will not benefit you personally. (D15)

Another IRB guideline also states that there will be no direct benefit, and suggests researchers add that there could be certain knowledge obtained by MTurk workers related to the research objectives:

> While there are no direct benefits to you, we hope to gain more knowledge on (State overall, anticipated benefit for conducting the research). (Add compensation, if applicable). (D5)

D15 and D5 are from a private university IRB and a public university IRB, respectively. Their statements on the research benefit of being non-monetary and separate from compensation are consistent with the majority of IRB respondents' opinions in my interviews. By contrast, it is noteworthy that no guideline from researchers or an academic association or community has stated anything particular about direct benefit in crowd work-based research.

Discussion

Based on my findings of ethics in payment issues in crowd work-based research so far, several questions deserve further reflection and discussion. First, where do these ethical considerations of payment come from? Second, what are the differences in these ethical concerns about payment, if any, between IRB respondents and researchers, between IRB respondents from different institutions, as well as between researchers from different disciplines? Third, what are the implications from these findings for crowd work-based research in the future? I will discuss these questions below.

The origin of ethical issues in payment

To start with, I argue that the origin of ethical issues regarding undue influence, fair payment, and benefit of payment is a central mismatch between MTurk workers' legal identity in

crowd work-based research, i.e., "independent contractors," and their perceived-identity, i.e., "employees" or "research subjects."

First, Amazon defines MTurk workers as independent contractors. By this definition, it indicates that a MTurk worker "will not represent yourself as an employee or agent of a Requester or Amazon Mechanical Turk" and "will not be entitled to any of the benefits that a Requester or Amazon Mechanical Turk may make available to its employees, such as vacation pay, sick leave, and insurance programs, including group health insurance or retirement benefits" ("Amazon Mechanical Turk," n.d.-b). Hence, MTurk workers' payment or benefit is completely at the discretion of requesters and is not protected by law. Second, the Belmont Report, which still guides IRBs and researchers, describes a research subject as "the reasonable volunteer…knowing that the procedure is neither necessary for their care nor perhaps fully understood, can decide whether they wish to participate in the furthering of knowledge (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part C, Section 1). In this sense, a research subject is essentially a volunteer and can make a reasonable decision to participate in research. Third, colloquially, MTurk workers are called "workers," and that could be easily being equated or confused with "employees" that are paid by a wage standard.

These three distinct identities are being applied and confused in the context of crowd work-based research, and I argue that it is an origin for the various ethical considerations in payment. For instance, if MTurk workers are only perceived as independent contractors in crowd work-based research, then some IRB respondents and researchers' arguments are valid that any benchmark of payment, such a federal or state minimum wage, does not apply to MTurk workers. Hence, some researchers' practice and research guidelines' recommendation of setting

up and proposing for a minimum wage is only their subjective interpretation of fair payment and unconvincing to generalize.

If MTurk workers are perceived as employees in crowd work-based research, the benchmark of a minimum wage would be fair, but a series of other ethical controversies would arise from my findings above. For example, exploitation would become possible if payment is much below a minimum wage to MTurk workers. In turn, some IRB respondents' denial of the existence or potential of exploitation would be questionable, and IRB guideline D15's recommended payment rate of 4.3 dollars per hour would seem exploitative. On the other hand, however, some IRB respondents' ethical concern that academic requesters might use their payment to manipulate MTurk workers would be viable as well because these academic requesters might assume that their MTurk subjects are hired by and dependent on them.

Finally, if MTurk workers are only perceived as research subjects, i.e., research volunteers, in crowd work-based research, the ethical concerns about payment would also be complicated. In this regard, IRB respondents' ethical concern about undue influence would be appropriate because an undue influence would intervene and damage the voluntariness of research participation. Their standpoint that compensation cannot be regarded as a benefit would also be well-grounded. However, perceiving MTurk workers only as research subjects would create a potential controversy that payment is arbitrary in crowd work-based research and could even be discouraged or dropped to ensure that all the research participants are only out of a volunteering motive. Also, it would contradict with many MTurk workers' motivation that is driven by money (e.g., Alkhatib et al., 2017; Durward et al., 2016; Lee et al., 2014) and plausibly exclude them from participating in research, and this, in turn, might cause certain ethical dispute of justice where research risks are only taking by MTurk workers who are volunteering for

research and not motivated by money while research benefits can distribute to the MTurk workers who choose not to participate because of the underpayment in the task.

To sum up, how IRBs and researchers perceive MTurk workers' identity makes a significant impact on their ethical concerns about payment issues in crowd work-based research. At present, the confusion and inconsistency in MTurk workers' different perceived identities could be traced as an origin to the contentious ethical considerations. The researchers' and IRB respondents' different framing and interpretations of MTurk workers' identity drives their different ethical considerations in crowd work-based research. Perhaps, to settle these ethical contentions about payment, IRBs and researchers need to reach a consensus on how to identify MTurk workers as a population in the context of crowd work-based research. For example, if they agree that crowd workers should be treated as research contractors, e.g., like research vendors hired by an academic institution, then they should emphasize more on paying them with a wage standard and emphasizing research obligation to them, e.g., ensure data quality, than expecting voluntariness and autonomy. If IRBs and researchers agree that crowd workers should rather be treated as research volunteers, such as the participants in citizen science projects, then they should emphasize more on voluntary participation and mitigating the undue influence of payment than calibrating a "fair" payment. However, if IRBs and researchers both agree that MTurk workers are "employees" of all the requesters on MTurk, then perhaps, they should not conduct academic research on a crowd work platform anymore because academia should not become a business activity.

Comparisons of ethical considerations from different entities

Though there is some disagreement in either camp, IRB respondents' ethical concerns about payment in crowd work-based research are distinctive from my interviewed researchers'.

Grossly speaking, most IRB respondents were concerned more about the undue influence of payment and a clear separation of research payment and research benefit. Such a rationale seems to be rooted in and can be traceable to the Belmont Report and the CITI training that focus more on the medical, clinical, and legal aspects of human subjects research than the social and behavioral aspects of it. In medical, clinical, and trial studies, undue influence and coercion is a prevailing and significant ethical concern, and some IRB respondents seemed to apply it to crowd work-based research directly.

On the other hand, most researchers were more concerned about fair payment. There could be two interpretations of it. In the first place, some researchers are conscious of a MTurk workers' "dual identity," that is, MTurk workers are both volunteers aiming to participate in research and "workers" aiming to earn money. In the context of crowd work-based research, these researchers are privileging MTurk workers' "worker identity" over their "research participant identity" and trying to figure out how to treat them ethically and fairly and balance the uneven power dynamics between MTurk workers and requesters. Second, I sense that some researchers were a bit guilty of their low payment to MTurk workers when they first started to use MTurk for academic research. Hence, they also advocated for a "fair" payment as an attempt and act to redeem their previous "exploitative" payment to MTurk workers.

As regards institutional or disciplinary differences related to the ethical concerns about payment, I did not find any significant ones. However, there does seem to be some discrepancy and vacancy in the research guidelines and papers that I have collected and analyzed. For example, even though some IRB respondents raised their concerns and opinions about fair payment, IRB guidelines for research on MTurk that I have collected don't have any notice or advice on it. Exploitation has been a topic in several publications that I have collected, but a

minimum wage's applicability or its potential undue influence on research subjects has not been mentioned in publications.

Implications for crowd work-based research in the future

Based on my findings and discussion on the ethical issues in payment, I propose three implications for crowd work-based research in the future. First, payment is an integral part of crowd work-based research, and it involves complicated, even contested ethical considerations. As such, neither IRBs nor academic researchers should take it lightly. I propose that frequent and healthy communication between IRBs and academic researchers to deliberate these ethical concerns about payment, such as undue influence, exploitation, and research benefit, would be constructive for crowd work-based research in the future. At present, to my best knowledge and observation, such communication is lacking between these two camps.

Second, particularly for academic researchers, I propose that a "fair payment" should not be arbitrarily decided or conveniently adopted from the other researchers or a minimum wage standard without a critical reflection on its validity. For example, I think several IRB respondents and researchers' concerns about the undue influence of using a minimum wage in different states of internationally, or the applicability of it to MTurk workers in the first place, deserve further discussion and empirical experimentations.

Third, particularly for IRBs, I propose that "exploitation" should not be casually treated or neglected. They need to acknowledge the nature of MTurk that MTurk workers value their time and labor like many traditional workers do. IRBs may expect MTurk workers to be research volunteers and motivated by the pursuit of knowledge, yet in reality, such an expectation might contradict many MTurk workers' motivation. Hence, IRBs may regard a low payment as

prevention of undue influence in academic research, while MTurk workers may perceive it as disrespecting and devaluing of their labor and time.

Fourth, I propose that "fair payment" and "exploitation" are both open questions in crowd work-based research. Karl Marx certainly could not imagine the prospering of the gig-economy in which the work relationships are distinctive from those between the proletariat and bourgeoisie in a factory. Crowd work and crowd work-based research, as the subset and the sub-subset of the gig-economy, will also challenge the conventional meaning of fair payment and exploitation in terms of their definitions and measurements. Thus, for both researchers and IRBs, it is important to avoid assertion and open their minds for deliberation on how to ensure due respect and treatment to crowd workers while also expecting the same attitudes and acts from crowd workers in return.

Finally, to Amazon and MTurk, I propose that they should take more responsibility in moderating crowd work-based research in the future. In this specific context of MTurk, research ethics and business ethics are closely interdependent and intertwined. Neither side could be intact if the other side is compromised. For example, a dynamic "fair" payment standard based on the task genre and the demand-and-supply provided by MTurk would be a great reference for all the entities: MTurk workers, researchers, and IRBs. Amazon owns such data and has the power to do the calculation. After all, being tagged as a "sweatshop" is not beneficial for MTurk and even Amazon's reputation in the long term.

Summary

Undue influence, fair payment, and benefit are three main themes in IRB respondents' and researchers' ethical concerns about payment academic research on MTurk. Undue influence is concerned because of the payment amount, the relative value of a payment, as well as the

interference of payment on research subjects' voluntary participation in research. Fair payment is concerned in terms of exploitation, the benchmark of using a minimum wage, and how to gauge the fairness. The research benefit is concerned in its relation to payment: some respondents regarded payment as a benefit in crowd work-based research; some held an opposite opinion, and some others perceived it as a gray area for negotiation. Some research guidelines and papers reflected these ethical concerns, but there are still discrepancies and vacancies in these documents. Finally, I discussed the origin of these ethical concerns and the comparisons of ethical concerns from different entities, and implications for future research.

CHAPTER 6 – ETHICS IN DATA ISSUES

Chapter 6 introduces various ethical concerns and practices in data issues in crowd work-based research. Data issues, such as data quality, have been a focus in crowd work and crowdsourcing more broadly. For example, scholars have raised concerns as to whether data collected from MTurk workers is of high quality (Buhrmester et al., 2016; Litman et al., 2015; Peer et al., 2014) and whether research using MTurk workers has satisfactory internal and external validity (Thomas and Clifford, 2017; Berinsky et al., 2012; Paolacci et al., 2010). However, these prior works explored these data issues via surveys or experiments and often lacked a qualitative interpretation from different perspectives. In this dissertation work, I identified ethical concerns and practices regarding data quality and validity issues as a significant theme in my interviews with IRB respondents and researchers. Also, I probed different opinions from academic researchers and IRBs on these data issues and explored the origin of them. Meanwhile, I found these ethical concerns with data quality and validity issues salient in documents pertaining to research guidelines and scholarly publications.

I analyze these identified data issues in crowd work-based research with reference to prior studies, such as those I listed above, as well as to my own research experiences on MTurk. My findings contribute to the current scholarship of data issues in crowd work in two main aspects. First, I inquired into IRB directors' and analysts' opinions about data issues in crowd work-based research, which have seldom been reported previously. Second, I found differences and commonalities between my interviewees and between my interview data and document analysis results. For instance, researchers from different disciplines held contrasting opinions on the effect of non-naïvety; both IRB respondents and researchers worried about the information diffusion in various MTurk forums; most published empirical research on MTurk did not reflect

my interviewees' ethical concerns about data quality and validity. I will enumerate these data issues and comparisons in detail below.

Ethical considerations in data quality

Data quality has been a primary ethical concern about crowd work-based research. Specifically, the IRB respondents and researchers I interviewed were concerned with poor data quality on MTurk due to random and fake inputs by cheaters. They also proposed various practices to ensure data quality, such as using attention and quality check questions and screening qualified MTurk workers to participate in the research based on a certain approval rate. Finally, some respondents discussed issues with the rejection of MTurk workers, blacklisting, and the importance of maintaining a reliable reputation system on MTurk in relation to data quality.

Even though several themes, such as the concern about cheaters, had been reported in previous survey or experimental studies (e.g., Difallah et al., 2012) on MTurk, my interviews further elucidate the reactions from the IRB respondents and researchers. My findings also revealed several specific quality concerns and quality control mechanisms that are under-reported in academia so far.

To begin with, numerous respondents were concerned about cheaters -- the MTurk workers who would scribble or give their responses without thought. These researchers worried that cheaters could damage data quality in research because they did not pay attention to the questions and would provide useless data. One researcher and one IRB respondent, in particular, regarded the problem of cheaters on MTurk to be quite severe that could impair experimental studies on MTurk and the reputation of MTurk more broadly as a platform for academic research. P28, the researcher, said:

> There is a minority [of MTurk workers], actually not a minority, can do lots of damage because these people are not paying attention to get work done but to maximizing their payoff for the minimum amount of work. And of course, given that they don't do any work, you could have a small number of participants that can really destroy a lot of experiments by submitting a lot of junk. (P28)

This researcher is a pioneer and prolific scholar in introducing MTurk to academia concerning its characteristics, such as worker demographics. However, he told me that he is not using MTurk actively anymore. When I followed up for the reason, he gave me the answer above. His first thought was that there was only a minority of cheaters on MTurk, but he immediately corrected it and said the cheaters were actually not a minority. He also speculated the motive of these cheaters, which was to maximize their compensation with the least amount of effort. Hence, these people were not attentive to questions and provided useless data that could damage the quality and validity of many research experiments. To my knowledge, the extant scholarship about MTurk workers' demographics and characteristics has no estimation about the percentage of cheaters on MTurk, but contrary to P28's viewpoint, the other researchers in my interviews posited that cheaters on MTurk were only a minority. Additionally, I also posit that the percentage of cheaters in a task may be contingent on the nature of the task, such as whether it is academic or non-academic and whether the task is meaningful or attractive to MTurk workers. In fact, a prior study by Eickhoff and De Vries (2012) found that "novel tasks that involve creativity and abstract thinking" were more likely to induce legitimate responses (p. 133).

Similar to P28, an IRB director also held a rather critical view on the cheating problem on MTurk and attributed this phenomenon to some MTurk workers' purpose to earn more money. Furthermore, she cast her doubt on MTurk as a platform for academic research:

> I am not in favor of using MTurk, and I don't know whether the data are reliable. You have people there doing it for the money and the more they can do, the more money they can make, so are they answering the questions, or are they understanding the questions, uh, I cannot think of a correct word, maybe "truthfully" or "with thought?" They may just write down the answer they want. (P13)

P13 has served in both public and private universities and is currently the IRB director in a private university. She explicitly told me that she would not recommend the researchers in her university to use MTurk. Her primary concern, like that of researcher P28, was that some MTurk workers purported to maximize earnings as fast as they can. Therefore, it is hard but important to discern whether some MTurk workers have thought through the questions carefully, or they just randomly give responses to get over the task and earn the compensation as soon as possible. In fact, P21, another researcher, gave a concrete example of how some MTurk workers would answer questions quickly without thought. He told me that some MTurk workers would use an online system for automatic translation rather than do it by themselves so that they could finish the researcher's translation task faster. In the end, to prevent such behavior of simply copying and pasting sentences in an online translation system, P21 had to use images containing the sentences to be translated so that at least the MTurk workers would have to know how to type in those sentences in a foreign language.

Apart from the concern about the cheaters on MTurk, another ethical concern about data quality is fraud. Unlike cheating that is related to scribbling or giving answers without thought, fraud means that some MTurk workers would deceive their qualifications and eligibility to take the task and know the answers, but in fact, their answers were unreliable. For example, one researcher told me that they inquired into Indian MTurk workers' behaviors and habits and got to

know that some of them would pretend to be MTurk workers in the U.S. and answer questions about U.S. politics:

> I actually don't believe any of the survey studies that happen on crowd work platforms… we have done interviews with MTurk workers in India, and a lot of times, they would take on those tasks where it was completely about U.S. politics. What they would do is that they would imagine that they were like a soccer mom in the U.S. and think about what she would respond. So, I think it's important to understand those types of biases. (P29)

P29 told me that she had interviewed multiple MTurk workers in both the U.S. and India to understand their behaviors in using MTurk. She was doubtful on survey studies on MTurk because, in her words, there could be biases in survey data on MTurk because of the fraud, as her example illustrated. However, it should be noted that biases, such as the social desirability bias that P29 implicated, can exist in any surveys and are not unique on MTurk. Also, the example of some Indian MTurk workers' fraud that P29 recounted is not so much related to biases but to the data quality issue. In addition, it could be the researcher's mistake or negligence that they accidentally recruited MTurk workers from India to their surveys about U.S. politics as well. It is indeed possible that some MTurk workers in the other areas would hide their real IP addresses and use a fake one to take MTurk tasks in the U.S. (Dennis et al., 2019). In fact, such fraud entails more ethical concerns about the "bot crisis" or what Dennis et al. (2019) called the "Virtual Private Servers (VPS) crisis" on MTurk, which I will detail below.

The bot crisis was first discovered by a Psychology graduate student Max Hui bai. He posted a message to a psychology researchers' group on Facebook on 07/07/2018, asking whether anyone that used MTurk had noticed "*any quality drop*" in the last few weeks. It

generated many responses in academia and beyond. *Wired* magazine published an article covering this incident and reported several researchers' discovery along with a quote from Kristy Milland, who has been the administrator for a large MTurk workers forum, Turk Nation, and a MTurk worker herself, that "There are a dozen people I know of personally who run bots, and they get away with it…This has been going on since the beginning of Mechanical Turk, since forever;" furthermore, she attributed it to the unfair payment on MTurk: "Mechanical Turk workers have been treated really, really badly for 12 years, and so in some ways I see this as a point of resistance…If we were paid fairly on the platform, nobody would be risking their account this way" (Dreyfuss, 2018).

Only two researchers in my interviews commented on this bot crisis in relation to their ethical concerns about data quality in research on MTurk. One of them, P24, a prolific scholar in crowd work-based research, noted:

> In terms of data quality…There is also a bot crisis on MTurk. Last summer, a bunch of people noticed a spike of data quality issues on MTurk. Those data quality issues seemed to be centered in responses coming from virtual private servers, which are essentially virtualized instances of desktop machines that live in a data center or somewhere. So, some scientists referred to it as a bot crisis. My knowledge is that it seemed to be most non-Americans who tried to get access to Americans because here we pay better, and there is very specific evidence of that. (P24)

P24 ascribed a payment-related reason to it and speculated that some non-U.S. based MTurk workers might have created these bots to automate their responses and earn a better payment on MTurk. Beside P24, another researcher, P20, briefly mentioned about the bot problem and proposed to screen MTurk workers with a higher approval rate: "you have to set your approval

rate above 95% to avoid the bots or something like this." (P30) Except for P24 and P30, no other researchers or IRB respondents mentioned this crisis.

Given these data quality problems, how did researchers and IRB respondents deal with them? They proposed a set of strategies and suggestions. First of all, screening MTurk workers based on their approval rate is the most common approach among my respondents. An approval rate is a MTurk workers' ratio of responses that have been approved by previous requesters (both academic and non-academic). For example, P24, the same researcher who reported the bot crisis above, told me that he would use the approval ratio of 95% as the screening criteria to recruit MTurk workers:

> Usually, I also restrict the HIT approval ratio above 95% or above; occasionally, that's not effective at filtering out poor quality workers anymore. At the minimum, when you try to restrict your sample, you try to make it less worse, so anybody who is below that [ratio] is certainly terrible, and so we will kick them out. (P24)

Screening with a 95% approval rate is a common practice among academic requesters in conducting research on MTurk. P24 acknowledged that using such a ratio to guarantee research data quality was not perfect, but at least it could prevent the quality from being even worse. He also contended that MTurk workers with a lower approval ratio than 95% were certainly not recruited to his research projects.

However, not every researcher in my interviews agreed with setting a 95% approval rate as the benchmark for data quality control. For example, another researcher, P22, raised a concern about the legitimacy of using the arbitrary approval rate of 95% and argued that no researcher has tested its validity. He told me that he was empathetic with many researchers' rationale of

having such a standard because they wanted to ensure the data quality and not waste their research money, but ethically, he was against using an approval rate. He explained:

> Well, I don't think that the psychology pool has good criteria [for screening]; it's just "Are you at the university and are you taking a class, and if so, we are thinking that you are appropriate and we can generalize to anyone based on what you have done." We see so much data from 18-20-year-old without working experience; is that useful? I don't know, and I am concerned about it. (P22)

He did not address the ethical question about MTurk directly but compared it with the traditional psychology research approach. He implied that there was no such screening criterion as an approval rate as that on MTurk in conventional psychology research, but he heard little criticism of it in academia. However, now on MTurk, researchers started to set up a specific screening criterion to recruit research subjects, and so it seemed to be a double standard to him.

P22's opinions deserve further examination. Though his critique on the arbitrariness of using the 95% approval rate is reasonable, prior research has indicated that hiring MTurk workers with a high approval rate is related to good data quality (Peer et al., 2014) (this empirical study used the 95% approval rate or higher as the benchmark for "high-reputation workers"). Hence, even though the percentage of 95% may be arbitrary, recruiting MTurk workers with such an approval rate seems to indicate good quality from their provided data. Second, many researchers, including a few psychologists in my interviews, are not indifferent about the validity of using the psychology pool for research. In fact, it was the very reason for P22's concern about the usefulness of data from the psychology pool, which motivated P30, a psychologist that I interviewed, to move from the psychology pool to MTurk to collect data for his research.

Still, not all the researchers were stick to a high approval rate to ensure data quality. For example, one researcher, P27, would intentionally use a low approval rate to recruit new MTurk workers to his research. He explained:

> I don't know if I should be telling you this. This is my own little secret. I like new workers that have completed almost no tasks [be]cause they're going to try hard. They're just like "I better do a good job." And they're clicking all the stuff, and they pay attention at the end. They're grateful whereas somebody who's taken a thousand surveys there, and they're more likely to know what I'm assessing, and know what I'm trying to test for. They're more used to all the attention checks, whereas a beginner is more likely to try hard. So far, it's worked out. Hopefully, it keeps working out. (P27)

P27 assumed that new MTurk workers would try harder and be more attentive than seasoned MTurk workers because they were still unfamiliar with various norms and tips in using MTurk. Thus, he contended that the data quality from new MTurk workers might be better than that from seasoned MTurk workers. This researcher's recruitment criteria and data quality perspective are unique among my interviewees; they are also particular to my knowledge of crowd work-based research. Despite it, I should still highlight here that several IRB respondents and researchers had their concerns about research validity in using the novice and seasoned research subjects on MTurk. For example, "trying too hard" in novice MTurk workers may distort a researcher's target population and expected sample. However, since these concerns were about research validity and not quality per se, I will discuss them later in the research validity part.

Apart from screening MTurk workers based on an approval rate, inserting attention check questions (ACQs) is also a common practice and idea among my interviewees to ensure good data quality on MTurk. The basic assumption is that if some MTurk workers failed the ACQs, it

would indicate that they have not paid sufficient attention to the questions either. Thus, their data input would be in poor quality. Several IRB respondents also proposed that ACQs could serve as a proof to reject MTurk workers or respond to their complaints. It is noteworthy, however, that one IRB director warned that an ACQ might "overkill" some responsible MTurk workers if the question was inappropriately designed. He told me a case where he received a complaint from a MTurk worker saying that he had been attentive to the survey even though he failed an ACQ. The IRB director later found that the ACQs was in fact too complicated for a mere "attention check." He recounted:

> We had someone [MTurk workers] who contacted the IRB because one of the screening questions on one of our MTurk studies was basically a little geometry math problem. The guy [the researcher] who wrote it thought it was super easy, and it turned out not to be super easy, and in fact, some MTurk workers did pay good attention to fill out the study got that geometry problem wrong. And one [MTurk worker] was reported to MTurk as an, what's called, an "inattentive subject," and then he was saying that it threatened my income…and we looked at this screening question, and we were like, actually it was not an obvious math question [laugh]. Then we made the PI change that, and we also made the PI un-report a couple of people who missed that screening question. (P23)

From this case, we can see that the design of an ACQ is delicate and can falsely screen out candid MTurk workers. An ACQ perceived easy by a researcher may be complicated for some MTurk workers, and those who fail an ACQ may not necessarily be inattentive. Therefore, a few researchers in my interviews proposed more specific and refined "gold standards" than mere ACQs to ensure data quality on MTurk. For example, one researcher set a gold standard with two thresholds to check data quality. He would approve all the respondents that were above the upper

threshold and reject all the ones that were below the lower threshold. For the respondents in between, he would reject in proportion to their failed quality check questions. He explained his strategy as follows:

> I'll have two threshold numbers. Let's say if they [MTurk workers] are getting 85% of the gold standard items correct, and I just approve everything. If they're performing like bad checks [below a threshold], then I'll reject everything because it's clear that they're just random clicking. And if they're performing somewhere in between, then I might reject only those items that they've got the gold standard wrong to just give a signal to them like, "Hey, if you want to do this task, you have to pay closer attention". (P21)

However, to reject or not to reject, it is still a difficult question. Most researchers in my interviews expressed their struggle in dealing with rejecting certain MTurk workers. On the one side, they were aware that cheats would damage their data quality and waste their research funding. On the other side, they did not feel it worthwhile of their time, energy, or stress to negotiate with MTurk workers about their rejections. Hence, some researchers chose to pay MTurk workers unconditionally, even if their data quality was bad so as to void hassles. For example, one researcher said:

> There are some cases when it was obvious that some workers did not use our plugin, and we had ways of tracking whether they did use it, and they basically wanted to get paid even though they did not use it. So, we sent them an email and said, ok, you didn't use it, so we were not going to pay you. Some of these workers mailed us again, and we did end up paying them, but it was more to just remove the hassle and not wanting to fight. (P29)

P29 had evidence that some MTurk workers cheated and planned to reject those MTurk workers originally but under the pressure of the complaints, they changed their mind and paid those

cheating MTurk workers at last. In fact, sometimes MTurk workers' complaints of rejection can be aggressive even though their complaints are not well-grounded. Such complaints could give researchers and IRBs much pressure. For example, an IRB director described his experience dealing with MTurk workers' complaints:

> I would say 75% of the complaints [of rejection] were aggressive…one that I received last week, he [a MTurk worker] missed an attention check, and the investigators responded that he did not respond [to] this particular benchmark. And he said, "I did," and the investigator said that "well, I can check it for you, but because it's de-identified, I will need a portion of your IP address to look it up." The investigator was just asking for the first few digits of the IP address, and this person [MTurk worker] got very upset and said, "your investigator was asking for PII, which is against Amazon's terms of conditions…I did not know that XXX University was sponsoring this kind of survey to obtain data for free"…and that's what we typically get. (P11)

From his account, we can see what "hassle and fight" would be like in dealing with a MTurk worker's complaint of rejection. Even though a researcher has solid evidence of a MTurk worker's poor-quality data input and endeavor to double-check the situation, a MTurk worker could still be antagonistic and threatening to IRB staffs and researchers. To a certain extent, P11's experience can explain why many researchers in my interviews, such as P29, would intentionally avoid negotiating with some MTurk workers' complaints of rejection and pay them unconditionally.

Despite it, not all my interviewed researchers agreed with no rejection of any MTurk workers even if their data quality is poor. One researcher, P24, in particular, devised a blacklist strategy. He would pay all the respondents in his project on MTurk but would put those with

poor data quality in a blacklist and not recruit them again in his future research projects. When I followed up with what motivated him to design such a blacklist and what was the distinction between a blacklist and rejection, he explained that first, he did not want to "fight with a stranger on the internet for a quarter" and "it's just easier that I give you [the MTurk workers who cheated] a quarter and we don't need to talk with each other again." This argument is similar with many other researchers, but P24 also provided his ethical consideration in using a blacklist instead of direct rejection:

> Then, there is this ethical thing that is not totally straightforward. On the one hand, people shouldn't be punished for refusing to participate in research, and that is the consideration that IRBs often think about. If people don't complete a task, should you not pay them? If you don't pay them, then not only are they not receiving money, then there is also this reputational thing where they are potentially being punished by losing access to other future tasks because they are not being paid. So, there is a consequence there you need to sort of have concern about. (P24)

P24 first shared an ethical concern that IRBs usually had that rejection should not contradict research subjects' voluntary participation. Hence, if some MTurk workers chose to quit a study on the way, they shouldn't be rejected. Then, he posited that rejection could be tricky not only related to the target MTurk workers' loss of money but also concerning their reputational risk, i.e., the decrease of their approval rate that could render them lose the eligibility to participate in other future tasks on MTurk. In fact, several researchers and IRB respondents also raised the concern about such a reputational risk on MTurk workers when they talked about human subjects protection, but here, P24 linked it to data quality issues more closely.

P24 further related his strategy of blacklisting and concern about rejection to his advocacy for maintaining a reliable reputation system to ensure good data quality in research on MTurk. Though his advocacy was single in my interviews, I regard it as a valid concern that deserve more attention in crowd work-based research. Specifically, P24 worried that if researchers continued the "no rejection" practice, the usefulness and validity of the reputation system on MTurk—the approval rate of MTurk workers—would deteriorate and collapse:

> It's not like there is a strong norm [on text] that says people who perform low quality work should be rejected. And other requesters are relying on the signals of people's HIT approval ratio to make decisions who to include or not in their work. That's where I consider a little tricky… part of the reason that the HIT approval ratio no longer works so well for signaling poor quality workers is that many researchers just approve everybody, whether they are mandated to by their IRB or because they don't want to deal with this hassle at all. It's kind of a "tragedy of the commons" where everybody for their self-interest makes the decision for just paying people and leaving them alone. Then collectively, it undermines the reputational mechanism that makes MTurk work. (P24)

In his view, academic researchers are collectively and interdependently relying on a reputation system on MTurk to screen qualified MTurk workers from unqualified ones. However, if every academic researcher decides not to reject any MTurk workers, it will keep their approval rates always high even if they are cheating. Then, the approval rates would become meaningless to screen qualified MTurk workers and encourage good data input. Academic researchers' decision on not to reject any respondent may be due to their IRB's mandate or their reluctance to deal with the hassle of rejection. However, such a decision out of their self-interest could collectively

undermine the reputation system. P24 continued to explain his idea about the "Tragedy of the commons" on MTurk:

> Requesters are sending works signals what they supposed to do or not, and they [the MTurk workers] follow those signals, and they respond to incentives. To the extent that it makes the data quality decline, then it means that everybody ends up paying more for participants to produce and increase their data quality. The money they spend are[sic] often public money, that's a consideration too. If more and more grant dollar is getting burnt out paying people for data that get rejected, so that's all our concern. (P24)

The "Tragedy of the Commons," according to P24, refers to the dilemma that if every academic requester is sending a signal to MTurk workers that they will be incentivized regardless of their data quality, then MTurk workers' approval rate will no more be a deterrent or prevention for cheating. Then, academic requesters would have to increase their payment to motivate better responses (here, an example could be spending extra money as a bonus to reward those who provide good quality data). Consequently, each academic requester's individual behavior of not rejecting any MTurk workers would lead to a "common tragedy" that all of them would have to pay more and burn more of their research funding.

Ethical considerations in research validity

Apart from data quality, research validity is also a focus of research ethics. Research validity can be further divided into internal validity and external validity. Internal validity means that certain conditions are shown to lead to other conditions, and their relationships are not due to spurious factors in a specific research context (Kidder and Judd, 1986). External validity means a study's findings can be generalized to a larger research context and population (Kidder and Judd, 1986). Both IRB respondents and researchers in my interviews have raised various

concerns about internal validity and external validity in conducting research on MTurk, which echoed as well as extended the existing findings on the validity issues in this context (e.g., Thomas and Clifford, 2017; Berinsky et al., 2012; Paolacci et al., 2010).

In the first place, several researchers and IRB respondents raised their concerns about the internal validity of conducting academic research on MTurk. Their concerns focused on two main aspects: the non-naïvety of some MTurk workers and the diffusion of information among MTurk workers off the MTurk platform. Non-naïvety has been detected in experimental research on MTurk, where some MTurk workers would take the same task multiple times (Chandler et al., 2012). However, in my interviews, such a concern about the non-naïvety associated with duplicate responses was uncommon. Perhaps, it is because many researchers have already known how to screen out duplicates and selectively recruit MTurk workers (Peer et al., 2012).

In my interviews, respondents expressed their concerns about non-naïvety because they observed that many MTurk workers had been seasoned in answering various research questions because they had become familiar with various survey questions and experimental treatments. Hence, they worried that MTurk workers' non-naïvety in the sense that they were savvy survey takers, and that could form a predisposition or prior knowledge to research questions and experimental treatments. This, then can become confounding factors and make the effect relationship spurious. For example, one researcher told me that researchers should never run a "Prisoner's Dilemma" experiment on MTurk anymore because so many MTurk workers had taken it and got used to it:

> I know from social science work on MTurk that there's almost like a running joke. Like you can't run a prisoner's dilemma experiments on mechanical Turk because so many

> people have done that on MTurk. The results will be skewed from what the natural population would do. (P21)

Prisoner's Dilemma (PD) is a classic game-theory problem about cooperation and defect, which was developed by mathematician Merrill M. Flood and Melvin Dresher at RAND in 1950. Basically, it describes a dilemma that if two prisoners cooperate, they will have a good mutual outcome; if one prisoner chooses to betray and the other still chooses to cooperate, the one who betrays will have the best individual interest, and the other will have the worst individual consequence; if two prisoners choose to betray each other, they both will have a worse consequence than if they both cooperate; the two prisoners cannot communicate with each other before and during the prosecution, and it's a one-shot decision. Back in 2011, Horton et al. (2011) ran one of the first PD experiments on MTurk. Most recently, Capraro et al. (2020) conducted a Stag-Hunt Game (SHG) experiment on MTurk, which is derived from PD. Thus, PD and its derivative experiments have been tested on MTurk for over a decade now, thus leading P21 to speculate that many MTurk workers are familiar with it.

In addition, two researchers shared the concern of non-naïvety on MTurk but offered two opposing opinions on its effect on their particular research fields. One researcher perceived that non-naïvety was detrimental to his research in politics; the other researcher, by contrast, posited that non-naïvety could actually be beneficial to his research in privacy. To start with, P23, a political researcher, stated:

> Another thing on MTurk, which is to some degree why I stop using MTurk, is [that] people may risk paying too much attention because they know that you are evaluating their work. You give them a task, they do it, and you pay them according to whether they complete it. I know that MTurk workers are very worried about the approval rate, and

> they are very careful about that. That doesn't serve the kind of work that I do in political science, whereas in the real world, people are not paying close attention to politics, so it's not a very accurate measure of what we are capturing [using MTurk workers]. (P23)

He noted that MTurk workers might pay too much attention to his questions about politics because they cared about their responses that would be evaluated by researchers for payment. They also didn't want their responses to be rejected because they are conscious of their approval rate. However, such attentive and serious attitude and behavior toward political questions did not embody people's attitudes and reactions to politics in the real world, which are more indifferent and casual. Hence, this type of non-naïvety was not helpful for his research in politics and partly motivated him to stop using MTurk.

On the other hand, P18, a privacy researcher, explained why non-naïvety might actually be helpful for his privacy research design:

> I and the others have been arguing for a while, that it's totally OK to keep on using MTurk even in privacy studies, so even if a participant made a lie in a privacy study, because in my experience, when I compare the results I get from MTurk to the results that I get from other samples, the results are almost always more conservative than the results that I get from the other samples. In other words, more conservative means, because the subjects may be a little less naïve, if I try to use a certain treatment, an experimental treatment, it's HARDER [emphasized by the respondent] to produce a statistically significant effect on MTurk samples than on non-MTurk samples. (P18)

He argued that precisely because many MTurk workers were prone to be non-naïve and more accustomed to various surveys and experimental treatments, it was harder to design a novel

treatment and obtain a statistically significant effect on them than that on the general public. Hence, MTurk workers' non-naïvety is useful to prevent the Type I error in research.

Here I will use an example that I created to illustrate P18's viewpoint on why MTurk workers being more conservative to privacy treatments would be a good thing to present Type I error. Suppose I hypothesize that a free mailbox with targeted ads is more acceptable than a paid mailbox without targeted ads, and I conduct an experiment on MTurk. In the experiment group, I ask MTurk workers their acceptance rate of a free mailbox with targeted ads; in the control group, I ask MTurk workers their acceptance rate of a paid mailbox without targeted ads. Because MTurk workers are less naïve to various privacy treatments and more prone to display their privacy concerns than the general public, they are more conservative in accepting a free mailbox with targeted ads that could compromise their privacy. Hence, if I still find that MTurk workers' acceptance rate in the experiment group is significantly higher than that in the control group, I can conclude that a free mailbox with targeted ads will be even more acceptable than a paid mailbox without targeted ads in the general public. In this experiment, the null hypothesis is that people's acceptance rate to a free mailbox with targeted ads and a paid mailbox with targeted ads is the same. Because MTurk workers are more likely to display their privacy concerns, thus they are less likely to accept a free mailbox with targeted ads and more likely to accept a paid mailbox without targeted ads than the general public. Therefore, the null hypothesis is harder to reject, and the Type I error is harder to occur.

Apart from non-naïvety, diffusion of information is also a common consideration of internal validity in research on MTurk. Specifically, several IRB respondents and researchers noted that there are some online forums where MTurk workers would share their tips and

experiences of research, which could threaten the internal validity of research on MTurk. For example, one IRB director said:

> You must make sure that they [MTurk workers] are not talking to one another. You must admit that limitation as well. We all know that mechanical Turk workers have their own groups that they are on all day long as they're working, and they're telling each other about the surveys that they're taking. And that is a huge threat to internal validity. (P12)

Essentially, this IRB director was concerned that MTurk workers were diffusing the information about research projects they participated in MTurk with each other, which would make some MTurk workers predisposed or equipped with prior knowledge before they take these projects. Such predisposition and prior knowledge could be spurious factors that affect the internal validity of the research. Another researcher gave a more concrete explanation of this threat to internal validity:

> [L]et's say, it's a study that only wants people who smoke. And the study says, "are you a smoker." I am not a smoker, but I read it on the forum where everybody says, in order to proceed to the second part of the study, you have to say that you are a smoker. And I will say that I am a smoker because I want the money [of research compensation]. But if I was not in the forum and was just in the study for my own knowledge, then there would no incentive for me to do that. (P30)

Though P30's primary concern was with the diffusion of information, he also implied the undue influence of payment. Another researcher, P31, explained the payment effect on internal validity more explicitly:

> I do believe it creates a power dynamic where "I better say what this requester thinks or wants in order to get my payment. I get to do what I think the requester is asking for in

order to just get paid" whereas if you are just volunteering for research, you don't have that same power dynamic. (P31)

P31 perceived that some MTurk workers would intentionally curate their responses to a requester's desirability rather than out of their real thoughts. Hence, it would impair the internal validity of the research. Meanwhile, he mentioned about the nature of MTurk workers as different from research volunteers, which I have analyzed and discussed in the previous chapter about ethics in payment. I will further examine the related issues of "workers" vs. "volunteers" in the next chapter about ethics in human subjects protection.

As regards the external validity of research on MTurk, my respondents' concerns centered on the MTurk population's overall representativeness and the sampling challenges such as the paradoxical effect of payment and the lack of statistical power.

To start with, MTurk workers' overall representativeness is multiple IRB directors' and researchers' major concerns about the external validity of research on MTurk. A 2018 survey on MTurk demographics and dynamics indicated that MTurk workers were indeed a bit wealthier than the U.S. population on average (Difallah et al., 2018). Another 2016 survey on MTurk demographics comparing with 2012 American National Election Studies (ANES) data revealed that the MTurk population was younger than the American population but not significantly less religious (Levay et al., 2016). An IRB director's concern reflected these disparities between the MTurk population and the U.S. population more broadly:

I will say that I worry a little bit about the MTurk worker population not being a representative sample. If you are a political scientist, for example, the MTurk worker population is not a representative sample of American voters. I think MTurk workers are quite bimodal in age distribution. That's like all people are young or old, and there is not

> too much[sic] in between. They are better educated than average Americans; they are probably a little bit wealthier than average Americans. I just guess that, because 1/3 of medical participants does not have access to a computer, so these guys [MTurk workers] are a little bit better off than 1/3 of the medical population because they got access to the internet. And they have been surveyed, and they are substantially less religious than most Americans. I don't think studies done on MTurk can be presented as a representative sample of Americans. (P17)

Resonating with P17's worry about the representativeness of the MTurk population, several researchers and IRB respondents expressed their concerns about the generalizability of the data from MTurk workers, which is not only related to the MTurk population but also related to a few sampling challenges.

First, the payment could have a paradoxical effect on the sampling of MTurk workers. For example, some researchers remarked that an insufficient monetary incentive could discourage certain people from participating. Thus, the exclusion of these people who chose not to participate because of money could render the research sample unrepresentative. On the other hand, another researcher argued that if the payment was not sufficient or there was no monetary incentive, yet some MTurk workers still chose to participate, then those participants might not constitute the right sample either. He used political research as an example and explained:

> There is a tradeoff, and imagine we did not pay anybody, then the only people that would participate would be the people who already love taking political surveys, and that is not the group we want to know more about. We want to know more in fact about people who maybe are not interested in politics. So, I think if we want to be more representative, we are still going to be in the position to compensate people. (P23)

P23 was not alone, and from the IRB respondents' camp, P10 gave a backup and said:

> People are motivated by different things, especially in politics, and you will see that a handful of people in politics [participate] without pay because they are interested in the outcome or they want to push their agenda. Like to [be] against Trump, they will just do the survey to make sure that people listen to them. (P10)

P10 pointed out that some participants might be motivated not by money but by pushing for some political agenda. However, according to P23, these people may not be representative of the target population. Thus, the research findings would not be generalizable, and the external validity may be compromised.

Second, one researcher raised his concern about the statistical power in research with a MTurk sample. He regarded it as a more severe problem than the sample characteristics of MTurk population per se:

> It's hard to get enough [statistical] power on MTurk study to do something really big like among "civic Republicans" because there are fewer of them. The fact that the sample characteristics are different is not really a big deal for me, because I can use methods to adjust that, but the problem comes to power. If I have something that I want to run among Republicans and I want it to break it up, Republicans old and young, and if I only have 200 Republicans in my sample [on MTurk], then it's basically impossible to do that analysis. (P20)

P20 noted that it was hard to get a large enough sample from a specific population on MTurk, and if he needs to further divide a sample into more specific demographic groups, then running such an analysis would be impossible to obtain sufficient statistical power.

Finally, a researcher advised that it was difficult to maintain both external validity and internal validity in research on MTurk, and he would compromise the former for the latter. He expounded:

> It's very hard to achieve both internal and external validity at the same time in experiments, and any experimental design is an exercise in a paradox. And for me, the most important goal is internal validity, which means that as I usually don't say, when we publish a study saying, "our results demonstrate that every human being will act the following way." Rather, we say, "with this kind of sample size, and the sample of the population we use, we found these results and because we believe that we did not have any confound, we stand by our result." And then it's the broader goal and mission of science to be able to replicate, to generalize, or to invalidate the results that I obtain, or the others obtain, using different samples. (P18)

P18 based his argument above in the context of conducting research on MTurk. He also told me that he would usually constrain his sample in the U.S. only for better data quality and internal validity, even though it would be at the expense of external validity. He later stressed that because he could maintain good internal validity on MTurk, he is still conducting research on this platform. By contrast, several researchers in my interviews, such as P20, P23, P30, are no longer using MTurk to conduct research because of their concerns about external validity, i.e., their findings from MTurk workers might not generalize to larger or the other populations.

Document analysis of ethical issues in data

To cross-reference the findings from my interviews, I also analyzed the crowd work-based research guidelines as well as my interviewed researchers' publications in terms of their

ethical considerations in data quality and validity issues. Data quality issues and guidance are neither prevalent nor explicit among the research guidelines, but four guidelines mention it.

First, a guideline for MTurk from a public university recommends researchers employ certain data integrity/quality control practices: "Normal data integrity/quality control practices should be employed." (D14) However, it does not provide any detailed advice further or explain what "normal" refers to in research on MTurk. Second, the M-Truk Guide from Dr. Michael Buhrmester (D10) mentions the correlation between the approval rating and data quality. It surmises that screening MTurk workers with a higher approval rating may delay the incoming of responses but not necessarily the data quality:

> I've personally moved the approval rating around between 50-99 and at least in my experience, the higher rating requirement seems to slow down the flow of incoming submissions without affecting data quality, but I've done no formal test of this. (D10)

This guideline emphasizes that the author has not formally tested the relation between the approval rating and the data quality on MTurk. On the other hand, a research publication that I have collected has confirmed that high-reputation MTurk workers (with an approval rate above 95%) would provide high-quality data without the need of using ACQs:

> We concluded that sampling high-reputation workers can ensure high-quality data without having to resort to using ACQs, which may lead to selection bias if participants who fail ACQs are excluded post-hoc. (D17)

D17 further reminds that ACQs may introduce selection bias that could have a negative effect on research validity.

Third, a private guideline from another researcher remarks that foreign MTurk workers outside of the U.S. tend to skip questions and enter short responses to questions. This guideline speculates that it is because foreign MTurk workers are more motivated by money:

> MTurk seems to have caught on in some Asian countries. So, as a researcher, ask yourself if you would like to sample from these countries. In my experience, some foreign Turkers tend to complete surveys quicker and are more likely to skip questions that require typed short-answer responses. This may be because they are more motivated by money than are workers from other countries (although this is certainly an empirical question). (D4)

Finally, "Guidelines for Academic Requesters 2.0" suggests that researchers should provide "reasonable time estimates." Because if a HIT's time turns out to be longer than MTurk workers' expectation or estimation, some MTurk workers may speed their responses and compromise their data quality:

> Turkers calculate estimated earnings based on time estimates, and their target earnings inform their choice of HITs. If a HIT takes longer than estimated, Turkers may speed through it to keep it to the requester-provided estimate, hurting quality and damaging requester reputation. (D15)

On the other hand, all the publications that I have collected for document analysis suggest that data quality from MTurk is better than or at least as good as that in the other research venues. Hence, it is interesting to see the disparity between the researchers' expressed concerns in my interviews about the data quality and validity in research on MTurk versus their confidence in the data quality and validity in research on MTurk in their publications. Despite it, it must be noted that these researchers published these papers several years ago, and their

opinions in these publications then might not accurately reflect their current opinions about data quality on MTurk now. In fact, several researchers who advocated for MTurk in their publications told me that they are not using MTurk anymore. Hence, their attitudes toward the data quality on MTurk in their publications should only be considered as a reference to their current opinions in my interviews.

In terms of internal or external validity, no research guideline raises a concern on crowd work-based research. Only one guideline makes a remark on it, but from a different perspective from the IRB respondents and researchers in my interviews. It proposes that crowd work-based research has the merit or potential to escape an especially Western and educated sample, and hence, improves its validity comparing to traditional psychology research:

> University samples have served scientists well and remain a cornerstone of psychological research. However, many have recognized that the behavioral sciences rely heavily on undergraduate samples, with the potential consequence of limited external validity. W.E.I.R.D samples are disproportionately western, educated, industrialized, rich and democratic participants. Crowdsourcing services may allow researchers to move beyond the W.E.I.R.D. and access more diverse and representative samples. (D1)

Even though this guideline does not discuss the potential threats to the validity of crowd work-based research, its comments on the limitation of the traditional sampling strategy in psychology research previously resonate with several researchers', such as P22's and P30's critiques and opinions.

As regards the publications that I have collected for document analysis, several analytical and review papers have discussed the research validity issues on MTurk. These papers are highly cited, but they were published many years ago who MTurk started to draw attention in academia.

For example, one pioneering paper found that MTurk workers were more representative of the U.S. population in terms of political views than in-person convenience samples but less representative than subjects recruited from internet-based panels or national probability samples. This finding was reflected in some of my interviewed researchers' opinions. For instance, P20 and P23 told me that they were not using MTurk so much as before and now they prefer to use internet-based panels such as Survey Sampling International (SSI).

Another publication argued that, to some extent, MTurk could strengthen internal validity because MTurk workers could not talk with each other during an experiment and avoid the influence from the experimenters that might trigger a reactance bias. This paper was published at the dawn of MTurk, and the situations have already changed since then. For instance, the authors proposed that MTurk would prevent MTurk workers from cross-talking with each other. However, as several researchers noted in my interviews, it is not as such anymore because there could be information diffusion among MTurk workers in various online forums, which was not feasible at the time of this publication. Many MTurk workers may also be more aware nowadays that they are participating in experiments. Just like P21 noted in the interview that researchers should not run the "Prisoners Dilemma" on MTurk anymore because many MTurk workers had done it repeatedly.

Finally, it is noteworthy that among the six publications about human subjects research that had recruited MTurk workers, only one had considered validity and reliability issues. They asked MTurk workers to answer several validated scales to measure their individual differences and social desirability. The other five had no discussion or reflection on their research validity.

Discussion

Based on my findings of ethics in data issues in crowd work-based research so far, several questions deserve further reflection and discussion. First, where do these ethical considerations of data come from? Second, what are the differences in these ethical concerns about data quality and validity, if any, between IRB respondents and researchers, between IRB respondents from different institutions, as well as between researchers from different disciplines? Third, what are the implications from these findings for crowd work-based research in the future? I will discuss these questions below.

The origin of ethical issues in data quality and validity

To start with, I argue that the ethical issues in data quality and validity originate from three sources: (1) researchers' overreliance on the convenience sampling and fast data collection from MTurk; (2) MTurk workers' prioritization of earning money over ensuring data quality, and (3) researchers' negligence of disciplinary differences and extrapolation of research validity from one discipline to another.

First, since MTurk's launch, researchers have positioned it as a paradigm of crowd work and praised its convenience, speed, and ease in data collection for research. In parallel, a few pioneering publications' findings that the data collected from MTurk could maintain good quality and validity (e.g., Paolacci et al., 2010; Buhrmester et al., 2011; Berinsky et al., 2012) made MTurk even more alluring to researchers. The researchers and IRB respondents in my interviews were also quite aware of it. Meanwhile, these early papers have been cited over and over again and been treated as the foundation for assessing the data quality and validity in academic research conducted on MTurk.

However, many researchers may have neglected that the ecosystem of MTurk has been evolving and changing. For example, ten years ago, Paolacci et al. (2010) argued that MTurk could improve internal validity because there was no crosstalk among MTurk workers. However, this claim is not true anymore. There have been numerous online forums where MTurk workers could share their tips and experiences in doing research. Such information leaking and diffusion, therefore, could damage the internal validity of a research project on MTurk, as several of my interviewed researchers and IRB directors pointed out. From my document analysis and literature review, few researchers in HCI research on MTurk would discuss sample representativeness or generalizability issues. They would rather use the platform for "exploratory findings." Also, few researchers discussed the potential effect of social desirability bias on the internal validity of their research findings but put it in the research limitations. Based on my document analysis and literature review so far, only one publication that includes both human subjects research and discussion about research validity.

Second, I agree with multiple IRB respondents' view that for many MTurk workers, earning money is their priority and primary purpose for doing tasks on MTurk. It has been debated for a long time how significant money could impact MTurk workers' participation and contribution quality (e.g., Fest et al., 2019; Buhrmester et al., 2016; Kaufmann et al., 2011). However, I argue that at least for those MTurk workers who are primarily motivated by money, they might compromise their data quality for time efficiency. Furthermore, as researcher P24 worried, if researchers continue to avoid rejection and pay MTurk workers regardless of their data quality, the reputation system would not be effective to deter cheating anymore. For those MTurk workers who aim for earning easy and quick money, without any control mechanism of rejection, their data input might be of even worse quality than that from their current behaviors.

Finally, I argue that the nature and characteristics of MTurk may impose different effects on the research validity in different disciplines. As researcher P18's and P23's opposing opinions demonstrate, non-naïvety of MTurk workers may be detrimental to internal validity in political research, but it could be beneficial for internal validity in privacy research. Thus, it would be problematic to neglect such disciplinary differences regarding validity issues. However, to my best knowledge, there has been little discussion on such disciplinary differences in the scholarship of crowd work-based research. Researchers would assume that the research validity on MTurk from a publication in one discipline many years ago would be extrapolatable and hold in their disciplines and projects on MTurk now. Such a blind assumption and extrapolation would defect the validity of using MTurk for academic research more broadly and in the long term. Hence, I propose that academic researchers should take an evolutionary view in crowd work-based research. I will discuss this point further in the discussion chapter.

Comparisons of ethical considerations from different entities

In general, my interviewed IRB respondents and researchers shared lots of common concerns about data quality and validity in research on MTurk. They resonated with each other about the cheating issues, fraud, information diffusion, non-naïvety, and MTurk workers' representativeness. Researchers seemed to be more conscious of some particular data issues such as the bot crisis, the reputation system, and the nuances in internal validity on MTurk than their IRB counterparts, which may be due to the fact that researchers are more familiar with the research design and data analysis.

On the other hand, as aforementioned, there is a noticeable disciplinary difference between political research and privacy research when considering the effect of non-naïvety on internal validity. Meanwhile, in terms of my document analysis findings, the current research

guidelines have not offered guidance or advice on data quality or validity issues on MTurk and in crowd work-based research more broadly. Also, the human subjects research papers that I have collected seldom discuss data quality or validity issues. These may be seen as a limitation of ethical consideration in the current research guidelines and a disparity between researchers' articulated concerns and their written considerations in their publications.

Implications for crowd work-based research in the future

Based on my findings and discussion on the ethical issues in data, I propose three implications for crowd work-based research in the future. First, researchers should be conscious of their individual decision's impact on their research community collective. More specifically, in the context of MTurk, individually speaking, it is about a researcher's decision whether to reject a MTurk worker based on a certain standard of data quality. Collectively speaking, it is about maintaining a reliable reputation system on MTurk that would be beneficial to all the researchers to obtain good-quality data on this platform. Second, researchers, as well as IRBs, should be conscious of the disciplinary differences, such as that between privacy research and political research, when examining validity issues. Third, guidelines for crowd work-based research should provide more specific instructions about data quality and validity issues. Finally, researchers should contemplate and report validity in more detail in their human subjects research publications.

Summary

Data quality and research validity are the two main ethical concerns in data issues in crowd work-based research on MTurk. Ethical concerns in data quality issues include cheating, gold standards for quality control, and the maintenance of a reliable reputation system. Ethical concerns in research validity can be divided into internal and external validity concerns. The

former includes the threat of non-naïvety and information diffusion; the latter includes the representativeness of MTurk workers and various sampling challenges. I also reported my findings in disciplinary differences and document analysis. Finally, I discussed the origin of ethics in data issues in research on MTurk, compared different views and disparities, and proposed implications for future research.

CHAPTER 7 – ETHICS IN HUMAN SUBJECTS ISSUES

Chapter 7 introduces various ethical concerns and considerations related to human subjects in crowd work-based research. In a broad sense, human subjects protection is the center of ethical research. The Belmont Principles and AoIR IREs for internet researchers have been pivotal in providing the ethical considerations related to human subjects research. Numerous others studies in crowd work have discussed related issues, such as dehumanization of crowd workers (Deng et al., 2016), commodification of labor (Bergvall-Kåreborn and Howcroft, 2014), empowerment of crowd workers (Salehi et al., 2015), privacy issues on MTurk (Xia et al., 2017), and justice in treating MTurk workers (Irani and Silberman, 2013). Power is an important concept related to ethics in crowd work-based research because crowd workers and requesters are usually in an uneven relationship where crowd workers are subordinate (Irani and Silberman, 2013), and an uneven power relationship is a core concern of feminism ethics (Suomela et al., 2019) that AoIR advocates to be applied in internet research (Franzke et al., 2020).

In this dissertation work, many IRB respondents and researchers talked about their ethical considerations and practices in human subjects projection in crowd work-based research. Some ethical considerations and practices are embodied in the research guidelines and publications that I have gathered for document analysis. My findings contribute to the current scholarship of human subjects issues in crowd work-based research in three aspects. First, I inquired into IRB directors' and analysts' opinions about ethics in human subjects in crowd work-based research, which have not been previously been compared or synthesized with researchers' perspectives in prior ethics scholarship. Second, I identified additional ethical concerns related to power dynamics on MTurk, which have not been discussed extensively in the current scholarship. Finally, I inquired into IRB respondents' and researchers' philosophical stances between

utilitarianism and Kantianism in ethics, which has not been studied empirically in crowd work-based research. In the pages that follow, I will unpack these three findings.

Ethical concerns and practices in *respect for persons*

Respect for persons is the first principle in the Belmont Report and has been adapted in AoIR IREs for internet researchers as well. In my interviews, many IRB respondents' and researchers' ethical concerns and practices echoed this principle in aspects such as informed consent, ensuring voluntariness, and empowering research subjects. In addition, a few ethical concerns and practices were especially germane to crowd work-based research, and in particular, to the context of MTurk. For example, the concerns with dehumanization of crowd workers and the practices of empowering MTurk workers.

To start with, concerns about dehumanization is a salient theme emerged from my respondents' concerns with the Belmont principle of respect for persons. Both IRB respondents and researchers noted that MTurk workers may be debased as a data pool instead of individual human beings that have their personal traits and voices that deserve respect and comprehension. Below, I will enumerate such concerns about dehumanization.

First, several IRB respondents raised concerns about dehumanizing MTurk workers because they might be merely treated as subjects for collecting research data rather than individual human beings. For example, one IRB director commented:

> There are too many researchers who really think of their participants as "subjects." They really don't think them as "people." If you think of people as subjects, you dehumanize them. If you think of them as people, you have to think, "well, what is the implication to the people at the receiving end on what I am asking?" Not "should I write my consent in

> a graduate-level language, but should I write the consent in the language that people [on MTurk] can actually read, which is 8th grade if we are lucky". (P15)

P15's concern above is not unique to research conducted on MTurk but can apply to internet research more broadly, where researchers do not have a visual cue or direct communication with research subjects. On MTurk, the dehumanization is also related to whether researchers would prioritize their needs to get the data from MTurk workers over MTurk workers' needs for money. As P15 further explained the dehumanization with compensation on MTurk:

> "Am I worrying about how much money I am giving or am I thinking that I would not give these people till they are done because they can drop out?" That's not particularly thoughtful about who the people are on the other side but rather about my needs as a researcher. So, when you think about them as subjects, you dehumanize them. It creates this monolith group out there who I am not really attending to; when you think about them as people, you have to think what my obligations are to another human being. (P15)

She noted that if researchers did not consider the compensation to MTurk workers thoughtfully and prioritize their research needs of data from these "subjects" over MTurk workers' needs of money as people who have contributed their time and effort, then the researchers have dehumanized the MTurk workers. P15 further contended that if researchers only perceived "MTurk workers" as a singular group term, then they might have obfuscated and neglected each MTurk workers' individual needs, traits, values, and circumstances. In fact, P15 inferred the problem of deindividuation, where an individual's personal identity and traits are shifted and obfuscated into a group and that individual is referred to with a group or social level identity (Tajfel, 1981; Bell and Khoury, 2011). Deindividuation is also considered as a potential precondition of dehumanization (Haslam, 2006; Bell and Khoury, 2011). However, if researchers

treated MTurk workers as individual human beings, they would be more thoughtful about their obligations to MTurk workers individually rather than as a group. For example, they would consider that each MTurk worker's time and effort in research should still be respected and compensated even if they decide to quit during participation.

In my subsequent interviews after P15, I used P15's viewpoint of dehumanization as a prompt to ask the other respondents. Another IRB director commented that dehumanization was not as much of a research issue per se but a common phenomenon online. However, she still perceived it as a real issue for IRBs to consider because it is related to the "protection and regulatory" standpoint of IRB to protect research subjects. Meanwhile, a researcher also mentioned her concern with dehumanizing MTurk workers where some researchers might just regard them as cheap laborers rather than human beings. However, she also assumed that many researchers might not be intentional or conscious about dehumanizing MTurk workers, and some of them might just be student researchers who did not have adequate money and hoped to obtain cheap data from MTurk. Another researcher described to me how he would educate his "student researchers" to treat MTurk workers as genuine human beings. His advice can be seen as how to humanize MTurk workers:

> I tell my students, "You need to be right there. Be ready to answer immediately. And when you answer, enter as though you work at a fancy store; answer as though this person is your customer and they are genuinely valued persons. You want to keep a good relationship with them, [so] answer politely." So, I sometimes get them to like CC me to make sure that they are treating them well. (P19)

P19 analogized MTurk workers to customers that are "genuinely valued persons" and deserve building a good relationship with researchers. Furthermore, P19 implicated a peculiar

relationship on MTurk between researchers and research subjects, where researchers are the servers and MTurk workers are to be served. It seems to be quite opposite to the defined relationship on MTurk where requesters are to be served by MTurk workers or even to that in conventional research contexts where researchers' needs are to be served by research subjects.

On the other hand, a few researchers took more positive and proactive action toward humanizing MTurk workers. They designed tools and websites to empower MTurk workers, for example, by cultivating their professional skills for their future career and uniting their voices to be heard by the general public and media. These researchers are not only treating MTurk workers as a data source but also as the target population to materialize their research agenda to help and empower crowd workers in the broad context of the gig-economy. For example, a researcher told me her approach to empower MTurk workers from letting them sort out and choose well-paid tasks to cultivating their professional skills:

> We build tools to help crowd workers to earn higher wages by helping them identify which tasks are going to pay them more. And also, we help them to realize what types of tasks will be more likely to take up a lot of their time, so they understand how much time they are losing in terms of their work…[also] we teach crowd workers how to better put their crowd work on their CVs, and how to present themselves to people outside of the platform of the gig-economy so that they can find their jobs later on. (P29)

Another researcher, P21, shared a similar view with P29 and further envisioned a "career ladder" in a crowd work platform like MTurk, where crowd workers could develop their professional skills and careers beyond doing the micro tasks that are usually simple and underpaid.

Finally, a researcher proposed a practice of empowering MTurk workers to unite them and facilitate their communication while also exposing their real state and conditions to the public and media. She explained:

> What our project does is to let workers help each other and survive better…[and] provoke attention to the issues because when we started it, crowdsourcing was seen as "Oh, [crowd] workers are doing it in their free time for spare change' or like 'They are wearing pajamas all day." And nobody is really asking workers [about what they were doing]. At the same time, we don't want to yell at it as journalists or computer scientists. So, we think if we make something that is useful based on workers' own critique of the system [MTurk]. It's always meant to be a media object, and to be more than just functional but also to tell the world that this thing is going on. (P26)

She explained that her project aims to help MTurk workers to be more united and autonomous to voice their working conditions, needs and critiques of the MTurk system. She also emphasized that the orientation of their project was not merely to create a website for MTurk workers to communicate but more importantly to draw media attention and publicity to cultivate comprehension and mitigate prior bias or misunderstanding among the general public and media.

Another major theme in respecting for persons in research on MTurk is the considerations of informed consent. To ensure informed consent is an edict in the Belmont principle of respect for persons. Informed consent requires researchers to give subjects the opportunity to choose in their best capability what shall or shall not befall them when they participate in research (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). In internet research, informed consent is often hard to achieve in Big Data projects, and AoIR IRE 3.0 has proposed to acquire informed consent not

at the initial stage of a research project but at the data dissemination stage of a project, e.g., when data sharing or publications occur (Franzke et al., 2020). In this dissertation work, many respondents expressed their ethical thoughts about the informed consent related to the transparency of research purpose, compensation, data collection as well as the transparency of data dissemination. One researcher, in particular, proposed a *bona fide* consent form to let MTurk workers preview the task instead of a mere briefing of research. Furthermore, a few IRB respondents related the transparency of consent to the autonomy of choice for MTurk workers.

In the first place, several IRB respondents and researchers suggested that researchers should ensure transparency about their research in the consent form to MTurk workers. The transparency should include the research purpose, duration, payment rate at the initial state of research and include the data sharing or publication plan pertaining to the dissemination stage of research. The two quotes below manifest the transparency of consent concerning to the initial stage and dissemination stage of research respectively:

> I think transparency is key. They [the MTurk workers] should know about the research duration, payment, the purpose of collecting the data. (P10)
>
> We want the PIs [Principal Investigators] to at least be clear in the consent about whether they are going to keep things anonymous or not; if they are planning to share the data, who they will be sharing with, and how and where are they going to store the data. We want them to be very clear with their subjects what they are going to do with that data before subjects agree to participate. (P2)

P2 highlighted the importance of transparency about data dissemination because even though the data collection was anonymous, the potential risk of data sharing and leakage could still be concerning. She used herself as an example to explain it:

> I know myself as a research participant, and if someone is going to take my data, even if it was anonymous, but [they] saved to a desktop that many people would have access to, then you know, [I would not consent]. I don't give permission on my data to be used by lots of other people in a public place if they did not say that in the consent form. (P2)

She argued from a research participant's perspective that if she had not been informed about the potential risk of data sharing and leakage, she would not consent to participate because the mere assurance of anonymity was not fully transparent about the potential risks in data leakage.

Apart from revealing various information about research and data sharing plan and protection, several respondents also emphasized the importance of avoiding jargons for transparency and keep the information concise and clear. For example, a researcher, P22, noticed that "consent forms [on MTurk] are usually unclear and pretty jargon-focused" and an IRB director emphasized the importance of using simple language in a consent form:

> We want to make sure that the person who participates in the study has all the information written in a language that the person could realistically understand to make the decision whether to participate in the research. (P15)

Another IRB director also emphasized the clarity and concision of a consent form to MTurk workers because many of them would only skim it or even not read it at all.

To ensure the transparency about research and avoid jargonistic language, one researcher, P19, proposed a *bona fide* consent form for MTurk workers to preview the task instead of detailing the research purpose, payment, data collection, and data dissemination as many IRB respondents in my interviews advocated. He explained:

> I think that ethically it's important to show them the task itself, not the consent form when they preview because that is the real informed consent. I don't know whether IRB

> would care whether I do that or not, but I know from the worker standpoint they should be able to see what the task is before they agree to do it. That's the bona fide informed consent. I just honestly don't think that [conventional] consent form has much real value for most workers, and I don't think they're going to read it. And I don't think that they need to read it. Uh, of course, they have a right to do so. (P19)

P19 was a bit skeptical about the IRB's understanding of MTurk workers' standpoint. In fact, many IRB respondents and researchers in my interviews mentioned that the consent form template for research on MTurk would be the same for that in the other venues for social and behavioral research. P19 told me that he had established a good relationship with MTurk workers, and as a recap, P19 is also the researcher that asked his students to treat MTurk workers as genuinely valued persons and as customers. But what could be a flaw in a traditional consent form? P19 provided an example to illustrate it:

> Let's say you have a study where when they [MTurk workers] click on something, and it adds new questions. If you click yes, then it adds another three sub-questions. I always insist that my students show the questions that might be added in gray because a worker can know what might be added dynamically. It's not fair to show somebody a three-question survey and offer them, you know, 30 cents to do it. And then, once they get into the survey, it adds ten more questions. And that's the thing that the IRB is not thinking about at all. But I think about it. (P19)

In addition, besides the transparency of consent, several IRB respondents stressed the autonomy of choice for MTurk workers related to the transparency of consent. Ensuring MTurk workers' autonomy means to treat them as autonomous agents that can make considered opinions and choices for themselves and act upon them (National Commission for the Protection

of Human Subjects of Biomedical and Behavioral Research, 1979; Franzke et al., 2020). Several IRB respondents' views reflected this stance when they considered informed consent and transparency. For example, one IRB director stated:

> We want information there to the participants and giving them the opportunity to proactively say "Yes, I want to participate in this research project," or "No, I don't want to participate." We are hoping that the subjects that are considering online participation would not participate if they don't feel comfortable to participate in it. (P8)

P8 remarked that the transparency of consent is related to the autonomy of MTurk workers' decision-making. Another IRB respondent, P13, emphasized the transparency and autonomy are the core for the principle of respect for persons: "respect for persons, under the Belmont Report, means that the investigators are giving the correct information to the participants and giving them the autonomy to make a choice." Finally, it should be noted that many respondents would reiterate their ethical concerns about autonomy related to the undue influence of payment and exploitation when I asked about their thoughts about respect for persons. Since I have already narrated these concerns in Chapter 5, I will not repeat them here.

Ethical concerns and practices in *beneficence*

Beneficence is the second principle in the Belmont Report, which emphasizes the obligation to avoid or minimize harm to subjects and maximize their possible benefits (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Many respondents perceived that the risks and benefits in crowd work-based research were both minimal, and research on MTurk is exempt in IRB review. The IRB respondents were more considerate of the privacy risks than the researchers but contended that their policies and regulations were adequate for privacy protection.

Many IRB respondents and researchers contended that the research risks on MTurk were minimum because the studies were usually benign surveys. When I asked my IRB respondents about how they had categorized their reviewed MTurk studies, all of them said that the majority of these studies went to the exempt category. Only occasionally, a MTurk study would need an expedited review, and no MTurk study would ever require a full board review. In fact, several IRB directors mentioned that with the IRB regulation revised since Jan. 2019, almost all crowdsourcing research would fall into the exempt category.

Meanwhile, researchers also claimed that most of their research on MTurk was of minimum risks. Thus, their research on MTurk would be exempt from IRB review. For example, one researcher said:

> We only do work that is deemed a minimum risk. So, we will not ask questions that would induce emotional distress; we will not ask questions about illegal behaviors that if someone told us something in the survey, they could be endangering their reputation. So, in that respect, there is no additional harm that befalls someone participating in the study beyond the risks associated with everyday life. (P23)

Even though many IRB respondents and researchers perceived the risks in research on MTurk to be minimal, some pointed out the reputational risk on MTurk workers, even though a few participants already mentioned such risks when they contemplated the ethical issues in data, as I have reported in Chapter 6. For example, one researcher described and explained the significance and difficulty to for MTurk workers to get a good reputation and the pressure to maintain such a good reputation:

> It is interesting on-site in terms of your qualification because what you want to do is to build a whole bunch of HITs to get a good acceptance rate. If you have blown 100

> [HITs], nobody is going to want you to do anything. If you are below like 90% or 95% [approval rate], nobody is going to want you to do anything. So, there is definitely the tension that you want to do well, and you don't want to be rejected, and those rejections actually hurt you in terms of what you can do. (P22)

A good reputation is reflected by a MTurk worker's approval rate, and as a rule of thumb, it needs to be above 95%. One rejection could have a lasting blemish on a MTurk worker's approval rate, and thus, on their reputation. Another researcher, P24, explained it more clearly and posited that the severity of the reputational risks on MTurk workers was higher than the consequence of a low payment:

> The more complicated issue that I think IRB is ignoring it or only indirectly addressing it is the issue of reputation. I think the rejection is the more important consequence than not getting the 50 cents or whatever. It is because it stays at your [a MTurk worker's] record forever, and there is no mechanism to make your old negative records go away. Let's say, like, you know, somebody has completed 100 HITs, and they had been rejected six times. I mean that shuts off pretty much any social science study for them, and it would take an unknown amount of work to push their reputation back above the threshold, and they would have to deal with the worst-possible requesters. (P24)

In his view, many IRBs may not have been sufficiently concerned or conscious about the reputational risks due to a rejection. A rejection will permanently stay at a MTurk worker's approval record and severely impair a MTurk worker's eligibility to take future academic tasks. A MTurk worker will have to choose tasks that do not require a high approval rate, which are more likely created by unreliable requesters. Prior research has found that many MTurk workers had encountered bad requesters that published phishing, scam, or spamming tasks on MTurk

(Xia et al., 2017). These tasks usually don't have any requirement for a MTurk worker's approval rate so that they could recruit (and affect) as many people as possible. Well-paid and reliable tasks, on the other hand, often require a high approval rate or even a master worker qualification. Hence, MTurk workers with low approval rates are more vulnerable to be victims of the "worst-possible" requesters.

Contrary to P24's surmise that the IRB was not conscious of the reputational risks on MTurk workers, the IRB respondents in my interviews were quite aware of such risks. For example, an IRB director said:

> I know that MTurk workers are also very concerned about their reputational scores and things like that. If we get complaints in a study from MTurk workers, oftentimes those complaints surround not necessarily the information that was collected, or "I didn't like the study," but [surround] their reputational scores that they keep on there to mark them as legitimate workers to actually perform tasks well. (P14)

In fact, several IRB respondents reminded me that junior researchers might not be familiar with the reputational risks on MTurk workers. There seems to be an information asymmetry between researchers and IRBs about each other's understanding of the reputational risks on MTurk.

Besides reputational risks, all the IRB respondents and researchers regarded the research projects that they had reviewed or conducted as with minimum risks. However, considerations in privacy and data confidentiality issues is also a main theme related to research risks. Overall, IRB respondents seemed to be more conscious of them than the researchers.

First, numerous IRB respondents pointed out that MTurk was not anonymous, and MTurk workers' identity could be deanonymized by parsing their worker ID with the link to their Amazon account. One IRB director explained:

> Well, the MTurk identification is linked to the Amazon account. And people are putting lots of information out there on their Amazon account, like their names, as well as other information about themselves. And so, it's very easy to connect the two and figure out who is participating in the research, and so it cannot be considered as anonymous. (P3)

Acknowledging that MTurk is not anonymous, several IRB directors told me that they would suggest researchers not to collect MTurk workers' IDs or the IP addresses. If they did, they would ask the researchers to delete MTurk workers' IDs and IP addresses as soon as possible or store them separately. Additionally, some IRB directors would recommend or even require researchers to use Qualtrics for data collection, which they perceived to be a more secure platform because it does not collect identifiable information. For example, one IRB director stated:

> We have them answered on Qualtrics because Qualtrics has a feature where they won't even collect the IP address, and so their data are completely anonymous. So, in most instances, unless the researchers can justify collecting any identifiers, we want the research to be done in an anonymous manner. I would REQUIRE [her emphasis] researchers to use Qualtrics to collect data from MTurk workers because Qualtrics does not collect identifiers. (P13)

Second, from the researchers' side, even though a few researchers mentioned privacy and data confidentiality issues on MTurk, they did not perceive them to be serious enough to deserve extra or special protective measures. Some researchers were aware of various privacy risks on MTurk, but he did not perceive the privacy risks to be unique in this context. Some researchers did not favor the privacy or anonymity argument on MTurk at all. For example, one researcher

argued that MTurk workers were like employees, so they are not supposed to be anonymous in a professional environment and their "employers" have the right to know who they are:

> I mean many people are putting their work information online and advertising their services. So, I am actually quite surprised with the whole obsession with anonymity. In a professional environment, what does it mean that you want to get anonymous? If you want to do work for Google, Microsoft, or for the government, your employer knows who you are, like your first name, last name, SSN, and they know a lot of things about you. I don't quite buy a privacy argument in a workplace. (P18)

Another researcher perceived that the various privacy risks in research on MTurk were not ethical issues per se but technical issues. Since IRB was not designed as a technical institution, he further argued that IRB is not well-structured to deal with privacy issues such as the data identification. He explained:

> If I am telling you that your answer is going to be anonymous, and I say it in good faith. And then, it turns out that someone can figure it out of your data, and that's a technical issue. IRB is not designed to be a technical [institution]. IRB is designed to make sure what principles should researchers operate under. And so, as a researcher, if I say, "I am ensuring anonymity." And it turns out that I am not, am I violating the principles of IRB where I said that I will [sic] not violate anonymity? I would say no, because I did not know that this could be something that could happen. (P20)

P20 defended IRB for its role in supervising research and dealing with "technical issues" related to privacy. He argued that IRB is a compliance committee, not a technical committee, and hence, IRB is not well suited to deal with privacy compromise due to technical problems. Also, P20 implicated a deontological ethical stance as opposed to a consequentialist stance, where he would

anchor the ethical value of an act, i.e., privacy protection of MTurk workers, on its motive rather than on its consequence.

Finally, a few researchers had a (false) perception that MTurk was anonymous, and therefore, data security and privacy issues were easier to deal with than that in the other research venues. For example, one researcher claimed:

> People may answer more honestly on an anonymous survey. That's a quite positive outcome if you don't need to know who these people are. And also, of cause, it makes data security and privacy issues much easier to deal with. You don't need to worry so much about the de-identification of data. (P30)

As can be seen from several IRB respondents' and researchers' quotes above, P30's assumption was wrong. De-identification of data is a definite privacy and data confidentiality risk on MTurk, which has been revealed by prior research as well (Lease et al., 2013; Xia et al., 2017)

As I have discussed in Chapter 5, most IRB respondents did not perceive compensation as a research benefit because it countered the IRB policy that research benefit cannot be monetary. However, some researchers and a few IRB respondents held that compensation was a benefit in reality because it was what most MTurk workers were actually looking for and rewarded from participating in research. Related to the theme about beneficence, my interviewed IRB respondents and researchers overwhelmingly focused on describing and preventing the potential risks in research on MTurk. In terms of benefits, however, either they have considered them associated with ethics in payment issues, or they regarded that there was little or no benefit at all.

Additionally, some IRB respondents and researchers posited that beneficence was about assessing the ratio of the research benefit versus the risk. Since both the research risks and

benefits on MTurk were nominal, the beneficence was acceptable. Meanwhile, several IRB respondents suggested that the focus of assessing the risk and benefit ratio in research on MTurk was to minimize risks rather than to analyze benefits. Finally, one IRB director directly linked beneficence to the assessment of compensation, but she focused on the equitable level of compensation rather than the amount of it:

> In terms of beneficence, you know, I would say that compensation levels, while not necessarily the only piece of the puzzle, making sure you are having equitable and fair compensation as a way that beneficence is considered with your research study. (P4)

Ethical concerns and practices in *justice*

Compared with my interviewees' ethical considerations and practices in respect for persons and beneficence, their responses to the third Belmont Principle, i.e., justice, were relatively sparse. However, there were still several themes about the distribution of research benefits and sampling biases. First, it could be unjust to conduct research on MTurk if MTurk workers is not the target population to receive the research benefits. Second, it could be unjust to sample a certain group of MTurk workers while ignoring the others. Third, it could be unjust conduct certain types of research on MTurk because MTurk workers is a vulnerable population. I will enumerate these themes in more detail below.

To start with, an IRB director argued that if MTurk workers could not represent a population that researchers aimed to target, yet the research could have benefits for that population, then conducting research on MTurk would be unjust to that population because they were not well represented in the research. He explained:

> I would imagine there are certain social economic groups in certain populations that are part of the MTurk system. So, if you are conducting research that has a benefit to a

> certain race or demographic, by studying it on MTurk, you may not be doing that specific population justice…The argument about justice is similar between college students and MTurk workers because it's essentially that you are more likely to find certain demographics or social economic groups that attend college, and there is a good argument there: are we doing populations that don't necessarily attend college? (P7)

Another IRB respondent, P4, developed from P7's viewpoint and argued that it was a conundrum to conduct research on MTurk, and the researchers had to balance between their research benefit to a specific population or the general "societal benefit" from their research:

> It is a conundrum, especially for an online platform. You are at the mercy of whoever is signing up to participate in that platform and does the potential loss of beneficence by using that platform outweigh the potential benefit of the research. While maybe using a platform like MTurk, you have limited scope of who you are recruiting from because there is a certain likelihood that you may not be targeting your audiences exactly for who you are recruiting from on MTurk. Because people are choosing you. You can screen for eligibility, but ultimately people are self-selecting by saying that "I'm interested in the research." So, it's ultimately a question of which is better: the research benefit for that targeted population or the societal benefit from your research. (P4)

The "conundrum," as referred by P4, could be interpreted as twofold. First, there is a conundrum that, not only researchers were selecting subjects based on their eligibility, but also subjects were choosing researchers and selecting themselves based on their willingness to participate. Hence, to some degree, researchers were "at the mercy" of subjects' choice and self-selection, hoping that the subjects who participate would be a representative sample of their target research population. However, the subjects that participated in the research might not always be

representative of the target population. Second, there is a conundrum between how researchers would frame their research benefit. If they framed it as pertaining to a specific population, then, they might meet with an injustice situation described by P7 above. If they framed their research benefit as a generic "societal benefit," according to P4, they might escape the injustice to a specific population.

Second, some other respondents associated their ethical concerns about justice to the potential sampling biases on MTurk. They argued that even if researchers framed their research benefits as societal, there could still be an injustice issue. For example, an IRB director, P15, argued that we know little about the MTurk population. Thus, our sample could be biased, and research findings could be unjust if they are generalized toward the general population or apply to other populations. She started her argument with an analogy to justice in medical research where drugs or medical devices were preponderantly tested on white men. Thus, we knew very little about their effects on the other populations, and when we used these drugs and devices on them, the results could be problematic and even dangerous. Then, P15 extrapolated her argument to the research on MTurk:

> The same thing would be true with MTurk. We are only generating understandings based on a population we don't really understand. I would like to ask that in both directions: do we have an over-representation of some populations, and my guess is yes; do we have an under-representation of some populations, and my guess is yes. So how insightful the data is [from MTurk workers] even though we have a tendency to call it "generalizable" because we have got a large enough sample size [of MTurk workers]? (P15)

P15 speculated that MTurk workers could either under-represent certain populations or over-represent certain populations. Hence, for either of these populations, it would be unjust to them if

we claimed that our research findings with MTurk workers were generalizable because our sampling had been intrinsically biased against these populations.

Finally, an IRB director posited that certain research could be unjust to the MTurk population because they were vulnerable and less privileged than other people. It is like piloting a risky surgery on prisoners because normal people who are more privileged than prisoners would not take it. She explained:

> We do have to remember that MTurk workers don't have all the benefits that people in the other professions would have. I don't know how we can protect them, though. The only thing they can do is that they can choose not to do the research. (P13)

She reminded researchers to be conscious that MTurk workers were less privileged and vulnerable than certain populations and implied that certain research might not be suitable to be conducted with MTurk workers only based on the assumption that they might be more willing to participate than the people off the MTurk platform.

Ethical choices between utilitarianism and Kantianism

Apart from inquiring into the IRB respondents' and researchers' empirical concerns and practices about ethics, I also asked them a metaphysical question to follow up my questions about the Belmont principles: How do they perceive and choose between two philosophical stances of ethics behind the Belmont report, i.e., utilitarianism and Kantianism, in research on MTurk? As I have reviewed in the literature, these two philosophical stances of ethics are opposed to each other, and I regard them as underlying various ethical principles, considerations, and practices. For instance, respecting and ensuring research subjects' autonomy in making a decision to participate in and withdraw from research embodies the Kantian thought of respecting an individual human being's rationality and freedom, i.e., their freedom from coercion

or manipulation. On the other hand, minimizing risks and maximizing benefits to secure research subjects' beneficence reflects a utilitarian consideration of maximizing the ratio of happiness over pain. Hence, I am curious on how my respondents would consider these two fundamental and popular stances of ethics concretely in the context of research on MTurk.

To start with, some researchers perceived that utilitarianism was more realistic than Kantianism in considering ethical issues on MTurk. For example, one researcher claimed:

> I think for something like the IRB, realistically that [utilitarianism] is the only way they can afford [apply and support]. How can we treat MTurk workers as ends in themselves and as moral agents? I mean you could say we are [Kantians] because I'm saying "You [MTurk workers] got two options: take my survey or don't. I don't care if you don't, somebody else will." So, we are treating them kind of as moral agents, but at the same time, they're also a means to our end, which Kant would say "that's not good." (P27)

P27's argument above deserves some elaboration. He argued that even though on the surface, researchers seemed to implement Kantian ethics by granting MTurk workers with the autonomy to participate in their research, they were still treating them as means to their end, which was to obtain research data from MTurk workers. Hence, Kantianism is not fully operated in this context since Kant requires us to treat other people not only as means but also ends in themselves. Hence, according to P27, utilitarianism is more realistic than Kantianism for IRB because research benefits and risks are more concrete to assess than evaluating whether collecting data from a MTurk worker is treating them only as a means for research.

Another researcher agreed that utilitarianism made more sense on MTurk; however, he posited that some researchers might have exaggerated their research benefits to justify that the benefits outweigh their research risks. He contended:

> I think utilitarianism makes sense in terms of what's going on [on MTurk]. I do think that it includes a lot of miscalculation of what the value is for the worker and for the benefit. We are supposed to look at the benefits over the costs, but there is not a lot of benefit beyond the money. I think some research studies pretend the research being published is giving off [sic] the community value in some way, but that is certainly not connected to the workers. (P22)

P22 pointed out that the research value and benefits were often miscalculated and exaggerated. Hence, the utilitarian ethics to maximizing benefits over risks masquerades in research on MTurk but in fact is not materialized or maintained because the benefits to MTurk workers are minimal, if not accounting for the monetary compensation, and are often disconnected to them even if the research gets published.

By contrast, some researchers still believed that there was a Kantian consideration in certain research on MTurk, especially when we consider the undue influence of payment. One researcher argued that the informed consent is fundamentally Kantian, where research subjects are treated as autonomous and rational individuals that can make decisions themselves. However, he then noted that Kantianism might become tricky when we conduct research on a "market" such as MTurk:

> Where it gets tricky is that any time you are involved in a market, in which you pay people to do things, you are under a potential of coercion. We imagine that at that world [a marketplace], they [e.g., MTurk workers] are choosing to do it, but we as an IRB or we as researchers still want to put us in a position where we are not being unduly coercive. I think that's where the important point is: If I was going to ask you to do something really dangerous, it would never be OK because I offer you lots of money. Even that would be

> respecting your autonomy, we think that's probably too utilitarian – you would choose it, but we think it's a bad idea. (P23)

He argued that some people might autonomously choose to participate in a dangerous research project because they wanted to receive lots of money. As such, is it right to pay these people a lot of money and recruit them in respecting or abiding by their autonomous choice? P23 said no because it would be too utilitarian by outweighing the present monetary benefits to the research subjects over the potential research risks to them. Thus, in his view, when conducting research on a market like MTurk, Kantianism might not always work, while too much utilitarianism should also be avoided.

Finally, an IRB director posited that the choice between utilitarianism and Kantianism is contingent on the level of research risks. First, he argued utilitarianism applied in research on MTurk because such research usually involves minimal risks and "usually the social benefit of getting this study done outweighs this tiny amount of risk." However, he then argued that, for research that requires full board review, which was extremely rare for a MTurk study, Kantianism is prioritized:

> But the stuff that the full IRB committee sees is above minimum risk, and often involves quite substantial risks to people. In that case, I would say that we are very Kantian. If you are talking about these MTurk studies that are below minimum risks, there is not much concern about the safety about the subjects. But if you talk about the studies like interviewing people in a refugee camp outside of Iraq who had been coerced as ISIS tax collectors, you know if that data were leaked in some way, then it might prevent those people from ever been able to immigrate. (P17)

He concluded his argument by pointing out a systematic difference in the IRB reviews when considering the ethical choice between utilitarianism and Kantianism: when research involves minimal risks, such as that on MTurk, a utilitarian thought of getting this study done outweighs MTurk workers' risks was prevalent and applicable; when research involves a high level of risks, however, a Kantian thought of protecting an individual research subject's autonomy and safety should be prioritized.

Document analysis of ethical issues in human subjects

To cross-reference the findings from my interview data, I also analyzed the crowd work-based research guidelines and my interviewed researchers' publications in terms of their ethical considerations in human subjects.

In the first place, several research publications that I have collected for document analysis raised concerns about respecting and empowering MTurk workers, which are resonant with some interviewees' opinions. For example, one academic publication proposes that a fair payment is a sign of respect for MTurk workers and the lack of a rating system on requesters is a sign of exploitation on MTurk workers. Corresponding to the issue of fair payment, one paper has explored a fair wage to MTurk workers and enable them to choose the tasks that have high hourly wages on MTurk. Another publication has introduced a system to let workers voice and share their experiences with each other and to evaluate requesters as a means to balance the power dynamics between MTurk workers and requesters.

On the other hand, not many research guidelines or templates for MTurk have emphasized empowerment. Only one guideline "Guidelines for Academic Requesters" from an academic community highlights respect and dignity by quoting a MTurk worker's verbatim:

> Treat your workers with respect and dignity. Workers are not numbers and statistics. Workers are not lab rats. Workers are people and should be treated with respect. - turker 'T'. (D15)

In terms of ensuring voluntary participation, many research guidelines and templates for MTurk research have included the clause regarding voluntary participation so as to ensure MTurk workers' autonomy to participate or quit the research. For example, one informed consent template from a public university, states:

> Your participation in this research is voluntary. If you decide to participate in the study, you may withdraw your consent and stop participation at any time without penalty. (D5)

Most consent form templates that I collected used a similar language as D5, and even though they all claim to ensure voluntary research participation and withdrawal, their specificities regarding payment and data policies vary. Some consent templates highlight that MTurk workers could withdraw from the research without penalty, but they do not specify whether MTurk workers who withdraw would still be paid in full or partially or would not be paid. Some are more specific about the policy about partial and full payment in relation to the incompletion and completion of a task. However, one guideline from a private university, D6, is a bit particular. It describes MTurk as a "voluntary place to earn money" for MTurk workers, and thus, the same ethical rules in traditional academic research or in a job don't apply. It states:

> MTurk is a voluntary place to earn money. The investment on the part of workers to get started on the site is extremely low, and they are free to come and go as they please. Basically, or so the argument goes, MTurk is not like a regular job, and therefore the same ethical rules don't apply. (D6)

In fact, this guideline reveals the paradox between a "voluntary place" for researchers and a "place to earn money" for MTurk workers. This is a paradox has led to various ethical issues and confusion. I will discuss it in more detail in the Discussion Chapter. The research publications that I have collected for document analysis don't have a special highlight on voluntary participation.

As regards research benefits and risks, several research guidelines and consent forms for MTurk explicitly claim that the research benefits and risks on MTurk research are none or minimal. The consent form templates would state that there is no benefit, or some state that there is a learning or societal benefit, but no template positions monetary incentive as a benefit to MTurk workers. On the other hand, the templates narrate research risks quite similarly to those on an internet study, and no template mentions reputational risks. As an example, a consent from template from a private university, D6, lists the risks and benefits in research on MTurk as follows:

> RISKS and BENFITS: The risks to your participation in this online study are those associated with basic computer tasks, including boredom, fatigue, mild stress, or breach of confidentiality. The only benefit to you is the learning experience from participating in a research study. The benefit to society is the contribution to scientific knowledge. (D6)

Only two research guidelines give some considerations and recommendations on rejection and reputation issues on MTurk. One guideline from a private university introduces the consequence of rejection on worker's eligibility and recommends not to reject workers but accept the HITs and throw out bad data. Specifically, it suggests:

HIT rejection case 1: Task was a 50-minute survey that paid $5; HIT rejected due to not checking a box on consent form; Three similar complaints about this study. I advised researcher to accept all HITs, pay the subjects

HIT rejection case 2: Photo/text matching task; Researcher believed subject wrote a computer script to complete the task quickly many times; Subject began harassing researcher and the IRB. I advised researcher to (1) Accept the HITs, (2) Pay the subject, (3) Throw out their data. (D2)

It is noteworthy that this guideline advises researchers not to reject MTurk workers at all and pay them unconditionally. Though this advice could avoid complaints or damage to workers' reputation, it might undermine the whole reputation system on MTurk, as researcher P24 described and worried in the chapter about ethics in data issues. The other guideline also reminds researchers about the consequence of rejections. However, unlike D2, it does not advise researchers to avoid rejection but recommends researchers to use rejection "only as a last resort." I will discuss such policies about rejection in more detail in the Discussion Chapter.

In terms of privacy and data confidentiality protection, the research guidelines and templates from IRBs all seem to be quite conscious about the fact that MTurk is not anonymous. Hence, they have provided various related suggestions. One outstanding suggestion is guideline D8's recommendation for researchers to differentiate between internal HITs and external HITs on MTurk. Internal HITs are based on a template provided by MTurk that would directly run on Amazon (Mason and Suri, 2011). External HITs are created outside of MTurk on platforms such as Qualtrics. In this case, MTurk is only used as a recruitment tool where a link to the external HIT is embedded in it, and D8 suggests that external HITs are better for privacy and data confidentiality protection.

Another research guideline is also worth highlighting because it is the only one that provides some advice for researchers' privacy protection on MTurk. It says:

> What about my [researcher's] privacy? Turkers who want to know can often figure out much of this information for an academic requester who doesn't provide it; however, this takes workers' time and effort, and burns their goodwill. (D15)

Neither D8's suggestion to differentiate between external and internal HITs nor D15's advice on protecting researchers' privacy had appeared in my interviewees' considerations, which could also reflect a kind of knowledge gap or information asymmetry between the existing research guidelines and IRBs as well as researchers.

Finally, justice related issues are not referenced in either the research guidelines and templates or the research publications that I have collected for document analysis. Only one guideline touches on it. It describes the sampling bias in crowdsourcing research, which relates more to a consideration of research validity.

Discussion

Based on my findings of ethics in human subjects issues in crowd work-based research so far, several questions deserve further reflection and discussion. First, where do these ethical considerations of human subjects issues come from? Second, what are the differences in these ethical concerns, if any, between IRB respondents and researchers, between IRB respondents from different institutions, as well as between researchers from different disciplines? Third, what are the implications from these findings for crowd work-based research in the future? I will discuss these questions below.

The origin of ethical issues in human subjects issues

The origin of ethical issues in human subjects in crowd work-based research on MTurk can be primarily attributed to two tensions. The first tension is between the dehumanization nature of Amazon and the expected autonomy of MTurk workers as individual and rational agents in research. The second tension is between perceiving MTurk research as only with minimal risks and benefits and excluding the effects of reputational risks and monetary benefits. I will explain these tensions in detail below.

First, there is a tension between the dehumanization of Amazon and the expectation of autonomy in research on MTurk, which would induce various ethical issues pertaining to respect for persons. Dehumanization of crowd workers could be seen as originating from the launch of MTurk when Jeff Bezos announced that "You've heard of software-as-a-service. Now, this is human-as-a-service" (Bezos, J. 2006). Since then, crowd workers are perceived to be treated as a "commodity" (Aloisi, 2015; Bergvall-Kåreborn et al. 2014) or as a "service" or merely an "API Call" (Irani and Silberman 2013; Silberman et al. 2010). I argue that dehumanization is engendered by the nature and purpose of Amazon and not so much by the requesters who use the service. Amazon treats crowd workers as independent contractors and "services" and provides no payment or data protection to them or any cues of who they are. Indeed, requesters, to some degree, are also dehumanized by Amazon because MTurk workers have little knowledge about who the requesters are, and the relationship between them is usually only bonded by data collection and monetary compensation. Hence, essentially, dehumanization is actuated by Amazon rather than by requesters.

Meanwhile, however, IRBs and researchers expect MTurk workers to be autonomous individuals on MTurk, who have the rationality and the right to make decisions without coercion

whether to participate in or withdraw from a study. I argue that this expectation is not only difficult to materialize but also incompatible with the dehumanizing nature of Amazon. Unlike traditional psychology pools or online survey panels where participants have a clear understanding of themselves as research subjects in those research venues, MTurk workers may not have such an awareness because they are dealing with both academic and non-academic tasks on MTurk and would have to switch between "research subjects" and "workers" constantly. Hence, it is somewhat unrealistic for IRBs to expect MTurk workers to always behave in a researcher subject's mode and rationalize before participating in a study about the influence of payment and transparency of research. Also, due to Amazon's dehumanization of MTurk workers, IRBs and researchers can never be sure about the motivations of those MTurk workers who participate in their study and how much such motivations are influenced by payment versus research. In reverse, due to Amazon's dehumanization of requesters, MTurk workers may perceive researchers only as data collectors like the other requesters on MTurk and ignore the importance and expectation of voluntary participation and autonomous decision-making in academic tasks versus the mere data collection in non-academic tasks.

Second, I propose that there is a tension between including or excluding monetary benefits and reputational risks when assessing the beneficence in research on MTurk. As can be seen in several IRB respondents' and researchers' quotes. Academic research on MTurk is usually perceived as with minimal risks and benefits, and thus, would usually be exempt from an in-depth IRB review. Such a perception is based on the assumptions that compensation should not be considered a monetary benefit and that reputational risks on MTurk workers, e.g., a rejection that would render them ineligible to take future tasks, are not counted as research risks. However, as can be seen in the previous quotes, many respondents in my interviews did not

support this assumption. Several researchers and even IRB respondents admitted that in the context of research on MTurk, monetary compensation is a benefit to MTurk workers because money is many MTurk workers' motivation and purpose to participate in an academic study, and the non-monetary research benefit to MTurk workers is usually none. On the other hand, several researchers and IRB respondents have also pointed out the consequences of reputational risks, such as a few rejections could render a MTurk worker to lose their eligibility to take well-paid and high-quality tasks and force them to deal with uninteresting or problematic tasks and unreliable requesters, even though these consequences are not directly related to any specific research project.

Therefore, the current assessment of research beneficence on MTurk seems to have been simplified from that in reality. If we include the evaluation of monetary benefits and reputational risks, several open questions would emerge in research on MTurk. First, what would be the criteria for such an evaluation, and can it be standardized in research guidelines or policies by any means? Second, how do these benefits and risks relate to the "traditional" research benefits and risks? For example, would minimum risk research be escalated to expedited review with a consideration of the reputational risks, or would high-risk research be leveled down due to an acknowledgment of the monetary benefits? Finally, how do crowd workers assess such risks and benefits when deciding whether to participate in a study? These questions deserve further exploration and discussion in the future.

Comparisons of ethical considerations from different entities

Based on my interview and document data, I identified several noticeable commonalities and differences in considering ethics in human subjects.

First, the IRB respondents and researchers in my interviews were quite agreed on issues with respect for persons. They both realized the potential of dehumanization on MTurk and proposed to ensure the transparency of research, voluntary participation, and the autonomy of MTurk workers. However, there is no consistent standard on the level of transparency. For example, a researcher, P19, critiqued his IRB consent form not transparent enough, and research guidelines and templates, such as D3, D5, and D6, display different specificities in their suggested research transparency.

Second, although the IRB respondents unanimously agreed that research on MTurk is with minimum risks, some of them and some researchers also pointed out the potential consequences of reputational risks on MTurk workers. These reputational risks are also reflected in the research guidelines and templates but are not considered as research risks per se. However, it is interesting to see that there seems to be a misunderstanding between IRBs and researchers. Either party assumed the other as somewhat unconscious of the reputational risks on MTurk yet each party in my interviews demonstrated their knowledge of such risks.

Third, the IRB respondents seemed to be more conscious and concerned about the privacy and data confidentiality risks in MTurk research, whereas the researchers seemed to care less. Some researchers even opposed the validity of an argument of privacy or made a false assumption of anonymity in this context. Such a disparity is also embodied in my collected documents. Research guidelines by the IRBs are conscious and considerate of privacy and data confidentiality issues, whereas research publications did not report nor consider such issues.

Finally, several IRB respondents expressed concerns about research justice on MTurk because of the limitation of sampling and the selection mechanisms on MTurk. Although many researchers did not explicitly raise their concerns about justice, their concerns about data quality

and validity issues in the previous chapter could be related. Moreover, research justice has not been considered in my collected documents except one guideline that touches upon it in relation to the sampling bias in research via crowdsourcing.

Implications for crowd work-based research in the future

Based on my findings and discussion on the ethical issues in data, I propose three implications for crowd work-based research in the future. First, the relation between dehumanization and autonomy should be investigated further in research on MTurk to consider the nature, purpose, and characteristics of this crowd work platform. From my current stance, the nature of MTurk has been sealed and dehumanized by Amazon as a platform of instant, accessible, and global transaction of money and piecework. Hence, to what extent the assumption that research subjects should be rational and independent-thinking human beings can still hold is still dubious. Meanwhile, the purpose of MTurk is not intended or designed for scientific research. Hence, how much can researchers or IRBs expect MTurk workers to behave as research subjects not only in terms of their data input quality but also in terms of their volunteering motive is also an open question. In addition, I propose that the characteristics of MTurk, such as its scalability of sampling, transience of task, and hands-off liability between MTurk workers and requesters (Xia et al., 2017) can also impact the dehumanization and autonomy of MTurk workers.

Second, whether or to what extent shall IRBs and researchers consider reputational risks and monetary benefits related research beneficence deserve further deliberation. As I have posed above, including reputational risks as research risks and monetary benefits as research benefits would induce many open questions such as the prevalent exempt research on MTurk may be considered expedited in the future and the research guidance and consent form templates may

need to be updated as well. Reputational risks and monetary benefits would also need deliberation on how to protect MTurk workers. Reputation risks and monetary benefits would also need deliberation on how to protect MTurk workers from these risks and benefits while ensuring a reliable reputation system on MTurk, where justifiable rejection is still necessary, and avoiding undue influence of payment even it is termed as a research benefit to MTurk workers.

Third, research justice on MTurk should be guided more. As my document analysis revealed, there has not been much guidance or regulation on how to ensure justice in research on MTurk specifically. Several IRB directors correctly pointed out in my interviews that we still know very little about MTurk population, and Amazon has concealed their demographic information to the public. Such a lack of knowledge hinders IRBs and researchers to make an informed assessment of research justice. Meanwhile, several researchers in my interviews also critiqued that the non-monetary benefits from research on MTurk are usually none or disconnected to MTurk workers. Thus, how to justify the validity and beneficence to conduct research on MTurk apart from *the benefits to researchers*, such as convenient sampling and decent data quality, is questionable.

Finally, different philosophical stances of ethics, e.g., utilitarianism and Kantianism, can be discussed and debated further in crowd work-based research. To my knowledge, such discussion and debate are still very rare in research about MTurk, crowd work, crowdsourcing, and the gig-economy more broadly. These discussion and debate are important and necessary because they can guide more specific ethical considerations and prevent policies about ethics to be solely from inductive reasoning. Last but not least, these philosophical stances of ethics are behind the essential ethical guidelines such as the Belmont Report and AoIR IRE, and thus,

deliberating with them will help adapt them to crowd work-based research or extrapolate or refine them to be more pertinent and instructive to this context.

Summary

In this chapter, I have traversed various ethical considerations and practices in my interviewed IRB respondents and researchers. These ethical issues can be categorized into four aspects: respect for persons, beneficence, justice, and different philosophical stances of ethics. Within each category, I have presented a diversity of ethical views and practices. Based on these findings, I have discussed their origins, compared between different entities, and proposed implications for future research.

CHAPTER 8 – DISCUSSION

Chapter 8 includes four discussion topics. First, I briefly recap the research questions that I have proposed and answer them with my research findings in the case of MTurk and beyond. Second, I reflect on the "original sin" of crowd work-based research, which gives rises to various origins of the ethical issues around payment, data, and human subjects issues. Third, I discuss the impacts of the original sin of crowd work-based research in academia. Finally, I discuss the limitations of the Belmont Report and the AoIR IRE 3.0 in guiding ethics in crowd work-based research.

The answers to the research questions in the case of MTurk and beyond

The first research question I proposed is: What do academic researchers perceive as the ethical issues with crowd work-based research? Through the interviews, I find that the researchers' ethical perceptions can be categorized into payment, data, and human subjects. Most researchers would stick to a federal or state minimum wage to pay MTurk workers and a few of them have raised ethical concerns about fair payment and the potential undue influence from a relatively high minimum wage rate comparing to the extremely low payment on MTurk normally. It is also noteworthy that a few researchers would regard monetary compensation as a research benefit to MTurk workers.

As regards data issues, several researchers raised their concerns with cheating yet still they would pay these cheaters. One researcher has been worried about such an unconditional payment to MTurk workers, which would undermine the reputation system on MTurk and deteriorate the overall data quality and MTurk workers' accountability on MTurk. Some researchers were also concerned about the validity of research on MTurk, as they assumed that many MTurk workers had been quite seasoned in responding survey questions and accustomed

to experimental treatments. However, it is noteworthy that two researchers held opposing views in MTurk workers' non-naïvety to the internal validly of their research in privacy and politics respectively. The privacy researcher regarded MTurk workers' non-naïvety as helpful to prevent him from falsely claiming alternative hypotheses, while the political scientist regarded it detrimental to his manipulations in people's political views.

Finally, in terms of human subjects issues, several researchers introduced their ideas of empowering MTurk workers and regarded it as an important act of respect for persons. One researcher, in particular, emphasized the limitations of the current consent form templates on MTurk, and designed his bona fide consent form that would allow MTurk workers preview the task instead of detailing the research risks and benefits. Some researchers also expressed their ethical concerns about the reputational risks on MTurk workers, and thus, they held a conservative view of rejecting MTurk workers. As regards justice, several researchers were concerned about the representativeness of MTurk workers and the potential sampling biases.

The second research question I proposed is: How do IRB directors and analysts interpret and enforce the federal government's research mandates in the context of crowd work-based research? I find that most IRB respondents in my interviews were quite abiding with the Belmont principles yet some of them also struggled with the applications of these principles. For example, numerous IRB respondents held a preemptive concern about the undue influence of payment in research on MTurk, even though they were conscious that the payment rate is usually extremely low on MTurk. A few IRB respondents were torn on whether payment on MTurk could be seen as a research benefit.

In terms of data issues, many IRB respondents were concerned about cheating and fraud, and one of them was particularly empathetic about researchers' waste of time and funding in

dealing with cheaters on MTurk. Many IRB respondents also raised doubt on the validity of research on MTurk because they did not regard MTurk workers as a representative population relative to many academic research projects. A few IRB respondents critiqued using MTurk for academic research because MTurk workers' primary motivation is money rather than research.

Finally, with human subjects issues, the IRB respondents were conformed to the Belmont principles. For example, most of them interpreted respect for persons as respecting MTurk workers' autonomy and protecting them from undue influence; most of them also clearly distinguished monetary payment with research benefit. Still, however, a few IRB respondents would contemplate ethics beyond the Belmont principles but more pertinent to MTurk. For example, they would consider the reputational risks on MTurk workers and an equitable payment as a sign of respect and justice to MTurk workers.

The third research question I proposed is: How do the existing guidelines for specific for crowd work-based research, such as those drafted by IRBs and researchers for MTurk, consider various ethical issues in crowd work-based research? Through a document analysis on the ethical guidelines and templates for research on MTurk, I find that they lack congruency between each other and exhibit various limitations. For example, some consent form templates proposed a below-minimum wage hourly rate; some guidelines advocate for paying "at least" a minimum wage; still, one guideline suggests unconditional payment to MTurk workers, which in the long term, could ruin the reputation system.

Although some guidelines resonated with the researchers' and IRB respondents' ethical concerns in my interviews, many ideas that I have discovered in the interviews were not reflected in these guidelines. For instance, the ethical concern about exploitation, undue influence, research validity, and reputational risks are all lacking in these guidelines and templates, which

suggested that many authors behind these guidelines were not quite conscious about or familiar with the particular ethical challenges and consequences in crowd work-based research.

The answers to my proposed research questions are obtained in a case study of MTurk; however, they apply to crowd work-based research not as an activity confined on a specific platform but more generally as a research practice. First, by definition, financial compensation is an essential ingredient of crowd work (Kittur et al., 2013), and as such, it is integral in crowd work-based research. Hence, beyond MTurk, so long as scholars collect research data from a crowd work platform, they will encounter the ethical issues of undue influence, fair payment, and research benefit. The only nuance would be the normalized payment rates may be different between crowd work platforms, and thus, the level of payment that would render an undue influence or exploitation may vary across these crowd work platforms for academic research.

Second, ethical issues around data quality and validity are also not only tightly bound to MTurk but also rooted in the nature of crowd work-based research. A crowd work task is normally transient (Kuek et al., 2015; Xia and Mckernan, 2020), and crowd workers are mostly motivated by monetary compensation to finish a task as fast as possible to earn more. As such, ensuring good data quality, detecting cheating, and screening qualified crowd workers are not pertaining to MTurk per se but are ethical challenges in academic research inherent in any crowd work platform. Furthermore, research validity issues such as the non-naïvety of crowd workers and the generalizability of research findings are also not particular to MTurk but to such a phenomenon of crowd work-based research because crowd workers may migrate from one crowd work platform to another (Katz, 2017) and their prior knowledge, their communication with their peers, and the their representativeness of certain populations would be with them in a new crowd work platform.

Third, my findings of the ethical issues with human subjects on MTurk are not particular to the platform of MTurk either. As I will discuss later, many human subjects issues stem from Amazon's "original sin" that has stained crowd work as a means to collect research data for academic purposes. Since many crowd work platforms are descendants or replicas of MTurk, they have also been infected by Amazon's "original sin" in terms of how they position crowd workers and to what extent they shoulder the responsibility to regulate payment, protect data quality, and respect crowd workers as human subjects in academic research.

The "original sin" of crowd work-based research

Converging the ethical problems and reflections together, I propose that they are all stem from the "original sin" of crowd work-based research, which includes Amazon's stance of "Human-as-a-service", Amazon's confusion of terminology, and Amazon's abdication of responsibilities. As the first and most popular crowd work platform, Amazon's MTurk passes on its original sin to the other crowd work platforms and sets the foundation and the origins of ethical issues in payment, data, and human subjects that I have discussed in Chapters 5, 6, and 7 respectively. Also, such original sin infects academic researchers who embrace it and continue to conduct academic research on this marketplace. Below, I critique Amazon's original sin as well as academia's complicity in it.

The stance of "Human-as-a-service"

Jeff Bezos's famous stance of "Human-as-a-service" has a lasting ethical impact on academic research. "Human-as-a-service" might sound exciting for business but could also be ethically problematic. Ever since its publication, it had been critiqued as commodification of MTurk workers (Aloisi, 2015; Bergvall-Kåreborn et al., 2014) and the origin of exploitation on

MTurk (Irani and Silberman, 2013; Silberman et al., 2010). However, I argue that it is also an origin of the ethical issues in academic research on MTurk.

First, philosophically speaking, it is a very utilitarian claim of MTurk. Positioning MTurk workers as a "service" implicates that they are a means to serve and satisfy requesters' "ends" such as their purposes of data validation and data collection for research. MTurk workers' own "ends," such as their self-fulfillment and welfare, are dismissed by Amazon. Hence, it is morally inappropriate from a Kantian perspective where each human being should be treated and respected as "ends in themselves." When requesters or researchers assume MTurk workers to be only a "service" to them, it would be hard for them to ponder what are the "ends" for the MTurk workers and to concretize the Belmont principle of respect for persons. Even from a utilitarian perspective, Amazon's stance of "Human-as-a-service" is still flawed because it has not considered MTurk workers' "utilities" such as their welfare and earnings. Such a negligence is obvious on MTurk's official website where the benefits of MTurk, i.e., optimize efficiency, increase flexibility, reduce cost, announced by Amazon are all about the *requesters'* benefits. Hence, it could also be morally wrong because the MTurk workers are more in number and have less advantage than the requesters, and the overall "pains" that MTurk creates may be higher than the overall "happiness" that it brings to requesters.

Second, practically speaking, "Human-as-a-service" would induce ethical concerns about MTurk workers' voluntary research participation and autonomy because it already inferred unbalanced power dynamics between MTurk workers and requesters, as if MTurk workers were servants and requesters were masters. Under such a power relationship, MTurk workers' voluntary participation and autonomous decision-making are at stake because in order to provide a good "service," servants presumably would incline to cater for masters' requests and consider

whether the compensation is fair for their service rather than whether such requesters or compensation may interfere their independent thinking and dignity. Reflected in this dissertation work, many IRB respondents raised ethical concerns about voluntariness and autonomy in research on MTurk; in comparison, many researchers worried about fair payment and wage benchmarks, but very few ever mentioned voluntary participation and autonomous decision-making. To some degree, I argue that it reflects the fact that academic researchers have more or less taken MTurk workers as a service for granted and would rather contemplate how to compensate and treat their "services" fairly than consider whether MTurk workers ought to be a sort of service for academic research.

Third, "Human-as-a-service" has also provoked ethical concerns and practices such as dehumanization and the empowerment of MTurk workers. These concerns and practices, such as the creation of Turkopticon (Irani and Silberman, 2013) and "We are Dynamo" (Salehi et al., 2015), would not have been so profound and influential if MTurk workers had not been dehumanized as a service at the beginning. In this dissertation work, some IRB respondents and researchers' ideas and practices of humanizing and empowering MTurk workers, such as treating them as "customers" and building tools to cultivate their professional skills also derive from the dehumanization origin of "Human-as-a-service." Furthermore, I argue that "Human-as-a-service" has degraded both MTurk workers and requesters to data. Cheney-Lippold (2017) argued that people are being categorized into different data templates, where their subjective experiences and individual traits are nullified and replaced by measurable types that can be defined, created, grouped, and searched just by a set of data entries – in a word, "we are data" nowadays. In a similar sense, Amazon has categorized MTurk workers into a "service" to provide data and researchers into "requesters" for data. In this context, I posit, humanization, and

mutual respect between two groups of human beings, i.e., MTurk workers and researchers, are not just hard to establish but perhaps even unrealistic to expect in the first place.

The confusion of terminology

The second part of the original sin is Amazon's confusion of terminology in terms of MTurk "workers," "independent contractors," "requesters," and "crowdsourcing marketplace." Amazon's confusing of these terms may be accidental because they were created over a decade ago at the dawn of crowdsourcing, but it could also be intentional because they have not been updated since their creation. I argue that the confusion of these terms also engenders various ethical disputes and conundrums.

First, Amazon calls the people who take tasks and get paid on MTurk as MTurk "workers," which seems to implicate that MTurk workers are in a sort of employment relationship and are earning wages. However, Amazon describes and positions MTurk workers as "independent contractors," which makes the term of "MTurk workers" confusing and embarrassing. Independent contractors are between employees and research volunteers. Thus, if researchers treat MTurk workers as if they are doing a job on MTurk, they would pay them with a wage standard, but some IRB would worry about undue influence and manipulation on MTurk workers. If researchers treat MTurk workers as volunteers, they will not pay them a minimum wage because that is not fit for an unemployment relationship, and also that's a relatively high rate on MTurk and could be coercive. If IRBs treat MTurk workers as "employees," then they would not approve most academic research on MTurk unless the research is about MTurk workers. If IRBs treat MTurk workers as research volunteers, they will discourage a minimum payment to them, but some researchers would worry about exploitation.

Meanwhile, it is unclear in academia whether research subjects being independent contractors necessitates particular policies and regulations that fit all research disciplines and IRBs, or such an issue can be left at individual researchers' discretion and contingent on different research projects. For instance, the Belmont report and IRB policies have considered the respect, protection, beneficence, and justice to vulnerable groups such as pregnant women, children, and prisoners. Would independent contractors also count as a vulnerable group if many of them will someday depend on working on MTurk or the other gig-economy platforms to make a living without any employment security or welfare? If partaking in academic research becomes a main source of income for many MTurk workers, how to balance between benefiting this group of independent contractors and avoiding any undue influence or coercion? Not the least, so far, it is unknown how MTurk workers would perceive their identity and position. For example, do they always perceive themselves as "workers" for business tasks or "volunteers" for academic tasks or sometimes as "workers" and "research volunteers?" What does "independent contractor" mean to them? I posit that questions would further perplex various ethical issues, which deserve further investigation in the future.

Second, Amazon's term of "requesters" is also confusing. On MTurk's website, Amazon describes it as "individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually" ("Amazon Mechanical Turk," n.d.-a). Perhaps the "businesses" part of this term is more understandable because Amazon clarifies it later as "MTurk enables *companies* to harness the collective intelligence, skills, and insights from a global workforce." But who are the "individuals?" Amazon has no further categorization or elucidation. Arguably, academic research is distinctive from business because it aims for knowledge generation and generalization instead of profit (though I notice that AoIR IRE 3.0 has

warned an ethical risk of commercialization of academic research). However, academic requesters are not simply "individuals" either. They are employed by academic institutions and supervised by IRBs. Thus, presumably, academic requesters are more regulated and organized than "individual" requesters. Academic requesters have been using MTurk since the conception of it and they constitute a large portion of requesters (Hitlin, 2016). Despite it, it seems that Amazon still hasn't acknowledged academic requesters' status comparable with that of business requesters on MTurk. It quite contrasts to the recently prospering Prolific that position scientific researchers as main task publishers on their platform ("Prolific | Online participant recruitment for surveys and market research," n.d.).

Amazon's obscure term of "requesters" triggers ethical issues in academic research on MTurk. On the one side, it obfuscates the essential difference between business requesters and academic requesters and further marginalize the significant role of the IRB. As I have discussed in Chapter 7, MTurk workers may not be able to swiftly switch between a "worker" mode and a "research volunteer" mode or differentiate between business tasks and academic studies. Academic researchers are also in an embarrassing position with Amazon's current description of requesters. On the other side, some researchers may feel obliged to pay MTurk workers with a fair wage, which can be seen an ethical act in a business context. However, such an act could be disputed by some IRBs because, as many IRB respondents in my interviews claimed, payment in academic research is essentially different from a wage, and a "fair wage" could potentially be unduly influential on MTurk workers. In addition, some researchers, e.g., P19 and P29 in my interviews, may feel skeptical and disconnected with the IRB's role in their research on MTurk because they assumed that IRBs did not understand the nature of research on MTurk. Nevertheless, I would side with the IRB here and argue that it is rather the researchers who have

been homogenized by Amazon's term of "requesters" and adapt their ethical concerns and practices to a business marketplace rather than stick to those in academia.

Finally, Amazon's term of MTurk as a "crowdsourcing marketplace" also confuses and misleads academic researchers and IRBs and evoke ethical problems. Amazon commercializes academic research on MTurk and transforms it from a process of scientific inquiry and knowledge exchange into a transaction. Every party has to obey the rules of a marketplace where everything has a price. Thus, not only does MTurk workers' labor becomes a commodity (Aloisi, 2015; Bergvall-Kåreborn et al., 2014), research data also becomes a commodity that must be screened and bought by researchers. MTurk workers with a high approval rate or a master worker badge are similar to the star sellers on Amazon that have received most positive reviews. In this sense, academic research has turned into a purchase behavior where many ethical issues emerge and center on the fairness of transaction, e.g., good data quality from the MTurk workers' side and an equitable payment from the researchers' side. In my interviews, the ethical debates of undue influence of payment, exploitation, beneficence, and even respect for MTurk workers and justice are all surround such a center.

Meanwhile, Amazon's term of MTurk as a "marketplace" completely distinguishes it from a volunteer's pool that many IRB respondents in my interviews would expect MTurk to be. Resonating with what this term implicates, several researchers, e.g., P21, would drop the expectation to recruit research volunteers from MTurk and suggest citizen science projects to fulfill such an expectation. They acknowledge MTurk to be a market and aim to make their academic projects fair on this market. However, such a market-oriented ideology of academic research has degraded researchers' criteria and altered their priority of ethics in academic investigation. In the era of the Belmont Report, academic research was ought to recruit only

volunteers and avoid any type of undue influence or coercion. In the era with a crowdsourcing marketplace, academic research no longer requires volunteers but can depend on a pool of "workers" that are primarily motived by compensation instead of research per se. Last but not least, a marketplace of "workers" poses ethical questions to the validity of research on it because MTurk workers' monetary motivation and opaque demographics hidden by Amazon make MTurk workers' representativeness dubious.

The abdication of responsibilities

The final part of the original sin from Amazon's MTurk, I argue, is its abdication of responsibilities, which includes the abdication of responsibilities between MTurk and MTurk workers, the abdication of responsibilities between MTurk and researchers, and the abdication of responsibilities between MTurk workers and requesters.

First, Amazon abdicates the responsibilities between MTurk and MTurk workers. MTurk only has a general policy and a weak warning to requesters: "You may not use Amazon Mechanical Turk for...collecting personal identifiable information; fraud; disrupting or degrading the operation of any website or internet service; direct marketing; spamming, etc." ("Amazon Mechanical Turk," n.d.-a). There is no mandatory prohibition to monitor or ban these activities, and MTurk workers report that many requesters have disrespected and disobeyed these policies (Xia et al. 2017). Meanwhile, Amazon places MTurk as an intermedia between requesters and MTurk workers and earns commission from requesters rather than MTurk workers. Hence, Amazon has a natural tendency to lean toward requesters and evade the responsibility to establish and mandate any rule to protect the MTurk workers' interests and welfare. Such an abdication of Amazon's responsibilities with MTurk workers makes MTurk workers more

vulnerable and potentially more skeptical toward requesters no matter they are from business or academia.

Second, Amazon also abdicates its responsibilities with academic requesters and support business requesters. Even though academic researchers have constituted a large portion of requesters on MTurk, Amazon still makes it quite clear that MTurk is purported for serving business requesters and has barely made any suggestion or policy for academic researchers. As such, academic requesters are like "undocumented immigrants" in a marketplace competing with protected citizens of MTurk. Amazon has developed a sandbox website to let requesters acquaint themselves with MTurk, but such a sandbox website has not been tailored to academic research or has any template for academic tasks. Academic requesters have to develop their own tools and sites, such as TurkPrime (Litman et al. 2017) to facilitate research on MTurk. They also have to brief the ethical issues such as fair payment and informed consent to MTurk workers, which can vary in details and emphases as my document analysis revealed, rather than utilizing any research task template that can be designed and launched by Amazon on MTurk (Xia and Mckernan, 2020).

Finally, Amazon also abdicates the responsibilities between MTurk workers and requesters. Requesters do not bear any responsibility on MTurk workers and can accept, pay, or reject MTurk workers at their discretion. Even though academic requesters are regulated and monitored by the IRB, non-academic requesters are quite free and independent to treat MTurk workers. As such, researchers have built third party tools such as Turkopticon to give MTurk workers more support to rate deter bad requesters (Silberman and Irani, 2013). On the other hand, except for an approval rate mechanism, MTurk workers do not charge much responsibility to ensure their data quality either. Amazon creates no screening mechanism to winnow good

quality data from spams, again, researchers have to build sites such as TurkPrime to refine the data screening process (Litman et al., 2017). As a result, MTurk, MTurk workers, and requesters have quite loose and aloof relationship and responsibility with each other, and such a situation renders mutual accountability between these parties brittle and frivolous. Thus, various ethical issues could occur on either MTurk workers' side, e.g., cheating and fraud, or requesters' side, e.g. exploitation and arbitrary rejection.

The impact of the "original sin" in academia

The original sin of crowd work-based research has also infected academia. I argue that such an infection is embodied in three aspects: (1) the negligence of different purposes of MTurk and academic research, (2) the ontological schism between researchers and IRBs in their ethical considerations, and (3) the flaws in the existing ethical guidelines for crowd work-based research.

The negligence of the teleological difference between MTurk and academic research

To start with, Amazon has made it quite clear that the telos, i.e., the ultimate purpose, of MTurk is *not* for academic research. However, many researchers disregard this fact and use MTurk for academic research from the inception of MTurk till this day. From a teleological perspective, such a negligence renders the foundation of crowd work-based research in academia questionable. The academic researchers can hardly blame MTurk for its unsupportiveness of academic research. Since the purpose of MTurk is for business crowdsourcing, and MTurk is the most popular and suited platform for such a purpose, MTurk is a *just* platform from a teleological perspective. As such, many academic researchers' continued launching of academic tasks on MTurk and their ethical complaints of MTurk are somewhat hypocritical. Adapting an Aristotelian analogy, it would be like forcing the best flutist to play a saxophone and blame her

performance, and then, try to train and transform her into a saxophonist. In comparison, some IRB respondents' suggestions to withdraw from MTurk for academic research and some researchers' behavior of leaving MTurk seem to be more ethical and appropriate. Thus, I argue that as long as MTurk keeps its purpose for business crowdsourcing and researchers keeps using MTurk for academic purposes, there will be a lasting ethical dilemma between MTurk's purpose and academic researchers' expectation.

By contrast, the telos of academic research, as proclaimed in the Belmont Report, is to "develop or contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships)" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979, Part A). This purpose is far and distinctive from MTurk's business goal for profit. Thus, can we obtain generalizable knowledge on a crowd work platform like MTurk? As I have reported in Chapter 6 and 7, many IRB respondents and researchers held ethical concerns about internal and external validity as well as the MTurk population's representativeness and sampling biases. I argue that these ethical concerns, to a large extent, originate from the disparity between the purposes of MTurk and academic research because MTurk is not designed for the exchange, development, or generalization of *scientific knowledge*. Hence, the ethical concerns about research validity on MTurk are hard to be appeased. Perhaps one way to escape this dilemma is as researcher P18 advised to compromise the external validity for the internal validity, and only focus the research lens on the MTurk population or maybe the crowd work population more broadly rather than claim for any knowledge generalizability to the other populations outside of the crowd work context.

Finally, because the purpose of MTurk is not for academic research but business, yet both academic researchers and business requesters are publishing on MTurk, it becomes essentially difficult to differentiate between academic tasks vs. non-academic tasks. Likewise, it also becomes extremely difficult to differentiate between MTurk workers who are primarily motivated research vs. MTurk workers who are primarily motivated by money. These mixes make ethical issues in MTurk research intertwined and muddy. First, MTurk workers might be hard to switch between business response mode and academic response mode, or they might treat all the tasks the same except considering different payment rates and reliability scores. Thus, to what extent MTurk workers who participate in an academic project on MTurk can be seen as research volunteers is dubious. Second, academic researchers would struggle with rejecting or not rejecting. As I have described in Chapter 6, some MTurk workers might just want fast money and cheat in a task, but a valid rejection based on an academic standard might stir a reputational risk from an ethical concern in business because it might affect those crowd workers' eligibility and welfare in the long term. Third, IRBs would struggle between their concerns about undue influence and the researchers' worry of fair payment because money is the dominant motivation, but it is also normally low on MTurk. They also have to deal with complaints from MTurk workers in a way as if MTurk workers were customers, researchers were business vendors, and IRBs were a customer service instead of a compliance committee.

The ontological schism between researchers and IRBs in their ethical concerns

Another impact of Amazon's original sin in academia is the ontological schism between researchers' and IRBs' ethical concerns in academic research on MTurk. Specifically, the ontological schism in the nature of MTurk workers and MTurk and how to conduct ethical research with these people and on this platform.

First, Amazon's stance of "Human-as-a-service" makes the ontology of MTurk workers obscure and ambiguous, which further renders researchers and IRBs distinct their ethical concerns about respecting MTurk workers as human beings. In the researchers' camp, respecting MTurk workers was instantiated by humanizing and empowering MTurk workers. For example, P19 required his student to treat MTurk workers as their customers and respond them promptly; P21 and P29 built tools to cultivate MTurk workers' professional skills; P26 aimed to unite MTurk workers' voices and expose them to requesters and the public. In publications, such an empowerment act of respecting MTurk workers is manifest in the works such as (Deng et al., 2016; Salehi et al., 2015; Silberman and Irani, 2013). In the IRBs' camp, respecting MTurk workers was less about empowering them but more about ensuring their autonomous decision-making without any undue influence of payment or coercion. The researchers' and IRBs' ethical concerns and practices of respecting MTurk workers can be interpreted as critiquing Amazon's stance of "Human-as-a-service." From the researchers' perspective, Amazon has dehumanized and debased MTurk workers, and thus respecting MTurk workers means re-humanizing and empowering them. From the IRB's perspective, Amazon has made MTurk workers vulnerable to the coercion of the "service-requesters" such as the researchers who outsource their tasks on MTurk. Thus, respecting MTurk workers means protecting their vulnerabilities and preventing them from undue influence or coercion. Though both rationales are reasonable, Amazon's stance of "Human-as-a-service" has departed researchers and IRBs from forming a more coherent and consistent view of "respect for persons" about MTurk workers.

Second, Amazon's confusing terminology of MTurk workers as independent contractors in between wage-earners and research volunteers makes researchers and IRB respondents hold contradictory views with regard to ethical issues in payment. The researchers camp would

default in positioning MTurk workers as wage-earners and paying them a minimum wage. Many expressed ethical concerns about fair payment. Such an act of paying a minimum wage and the concerns about fair payment also were referenced in publications such as (Silberman et al., 2018; Hara et al., 2018). Furthermore, several researchers, as I reported in Chapter 6, would pay some MTurk workers unconditionally even though they had evidence and were sure that some MTurk workers' data input was in poor quality and useless. The IRB camp, on the other hand, would default in positioning MTurk workers as research volunteers and concerning about the undue influence of payment, even though they were conscious about how low the payment is normally on MTurk. They worried about paying MTurk workers a sort of minimum wage precisely because the payment is normally low, and a minimum wage could be relatively high and become an undue influence on some MTurk workers. Some IRB respondents even argued that money should not be the motivation to research participation at all, and therefore, a low payment might be more justifiable on MTurk. However, a few researchers, such as P29, would critique that the IRB did not comprehend the significance of a fair payment to MTurk workers' wellbeing. Thus, Amazon's confusing terminology of MTurk workers makes researchers and IRBs hold contrary opinions about ethics in payment issues.

Finally, Amazon's position of MTurk as a "crowdsourcing marketplace" also makes the researchers and IRB a bit torn in their overall ethical concerns with MTurk. Both researchers and IRBs have been aware of MTurk as a "marketplace" in publications as well as in my interviews. For example, Litman et al. (2017) clearly pointed out that MTurk was a business crowdsourcing platform not intended to be an academic research tool; D2, a MTurk research guide drafted by a private university IRB, which I collected and analyzed, also acknowledges that "MTurk was designed as a crowdsourcing platform for business –not a research platform" and this fact leads

to "unique challenges." However, the researchers' and IRBs' reactions to this common awareness and unique challenges of MTurk differ. Researchers, to certain extent, have adapted their ethical concerns from academia to a business setting. For example, some researchers in my interviews were more concerned about fair payment and exploitation rather than undue influence; or as TurkPrime represented, some researchers tried to transform MTurk to be more suited for academic research (Litman et al., 2017). In comparison, several IRB respondents in my interviews were skeptical and even against of using MTurk for academic research because they perceived that such a marketplace could not guarantee either voluntary participation or data quality. Such a disparity further creates a delicate tension between some researchers and IRBs. Several researchers in my interviews, e.g., P19, P27, P29, regarded themselves as the protector of MTurk workers while the IRB might not have adequate knowledge about their protection. In the other camp, several IRB respondents, e.g., P11, P14, regarded themselves as the protector of *researchers* because some MTurk workers were irresponsible and researchers might not know how to deal with them. Hence, Amazon's term of MTurk as a marketplace diverge from researchers' and IRB respondents' overall reactions to MTurk and created a tension among them.

The flaws in the existing ethical guidelines for crowd work-based research

A third impact of Amazon's original sin in academia is that it originates various flaws in the existing ethical guidelines for crowd work-based research. I have discussed many such flaws in the three findings chapters. For instance, in Chapter 5, some research guidelines for research on MTurk held arbitrary recommendations of payment rate, which is largely due to Amazon's deficiency of a transparent payment standard. In Chapter 6, some guidelines struggled on the relation between data quality and MTurk workers' approval rating, which can be attribute to Amazon's lack of policy and mechanism in monitoring cheating and controlling data quality. In

Chapter 7, some guidelines cannot offer a congruent policy of voluntary participation in research on MTurk, such as how their voluntary withdraw from research is linked to the policy of payment. Below, I will discuss the impact of Amazon's origin sin on three important but flawed recommendations from my collected ethical guidelines for crowd work-based research.

The first flaw is a paradoxical position of MTurk as a "voluntary place to earn money," which is in Document D6's verbatim. Such a depiction of MTurk can be confusing to both researchers and IRBs. If MTurk is a voluntary place, then it should be ideal for scientific research that is based on voluntary participation. As such, MTurk would be similar to citizen science sites such as Zooniverse, and MTurk users who participate in academic projects would be hobbyists or volunteers. However, D6 adds that MTurk is a voluntary place "to earn money." Then, MTurk workers cannot be merely seen as research volunteers or hobbyists, and their motivation to participate in an academic project is dominated or at least influenced by monetary incentive. In this regard, whether MTurk should be taken as an academic research platform becomes doubtful, and D6's message to guide research on MTurk becomes ambiguous. Despite it, D6 is not my real critique target because its confusing definition of MTurk reflects and originates from Amazon's confusing terminology of MTurk as a "crowdsourcing marketplace." Amazon's original definition of MTurk has obscured the boundary between voluntary participation in crowdsourcing and motivation of payment in a workplace. As a consequence, guidelines such as D6 are further delivering this ambiguous message to researchers and IRBs.

The second flaw is represented in Document D2's suggestion "never reject a MTurk worker." As I described in Chapter 7, D2 presented two scenarios: (1) MTurk workers failed the ACQs, and (2) MTurk workers used automatic script to complete the task. In both scenarios, rejections of these MTurk were viable, but D2's final suggestion was "to accept the HITs and

pay the subject(s)." Hence, D2 suggested never reject a MTurk worker (I could not think of any worse scenario to not reject a MTurk worker). However, it is a flawed recommendation. In fact, researcher P24 had explained it thoroughly in Chapter 6. If every researcher never rejects a MTurk worker no matter how bad their input data quality is, the reputation system on MTurk, which somewhat deters irresponsible behaviors of MTurk workers, will collapse because every MTurk worker will have a perfect reputation which may not reflect their real accountability. As a consequence, every researcher will lose a screening mechanism and their data cleansing cost and effort will mount up. Thus, D2's recommendation of never rejecting a MTurk worker is also flawed, but again, D2 is not the real target of my critique. D2's flaw is also due to Amazon's abdication of responsibility between MTurk, MTurk workers, and requesters. Just as a MTurk worker's arbitrary responses can ruin a requester's task; a requester's arbitrary rejection can also ruin a MTurk worker's reputation. However, MTurk offers no regulation or control over either arbitrary behavior.

The third flaw is about payment, which is reflected in Document D10 that writes "Simply pay more." D10's recommendation to researchers to simply compensate more to MTurk workers may appear to be a fair and noble act in a conventional employment, but it is flawed in crowd work-based research. As I described in Chapter 5, many IRBs would worry about the undue influence of payment in research on MTurk, and as researchers pay more, the risk of undue influence also increases. Thus, "simply pay more" could further aggravate the motivation of money to MTurk workers and eclipse their motivation for doing research. Meanwhile, this edict offers no payment standard, and different researchers can interpret it in their own ways. In turn, MTurk workers experienced varied payment rate can display different levels of social desirability bias and accountability of behaviors. Hence, "simply pay more" is also a flawed

recommendation. Similar to my arguments above, I critique Amazon's original sin instead of D10 for such a flaw. Amazon's stance of "Human-as-a-service" creates such a motive for researchers to avoid dehumanization of MTurk workers and treat them fairly, as that in the payment. However, for voluntary-based academic research, researches' good motive of paying more to MTurk workers may end up undermining the validity of academic research. Thus, Amazon's original stance, rather than the researchers, should be blamed on it.

The limitations of the Belmont Report in crowd work-based research

Apart from the impacts of Amazon's original sin in academia, I further argue that it renders the Belmont Report to have multiple limitations in guiding crowd work-based research. Below, I will reflect on these limitations.

The first limitation of the Belmont Report is its rigid delineation of the boundary between research and practice, which may not apply on MTurk. The Belmont Report originally differentiated practice from research and defined it as means to enhance *patients'* well-being with a reasonable expectation of success (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Friesen et al. (2017) critiqued the Belmont Report's demarcation of practice and research to be too stringent and argued that there was much overlap of research and practice in a biomedical context (Friesen et al., 2017). I posit that this critique also applies to academic research on MTurk because the boundary between research and practice is also ambiguous on MTurk. Scholars can collect data from MTurk workers to contribute to the generalization of scientific knowledge such as that in the field of clinical psychology (Chandler and Shapiro, 2016). As such, these studies use MTurk as a platform for data collection to generalize scientific knowledge.

There is a nuance, however, when scholars collect data from MTurk workers not for generalizing scientific knowledge but for *enhancing MTurk workers' well-being*. These scholars' investigations are more *about* MTurk workers than for knowledge generalization to the other contexts or populations. For example, Saito et al. (2019) and Hara et al. (2018)'s work aimed to help MTurk workers select well-paid tasks and manage their time more efficiently, but their work was not so much for generalizing their findings to the other crowd work platforms or to the other populations outside of crowd work. From the Belmont Report's original stance of practice and research, Saito et al. (2019) and Hara et al. (2018)'s studies would be *practice* rather than *research* because MTurk workers are comparable with patients, and their work was more about improving their well-beings. Hence, echoing Friesen et al.'s (2017) critique that the Belmont Report's original delineation of research and practice is too rigid in medical research, I also posit that it is too stringent in research on MTurk. Furthermore, I argue that this limitation of the Belmont Report is partly due to Amazon's original sin that has made MTurk workers as a service that can contribute to generalizing scientific knowledge but also in an exploitative and below well-being situation, who need humanization and fair treatment.

The second limitation of the Belmont Report is related to Friesen et al.'s (2017) critique of its deductive relationship between the ethical principles and their applications. Friesen et al. (2017) argued that the Belmont principles were set as the ethical canons in academic research and the subsequent applications of these principles introduced in the Belmont report were deduced from the Belmont principles. For example, the acquisition of informed consent from research subjects is deduced from the Belmont principle of respect for persons; the analysis of the ratio between risks and benefits is deduced from the Belmont principle of beneficence. In medical research, such a deductive relationship often falls short and limited, for example, the

principle of justice may fail to account for social justice related to historical oppression (Friesen et al., 2017; Shore, 2006) or compensatory justice related to those subjects injured from research (Friesen et al., 2017; King, 2005). In research on MTurk, I argue that the Belmont Report's deductive limitation is also evident. For example, as I reported in Chapter 7, respect for persons not only includes informed consent and autonomy but also encompasses the ethical concerns and practices of dehumanization and empowerment, which the Belmont Report has not considered originally; beneficence is limited when monetary benefits and reputational risks are excluded from benefit/risk assessment; justice also needs to extend beyond the subject selection as the only emphasis originally to the contemplation of the overall fitness of MTurk population for different research purposes. Thus, the deductive relationship between Belmont principles and Belmont report's proposed applications of these principles cannot interpret and encompass all the ethical issues with human subjects in research on MTurk.

Meanwhile, as I have described and discussed in Chapter 5 and 6, researchers and IRBs have particular ethical concerns about payment issues and data issues on MTurk, such as exploitation, paying a fair wage, cheating, and rejection, which are not readily explicable using the Belmont Report's three ethical principles. In this sense, the Belmont Report exhibits its limitation of not representing the complicated ethical issues and acts in research on MTurk. However, I regard it is also partly due to Amazon's original sin on MTurk. Payment and data issues such as exploitation, cheating, and reputational risks are not prevalent ethical issues in traditional medical and social behavioral research, but they are common on MTurk. Amazon creates MTurk that has embedded a monetary incentive system, but at the same time, it normalizes the incentive rate to be extremely low. Hence, on the one side, MTurk workers are motivated by the monetary incentive to take tasks, on the other side, they have to do a large

number of tasks to make their earnings tangible. In turn, such a fact yields the ethical issues of fair payment to MTurk workers and data quality control for some cheaters on MTurk, who randomly scribble through tasks or even create automatic scripts and "bots" to maximize their speed, the number of tasks they take, and their overall earnings.

The third limitation of the Belmont Report is its limited scope of protection that may not include the diverse ethical issues in research on MTurk. The Belmont Report focuses on the protection of vulnerable groups, such as pregnant women, children, and institutionalized people, from being exploited and coerced in research, but its definition or characterization of "vulnerable groups" is obscure (Friesen et al., 2017; Rogers and Lange, 2013), which led to a widespread assumption that every individual in a vulnerable group lacks autonomy (Rhodes, 2010). Such a limitation of ambiguous definition of a vulnerable group is also revealed in evaluating ethics in research on MTurk. For example, several researchers and IRB respondents, such as P29, P6, perceived MTurk workers to be a vulnerable group, but they were unable to refer to the Belmont Report to reason why MTurk workers are vulnerable and whether all MTurk workers are at risk of undue influence or coercion. Meanwhile, a few IRB respondents, such as P3 and P11, also pointed out that researchers, especially junior researchers, could be "vulnerable" in research on MTurk as well when they encountered irresponsible and fraudulent MTurk workers that would deteriorate their research data quality, waste their funding money, as well as create hassles and pressure on them. However, the Belmont Report has not considered the protection of researchers or considered them to be potentially vulnerable in certain research conditions or contexts.

Finally, the Belmont Report's protection is also limited as regards ethics in data issues in research on MTurk. Vitak et al. (2016) have listed numerous ethical challenges, practices and beliefs that went beyond the Belmont Principles when concerning online data collection,

management, and dissemination. For example, the Belmont Report does not have a specific recommendation on the sharing of raw research data with the other stakeholders. Hence, one of ethics heuristics for online data research that Vitak et al. (2016) proposed for researchers is to be cautious about sharing research data in and outside of academia and to consult with the IRB and colleagues throughout the research process. Smith et al. (1996) also found that unauthorized secondary data sharing is also an essential privacy problem on the internet. Several respondents in my interviews, such as P2 and P10, also raised such concerns about data sharing in research on MTurk and regarded them as essential in informed consent to MTurk workers. However, the Belmont Report's scope of protection does not account for such data ethical issues. Like the other limitations of the Belmont Report that I discussed above, I also attribute the limitation of protection scope partly to the impact of Amazon's original sin on MTurk. Data privacy issues pose particular challenges and consequences on MTurk due to its transient data collection, scalable data compromise, and disguised offenders (Xia et al., 2017; Xia and Mckernan, 2020), and I argue that these problems stem from Amazon's abdication of responsibilities between MTurk, MTurk workers, and requesters.

The limitations of the AoIR IRE 3.0 in crowd work-based research

Besides the Belmont Report's limitations, AoIR IRE also exhibits several limitations when considering it relative to crowd work-based research. Since AoIR IRE 3.0 is its most recent version, I will focus on it and reflect on its limitations in crowd work-based research.

The foremost limitation of the AoIR IRE 3.0 is its overall negligence of the ethical issues and concerns in crowd work-based research that has become a prevalent and popular form of internet research. Although AoIR IRE 3.0 mentioned research on crowdsourcing as having received "some attention" and listed a few ethical issues such as paying a living wage and data

quality, its treatment of ethics in research on crowdsourcing is rather marginal. As this dissertation work reveals, there are much more ethical issues in research on crowdsourcing, or in what I called crowd work-based research, than fair payment and data quality that deserve more ethical contemplation and deliberation. Below, I will pose a few challenges on AoIR IRE 3.0 regarding ethics in research on MTurk, which it has not amply addressed.

First, AoIR IRE 3.0 emphasized the importance of obtaining informed consent and suggests that if it is difficult to obtain it in the initial stage of research, researchers should try to obtain it with their subjects at the data dissemination stage of research, e.g., when researchers plan to publish or share data with other stakeholders (Franzke et al., 2020). However, the informed consent is hard to ensure or obtain with MTurk workers in both the initial stage and the dissemination stage of research because many MTurk workers are not likely to review a consent form closely and they may not be connected or contactable after a participating a short-term research project on MTurk. As revealed in Chapter 6, numerous researchers and IRB respondents were aware and concerned that many MTurk workers would not read a consent form. MTurk workers have a pressure to complete a task as fast as possible so that they can take more tasks and earn more in a given time slot. Hence, it is difficult to ensure informed consent in the initial stage of research on MTurk. Meanwhile, the connection between researchers and the MTurk workers is usually transient and limited within the launch and completion of a HIT (Xia and Mckernan, 2020). Therefore, it is also unlikely to obtain informed consent at the dissemination stage because it is hard to identify and find the original batch of MTurk workers who have participated in and completed a research project. As such, AoIR IRE 3.0's suggestion to obtain informed consent in different stages of a research project is hard to materialize on MTurk and is less likely to be approved by the IRB to conduct research on MTurk in the first place.

Second, AoIR IRE 3.0 made a good call to protect researchers in their public identity, personal safety, and psychological health and well-being when encounter research subjects' ideological reaction, direct threats and retaliation, as well as research content's violent or uncensored scenes (Franzke et al., 2020). However, I have found and thus argue that in research on MTurk, protecting the researchers is also necessary but in the aspects of data quality, research funding, and academic reputation. As I have reported in Chapter 6, numerous researchers and IRB respondents were concerned about poor data quality and fraud on MTurk; P11 explicitly worried that spamming data was a waste to researchers' funding and derogatory comments from some MTurk workers were harmful to an IRB and even institution's reputation and public image. Prior research also indicated that MTurk workers' cheating and carelessness were not uncommon on MTurk (Difallah et al. 2012; Aruguete et al., 2019). These data quality and responsibility-related factors are not directly threatening to researchers' physical or psychological status, and they are not covered by AoIR IRE 3.0. However, they still deserve serious guidance to be tackled so as to protect researchers' time, funding, and reputation.

Third, AoIR IRE 3.0 highlights the legal aspects of ethical guidance for internet research and noted that "it remains debatable as to whether following a website's terms and conditions is a legal requirement for academic researchers whose work benefit the knowledge level of society at large" (Franzke et al., 2020, p. 15). Based on my research findings and contemplations in this dissertation work, I venture to propose and push to debate whether researchers are obliged to challenge, contribute to, or improve a website's terms and conditions (TOC) if this TOC is ethically flawed or problematic rather than merely contending over whether to follow it or not. For example, many researchers in my interviews were proposing and paying a minimum wage to MTurk workers as a benchmark of fair payment. However, there is no mention of any minimum

wage in Amazon MTurk's TOC ("Amazon Mechanical Turk," n.d.-b). Hence, should researchers challenge and propose to Amazon to establish and normalize a minimum wage system in their TOC? Also, Amazon MTurk declared:

> Unless we are participating on the Site as a Requester, we are not involved in the request or the performance of Tasks, and have no control over the quality, safety, or legality of Tasks or consideration for Tasks, the ability of Workers to perform Tasks to Requesters' satisfaction, or the ability of Requesters to pay for Tasks. ("Amazon Mechanical Turk," n.d.-b)

This is an aloof statement of Amazon's role, and to some degree, represents Amazon's abdication of responsibilities. As such, should researchers push Amazon to be more involved in monitoring data quality and both requesters' and MTurk workers' responsibility or just following Amazon's Participation Agreement as it is? I propose such questions to deserve more discussion in AoIR and in academia more broadly.

CHAPTER 9 – CONCLUSION

Chapter 9 is the final chapter of this dissertation. First, I present the research implications and suggestions for scholars, system designers, and policymakers. Then, I introduce the theoretical and empirical contributions from this dissertation work. Afterward, I reflect on the research limitations and look into future work. Finally, I wrap up my dissertation with a conclusion.

Research implications

The research implications of this dissertation work are threefold. First, scholars and IRBs should take an evolutionary perspective on crowd work-based research. Second, the communication between crowd workers, researchers, and IRBs should be enhanced. Third, scholars should explore alternative research venues beyond MTurk and crowd work platforms and contemplate whether the motive of data collection fit the purpose of research. Below, I describe these three research implications in detail.

The first research implication is to take an evolutionary perspective on crowd work-based research. To start with, the features of a crowd work platform can change and evolve. Take MTurk as an example, even though its purpose and target customers remain to be business, its characteristics for academic research have evolved. Silberman et al. (2018) have pointed out that the normative payment on MTurk has increased in the last decade, and thus, researchers should not stick to their research about payment on MTurk a decade ago to corroborate their payment standard on MTurk today. Meanwhile, Amazon had estimated that the MTurk population size had grown from five hundred thousand to one million in 2019. Pew Research also forecasted that an increasing number of academic studies would be conducted on MTurk (Hitlin, 2016). All these changes require scholars to perceive MTurk not as a static pool but an evolving platform.

Second, more types of academic research can be conducted on a crowd work platform (whether it ought to be is a different question here). For example, the pioneering academic studies on MTurk were primarily about language translation or annotation (e.g., Callison-Burch, 2009; Snow et al., 2008), but now, the complicated clinical and developmental psychology research is being conducted on MTurk workers (e.g., Strickland and Stoops, 2019; Engle et al., 2019). Hence, MTurk's potential for scholars to test different types of research has also been developing so that scholars can pilot their research ideas and imaginations with the MTurk's flexible functions and diverse populations. Also, IRBs should take an evolutionary perspective in approving and supervising various novel research studies on MTurk, which may not be commonly seen in previous MTurk research applications.

Third, the ethical issues in crowd work-based research are evolving as well. The scope of ethical concerns in my interviewees and the limitations of some research guidelines that I have highlighted can be evidence. Here, I can use a specific ethical topic to demonstrate such an evolution. When scholars started to use MTurk, they assumed MTurk workers to be totally anonymous. Several years later, Lease et al. (2013) discovered that MTurk workers' identity could be deanonymized because their MTurk ID was linked with their Amazon account, and their real names would be disclosed in email communication with requesters. However, privacy in crowd work was still not considered as a serious problem. Then, in 2017, my empirical surveys on MTurk revealed that privacy-intrusive experiences and concerns were extensive on MTurk in different areas in the world (Xia et al., 2017). More recently, my review has demonstrated that privacy in crowd work has particular threats and challenges that differ from those in other contexts (Xia and Mckernan, 2020). Thus, scholars and IRBs should be cautious and conscious of the evolution of various ethical issues in research on MTurk.

Fourth, MTurk is not a unique crowd work platform anymore, and several innovative crowd work and crowdsourcing platforms are learning from MTurk's limitations and challenging the dominance of MTurk. In particular, Prolific, which is originally created by academic researchers, has earned a growing reputation and popularity in academia. For example, Prolific has a lower commission fee than MTurk to charge requesters and claims to have a larger participant pool than MTurk workers; also, Prolific claims to have a more diverse, "naïve," and representative samples than MTurk and offers a more "ethical" reward to crowd workers ("Prolific | The alternative to MTurk for online survey research," n.d.). Hence, scholars and IRBs should take an evolutionary perspective on the broad landscape of the crowd work market and, as I will argue later, to keep an open mind to try alternative crowd work or crowdsourcing platforms beyond MTurk to conduct academic research.

The second research implication is to enhance communication between crowd workers, researchers, and IRBs. During my dissertation work, I noticed that the communication between crowd workers, researchers, and IRBs is neither frequent nor constructive enough. First, apart from a few scholars whose research is about the empowerment of MTurk workers or concerns about their welfare and risks, many scholars don't have much direct experience or knowledge of MTurk workers. They would treat them as a data source but find no necessity or curiosity to know more about their motivation, need, or perceptions of academic researchers. On the contrary, they would avoid direct communication or "fight" with some MTurk workers. Hence, I propose that scholars need to be more proactive and tolerant to communicate with MTurk workers to enhance mutual understanding and trust. Scholars' time in research is certainly valuable, but I posit that communication between scholars and MTurk workers out of mutual respect instead of fear or indifference is never wasted.

Second, I propose that it would be beneficial if IRBs be more familiar with how MTurk operates and what ethical dilemmas an academic research project on MTurk could face. Knowledge about these aspects of research on MTurk can be further shared and communicated across different IRBs in the U.S. so that each IRB's guidance for research on MTurk would be more consistent and transparent. In this dissertation work, I have noticed that some IRBs have fewer MTurk research applications in their institutions than the others, and the IRB respondents' comprehension and acquaintance with MTurk varied. In the future, it would be helpful to enhance the communication between different IRBs and even their knowledge level of MTurk and crowd work-based research.

Finally, I noticed that scholars and IRBs were not always in agreement with each other or shared the same perspectives on how to respect and be just to MTurk workers as research participants. For example, several scholars in my interviews told me that they did not trust their IRBs' knowledge about MTurk, for example around the reputational risks on MTurk workers. In fact, the IRB respondents in my interviews were quite aware of such risks but they assumed that junior researchers did not know about them. Another example is that some scholars were skeptical about IRBs' awareness of protecting MTurk workers in terms of their earning and informed consent, while some IRB respondents reported their knowledge and efforts to protect workers in these aspects. Even though these examples were not prevalent in my dissertation work, they did reflect a lack of mutual understanding and transparent communication between scholars and IRBs. Hence, I propose that the communication between scholars and IRBs should be enhanced, and it should even be beyond Vitak et al.'s (2016) survey study to the IRBs.

The third research implication is to explore alternative research venues beyond Amazon MTurk and ponder whether collecting data from crowd workers legitimately serves the research

purpose beyond the logistical reasons of fast, convenient, and cheap. First, this implication relates to my arguments about the "original sin" of crowd work-based research. Take MTurk as an example, I encourage scholars to ponder -- besides its convenience for sampling, its "instant-on" nature of data collection, and its cheap labor--is there any other advantage or affordance that would make conducting research on MTurk necessary and meaningful? Further, even if some scholars deny the impact original sin of Amazon or claim that they can eliminate it, I argue that the original sin of Amazon will impact crowd work-based research continuously because the teleological disparity between MTurk and academic research will be ever-lasting and irreconcilable.

Second, as I reported previously, some scholars in my interviews had already transferred to the other more academic or research-oriented platforms, and scholars are also proposing and exploring alternative platforms to MTurk (Chandler et al., 2019; Peer et al., 2017; Vakharia and Lease, 2015). For example, Prolific has a variety of advantages over MTurk, and it is more tuned for academic research ("Prolific | The alternative to MTurk for online survey research," n.d.; Palan and Schitter, 2018). I advocate scholars to explore such alternative venues of MTurk and investigate the quality and validity of these alternative venues further.

Finally, a more fundamental question is whether crowd work or crowdsourcing more broadly should be a suitable means for academic research at all. Their advantages of convenience, speed, ad diversity is obvious, but their implications for the ethical issues in academic research, as partly revealed by this dissertation work, may not be so obvious and deserve further deliberation and research. Here, I venture to propose that, on the one hand, for the type of research that aims for generalizability of research findings to the other contexts or populations, crowd work may not be an ideal or appropriate means to collect research data

because crowd workers may not be representative of many populations. Hence, scholars conducting this type of research should consider departing crowd work and seeking a more representative population for data collection even though it would be at the expense of forsaking the convenient sampling and fast data return via crowd work. On the other hand, for the type of research that focuses on studying the population of crowd workers or the phenomena of crowd work and the gig economy, collecting data from crowd workers and harnessing the advantages of crowd work are legitimate.

Design suggestions

Besides the research implications and based on a previous study about privacy in crowdsourcing (Xia and Mckernan, 2020), I synthesize and propose two design suggestions for crowd work platforms to better support academic research, even though they were not purposed for academic purposes originally.

The first design suggestion is to mark out academic vs. non-academic tasks differently on the platform. This design suggestion was originally proposed by an IRB director, P11, in the case of MTurk. He proposed that to solve the paradox of volunteer-oriented academic research and remuneration-oriented MTurk workers, MTurk could provide some options for MTurk workers to choose from. For those who are motivated by participating in various academic research, they could choose the HITs with a mark of "academic projects;" for those who aim to earn money on MTurk, they could choose the HITs without such a mark. As such, researchers may not need to pay their participants because the MTurk workers who choose to participate would be the ones that are not motivated by the monetary incentive. In his own words, P11 explained:

> I think what might make the most sense is if you [MTurk] have options. If you have an option of "we are doing these 'work activities' over here" and over here "academic

> projects." So, if you give people a choice, they know what they are getting into. You can go here, and here is what being a worker means; when you go over here, you can see what being a research participant means. In the academic land, it should be clear from the start that there is a disclaimer that "You are voluntary, and it is not work, and it is being a research participant." I think if that is clear, that might help the process (P24).

He proposed that such a mechanism of differentiating academic tasks and non-academic tasks on MTurk could maximize the probability of recruiting only research volunteers and avoid cheaters, professional survey takers, or money seekers. I posit that this design suggestion is extrapolatable to the other crowd work platforms similar to MTurk, even though they may not be as popular as MTurk for academic scholars to collect research data. Meanwhile, another limitation of this design would be a potential decrease in the research subjects pool size and diversity on a crowd work platform for academic scholars to choose from because as I have reviewed in Chapter 3, prior research has indicated that most crowd workers are money seekers.

The second design suggestion is to implement a bi-directional and double screening mechanism on to facilitate researchers and crowd workers' screening of each other. This design suggestion is based on Silberman and Irani's (2015)'s description of the basic labor process in MTurk, which is largely the same in the other crowd work platforms. They divide the basic labor process into three main parts: (1) requesters post tasks on MTurk, (2) workers select and do the tasks, and (3) requesters review and approve the completed tasks submitted by the crowd workers (Silberman and Irani, 2016). Based on this model, I propose to implement a double and bi-directional screening within this process to protect both crowd workers and requesters in research on a crowd work platform. From the requesters' direction, the screening should be on both the crowd workers and their submitted tasks, and the screening of crowd workers could be

based on their past performance. A second layer of screening from the requester's direction is to eliminate spamming tasks with certain criteria, such as by implementing attention check questions and random perturbations (Varshney et al., 2014) or by screening based on crowd workers' reputation (Peer et al., 2014).

From the crowd workers' direction, there also should be a double screening mechanism. The first screening is on the requesters. Some third-party tools, such as Turkopticon (Irani and Silberman, 2013), has already demonstrated its promise to deter some unethical requesters and the surreptitious requests (Silberman and Irani, 2016). I advocate that such a reputation system should be embedded in MTurk as well as the other crowd work platforms. The second layer of screening could be on the tasks. For example, this screening mechanism could allow crowd workers to search with some key words for their most interested research tasks such as let them choose between survey studies and information identification tasks.

Policy recommendations

Apart from the research implications and design suggestions, below I propose a few policy recommendations for both academia and industry stakeholders to better support and guide crowd work-based research.

My first policy recommendation is to refine and revise guidance for crowd work-based research. As can be seen in my dissertation findings and discussion, there are various defects in the existing guidance for crowd work-based research. For example, some suggestions such as "simply pay more" and unconditional payment are not accountable for data quality and research validity in the long term. Some consent form templates for MTurk research have different policies about research transparency and about the rejection of unqualified MTurk workers. The research guidelines from researchers, IRBs, academic communities, and organizations also lack

coherence and consistency. Last but not least, the existing guidelines for crowd work-based research are very rare and scant, and many institutions don't have any such guidelines. As such, I propose the following recommendations in more detail.

To start with, research guidelines should be renewed and refined. Several guidelines I found, such as D11 and D12, were published in 2013 and quite dated now. As I proposed above, crowd work platforms and crowd work-based research are both evolving, and IRBs should keep on refining their guidance for crowd work-based research or crowd work-based research more broadly abreast of such evolutions. Second, as I discussed earlier, several guidance items are not accountable and should be revised. The findings from this dissertation can serve as guidance for such revisions, and I plan to propose concrete policy refinements in my future work following this dissertation research, Third, the development of guidelines should involve and incorporate different stakeholders, such as scholars, IRBs, academic communities, and organizations, and even crowd workers where necessary, so that the guidelines are comprehensive and accepted by different stakeholders. Finally, I propose that guidelines from different institutions and entities can vary in some specificities but should remain coherent and consistent in their principles, such as an agreed principle on the transparency of the consent form and the policy about when rejection should be enacted.

My second policy recommendation is to encourage crowd work platforms to develop policies for academic research. Even though a crowd work platform is not designed for academic research, and even though, to some extent, I hold a skeptical and critical attitude on crowd work-based research, yet undeniably, crowd work-based research is still on its rise. Hence, to make it more compatible with and supported by crowd work platforms, I propose that scholars should encourage crowd work platforms to develop their specific policies and regulations for academic

research. For example, although many academic studies are conducted on MTurk, MTurk has very little policy for academic research. This lack of guidance for academic research would render crowd workers to perceive all the tasks on MTurk as if they were the same in nature. Having specific policies for academic research would give crowd workers more knowledge about academic tasks in comparison with non-academics, and also give scholars more support.

To achieve it, as my design suggestions demonstrate, crowd work platforms could consider marking out academic tasks vs. non-academics explicitly on its platform. It could also build sandboxes and templates that simulate academic task features such as the consent form, IRB's role, compensation policy, and academic research purpose. Accordingly, I propose that a crowd work platform can build policies and regulations corresponding to these design ideas when they are developed and materialized. Finally, scholars should ensure the consistency between crowd work platforms' policies for academic research and their guidelines for crowd work-based research. Thus, the whole ecosystem for crowd work-based research will have coherent regulations.

Contributions

This dissertation work makes several contributions to academic research in crowd work and ethics, as well as to the industry of crowd work platforms. Below, I explain the specific theoretical and empirical contributions from this dissertation work.

First, I have identified and discussed the "origin sin" of various ethical issues in crowd work-based research. Prior research has investigated certain aspects of ethics, e.g., privacy, in academic research on MTurk, but has not analyzed or explored the origin of various ethical issues in this research context. This dissertation work can shed light on this area and inspire future theoretical research in the other contexts using a teleological lens, i.e., what are the

implications of the purpose of a specific context or phenomenon for the ethical issues within that context or phenomenon. Meanwhile, this discovery can also guide subsequent empirical studies to explore crowd workers' attitudes toward ethical issues and help crowd work platforms to develop specific policies and design features that can better support academic research. It also reminds scholars to reflect their motive in conducting research on a crowd work platform.

Second, I conducted a pioneering qualitative inquiry into ethics in crowd work-based research. Though Vitak et al. (2016) had surveyed IRB directors and analysts for their ethical considerations in social computing more broadly, my work focused on interviewing IRB directors and analysts for their ethical concerns about academic research in crowd work, which can obtain richer descriptions and a more focused topic about ethics in crowd work-based research from the IRB's perspectives. Meanwhile, to my best knowledge, this dissertation is also pioneering in studying scholars' ethical concerns in this research context and compared them with the IRBs' perspectives. In addition, this dissertation also pioneers in collecting and analyzing the guidelines and templates for academic research on MTurk, which were largely neglected by academia.

Third, through this qualitative inquiry, I have identified the dimensions of the ethical issues in crowd work-based research, which include ethics in payment issues, ethics in data issues, and ethics in human subjects issues. Within each dimension, I have also identified a set of ethical topics, for example, the undue influence and fairness issues in payment ethics, the quality and validity issues in data ethics, and the respect and power dynamics issues in human subjects ethics. These ethical dimensions and topics in crowd work-based research can guide empirical researchers to delve into these ethical issues with crowd workers or within their specific research

discipline. They can also help theorists develop a taxonomy of ethics in crowd work in the future.

Finally, based on my research findings and discussions, I have proposed a set of research implications, design suggestions, and policy recommendations. These implications can help researchers, system designers, and policymakers to refine their practices and develop solutions that will make crowd work-based research more accountable.

Limitations and future work

There are a few limitations in this dissertation work. First, I used MTurk as the epitomic case of crowd work platforms and academic research on MTurk as the representation of crowd work-based research. Even though MTurk is the first and the most representative crowd work platform, and most scholars are choosing it for research, there are still some other crowd work platforms that have been drawing research attention, such as Prolific and Figure. In the future, I will explore the ethical issues in the other growing crowd work platforms and validate my research findings and implications from this dissertation work.

Second, I did not interview MTurk workers to probe their ethical considerations in academic research. For one thing, it is a logistical limitation because researchers are not allowed (though not mandated by MTurk) to collect MTurk workers' personal information such as their email or phone number. For the other, I contended that scholars and IRBs are more familiar with ethical issues in academic research than MTurk workers. In fact, I assume that MTurk workers cannot easily differentiate academic tasks vs. non-academic tasks on MTurk, and several scholars and IRB respondents shared this viewpoint. However, in the future, I still plan to investigate MTurk workers' knowledge and experience with ethical issues in academic research. Some academic researchers have tried to get around with collecting personal information by

inviting MTurk workers to an online chat room to conduct a text-based interview (Sannon and Cosley, 2019), I may reference this practice to interview MTurk workers. I may also administer a survey with a series of open-ended questions to MTurk workers as I did previously in a privacy study (Xia et al., 2017).

Third, many researchers and IRB respondents were not from the same institution in this dissertation work. As I described in the Methodology Chapter, I had tried to recruit scholars and IRB respondents in the same institutions, but many of them never responded to my invitations. Interviewing scholars and IRB staffers at the same institution may generate more institutionally specific ethical considerations and comparisons. However, so far, I did not see such a pattern either among the scholars or the IRB respondents. In the future, I plan to conduct a survey study to both scholars and IRB staffers in the same institutions to explore institutional differences in interpreting ethics in crowd work-based research.

Last but not least, ethics is complicated in nature. By no means, I intend to claim that my research findings of ethics in crowd work-based research are authoritative or prescriptive. Rather, I wish my findings to be illuminating and reflexive. I quite agree and follow AoIR IRE 3.0's suggestion that "each and every point is open for debate and ethics is an ongoing process" (Franzke et al., 2020, p. 2). The findings and implications from this dissertation work about ethics in crowd work-based research are also open for deliberation. In the future, I plan to extend my research lens in two directions. First, I plan to extend it from academic ethics to business ethics, and second, I plan to extend it from crowd work to the gig-economy.

Conclusion

Ethics in academic research has been a center of deliberation for decades. The Belmont Report, AoIR IRE, and the ethical codes in various research disciplines such as ACM and APA

have all embodied and contributed to such ethical deliberation. Ethics in crowd work-based research, such as that in the context of MTurk, by contrast, has not been studied extensively. To fill this gap, I interviewed 17 IRB directors and analysts and 15 scholars in the U.S. I also analyzed 15 research guidelines and 14 publications about academic research on MTurk. Through an inductive and deductive analysis of these data, I identified three dimensions of ethical issues in crowd work-based research: payment, data, and human subjects. Within each dimension, I also identified a set of ethical topics such as the undue influence and fairness issues in payment ethics, the quality and validity issues in data ethics, and the respect and power dynamics issues in human subjects ethics.

Based on these research findings, first, I discussed the origin and implication of each ethical dimension in crowd work-based research. Then, I discussed my research questions about how scholars and IRB respondents translated the Belmont principles into their ethical considerations and practices in crowd work-based research. Finally, I identified the "original sin" of ethics in this context, which includes Amazon's fundamental problematic stance of crowd workers, its confusion of terminology, as well as its abdication of responsibilities between MTurk, MTurk workers, and researchers.

My dissertation makes both theoretical and empirical contributions. Theoretically, it has identified and discussed the original problems of various ethical challenges in crowd work-based research and discussed the ontological schisms between IRBs and researchers and different philosophical stances of ethics in this context. Empirically, it has identified the dimensions and specific topics of ethics in crowd work-based research and proposed a set of research, design, and policy implications for scholars, IRBs, crowd work system designers, and policymakers to

reference. Meanwhile, my dissertation is a pioneering qualitative inquiry into the ethical issues in crowd work-based research.

APPENDIX

Interview Questions on Ethical Issues in Crowd Work-based Research

(To Academic Researchers)

General Questions

1. Could you introduce a bit more about yourself?

    Probing: how do you define your research community, e.g., conferences you go to?

2. When did you start to use MTurk to gather research data?

3. What motivated you to use MTurk to gather research data?

Questions related to the IRB review application for research on MTurk

1. What are your inclusion and exclusion criteria to recruit MTurk workers?

    Probing: Why do you choose these inclusion and exclusion criteria?

2. How do you ensure that your research participants on MTurk have been informed about your study?

    Probing: have you used any strategy to ensure MTurk workers read consent form?

3. How much did you pay the MTurk workers in this study?

4. What was your payment standard?

Probing: Why did you choose this standard?

5. How do you ensure that your payment would not coerce or impose undue influence on MTurk workers to participate in your study?

6. How do you perceive privacy issues in research on MTurk?

8. What are the risks to MTurk workers in research on MTurk?

9. What are the benefits to MTurk workers in research on MTurk?

Probing 1): How do you calculate or assess the balance or ratio of research risks and benefits to MTurk workers?

Probing 2): Are there any uncommon risks or benefits to MTurk workers that you have ever encountered? If so, what are they?

10. Has the IRB office or review board in your institution ever changed or reshaped your research plan in any of your study on MTurk?

If yes, how has it reshaped or changed your research plan?

Questions beyond the IRB and MTurk

1. What ethical issues do you see as a potential challenge for crow work-based research?

2. How do you compare academic research ethics in general and crowd work-based research ethics in particular? For example, what are their commonalities and differences?

4. What is your fundamental ethical stance in conducting research on MTurk?

Probing: for example, how do you compare and evaluate between Kantian ethics and utilitarian ethics in the context of MTurk?

Interview Questions on Ethical Issues in Crowd Work-based Research

(To IRB Directors and Analysts)

General Questions

1. How long have you been on the IRB committee at this institution?

2. How often do you review crowd work-based research, such as that is conducted on Amazon Mechanical Turk (MTurk)?

Probing: are they exempt, expedited, or full board review?

Probing: what research disciplines (departments) do they usually come from?

Questions related to the IRB review application for research on MTurk

1. What are your criteria to decide whether crowd work-based research (such as that is conducted on MTurk) should be exempt, versus expedited, versus full-board reviewed?

2. How did you evaluate the payment in a crowd work-based study and what was your standard or benchmark for approving this payment?

3. What measures did you require an application for crowd work-based research to take to protect crowd workers, such as MTurk workers' privacy?

4. What is the possibility of coercion or undue influence for a study on MTurk?

Probing: how do you assess whether a monetary incentive in a study on MTurk is coercion (as paying too much) or exploitation (as paying too little)

5. How do you calculate or assess the balance between the risks and benefits to MTurk workers in a study conducted on MTurk?

Questions about the Belmont principles and beyond

1. How do you compare academic research ethics in general and crowd work-based research ethics in particular, for example, what are their commonalities and differences?

2. How do you think about the IRB principle of *respect for persons* to crowd work-based research such as that on MTurk

3. How do you think about the IRB principle of *beneficence* to crowd work-based research such as that on MTurk

4. How do you think about the IRB principle of *justice* to crowd work-based research such as that on MTurk

5. What is your fundamental ethical stance in conducting research on MTurk?

Probing: for example, how do you compare and evaluate between Kantian ethics and utilitarian ethics in the context of MTurk?

REFERENCES

ACM, C. M. (1992). ACM code of ethics and professional conduct. *Code of Ethics*.

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce* (pp. 21-29). ACM.

Alam, S. L., & Campbell, J. (2012, January). Crowdsourcing motivations in a not-for-profit GLAM context: the Australian newspapers digitization program. In *ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012*. ACIS, 1-11.

Alkhatib, A., Bernstein, M. S., & Levi, M. (2017, May). Examining crowd work and gig work through the historical lens of piecework. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4599-4616).

Aloisi, A. 2015. Commoditized Workers. The Rising of On-Demand Work, a Case Study Research on a Set of Online Platforms and Apps. In *4th Conference of the Regulating for Decent Work Network*.

Amazon Mechanical Turk. (n.d.-a). Retrieved April 20, 2020, from https://www.mturk.com/worker/help?helpPage=policies

Amazon Mechanical Turk. (n.d.-b). Retrieved April 20, 2020, from https://www.mturk.com/worker/participation-agreement

Amazon Mechanical Turk. (n.d.-c). Retrieved April 21, 2020, from https://www.mturk.com

Amazon Mechanical Turk "Workers" are not anonymous. (2013). Retrieved February 26, 2020, from https://research-compliance.umich.edu/sites/default/files/resource-download/irb-hsbs_newsletter_spring_2013.pdf

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American psychologist*, *57*(12), 1060-1073.

Application for Review of Research Involving Human Subjects. (2013). Retrieved February 26, 2020, from http://mdbailey.ece.illinois.edu/irb/usb_irb_initial.pdf

Aruguete, M. S., Huynh, H., Browne, B. L., Jurs, B., Flint, E., & McCutcheon, L. E. (2019). How serious is the 'carelessness' problem on Mechanical Turk?. *International Journal of Social Research Methodology*, *22*(5), 441-449.

Bailey, M., Dittrich, D., Kenneally, E., & Maughan, D. (2012). The Menlo report. *IEEE Security & Privacy*, *10*(2), 71-75.

Bailey, A. F. (2017). Amazon Mechanical Turk IRB Considerations. Retrieved from https://www.uwsp.edu/acadaff/orsp/Documents/Bailey_SBER-mTurk-slides-Stanford-University.pdf

Beecher, H. K. (1966). Ethics and clinical research. In *Biomedical ethics and the law* (pp. 215-227). Springer, Boston, MA.

Bell, C. M., & Khoury, C. (2011). Organizational de/humanization, deindividuation, anomie, and in/justice. In S. W. Gilliland, D. D. Steiner, & D. P. Skarlicki (Eds.), *Research in social issues in management. Emerging perspectives on organizational justice and ethics* (p. 167–197). IAP Information Age Publishing.

Bentham, J., & Mill, J. S. (2004). *Utilitarianism and other essays*. Penguin UK.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2011). Using Mechanical Turk as a subject recruitment tool for experimental research.

Bergvall-Kåreborn, B., & Howcroft, D. (2014). Amazon Mechanical Turk and the commodification of labour. *New Technology, Work and Employment*, *29*(3), 213-223.

Bezos, J. 2006. Opening Keynote. MIT Emerging Technologies Conference. (2006). http://video.mit.edu/watch/opening-keynote-andkeynote-interview-with-jeff-bezos-9197/

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., ... & Yeh, T. (2010, October). VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (pp. 333-342). ACM.

Binik, Y. M., Mah, K., & Kiesler, S. (1999). Ethical issues in conducting sex research on the Internet. *Journal of Sex Research*, *36*(1), 82-90.

Boser, S. (2007). Power, ethics, and the IRB: Dissonance over human participant review of participatory research. *Qualitative Inquiry*, *13*(8), 1060-1074.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Brabham, D. C. (2010). Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, *13*(8), 1122-1145.

Brabham, D. C. (2012). Motivations for participation in a crowdsourcing application to improve public engagement in transit planning. *Journal of Applied Communication Research*, *40*(3), 307-328.

Brabham, D. C. 2013. Crowdsourcing. MIT Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3(2),* 77-101.

Brewer, R., Morris, M. R., & Piper, A. M. (2016, May). Why would anybody do this? Understanding older adults' motivations and challenges in crowd work. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2246-2257.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science, 6(1),* 3-5.

Buhrmester, M. D. (2018). M-Turk Guide. Retrieved from https://michaelbuhrmester.com/mechanical-turk-guide/

Busarovs, A. (2013). Open innovation: current trends and future perspectives. *Humanities and Social Sciences*, *21*(2), 103-119.

Callison-Burch, C. (2009, August). Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 286-295.

Capraro, V., Rodriguez-Lara, I., & Ruiz-Martos, M. J. (2020). Preferences for efficiency, rather than preferences for morality, drive cooperation in the one-shot Stag-Hunt Game. *Journal of Behavioral and Experimental Economics*, https://doi.org/10.1016/j.socec.2020.101535.

Carlson, R. V., Boyd, K. M., & Webb, D. J. (2004). The revision of the Declaration of Helsinki: past, present and future. *British journal of clinical pharmacology*, *57*(6), 695-713.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156-2160.

Chandler, J., Mueller, P., & Paolacci, G. (2012). Non-naïvety Among Experimental Participants on Amazon Mechanical Turk. ACR North American Advances.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual review of clinical psychology*, *12*.

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior research methods*, *51*(5), 2022-2038.

Chen, W. C., Suri, S., & Gray, M. L. (2019, July). More Than Money: Correlation among Worker Demographics, Motivations, and Participation in Online Labor Market. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 13, No. 01, pp. 134-145).

Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. NYU Press.

Code, N. (1947). Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10. vol. 2. Washington, DC: US Government Printing Office, 1949: 181–2.

Committee for Protection of Human Subjects. (2020). Mechanical Turk (MTurk) for Online Research. Retrieved from https://cphs.berkeley.edu/mechanicalturk.pdf

Consent & Recruitment for mTurk population. (2015). Retrieved from https://www.research.colostate.edu/ricro/irb/templates/recruitment/

Crawley F. P. (2003). The Limits of the Declaration of Helsinki. Address to Scientific Session, World Medical Association General Assembly, September 2003, Helsinki.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice, 39(3),* 124-130.

De Stefano, V. (2015). The rise of the just-in-time workforce: On-demand work, crowdwork, and labor protection in the gig-economy. *Comp. Lab. L. & Pol'y J.*, *37*, 471.

Deng, X. N., & Joshi, K. D. (2013). Is crowdsourcing a source of worker empowerment or exploitation? Understanding crowd workers' perceptions of crowdsourcing career.

Deng, X., Joshi, K. D., & Galliers, R. D. (2016). The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. *Mis Quarterly*, *40*(2), 279-302.

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2019). Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures. *Behavioral Research in Accounting*.

Difallah, D. E., Demartini, G., & Cudré-Mauroux, P. (2012, April). Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms. In *CrowdSearch*. 26-30.

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

Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February). Demographics and dynamics of mechanical Turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*, 135-143.

Dreyfuss, E. (2018, November 28). A Bot Panic Hits Amazon Mechanical Turk. Retrieved from https://www.wired.com/story/amazon-mechanical-turk-bot-panic/

Durward, D., Blohm, I., & Leimeister, J. M. (2016, July). Is There PAPA in Crowd Work?: A Literature Review on Ethical Dimensions in Crowdsourcing. In *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress, 2016 Intl IEEE Conferences* 823-832.

Edgar, H., & Rothman, D. J. (1995). The institutional review board and beyond: future challenges to the ethics of human experimentation. *The Milbank Quarterly*, 489-506.

Eickhoff, C., & De Vries, A.P. (2013). Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval* 16, 121-137. https://doi.org/10.1007/s10791-011-9181-9

Engle, K., Talbot, M., & Samuelson, K. W. (2019). Is Amazon's Mechanical Turk (MTurk) a comparable recruitment source for trauma studies?. *Psychological Trauma: Theory, Research, Practice, and Policy*.

Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, *38*(2), 189-200.

Ess, C. (2002). Ethical decision-making and internet research: Recommendations from the aoir ethics working committee. *Readings in virtual research ethics: Issues and controversies*, 27-44.

Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *BMJ*. *323*(7321), 1103-1105. Doi: 10.1136/bmj.323.7321.1103

Faraj, S., & Azad, B. (2012). The materiality of technology: An affordance perspective. *Materiality and organizing: Social interaction in a technological world*, *237*, 258.

Fest, S., Kvaloy, O., Nieken, P., & Schöttner, A. (2019). Motivation and incentives in an online labor market. CESifo Working Paper, No. 7526, Center for Economics Studies and Ifo Institute (CESifo), Munich.

Ford, J. B. (2017). Amazon's Mechanical Turk: a comment. *Journal of Advertising*, *46*(1), 156-158.

Frankel, M. S., & Siang, S. (1999). Ethical and legal aspects of human subjects research on the internet. *Published by AAAS online*.

Franzke, Aline Shakti, Bechmann, Anja, Zimmer, Michael, Ess, Charles and the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0*. https://aoir.org/reports/ethics3.pdf

Friesen, P., Kearns, L., Redman, B., & Caplan, A. L. (2017). Rethinking the Belmont report?. *The American Journal of Bioethics*, *17*(7), 15-21.

Gadiraju, U., Kawase, R., & Dietze, S. (2014, September). A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 218-223).

Gadiraju, U., Kawase, R., Dietze, S., & Demartini, G. (2015, April). Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1631-1640.

Gardner, R. M., Brown, D. L., & Boice, R. (2012). Using Amazon's Mechanical Turk website to measure accuracy of body size estimation and body dissatisfaction. *Body image*, *9*(4), 532-534.

Gibbs, G. R. (2007). Analyzing qualitative data. Sage.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213-224.

Gottlieb, L., Choi, J., Kelm, P., Sikora, T., & Friedland, G. (2012, October). Pushing the limits of mechanical turk: qualifying the crowd for video geo-location. In *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia*. ACM, 23-28.

Graber, M. A., & Graber, A. (2012). Internet-based crowdsourcing and research ethics: the case for IRB review. *Journal of medical ethics, medethics*. 2012.

Guidelines: Human Research Protections and Amazon mTurk. (n.d.). Retrieved February 27, 2020, from https://uindy.edu/human-research-protections-program/files/guidelines_for_mturk_at_uindy_092418.pdf

Guidelines for Academic Requesters. (n.d.). Retrieved February 27, 2020, from https://wearedynamo.fandom.com/wiki/Guidelines_for_Academic_Requesters

Gupta, N., Martin, D., Hanrahan, B. V., & O'Neill, J. (2014, November). Turk-life in India. In *Proceedings of the 18th International Conference on Supporting Group Work* (pp. 1-11). ACM.

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, *57*(5), 664-688.

Hansson, K., Aitamurto, T., Ludwig, T., Muller, M., & Hansson, I. K. (2016). From abdication to relation: Modes of production in crowd work. In: *Proceedings of the CSCW 2016 – Workshop: Toward a Typology of Participation in Crowdwork,* 13(1), 13-22.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018, April). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* 1-14.

Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63-76.

Harris, C. G. 2011. Dirty deeds done dirt cheap: a darker side to crowdsourcing. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. IEEE, 1314–1317.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and social psychology review*, *10*(3), 252-264.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. Suny Press.

Haug, M. C. (2018). Fast, Cheap, and Unethical? The Interplay of Morality and Methodology in Crowdsourced Survey Research. *Review of Philosophy and Psychology*, *9*(2), 363-379.

Higgins, C., McGrath, E., & Moretto, L. (2010, June). MTurk crowdsourcing: a viable method for rapid discovery of Arabic nicknames? In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, 89-92.

Hitlin, P. (2016). Research in the crowdsourcing age, a case study. Pew Research Center. http://www.pewinternet.org/2016/07/11/research-in-the-crowdsourcing-age-a-case-study.

Ho, C. J., Slivkins, A., Suri, S., & Vaughan, J. W. (2015, May). Incentivizing high quality crowdwork. In Proceedings of the 24th International Conference on World Wide Web(pp. 419-429).

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, *14*(3), 399-425.

Horton, J. J., & Chilton, L. B. (2010, June). The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*. ACM, 209–218.

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

IRB Consent Example_Amazon MTurk study. (n.d.). Retrieved February 25, 2020, from https://www.furman.edu/institutional-review-board/information-for-researchers/consent-forms/

IRB Guidelines and Suggestions for Using Mechanical Turk (MTurk) for Social/Behavioral Research Projects. (2015). Retrieved February 25, 2020, from https://research.utexas.edu/wp-content/uploads/sites/3/2015/10/mechanical_turk.pdf

Ipeirotis, P. G. (2010a). Demographics of mechanical turk.

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

Ipeirotis, P. G., Provost, F., & Wang, J. (2010, July). Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*. ACM, 64–67.

Irani, L. C., & Silberman, M. (2013, April). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 611-620). ACM.

Marx, K. (1867). 1976. *Capital: A critique of political economy*, *1*.

Kandappu, T., Friedman, A., Sivaraman, V., & Boreli, R. (2015). Privacy in crowdsourced platforms. In *Privacy in a Digital, Networked World* (pp. 57-84). Springer, Cham.

Kandappu, T., Sivaraman, V., Friedman, A., & Boreli, R. (2014, January). Loki: A privacy-conscious platform for crowdsourced surveys. In *COMSNETS* (pp. 1-8).

Kant, I. (1964). Groundwork of the metaphysic of morals (HJ Paton, Trans.). *NY: Harper & Row*.

Kant, I. (1995). Kant: Ethical Philosophy: Grounding for the Metaphysics of Morals, and, Metaphysical Principles of Virtue, with, "On a Supposed Right to Lie Because of Philanthropic Concerns". *Hackett Publishing*.

Kaplan, T., Saito, S., Hara, K., & Bigham, J. P. (2018, June). Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

Katz, M. (2017). Amazon's Turker Crowd Has Had Enough. Retrieved April 22, 2020, from https://www.wired.com/story/amazons-turker-crowd-has-had-enough/

Kaufmann, N., Schulze, T., & Veit, D. (2011, August). More than fun and money. Worker Motivation in Crowdsourcing-A Study on Mechanical Turk. In *AMCIS (Vol. 11, No. 2011,* pp. 1-11).

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017a). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141-155.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017b). Reply to "Amazon's Mechanical Turk: A Comment". *Journal of Advertising*, *46*(1), 159-162.

Keller, H. E., & Lee, S. (2003). Ethical issues surrounding human participants research using the Internet. *Ethics & behavior*, *13*(3), 211-219.

Kidder, L., & Judd, C. (1986). Research methods in social science. New York: CBS College

Publishing.

King, P. A. 2005. Justice beyond Belmont. In Belmont revisited: Ethical principles for research with human subjects, ed. J. F. Childress, E. M. Meslin, and H. T. Shapiro, 136–47. Washington, DC: Georgetown University Press.

Kittur, A., Chi, E. H., & Suh, B. (2008, April). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453-456.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., ... & Horton, J. (2013, February). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1301-1318). ACM.

Klitzman, R. (2013). How IRBs view and make decisions about coercion and undue influence. *Journal of Medical Ethics*, *39*(4), 224-229.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American psychologist, 59(2), 105*.

Kuek, S. C., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., Ipeirotis, P., Pina, P., & Singh, M. (2015). *The Global Opportunity in Online Outsourcing*. Washington: World Bank.

Largent, E., Grady, C., Miller, F. G., & Wertheimer, A. (2013). Misconceptions about coercion and undue influence: reflections on the views of IRB members. *Bioethics*, *27*(9), 500-507.

Lasecki, W. S., Teevan, J., & Kamar, E. (2014, February). Information extraction and manipulation threats in crowd-powered systems. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 248-256). ACM.

Lasecki, W. S., Gordon, M., Leung, W., Lim, E., Bigham, J. P., & Dow, S. P. (2015a). Exploring privacy and accuracy trade-offs in crowdsourced behavioral video coding. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1945-1954). ACM.

Lasecki, W. S., Teevan, J., & Kamar, E. (2015b). The cost of asking crowd workers to behave maliciously. In *Proc. the AAMAS Workshop on Human-Agent Interaction Design and Models*.

Lease, M., Hullman, J., Bigham, J., Bernstein, M., Kim, J., Lasecki, W., ... & Miller, R. (2013). Mechanical turk is not anonymous. *SSRN Electronic Journal.* Retrieved from https://ssrn.com/abstract=2228728.

Lee, K., Webb, S., & Ge, H. (2014). The dark side of micro-task marketplaces: Characterizing fiverr and automatically detecting crowdturfing. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *Sage Open*, *6*(1), 2158244016636433.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry (Vol. 75). Sage.

Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among U.S.-and India-based workers on Mechanical Turk. *Behavior research methods*, *47*(2), 519-528.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods*, *49*(2), 433-442.

Liu, D., Bias, R. G., Lease, M., & Kuipers, R. (2012). Crowdsourcing for usability

testing. *Proceedings of the Association for Information Science and Technology, 49(1),* 1-10.

Lovett, M., Bajaba, S., Lovett, M., & Simmering, M. J. (2018). Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of amazon's mechanical turk masters. *Applied Psychology*, *67*(2), 339-366.

Macklin R. (2003). Future challenges for the Declaration of Helsinki: Maintaining Credibility in the Face of Ethical Controversies. Address to Scientific Session, World Medical Association General Assembly, September 2003, Helsinki.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. *MIT Sloan Management Review*, *51*(3), 21.

Markham, A., Buchanan, E., & AoIR Ethics Working Committee. (2012). Ethical decision-making and Internet research: Version 2.0. Association of Internet Researchers.

Marge, M., Banerjee, S., & Rudnicky, A. I. (2010, March). Using the Amazon Mechanical Turk for transcription of spoken language. In *Processing (ICASSP), 2010 IEEE International Conference on Acoustics Speech and Signal*. IEEE, 5270-5273.

Marshall, C., & Rossman, G. B. (1989). Designing qualitative research. Sage publications.

Martin, D. G. (2007). Bureaucratizing ethics: Institutional review boards and participatory research. *ACME: An International E-Journal for Critical Geographies, 6*(3), 319-328.

Mason, W., & Watts, D. J. (2009, June). Financial incentives and the performance of crowds. In *Proceedings of the ACM SIGKDD workshop on human computation*. ACM, 77-85.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, *44*(1), 1-23.

Merriam, S. B. (1988). Case study research in education: A qualitative approach. Jossey-Bass.

Mohammad, S. M., & Turney, P. D. (2010, June). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics, 26-34.

MTurk Guidance. (2019). Retrieved from https://www.umass.edu/research/guidance/mturk-guidance

Murray, C. D., & Sixsmith, J. (1998). E-mail: a qualitative research medium for interviewing? *International Journal of Social Research Methodology, 1*(2), 103-121.

Nagrale, P. (2018). Top 12 Crowdsourcing Sites Like MTurk to Find High Paying Micro Jobs. Retrieved May 22, 2020, from https://moneyconnexion.com/top-10-crowdsourcing-sites-like-mturk.htm

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *Report and Recommendations, Institutional Review Boards*. U.S. Government Printing Office.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: ethical principles and guidelines for the protection of human subjects of research* (Vol. 1). Department of Health, Education, and Welfare, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: ethical principles and guidelines for the protection of human subjects of research* (Vol. 2). Department of Health, Education, and Welfare,

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Bethesda, MD.

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27.

Palmer, J. C., & Strickland, J. (2016, June). A beginner's guide to crowdsourcing Strengths, limitations and best practices for psychological research. Retrieved from https://www.apa.org/science/about/psa/2016/06/changing-minds

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Peer, E., Paolacci, G., Chandler, J., & Mueller, P. (2012). Selectively recruiting participants from Amazon Mechanical Turk using qualtrics. *Available at SSRN 2100631*.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods*, *46*(4), 1023-1031.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163.

PERVADE. (2017). Retrieved April 20, 2020, from https://pervade.umd.edu

Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological bulletin*, *123*(3), 238.

Prolific | Online participant recruitment for surveys and market research. (n.d.). Retrieved April 20, 2020, from https://www.prolific.co

Prolific | The alternative to MTurk for online survey research. (n.d.). Retrieved April 20, 2020, from https://www.prolific.co/prolific-vs-mturk/

Quinn, A. J., & Bederson, B. B. (2011, May). Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1403-1412). ACM.

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of business ethics*, *10*(11), 805-817.

Rawls, J. (1971). A theory of justice. Cambridge. Mass.: Harvard University.

Rawls, J. (2009). A theory of justice: Revised edition. Harvard university press.

Reid, E. (1996). Informed consent in the study of on-line communities: A reflection on the effects of computer-mediated social research. *The Information Society, 12*(2), 169-174.

Rhodes, R. (2010). Rethinking research ethics. *The American Journal of Bioethics*, *10*(10), 19-36.

Rogers, W., & Lange, M. M. (2013). Rethinking the vulnerability of minority populations in research. *American Journal of Public Health*, *103*(12), 2141-2146.

Russell, B. (2013). *History of western philosophy: Collectors edition*. Routledge.

Salehi, N., Irani, L. C., Bernstein, M. S., Alkhatib, A., Ogbe, E., & Milland, K. (2015, April). We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1621-1630). ACM.

Saito, S., Chiang, C. W., Savage, S., Nakano, T., Kobayashi, T., & Bigham, J. P. (2019, May). Turkscanner: Predicting the hourly wage of microtasks. In *The World Wide Web Conference* (pp. 3187-3193).

Sandel, M. J. (2010). *Justice: What's the right thing to do?*. Macmillan.

Sannon, S., & Cosley, D. (2018, April). It was a shady HIT: Navigating Work-Related Privacy Concerns on MTurk. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM.

Sannon, S., & Cosley, D. (2019, May). Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*, (1), 93-107.

Schmidt, G. B., & Jettinghoff, W. M. (2016). Using Amazon Mechanical Turk and other compensated crowdsourcing sites. *Business Horizons*, *59*(4), 391-400.

Schulze, T., Seedorf, S., Geiger, D., Kaufmann, N., & Schader, M. (2011). Exploring task properties in crowdsourcing–An empirical study on Mechanical Turk. In: *ECIS 2011 Proceedings of the 19th European Conference on Information Systems*.

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs, 85(1),* 140-156.

Sheehan, K. B., & Pittman, M. (2016). *Amazon's Mechanical Turk for academics: The HIT handbook for social science research*. Melvin & Leigh, Publishers.

Shore, N. 2006. Re-conceptualizing the Belmont report: A community-based participatory research perspective. *Journal of Community Practice 14 (4)*, 5–26.

Shuster, E. (1997). Fifty years later: the significance of the Nuremberg Code. *New England Journal of Medicine*, *337*(20), 1436-1440.

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

Silberman, M., & Irani, L. (2016). Operating an employer reputation system: lessons from Turkopticon, 2008-2015. *Comparative Labor Law & Policy Journal, Forthcoming*.

Silberman, M. S., Tomlinson, B., LaPlante, R., Ross, J., Irani, L., & Zaldivar, A. (2018). Responsible research with crowds: pay crowdworkers at least minimum wage. *Communications of the ACM*, *61*(3), 39-41.

Singer, E., & Couper, M. P. (2008). Do incentives exert undue influence on survey participation? Experimental evidence. *Journal of Empirical Research on Human Research Ethics*, *3*(3), 49-56.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: measuring individuals' concerns about organizational practices. *MIS quarterly*, 167-196.

Snow, R., O'connor, B., Jurafsky, D., & Ng, A. Y. (2008, October). Cheap and fast–but is it good? evaluating non-expert annotations for natural language tasks. In *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 254-263).

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods*, *43*(1), 155-167.

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in cognitive sciences*, *21*(10), 736-748.

Strickland, J. C., & Stoops, W. W. (2019). The use of crowdsourcing in addiction science research: Amazon Mechanical Turk. *Experimental and clinical psychopharmacology*, *27*(1), 1.

Suomela, T., Chee, F., Berendt, B., & Rockwell, G. (2019). Applying an Ethics of Care to Internet Research: Gamergate and Digital Humanities. *Digital Studies/Le champ numérique*, *9*(1).

Surowiecki, J. 2005. *The wisdom of crowds*. Anchor.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cup Archive.

The model MTurk consent form. (2019). Retrieved from https://campusirb.duke.edu/resources/samples/mturk-and-other-online-consent

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184-197.

Thomson Reuters Foundation. (2019, August 2). India passes "historic" minimum wage law amid activist worries. Retrieved from https://news.trust.org/item/20190802170845-5q2uq/

TurkerNation. (n.d.). Retrieved April 20, 2020, from http://www.turker-nation.com/

US Department of Health and Human Services. (2009). Code of Federal Regulations Title 45 Public Welfare Department of Health and Human Services Part 46: Protection of Human Services.

Use of Amazon Mechanical Turk. (2020). Retrieved from https://www.research.iastate.edu/use-amazon-mechanical-turk-iowa-state-university-research/

Vakharia, D., & Lease, M. (2015). Beyond Mechanical Turk: An analysis of paid crowd work platforms. *Proceedings of the iConference*.

Varshney, L. R., Vempaty, A., & Varshney, P. K. (2014, February). Assuring privacy and reliability in crowdsourcing with coding. In *Information Theory and Applications Workshop (ITA), 2014* (pp. 1-6). IEEE.

Vaughan, J. W. (2017). Making better use of the crowd: How crowdsourcing can advance machine learning research. *The Journal of Machine Learning Research*, *18*(1), 7026-7071.

Vitak, J., Shilton, K., & Ashktorab, Z. (2016, February). Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 941-953.

Von Ahn, L. (2008, April). Human computation. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering* (pp. 1-2). IEEE Computer Society.

Walsham, G. (1996). Ethical theory, codes of ethics and IS practice. *Information Systems Journal, 6(1),* 69-81.

Walther, J. B. (2002). Research ethics in Internet-enabled research: Human subjects issues and methodological myopia. *Ethics and information technology*, *4*(3), 205-216.

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2012, April). Serf and turf: crowdturfing for fun and profit. In *Proceedings of the 21st international conference on World Wide Web* (pp. 679-688). ACM.

Williamson, V. (2016). On the ethics of crowdsourced research. *PS: Political Science & Politics, 49(1),* 77-81.

Wolff, J. (1999). Marx and exploitation. *The Journal of Ethics*, *3*(2), 105-120.

Workers in Majority of U.S. States to See an Increase in Minimum Wage in 2020. (2019). Retrieved March 5, 2020, from https://wolterskluwer.com/company/newsroom/news/2019/12/workers-in-majority-of-us-states-to-see-an-increase-in-minimum-wage-in-2020.html

World Medical Association. (1996). Declaration of Helsinki (1964). *BMJ*, *313*(7070), 1448-1449.

Xia, H., Huang, Y., Wang, Y., and Anuj, S. 2017. "Our Privacy Needs to Be Protected at All Costs": Crowd Workers' Privacy Experiences on Mechanical Turk. *Proceedings of ACM Human-Computer Interaction 1, 2 (2017).*

Xia, H., Mckernan, B. (2020), Privacy in Crowdsourcing: A Review of the Threats and Challenges. *Computer-Supported Cooperative Work and Social Computing (CSCW):1-39. Springer*. Doi: https://doi.org/10.1007/s10606-020-09374-0

Yin, M., Chen, Y., & Sun, Y. A. (2014, September). Monetary interventions in crowdsourcing task switching. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

VITA

HUICHUAN XIA

Ph.D. in Information Science and Technology
Email: hxia@syr.edu
Personal website: https://huichuanxia.com

Research Interests: Ethics, Privacy, Crowd work, Gig-economy, Computer-Supported Cooperative Work (CSCW), Human-Computer Interaction (HCI), Intelligent Analysis.

EDUCATION

Ph.D., Information Science and Technology (*2013 - 2020*)
Syracuse University, Syracuse, NY, USA

- *Dissertation:* A Study of Ethics in Crowd Work-based Research
- *Committee:* Dr. Jennifer Stromer-Galley (Advisor), Dr. Caroline Haythornthwaite, Dr. Ingrid Erickson, Dr. Carsten S. Østerlund, Dr. Jeffrey M. Stanton (Internal Reader), Dr. Michael Zimmer (External Reader), Dr. Lael Schooler (Chair)

M.S., Information Management (*2008 - 2010*)
Peking University, Beijing, China

- *Thesis:* Case Study of Knowledge Sharing in Online Communities of Practice
- *Committee:* Dr. Yanli Qi (Advisor), Dr. Jun Wang, Dr. Yanfei Wang, Dr. Xueshan Yan

B.S., Computer Science (*2002 - 2006*)
Central China Normal University, Wuhan, China

- *Thesis: A Web-Crawler System based on Lucene*
- *Advisor:* Dr. Tingting He

PROFESSIONAL EXPERIENCE

Sep. 2019 – May 2020
Adjunct, *School of Information Studies, Syracuse University*

Independently teach IST 343 *Data in Society* to undergraduate students from different schools and departments across Syracuse University

Jun. 2019 – Aug. 2019

Research Intern, *Privacy Research Team, Facebook, Inc. Menlo Park*

- First UX Research Intern in the Privacy Research Team at Facebook
- Independently conducted a user study to investigate Facebook users' mental models of privacy notifications
- Independently devised a framework for privacy education on Facebook users.

Jun. 2012 – Aug. 2013

Market Intelligence Assistant Manager, *Renault (China) Automotive Co., Ltd.*

- Collected and analyzed competitive intelligence in the auto market in China
- Conducted research on Renault models' marketing plan, e.g., new car launch analysis

Jun. 2011 – Jun. 2012

Market Research Project Manager, *Beijing Business-Wing Management Consulting Co., Ltd.*

- Conducted survey, in-depth interviews, and focus groups with 20 Volkswagen dealers
- Controlled project budget, time, and quality

Jun. 2010 – Jun. 2011

Senior Trainee, *Beijing Bester Advertisement & Media Co., Ltd.*

- Director's trainee, rotated in sales, media buying and research departments
- Involved in many large projects e.g. Chang-An Auto's yearly advertising plan

Jan. 2010 – Apr. 2010

Market Intern, *MOMA (Beijing) Co., Ltd.*

- Reported directly to the company's director
- Conducted research on the real-estate market in China

Jun. 2006 – Jun. 2008

Salesman, *China Taiping Life Insurance Company (Beijing).*

- Made cold call and visit
- Sold life insurance products

AWARDS & RECOGNITIONS

* Recipient of the Inaugural Summer Dissertation Fellowship from Graduate School of Syracuse University (2019) **[Top 5%]**

* Graduate Assistantship (GA) in Syracuse University (2013 – 2020)
* Graduate Student Organization (GSO) Travel funding (2018)
* Katzer Research Funding for Summer from iSchool, Syracuse University (2014 - 2019)
* Best Paper Nominee in iConference (2015)
* Rank #1 in the graduate entrance test to the iSchool at Peking University (2008)
* Scholarship and tuition waiver recipient in the iSchool at Peking University (2008-2010)
* Scholarship recipient in the CS Department at Central China Normal University (2003-2004)
* Teaching assistant in the CS Department at Central China Normal University (2003-2005)

PUBLICATIONS

Refereed Journal Publications

[J6] Xia, H., Mckernan, B. (2020), Privacy in Crowdsourcing: A Review of the Threats and Challenges. *Journal of Computer-Supported Cooperative Work and Social Computing (JCSCW). Springer*. DOI: 10.1007/s10606-020-09374-0

[J5] Wang, Y., **Xia, H.**, Yao, Y., Huang, Y. (2016), "Flying Eyes and Hidden Controllers: A Qualitative Study of People's Privacy Perceptions of Civilian Drones in the US," *Proceedings on Privacy Enhancing Technologies (PoPETS)*, 3, 172-190.

[J4] Huang, Y., White, C., **Xia, H.**, Wang, Y. (2016), "A Computational Cognitive Modeling Approach to Understand and Design Mobile Crowdsourcing for Campus Safety Reporting," *International Journal of Human-Computer Studies (IJHCS)*, 102, 27-42.

[J3] Huang, Y., Shema, A., **Xia, H.** (2016), "A Proposed Genome of Mobile and Situated Crowdsourcing and Its Design Implications for Encouraging Contributions," *International Journal of Human-Computer Studies (IJHCS)*, 102, 69-80.

[J2] Xia, H. (2010), "A Literature Review of Knowledge Management in the Taiwan District," *Journal of Intelligence*, volume 2, 103-107. (Journal of Intelligence is a core journal in Information Science in China)

[J1] Qi, Y., Dou, X., **Xia, H.** (2009), "Limitations of SCI Academic Quantitative Evaluation and Its Modifications," *Information Studies: Theory and Application, 32(9), 48-52*. (Journal of Information Studies: Theory and Application is a core journal in Information Science in China)

Refereed Conference Papers

[C8] Xia, H., Østerlund, C, Mckernan, B., Folkestad, J., Rossini, P., Boichak, O., Robinson, J., Kensiki, K., Myers, R., Clegg, B. A., Stromer-Galley, J. (2019), TRACE: A Stigmergic Crowdsourcing Platform for Intelligence Analysis. *The Proceeding of the 52th Hawaii International Conference on System Sciences (HICSS)*.

[C7] Xia, H., Wang, Y., Huang, Y., Shah, A. (2017), "Our Privacy Needs to be Protected at All Costs: Crowd Workers' Privacy Experiences on Mechanical Turk," *Proceedings of the ACM: Human-Computer Interaction (PACMHCI): Volume 1: Issue 2: Computer-Supported Cooperative Work and Social Computing (CSCW). (The inaugural volume of PACMHCI)*.

[C6] Yao, Y., **Xia, H.**, Huang, Y., Wang, Y. (2017), "Privacy Mechanisms for Drones: Perceptions of Drone Controllers and Bystanders," *In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 6777-6788*. Denver, CO, USA, 6-11, May.

[C5] Yao, Y., **Xia, H.**, Huang, Y., Wang, Y. (2017), "Free to Fly in Public Spaces: Drone Controllers' Privacy Perceptions and Practices," *In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 6789-6793*. Denver, CO, USA, 6-11, May.

[C4] Huang, Y., Debreski, B., **Xia, H.** (2017), "Human Library: Understanding Experience Sharing as a Participatory Service," *In Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work (CSCW), 1152-1165*. Portland, OR, USA, 25 February – 1 March.

[C3] Wang, Y., **Xia, H.**, Huang, Y. (2016), "Examining American and Chinese Internet Users' Contextual Privacy Preferences of Behavioural Advertising," *In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work (CSCW), 539-552*. San Francisco, CA, USA, 27 February – 2 March.

[C2] Tan, E., **Xia, H.**, Ji, C., Joshi, R. V., Huang, Y. (2015), "Designing a Mobile Crowdsourcing System for Campus Safety," *iConference 2015 Proceedings*. **[Best paper nominee]**. Newport Beach, CA, USA, 24-27 March.

[C1] Huang, Y., White, C., **Xia, H.**, Wang, Y. (2015), "Modelling Sharing Decision of Campus Safety Reports and Its Design Implications to Mobile Crowdsourcing for Safety," *In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI), 400-409*. Copenhagen, Denmark, 24-27 August.

Refereed Short Papers

[W6] Xia, H. (2019), "The Three Phases of Theoretical Development in HCI*" In Workshop of the ACM Conference on Human Factors in Computing Systems (CHI) 2019*. Glasgow, UK, 4-9 May.

(Workshop paper, Presenter).

[W5] Xia, H. (2017), "Privacy Experiences and Ethical Challenges with Crowd Workers," *Future of Privacy Forum*. Facebook at New York City, NY, USA, 2-3 November. (Workshop paper, Presenter).

[W4] Xia, H., Huang, Y., Wang, Y. (2017). "Victim Privacy in Crowdsourcing Based Public Safety Reporting: A Case Study of LiveSafe," *Workshop on Inclusive Privacy and Security (WIPS) in the Symposium On Usable Privacy and Security (SOUPS)*. Santa Clara, CA., USA, 12-14 July. (Workshop paper, Presenter)

[W3] Xia, H., Huang, Y., Stromer-Galley, J. (2016), "A Collective Action Inspired Motivation Framework for Crowdsourcing," *In Workshop of The 19th ACM Conference on Computer-Supported Cooperative Work (CSCW)*. San Francisco, CA, USA, 27 February – 2 March. (Workshop paper, Presenter).

[W2] Xia, H. (2015), "Towards Universal Authentication: Ability-Based Design, Crowdsourcing, and Privacy-Preserving Biometrics," *Workshop on Inclusive Privacy and Security (WIPS) in the Symposium On Usable Privacy and Security (SOUPS)*. Ottawa, Canada, 22-24 June. (Workshop paper, Presenter).

[W1] Stromer-Galley, J., Kenski, K., Hemsley, J., Bryant, L., **Xia, H.**, Semaan, B. (2015), "Adoption and Adaptation: Diachronic Perspectives on the Growing Sophistication of Social Media Uses in Elections Campaigns," PANEL paper in *Association of Internet Researchers (AoIR)*. Phoenix, AZ, USA, 21-24, October.

Refereed Posters

[P5] Yao, Y., **Xia, H.**, Kaushik, S., Wang, Y. Design and Evaluation of an Information-Based Web Tracking Blocking Mechanism. *The Federal Trade Commission PrivacyCon*, Washington D.C., February 2018.

[P4] Xia, H., Huang, Y., Wang, Y. (2017), "Victim Privacy in Crowdsourcing Based Public Safety Reporting: A Case Study of LiveSafe," *13th Symposium On Usable Privacy and Security (SOUPS)*. Santa Clara, CA., 12-14 July. Santa Clara, CA., USA, 12-14 July. (Poster).

[P3] Yao, Y., **Xia, H.**, Kaushik, S., Wang, Y. Design and Evaluation of an Information-Based Web Tracking Blocking Mechanism. *Great Lake Security Day*, Rochester, NY, 2017

[P2] Huang, Y., **Xia, H.**, Shema, A. (2016), "The Mobile Crowdsourcing Genome," *Collective*

Intelligence Conference. New York, NY, USA, 1-3 June. (Poster).

[P1] Wang, Y., **Xia, H.**, Yao, Y., Huang, Y. Flying Eyes and Hidden Controllers: A Qualitative Study of People's Privacy Perceptions of Civilian Drones in the US. *The 12th Symposium on Usable Privacy and Security* (SOUPS 2016)

Doctoral Colloquium:

[D2] Xia, H. (2018), Ethical Issues in Crowd work-based Research, *the 68th Annual Conference of International Communication Association (ICA)*. Prague, Czech Republic, 24-28 May.

[D1] Xia, H. (2017), Privacy Concerns and Effects in Crowdsourcing, *iConference*. Wuhan, China, 22-24 June.

Book and Book Chapters

[B3] Qi, Y. and **Xia, H.** (2010), "Selected Readings in Classic Literature of Library and Information Science," Ocean Press of China.

[B2] Chapter 3: Economic Information Retrieval, *in "Social Science Information Retrieval", pp 240-249,* Press of Tsinghua University, 2012.

[B1] Chapter 2: List, Chapter 3: Stack and Queue, and Chapter 6: Graph, in "Practical Reference Book on Data Structure", pp 38-83, pp 156-204, Press of Huazhong University of Science and Technology, 2004.

Invited Talk

[T1] Xia, H. (2017), Challenges of Privacy and Security Protection in Crowdsourcing, the inaugural Great Lake Security Day (GLSD) in Rochester. (*Peer-reviewed, acceptance rate: 50%*)

RESEARCH POSITIONS

2017 – 2019 **Research Assistant,** *Intelligence Advanced Research Projects Activity. Trackable Reasoning and Analysis for Collaboration and Evaluation (TRACE)*. PI: Prof. Jennifer Stromer-Galley. Syracuse University

- Duties: paper-based system prototyping; Ethnographical observation of user testing; Content analysis of survey data on user feedback; Statistical analysis of TRACE in comparison with the other analytical tools; Literature

review
- Outcome: Paper (first-authored) accepted in The Proceeding of the 52th Hawaii International Conference on System Sciences (HICSS).

2016 – 2017 **Research Assistant,** *Examining and Addressing Privacy Issues in Crowdsourcing*
PI: Prof. Yang Wang, Syracuse University
- Duties: Research design; Data collection, cleaning, and analysis (*survey*); Literature review; Paper writing and revision
- Outcome: Paper (first-authored) accepted in the Proceedings of the ACM Human-Computer Interaction (PACMHCI): Volume 1: Issue 2: Computer-Supported Cooperative Work and Social Computing (CSCW). (The inaugural volume of the PACMHCI Journal)

2016 – 2017 **Research Assistant,** *Users' Mental Models of Smartphone Privacy*
PI: Prof. Yang Wang, Syracuse University
- Duties: Research design; Data collection and analysis (*interview*); Literature review; Paper writing and revision
- Outcome: Paper (first-authored) submitted to Computer-Supported Cooperative Work and Social Computing (CSCW). (Under review).

2016 – 2017 **Research Assistant,** *Privacy Issues with Drone*
PI: Prof. Yang Wang, Syracuse University
- Duties: Data collection, cleansing, and analysis (*interview and survey*); Literature review; Paper writing and revision
- Outcome: One paper (co-author) accepted to the Proceedings on Privacy Enhancing Technologies (PoPETS) 2016; Two papers (co-author) accepted to the ACM Conference on Human Factors in Computing Systems (CHI) 2017

2016 – 2017 **Research Assistant,** *Mobile and Situated Crowdsourcing and Design Implications for Encouraging Contributions*
PI: Prof. Yun Huang, Syracuse University
- Duties: Literature review; Data collection and analysis (*interview*); Paper writing and revision
- Outcome: Paper (co-author) accepted in the International Journal of Human-Computer Studies (IJHCS) 2016.

2016 – 2017 **Research Assistant,** *Human Library: Understanding Experience Sharing as a Participatory Service*
PI: Prof. Yun Huang, Syracuse University

- Duties: Literature review; Data collection and Analysis (interview); Paper writing and revision
- Outcome: Paper (co-author) accepted in the ACM Conference on Computer-Supported Cooperative Work (CSCW) 2016.

2015 – 2017 **Research Assistant,** *Mobile Crowdsourcing Based Public Safety Reporting System Design*

PI: Prof. Yun Huang, Syracuse University

- Duties: Data collection and analysis (*interview and prototype study*); Literature review; Paper writing and revision
- Outcome: One paper (co-author) accepted in the iConference 2015 (Best paper nominee); One paper (co-author) accepted in the International Journal of Human-Computer Studies (IJHCS) 2016; One paper (co-author) accepted in the International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI) 2016.

2014 – 2016 **Research Assistant,** *Towards a Universal Authentication System Design for People with Disabilities*

PI: Prof. Yang Wang, Syracuse University

- Duties: Research design and proposal; Literature review; Paper writing and revision
- Outcome: Workshop paper (single author) accepted in the Symposium On Usable Privacy and Security (SOUPS) 2015 (Presentation and panel discussion).

2014 – 2015 **Research Assistant,** *A Collective Action Inspired Motivational Model for Crowdsourcing*

PI: Prof. Jennifer Stromer-Galley, Syracuse University

- Duties: Research design and proposal; Literature review; Paper writing and revision
- Outcome: Workshop paper (first author) accepted in the ACM Conference on Computer-Supported Cooperative Work (CSCW) 2016.

2013 – 2015 **Research Assistant,** *Social Media Usage in Public Campaigns*

PI: Prof. Jennifer Stromer-Galley, Syracuse University

- Duties: Data collection and analysis; coder training
- Outcome: Panel paper (co-author) accepted in the Conference of the Association of Internet Researchers (AoIR) 2016.

2013 – 2015 **Research Assistant,** *Investigating Internet users' contextual privacy concerns*

about Behavioral Advertising (OBA)
PI: Prof. Yang Wang, Syracuse University

- Duties: Data collection, cleaning, and analysis (*interview and survey*); Literature review; Paper writing and revision
- Outcome: One paper (co-author) accepted to the ACM Conference on Computer-Supported Cooperative Work (CSCW) 2015.

2013 – 2014 **Research Assistant,** *The Design for a Gamification Based Educational System*
PI: Prof. Jeffrey M. Stanton, Syracuse University

- Duties: Literature review; Report writing and presentation
- Outcome: Final report presented to Vikki Brackens, the CEO of Cheddar Bowl, an educational gamification company.

2009 – 2010 **Research Assistant,** *The Enhancement of the Commercialization of Scientific and Technological Findings in Beijing*
PI: Prof. Dachen Kang, Institute of Policy & Management of Chinese Academy of Science (CAS)

- Duties: Constructed a model on a scientific information system for Beijing Academic Committee
- Outcome: Final report submitted to the city government of Beijing.

2009 – 2010 **Research Assistant,** *A Study on Online Academic Communication Platforms – Presented to the Beijing Academic Committee (BAC)*
PI: Prof. Maosheng Lai, Peking University

- Duties: Made a case study on http://www.ScienceNet.cn/ and examined the knowledge communication within its various channels such as forum, blog, micro blog and group
- Outcome: Final report submitted to the Beijing Academic Committee (BAC).

2008 – 2009 **Research Assistant,** *The Guidelines on Data Standardization Application – Submitted to the China National Institute of Standardization (CNIS)*
PI: Prof. Maosheng Lai, Peking University

- Duties: wrote a chapter on The Construction of Data Standardization Architecture
- Outcome: Final report submitted to the China National Institute of Standardization (CNIS).

2008 – 2009 **Research Assistant,** *Translation on the Classic Readings in the Library and Information Science*

PI: Prof. Yanli Qi, Peking University

- Duties: Translated several classic literatures in the library and information science, e.g., the first chapter of "The Text Retrieval Conference" by Ellen M. Voorhees and Donna K. Harman.
- Outcome: Co-author of a book "Selected Readings in Classic Literature of Library and Information Science" published by the Ocean Press of China.

TEACHING EXPERIENCE

Sep. 2019 – May 2020

Adjunct, *School of Information Studies, Syracuse University*

Independently teach IST 343 *Data in Society* to undergraduate students from different schools and departments across Syracuse University

Jan. 2019 – May 2019

Co-teach Instructor, *IST 343: Data and Society*
Instructor/Supervisor: Prof. Jennifer Stromer-Galley, Syracuse University
Level / mode: Undergraduate / In-person
Duties: Co-teach the course with my advisor Dr. Jennifer Stromer-Galley; give lectures; organize in-class discussion and activities; grade assignments

Jan. 2015 – May 2015

Teaching Assistant, *IST 649: Human Computer Interaction (HCI) Theories*
Instructor: Prof. Bryan Semaan, Syracuse University
Level / mode: Graduate / In-person
Duties: Discussion coordination, grading, lecture on "universal design," and class organization

Sep. 2014 – Dec. 2014

Teaching Assistant, *IST 486: Social Media with Enterprise*
Instructor: Prof. Jennifer Stromer-Galley, Syracuse University
Level / mode: Undergraduate / In-person
Duties: Syllabus design, class organizing, discussion participation and coordination, grading

Jan. 2014 – May 2014

Teaching Assistant, *IST 631: Theory of Classification and Subject Representation*
Instructor: Prof. Barbara Kwasnik, Syracuse University
Level / mode: Graduate / Online

Duties: Online participation monitoring, online discussion coordination, lecture on the topic of Ontology

Jan. 2014 – May 2014

Teaching Assistant, *IST 654: Information System Design and Analysis*
Instructor: Prof. Ping Zhang, Syracuse University
Level / mode: Graduate / In-person
Duties: Held office hour for Q&A, group project coordination, lecture on information system maintenance and installation in organization

---

SERVICE TO DISCIPLINARY AND PROFESSOIONAL SOCIETIES

Peer Reviewer for:

* Journal of Information and Organization
* Journal of the Association for Information Science and Technology (JASIST)
* Journal of Sage Open
* Proceedings on Privacy Enhancing Technologies (PoPETS)
* ACM Conference on Computer-Supported Cooperative Work (CSCW)
* ACM Conference on Human Factors in Computing Systems (CHI)
* ACM International Conference on Supporting Group Work (GROUP)
* ACM Conference on User Modeling, Adaptation and Personalization (UMAP)
* CHINESE Conference on Human Factors in Computing Systems (Chinese CHI)
* Symposium On Usable Privacy and Security (SOUPS)
* IEEE Conference on Privacy & Security
* iConference

Committee Member in:

Faculty Search Committee of iSchool, Syracuse University (2017): helped recruiting and assessing new faculty applicants.

Personnel Committee of iSchool, Syracuse University (2016): helped assessing and presenting new faculties' evaluation reports.

Doctoral Committee of iSchool, Syracuse University (2013, 2015): helped recruiting and interviewing new Ph.D. students.

STUDENT VOLUNTEER IN CONFERENCES

ACM Conference on Computer-Supported Cooperative Work (CSCW) (2016)
ACM Conference on Human Factors in Computing Systems (CHI) (2016)
iConference (2015)

OTHER MEMBERSHIP

2013 – 2020 **Fellow,** *Syracuse University Oratorio Society*

2016 – 2020 **Member,** *Syracuse University French & Francophonie Club*

2015 – 2020 **Member,** *Syracuse University Badminton Club*

2015 – 2017 **Chair of the Academic Branch,** *Syracuse University Chinese Student & Scholar Association (SU-CSSA)*

Academic Associations

* Professional member of the ACM (Association of Computing Machinery)
* Professional member of the AoIR (Association of Internet Researchers)