

Seeing Humans in the Data: Ethical Blind Spots of Taiwan Academic Researchers in the Era of Behavioral Big Data

Jan Fell
Institute of Service Science,
National Tsing Hua University /
Taiwan Assessment and
Evaluation Association
jan.fell@iss.nthu.edu.tw

Galit Shmueli
Institute of Service Science,
National Tsing Hua University
galit.shmueli@iss.nthu.edu.tw

Travis Greene
Institute of Service Science,
National Tsing Hua University
travis.greene@iss.nthu.edu.tw

Jyun-Cheng Wang
Institute of Service Science,
National Tsing Hua University
jcwang@iss.nthu.edu.tw

Soumya Ray
Institute of Service Science,
National Tsing Hua University
soumya.ray@iss.nthu.edu.tw

Shu-Yuan Wu
Taiwan Assessment and
Evaluation Association
wusu@twaea.org.tw

Abstract

The advent of Behavioral Big Data (BBD) has profoundly impacted research ethics. At the same time, academic disciplines with no experience in human subjects research increasingly make use of BBD datasets. In this first-of-its-kind study, we evaluate Taiwan academic researchers' knowledge and awareness of data ethics using a series of four BBD-based hypothetical research scenarios. We uncover several data ethics blind spots affecting academic researchers. Through the results of this research we hope to strengthen academic researchers' data ethics awareness and knowledge in the context of BBD, and provide suggestions for improving the ethics training of academic researchers conducting BBD studies. We also contribute a re-conceptualization of data ethics encompassing both traditional human subjects research ethics and new paradigms for the regulation of personal data, such as the General Data Protection Regulation (GDPR).

1. Introduction

The new realm of big data has made large and rich micro-level data on the behaviors, actions and interactions of individuals available to governments, industry, and researchers in diverse academic fields [1-3]. The resulting *Behavioral Big Data* (BBD) is not only generated by large organizations, such as telecommunications, retail, government, financial institutions and insurance companies. BBD is also generated ubiquitously in online platforms, social media, wearable technologies, and mobile apps. BBD captures human actions and interactions, self-reported

opinions, thoughts, and feelings pertaining to day-to-day life [4]. Consequently, BBD has impacted academic research in almost every discipline, from natural sciences, engineering, life sciences, and management, to the arts & humanities [5] and challenges orthodox epistemologies across disciplines [6]. The advent of BBD research has been paralleled by the emergence of entirely new questions in the realm of research ethics [7-9]. For example, the *Facebook Social Contagion Experiment* [10], which involved manipulating the digital environment of several hundred thousand unwitting participants in a large-scale field experiment, raised ethical questions about informed consent and scale [11], stirring public outcry [12]. Recent far-reaching legislation on how personal data may be collected, processed, and used – such as the 2018 *General Data Protection Regulation* (GDPR) of the European Union (EU) – adds the further conundrum of legal compliance to research ethics in studies involving BBD [13].

In the present article, our major concern lies with the ethical problems posed by the collection, analysis, and other uses of Behavioral Big Data in academic research. A subset of research ethics, we subsume these data-specific human subjects research ethics issues under the term *data ethics*, which is defined as follows [14]: “*A new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes).*”

In the context of data ethics in BBD research, we are motivated by two major issues raised in the recent literature. First, there are concerns about the pertinence of traditional research ethics paradigms – embodied in institutional review through IRB and notions of informed consent, beneficence, respect for persons, and justice – and the often unforeseen and indirect privacy harms to subjects (and classes of subjects) that can arise during secondary analyses of data [4, 15]. There exist several conceptual gaps in how established human subjects research ethics principles should be applied in the context of BBD research [16].

A second issue is the degree to which BBD research even constitutes human subjects research. As Metcalf and Crawford [7] note, researchers in the precursor disciplines of data science (e.g., statistics and computer science) have tended to view BBD as a substrate for testing systems, not the object of interest in itself. As a result, data ethics considerations have historically not played a major role in those fields. In this vein, *ethicality* has been identified as one of the major problems of research for academia and industry using Big Data [e.g., 17, 18, 19].

Given this growing importance of research ethics in the context of BBD studies, the objective of the present article is to empirically investigate whether, and to what extent, academic researchers might exhibit data ethics *blind spots*. Broadly, *ethical blind spots* refer to a lack of awareness and/or knowledge impairing the capacity to recognize the ethical implications of a situation [20]. Towards this goal of revealing data ethics blind spots, we conducted a survey consisting of BBD research scenarios involving 431 academic researchers. Our findings shed light on the challenges faced by academic researchers conducting BBD studies. While the existing literature has proposed a number of ethical challenges innate to the conduct of BBD studies [e.g., 4, 7, 8, 15, 16, 21], to our knowledge the present research is the first to empirically examine the ethical awareness and knowledge of researchers in a Behavioral Big Data context. In addition to shedding light on ethical blind spots of researchers conducting these studies, we hope our findings will help higher education institutions and policy makers on how to improve research training for the BBD era.

2. Literature Review

2.1. Research ethics in BBD studies

Data ethics, the overarching principle under which we subsume research ethics in studies involving BBD, encompasses the ethical problems presented by

the collection and processing of large datasets in scientific research [14]. As people interact with real and virtual environments, they leave traces of their behavior in manifold instances, leading to the generation and collection of ever-increasing amounts of behavioral data [22]. BBD is different from *Inanimate Big Data* (IBD), which is collected on objects and products, as the behavioral aspect relates to human actions and interactions, self-reported opinions, thoughts, and feelings that pertain to people’s day-to-day life [4].

A closer examination of the pertinent human subjects research regulation, namely the *Common Rule* [23], further substantiates this interpretation of all BBD research constituting human subjects research: “*Human subject means a living individual about whom an investigator [...] conducting research: Obtains information [...] through intervention or interaction with the individual, and uses, studies, or analyzes the information [...]; or obtains, uses, studies, analyzes, or generates identifiable private information [...].*”

If indeed BBD research is human subjects research, then the foundational principles of ethical research, espoused in the 1947 *Nuremberg Code*, the 1964 *Helsinki Declaration*, the 1979 *Belmont Report*, and various locally applicable national laws and guidelines, apply to BBD studies. In other words, a researcher conducting a BBD study would need to consider the same issues and dilemmas involved in designing a “small data” lab or field study involving human participants [4].

In addition to human subjects research regulation in the realm of academia, BBD studies are also subject to the legal regulation of personal data in the form of local and global regulations and rules, for example the GDPR. The European Union’s GDPR, which took effect on May 25, 2018, not only holds institutions within the EU accountable for the collection and processing of personal data, but equally applies to non-EU institutions when the personal data concerned is about European Union residents (referring to anyone physically located in the EU irrespective of their immigration status). The monetary sanctions for non-compliance with the GDPR are severe [24]. The applicability of the GDPR varies depending on the data in question: for example, it does not apply to anonymous data, and utilizing pseudonymous data lawfully collected by a data controller and subsequently shared with a researcher is subject to fewer storage and processing restrictions. Nonetheless it is essential to emphasize that when the researcher is the data controller – who by design would intervene and interact with data subjects during BBD studies – the GDPR applies equally to research purposes [13].

In sum, the practice of ethical research in the era of BBD and its concomitant notion of *data ethics* therefore comprise two main subfields, namely the existing array of human subjects research guidelines and new legal regulation of data, such as the GDPR.

2.2. The increasing importance of data ethics

Breaches of data ethics increasingly carry far-reaching implications for authors and institutions alike. For instance, the journal *Management Science* has adopted a data provenance policy [25] which states: “Recently, we have seen a few submitted papers that raise legal and ethical questions. For example, papers that apply scraped data from a variety of websites, some of which ban such practice; or papers that use fake accounts to generate data. In response to these developments, *Management Science* has modified its data provenance policy. Beyond the formal wording of the policy, it is important to clarify its intent. The objective of the policy is for authors to not use data obtained by means that materially harms individual, business, public sector, or societal interests.”

Academic and professional bodies are also following this trend. One example is the recently-updated ethics code of the Association for Computing Machinery (ACM) [26], which states: “*Computing professionals should only use personal information for legitimate ends and without violating the rights of individuals and groups. This requires taking precautions to prevent re-identification of anonymized data or unauthorized data collection, ensuring the accuracy of data, understanding the provenance of the data, and protecting it from unauthorized access and accidental disclosure.*” Here again, the concepts of data provenance and protecting the rights of societal sectors found in the aforementioned *Management Science* policy are echoed.

These revised ethics guidelines of journals as well as learned and professional societies are an indication of the growing sensitivity surrounding data ethics across academic disciplines and communities. Data ethics is no longer only a desirable notion in literature, but has also emerged as a vital consideration for researchers seeking to publish their BBD studies.

2.3. Data ethics blind spots

The *physiological blind spot* is the part of the retina where the axons of the ganglion cells converge to form the optic nerve. The blind spot is insensitive to visual stimulation, and the brain compensates for this insensitivity based on surrounding details and

information from the other eye, leaving us perceptually unaware of the process [27-29].

Figuratively, *ethical blind spots* refer to a lack of awareness and/or knowledge impairing the capacity to recognize the ethical implications of a situation [20]. Ethical blind spots pose a serious hazard, as they not only denote a failure to perceive ethical problems, but one also remains unaware of this failure [30]. They are, in effect, *unknown unknowns*. In other words, a researcher exhibiting data ethics blind spots might engage in unethical conduct without ever knowing, potentially harming human subjects along the way.

The existing literature has identified *ethicality* as a major challenge for BBD research [18]. And while the revised ethics guidelines of journals and professional societies are an indication of the growing importance placed upon data ethics, they are in no way a guarantee for the ethical conduct of researchers conducting BBD studies. These guidelines are often general in their claims [e.g., 26, 31], or only address a single very specific instance of unethical behavior [25, 32]. Others are concerned with the implications of research for society *outside the lab* once research results are disseminated, but not with what happens with data *inside the lab* during the research [e.g., 33, 34]. The immediate positive effect of these guidelines and professional codes on researchers’ behavior is thus unclear.

We therefore posit that academic researchers conducting BBD research will exhibit blind spots reflecting the ethical challenges innate to BBD research as a category of human subjects research. Specifically, the extant conceptual literature has identified the application of existing human subjects research guidelines to BBD as one major challenge [4, 15]. That is, academic researchers might not know how to apply these established guidelines to BBD research, or might not even be aware that they are conducting human subjects research involving real humans. At the same time, researchers in disciplines now using data science methodologies, such as computer science, statistics, mathematics, engineering [7, 15], or physics [34], are often not experienced in traditional social science research and the ethics of conducting research on and with human subjects. Indeed, prior research has shown that “many people treat unidentified individuals (and in particular statistical individuals) as fictional characters” [35]. Lastly, prominent moral philosophers have warned against the “objectifying” effects of statistical abstraction on our sense of moral duties towards individuals [36, 37].

A further area susceptible to potential ethical blind spots pertains to the new legal regulation of personal data. The GDPR, enacted in May 2018, has

far-reaching implications for industry and academic researchers utilizing BBD [13]. While the GDPR is a European Union regulation, it has a potentially global reach. Non-EU-based researchers may thus be unaware of how the GDPR applies to their studies, if they are even aware of the GDPR at all.

2.4. Awareness and knowledge

Ethical behavior is the culmination of a multistage process that begins with the *awareness* of ethical problems [38]. Awareness refers to the ability to recognize ethical problems when conducting research [39, 40]. In other words, ethical research behavior is conditional on the researcher being aware of the ethical implications of a proposed study. In this vein, *awareness* is a commonly measured construct in studies pertaining to the ethical conduct of research [e.g., 41, 42]. In addition, awareness for ethical issues in research among academia is an outcome desired by policy makers [43].

Beyond awareness of the ethical implications, a researcher also requires a working knowledge of research ethics guidelines to ultimately engage in the ethical conduct of research [44]. *Knowledge* is a routinely-measured construct in empirical studies to assess the effectiveness of research ethics training [e.g., 45, 46, 47], as well as to examine the adherence to principles of research ethics within a population of academic researchers [e.g., 48, 49].

3. Methodology

3.1. Scenario approach

Ideally, we would study academic researchers' data ethics in situ, not simply assess their data ethics knowledge and awareness. Nevertheless, it is not practically feasible to observe researchers in the field to measure the extent to which they engage in behaviors deemed appropriate in data ethics. Furthermore, experimentally manipulating situations to observe researchers' reactions would not be ethically acceptable (e.g., offer researchers knowingly hacked datasets in the treatment condition and non-hacked datasets in the control, and compare the proportions of researchers in the disciplines that agree to do research using the hacked datasets).

In consideration of these constraints, we selected the *scenario approach*. The scenario approach is often the favored approach to measure research ethics knowledge, awareness, and perceptions of researchers and IRB committee members in the biomedical field [e.g., 50, 51, 52], business and management [e.g., 53],

social and behavioral sciences [e.g., 54], and engineering fields [e.g., 55].

3.2. Scenario Development

We assessed respondents' data ethics awareness and knowledge through four original research scenarios with a total of nine iterations. Iterations are modified version of a scenario in which ethically relevant details are modified. Each of the scenarios described the conduct of a hypothetical academic researcher facing a BBD-related data ethics dilemma.

Each scenario iteration was followed by a 5-point Likert scale question and an open-ended question. Harris, et al. [56] provide a rationale for this approach: they argue that the act of thinking through an ethical dilemma can better equip one to deal with similar such situations in the future. *Knowledge* was measured through agreement with positive statements, such as "Professor Lin needs to apply for IRB review before she can use the data in her research" or "the GDPR applies to the proposed study." The preceding statements acted as proxies for knowing that a proposed study constituting human subjects research, or that the GDPR applies, respectively, as well as "doing the right thing." *Awareness* was measured through the open-ended item, which served to provide respondents with the opportunity to detail their ethical reasoning.

As data ethics in the era of BBD consists of both human subjects research guidelines and new data regulation such as the GDPR, and keeping with the blind spots we intend to shed light on, we have devised the scenarios (Table 1) to address both these constituents of data ethics. English and Chinese language versions of the full scenarios and iterations are available from the authors.

3.3. Sample selection

Study participants were recruited from a list of circa 1,700 academics in Taiwan, who had attended institutional research workshops and participated in higher education peer-evaluations commissioned by the Ministry of Education in Taiwan within the previous 10 years. The list included faculty of all academic ranks (i.e., assistant, associate and full professors, and research fellows) and disciplines affiliated with Taiwanese higher education institutions engaged in research. Given the qualitative components of our survey, the possibility of invalid responses, and the exploratory nature of this study, all 1,700 academics on our list were invited.

Table 1. Summary of BBD research scenarios and iterations.

Scenario	Iteration	Data Ethics Issues	Ethics Dimension(s)
1. Prisoner data	1	Public data released by government; vulnerable population; identifiable subjects	Existing human subjects research guidelines
	2	Public data of unknown source; vulnerable population; identifiable subjects	Existing human subjects research guidelines
	3	Public data of unknown source; vulnerable population; identifiable subjects; GDPR	Existing human subjects research guidelines, new legal regulation of data
2. Reddit.com marketing research	1	Observational studies on online platforms	Existing human subjects research guidelines
	2	Intervention studies on online platforms	Existing human subjects research guidelines
	3	GDPR (subjects located in EU)	New legal regulation of data
	4	GDPR (subjects located outside of EU, researchers located in EU)	New legal regulation of data
3. Leaked hospital data	1	GDPR (EU citizen outside of EU uses non-EU data for research)	New legal regulation of data
4. Hacked dating app	1	Hacked dataset	Existing human subjects research guidelines

3.4. Data collection

We collected data in a four-month period from June 2nd to August 6th, 2019 using a computer-administered survey instrument in Chinese. The present research has received approval through an institutional ethics review (IRB) at National Tsing Hua University.

4. Results

4.1. Descriptive statistics

We collected a total of 431 responses. 157 participants (36.4%) were female. 93 (21.6%) respondents were affiliated with biomedical and life science departments, 143 (33.2%) were affiliated with social science, management, and economics departments, and 83 (19.3%) were affiliated with other departments, such as humanities and arts. 88 (20.4%) were affiliated with departments in the precursor disciplines of data science, such as computer science, physics, statistics, mathematics, or engineering [7, 15], and other fields not trained in behavioral science methods [4]. These fields represent the Science, Technology, Engineering and Mathematics (STEM) disciplines except for those biomedical and life sciences routinely trained in human subjects research ethics. We refer to these fields as *inanimate STEM*, or *ISTEM* in short. 24 (5.5%) respondents did not provide information on their academic discipline.

4.2. Data ethics awareness and knowledge

The results of our scenario-based survey assessing the data ethics *knowledge* (scale items) and data ethics *awareness* (open-ended items) are presented as follows. We converted our 5-point scale Likert scales into *incorrect*, *unsure*, and *correct* (*Incorrect* [1, 2]; *unsure* [3]; *correct* [4, 5]; scales were flipped as needed to reflect the correctness sequence). Figure 1 presents the responses to all scenarios and iterations. As shown in Figure 1, the results varied. While over 50% of respondents answered correctly in seven out of nine iterations, participants exhibited high levels of uncertainty on several iterations, and for two iterations we recorded a majority of incorrect responses (S2_I1; S3_I1). A linear regression indicated that results did not correlate with academic discipline, location of doctoral training, prior ethics training, gender, or prior BBD research experience.

Furthermore, respondents exhibited varying view-points, which indicates varying degrees of awareness. While respondents who answered correctly on the scale items measuring knowledge generally also demonstrated awareness for the issues underlying their correct response, we observed a sizeable number of incorrect responses signaling low awareness or unawareness.

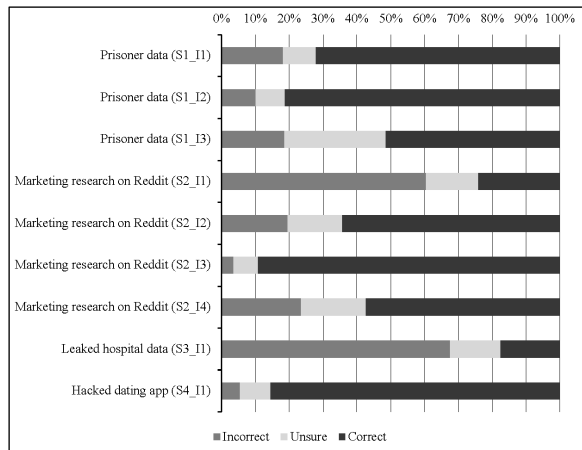


Figure 1. Data ethics knowledge.

5. Discussion

Our survey measuring data ethics knowledge and awareness revealed a number of *blind spots* related to BBD research. These blind spots pertain to both the application of existing research ethics guidelines to the BBD context, as well as the new legal regulation of data.

Publicly available data. One blind spot concerns the potential identifiability of subjects in publicly available data. That is, the data allows for the identification of subjects either directly or through identifiers linked to the subjects' data. An ancillary blind spot is related to the recognition of vulnerable populations in publicly available data.

Open data initiatives worldwide have led to a proliferation of datasets released by governments, ranging from census data and fiscal information to sex offender registries and parole data. As academics can obtain these datasets easily and typically free of charge, they present themselves as ideal opportunities to conduct BBD research. While in most cases such public data does not require IRB review, publicly available data on subjects belonging to vulnerable populations and/or that allows for subject identification is not exempt from ethics review [57]. While our survey revealed that almost three quarters of respondents knew this and were aware of the data being comprised of prisoners as well as allowing for subject identification, 18% did not know this and another 9% of respondents were unsure in their response. In other words, more than a quarter of academic researchers surveyed would conduct research using publicly available data on vulnerable populations or that allows for subject identification without grasping the ethical implications involved. Vulnerable populations go beyond the incarcerated. Consider that in 2017 more than half of children aged

8 or younger in the United States used mobile apps [58], and 61% of 14-year-olds in France had Snapchat accounts in 2018 [59]. Given the ongoing and ever-increasing popularity of using public data in BBD research, this finding highlights an ethical blind spot not to be overlooked.

Seeing the humans in the data. Another blind spot concerns interaction with subjects on online platforms. Field experiments are a common and increasing trend in BBD research [e.g., 60]. While about two thirds of respondents correctly identified interaction with and manipulation of online platform users as human subjects research requiring IRB review, 20% and 16% did not or were unsure. BBD field experiments often include gifting features or non-tangible virtual objects as the manipulation. However, the non-tangible virtual quality of the manipulation in scenario 1 iteration 2 (i.e., virtual currency) was by a number of respondents given as the rationale why they did not see any potential ethical issues in the scenario. In other words, our findings imply that the *virtualness* and *intangibility* of interactions with subjects in BBD can cloud the researcher's awareness that they are dealing with an actual human being. This echoes findings in the existing literature. For instance, the treatment of MTurk workers recruited for experiments is described as "the equivalent of undergraduate students happy to participate in research for free pizza", who are only identified by a worker ID, stripping away "everything that makes a human a person, such as their beliefs, attributes, and experiences" [61]. Moreover, Żuradzki [35] found that "many people treat unidentified individuals (and in particular statistical individuals) as fictional characters." A major blind spot is therefore that the researcher may not see the humans in the data.

Legal regulation of data. A further blind spot relates to the legal regulation of data. Survey respondents were not knowledgeable and aware of key GDPR concepts as they relate to the conduct of BBD studies. While most participants knew that the GDPR applies to subjects located in the EU, only little more than half knew that the GDPR also applies when only the researcher is located in the EU. More than half of participants did not know that the citizenship of either researcher or subjects is entirely irrelevant for the applicability of the GDPR. Given that we collected data in the Summer of 2019, a year after the enactment of the GDPR, our findings imply that little is known in academia about a central piece of data legislation with a potentially global reach beyond the European Union.

Chilling effects for BBD research. While the preceding blind spots highlight potentially unethical conduct of researchers, a lack of knowledge can also lead to the opposite, i.e., researchers falsely assuming

a proposed study constitutes human subjects research and thus requires IRB review. Only 24% of respondents knew that observational studies on publicly accessible online platforms do not require ethics review as observational of nonidentified persons in a public setting is not defined as human subjects research under the Common Rule. Conversely, 60% of participants responded incorrectly, and 15% were unsure. In other words, a large number of participants assumed a higher standard of ethics scrutiny than what is actually required. Such false data ethics perceptions can then lead to *chilling effects*. That is, researchers incorrectly assume a proposed research is unethical or could be scrutinized during IRB review. Fearing bureaucratic effort applying for review and possible rejection, researchers thus do not conduct the proposed research. This *chilling effect* of ethics review through IRB is a common theme in critiques of the current state of human subjects research ethics review [62]. As a consequence, society and academia could lose the opportunity to benefit from potentially valuable research.

6. Conclusion

6.1 Theoretical contributions

The present research assessed the data ethics awareness and knowledge of academic researchers through research scenarios. The scenario-based approach has been utilized for research ethics training and assessment in multiple fields. To our knowledge, the present study is the first to apply a BBD-relevant *scenario-based* survey approach. Therefore, the present research contributes to the development of instruments to measure research ethics. Our study also provided rich insights into the moral reasoning of respondents in the new realm of BBD.

The challenges and hazards of data ethics and studies involving BBD have been discussed in the conceptual literature [e.g., 4, 7, 8, 15, 21]. We contribute to this stream with the first empirical study shedding light on the ethical blind spots of academic researchers. Our results affirm a number of propositions in the conceptual literature, such as the fundamental difficulty of transferring existing guidelines into new BBD research situations.

Our research also contributes to the understanding of ethical blind spots in human subjects research. The findings of our survey indicate that blind spots not only enable potentially unethical behavior, but can also lead to chilling effects, where researchers mistakenly believe a proposed study is unethical or requires a tedious review process.

6.2 Policy implications

Our study has revealed a number of data ethics blind spots. At the same time, we did not find that ethics training led to better performance in our scenarios. Ethics training is increasingly compulsory for researchers and students; higher education institutions invest substantial resources in implementing training curricula. Our results suggest higher education decision makers must begin to include BBD-relevant content in ethics training curricula.

This need is underlined not only by the potential lapses revealed through our survey, but also to the chilling effect on research activity if researchers incorrectly assume a proposed research is unethical and thus do not pursue it. Thus, in addition to augmenting research ethics curricula for BBD relevance, decision makers need to better inform researchers about the review process and what types of research are exempt or qualify for an expedited review to avoid chilling effects on benign and useful BBD research activity.

Related to the IRB review process is a further aspect, namely, that of IRB members. While the specific composition of IRBs varies by jurisdiction, they generally include a number of academic researchers affiliated with the higher education institution in question or academia in general. Therefore, the data ethics blind spots revealed in our study affect the *conduct* of research, as much as they affect the *review* of research. Promoting data ethics among academic researchers is thus also an issue of quality control for the IRB process.

6.3 Limitations and future research

As with any empirical research, this study suffers from a number of methodological limitations. While we have focused on academic researchers, we note that BBD is also generated and used by firms and other organizations with far-reaching effects on society as a whole, as well as individuals. Our findings might not be generalizable to such contexts, and this poses an important and interesting direction for future research.

The institutionalization of research ethics in the United States, goes back 70 years, whereas in most East Asian contexts this institutionalization is still developing [63]. Taiwan may be an outlier in this regard. Formal ethics review became mandatory for biomedical research in 2002 and for all other disciplines conducting research involving human subjects in 2007 [64]. More recently, research ethics has been added to the rating criteria for publicly funded research projects [65], and in 2019 the Ministry

of Science and Technology in Taiwan added research guidelines for ethical AI research [66]. Thus, the institutionalization of research ethics in Taiwan can be seen as fairly well-entrenched. Outside of Taiwan, however, our results might not be generalizable across all countries and territories in East Asia. We envision future research to take into consideration the experiences of academic researchers in other regions and contribute to a better understanding of ethical blind spots across countries. As a useful extension of our research we envision a replication in other countries and cultural settings.

Another limitation concerns the generality of our survey questions. Rather than asking the specific consequence GDPR applicability would have on the use of a dataset, we only asked respondents whether the GDPR applied at all or not. In other words, the present study is limited in that we only measured whether respondents were aware of the GDPR's very existence and its applicability to the scenario. The GDPR is a complex legal framework: asking for specific consequences would have exceeded our focus on the notion of data ethics in general. Given its global reach and the serious potential consequences for violations [13], the knowledge of GDPR principles and what they mean for the use of data in studies require further examination among the wider population of academic researchers. We therefore envision future research to go beyond the mere applicability of the GDPR and to focus more on the content of its principles.

In sum, the present research is the first empirical study on data ethics blind spots of academic researchers. We hope the present study helps academic researchers to become more aware of potential data ethics challenges, and aids higher education decision makers in augmenting existing research ethics curricula.

7. References

- [1] danah boyd and K. Crawford, "Critical Questions for Big Data," *Information, Communication & Society*, vol. 15, no. 5, pp. 662-679, 2012.
- [2] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165-1188, 2012.
- [3] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Communications of the ACM*, vol. 57, no. 3, pp. 78-85, 2014.
- [4] G. Shmueli, "Analyzing Behavioral Big Data: Methodological, practical, ethical, and moral issues," *Quality Engineering*, vol. 29, no. 1, pp. 57-74, 2017.
- [5] H. V. Jagadish, "Big Data and Science: Myths and Reality," *Big Data Research*, vol. 2, no. 2, pp. 49-52, 2015.
- [6] R. Kitchin, "Big Data, new epistemologies and paradigm shifts," *Big Data & Society*, vol. 1, no. 1, pp. 1-12, 2014.
- [7] J. Metcalf and K. Crawford, "Where are human subjects in Big Data research? The emerging ethics divide," *Big Data & Society*, vol. 3, no. 1, pp. 1-14, 2016.
- [8] J. Metcalf, "Big data analytics and revision of the common rule," *Communications of the ACM*, vol. 59, no. 7, pp. 31-33, 2016.
- [9] A. Zwitter, "The Network Effect on Ethics in the Big Data Age," in *Big Data Challenges: Society, Security, Innovation and Ethics*, A. Bunnik, A. Cawley, M. Mulqueen, and A. Zwitter Eds. London: Palgrave Macmillan UK, 2016, pp. 23-34.
- [10] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, p. 8788, 2014.
- [11] I. M. Verma, "Editorial Expression of Concern: Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 29, pp. 10779-10779, 2014.
- [12] New York Times. "Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry." <https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html> (accessed September 1, 2020).
- [13] T. Greene, G. Shmueli, S. Ray, and J. Fell, "Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers," *Big Data*, vol. 7, no. 3, pp. 140-162, 2019.
- [14] L. Floridi and M. Taddeo, "What is data ethics?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2083, pp. 1-5, 2016.
- [15] G. Shmueli, "Research Dilemmas with Behavioral Big Data," *Big Data*, vol. 5, no. 2, pp. 98-119, 2017.
- [16] M. Zimmer, "Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity," *Social Media + Society*, vol. 4, no. 2, pp. 1-11, 2018.

- [17] I. Anagnostopoulos, S. Zeadally, and E. Exposito, "Handling big data: research challenges and future directions," *The Journal of Supercomputing*, vol. 72, no. 4, pp. 1494-1516, 2016.
- [18] A. Abbasi, S. Sarker, and R. H. Chiang, "Big data research in information systems: Toward an inclusive research agenda," *Journal of the Association for Information Systems*, vol. 17, no. 2, pp. i-xxxii, 2016.
- [19] R. Vidgen, G. Hindle, and I. Randolph, "Exploring the ethical implications of business analytics with a business ethics canvas," *European Journal of Operational Research*, 2019.
- [20] D. Chugh, M. H. Bazerman, and M. R. Banaji, "Bounded Ethicality as a Psychological Barrier to Recognizing Conflicts of Interest," in *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman Eds. Cambridge: Cambridge University Press, 2005, pp. 74-95.
- [21] D. J. Hand, "Aspects of Data Ethics in a Changing World: Where Are We Now?," *Big Data*, vol. 6, no. 3, pp. 176-190, 2018.
- [22] A. Paxton and T. L. Griffiths, "Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets," *Behavior Research Methods*, vol. 49, no. 5, pp. 1630-1638, 2017.
- [23] "Protection of Human Subjects," 45 C.F.R. §46.102, 2009.
- [24] G. Buttarelli, "The EU GDPR as a clarion call for a new global digital gold standard," *International Data Privacy Law*, vol. 6, no. 2, pp. 77-78, 2016.
- [25] INFORMS. "Management Science: Progress and Challenges." <https://www.informs.org/Blogs/ManSci-Blogs/From-the-Editor/Management-Science-Progress-and-Challenges> (accessed May 7, 2019).
- [26] ACM. "ACM Code of Ethics and Professional Conduct." <https://www.acm.org/code-of-ethics> (accessed May 8, 2019).
- [27] M. F. Land, *The Eye: A Very Short Introduction*. Oxford: Oxford University Press, 2014.
- [28] S. A. Rathus, *Psychology: Concepts and Connections*, 10 ed. Belmont, CA: Wadsworth, 2010.
- [29] J. W. Kalat, *Biological Psychology*, 10 ed. Belmont, CA: Wadsworth, 2008.
- [30] P. H. Werhane, L. P. Hartman, C. Archer, E. Englehardt, and M. S. Pritchard, "Obstacles to Ethical Decision-Making in the Perception of Ethical Context," in *Systems Thinking and Moral Imagination: Rethinking Business Ethics with Patricia Werhane*, D. J. Bevan, R. W. Wolfe, and P. H. Werhane Eds. Cham: Springer, 2019, pp. 359-377.
- [31] IEEE. "Ethically Aligned Design." <https://ethicsinaction.ieee.org/> (accessed May 8, 2019).
- [32] INFORMS. "Management Science Data Provenance Policy January 2019." <https://pubsonline.informs.org/pb-assets/filesmnsdc/DataProvenancePolicyFinal-1566575064017.pdf> (accessed September 1, 2020).
- [33] NeurIPS. "NeurIPS 2020 Call for Papers." <https://nips.cc/Conferences/2020/CallForPapers> (accessed September 1, 2020).
- [34] S. Thais, "Physicists Must Engage with AI Ethics, Now," *Physics*, vol. 13, no. 107, pp. 1-2, 2020.
- [35] T. Żuradzki, "The normative significance of identifiability," *Ethics and Information Technology*, vol. 21, no. 4, pp. 295-305, 2019.
- [36] P. F. Strawson, *Freedom and Resentment and other Essays*. Oxon: Routledge, 2008.
- [37] P. Singer, *Practical Ethics*, 3 ed. Cambridge: Cambridge University Press, 2011.
- [38] O. C. Ferrell, L. G. Gresham, and J. Fraedrich, "A Synthesis of Ethical Decision Models for Marketing," *Journal of Macromarketing*, vol. 9, no. 2, pp. 55-64, 1989.
- [39] J. R. Rest, *Moral Development: Advances in Research and Theory*. New York, NY: Praeger, 1986.
- [40] J. R. Rest, *Development in Judging Moral Issues*. Minneapolis, MN: University of Minnesota Press, 1979.
- [41] N. Kandeel *et al.*, "A Multicenter Study of the Awareness and Attitudes of Egyptian Faculty towards Research Ethics: A Pilot Study," *Journal of Empirical Research on Human Research Ethics*, vol. 6, no. 4, pp. 99-108, 2011.
- [42] E. Löfström, "Students' Ethical Awareness and Conceptions of Research Ethics," *Ethics & Behavior*, vol. 22, no. 5, pp. 349-361, 2012.
- [43] B. Williams-Jones and S. Holm, "A University Wide Model for the Ethical Review of Human Subjects Research," *Research Ethics*, vol. 1, no. 2, pp. 39-44, 2005.
- [44] N. Palmer and R. Forrester-Jones, "Research Ethics Training: Using a Virtue Ethics Approach to Training to Support Development of Researcher Integrity," in *Virtue Ethics in the Conduct and Governance of Social Science*

- Research, N. Emmerich Ed., (Advances in Research Ethics and Integrity, no. 3). Bingley: Emerald Publishing, 2018, pp. 65-82.
- [45] A. J. Ajuwon and N. Kass, "Outcome of a research ethics training workshop among clinicians and scientists in a Nigerian university," *BMC Medical Ethics*, vol. 9, no. 1, p. 1, 2008.
- [46] H. Davies, F. Wells, and C. Druml, "How can we provide effective training for research ethics committee members? A European assessment," *Journal of Medical Ethics*, vol. 34, no. 4, p. 301, 2008.
- [47] A. Halkoaho, M. Matveinen, V. Leinonen, K. Luoto, and T. Keränen, "Education of research ethics for clinical investigators with Moodle tool," *BMC Medical Ethics*, vol. 14, no. 1, p. 53, 2013.
- [48] H. A. Taylor, N. E. Kass, J. Ali, S. Sisson, A. Bertram, and A. Bhan, "Development of a research ethics knowledge and analytical skills assessment tool," *Journal of Medical Ethics*, vol. 38, no. 4, p. 236, 2012.
- [49] O. A. Ogunrin, F. Daniel, and V. Ansa, "Knowledge of the Nigerian Code of Health Research Ethics Among Biomedical Researchers in Southern Nigeria," *Journal of Empirical Research on Human Research Ethics*, vol. 11, no. 5, pp. 397-407, 2016.
- [50] T. M. Reynolds, "Down's syndrome screening is unethical: views of today's research ethics committees," *Journal of Clinical Pathology*, vol. 56, no. 4, p. 268, 2003.
- [51] K. B. Schmaling and A. W. Blume, "Ethics Instruction Increases Graduate Students' Responsible Conduct of Research Knowledge but not Moral Reasoning," *Accountability in Research*, vol. 16, no. 5, pp. 268-283, 2009.
- [52] C. M. Ulrich and S. J. Ratcliffe, "Hypothetical Vignettes in Empirical Bioethics Research," in *Empirical Methods for Bioethics: A Primer*, vol. 11, (Advances in Bioethics, no. 11). Bingley: Emerald Publishing, 2007, pp. 161-181.
- [53] M. Wallace and N. Sheldon, "Business Research Ethics: Participant Observer Perspectives," *Journal of Business Ethics*, vol. 128, no. 2, pp. 267-277, 2015/05/01 2015.
- [54] C. B. Fisher and T. L. Kuther, "Integrating research ethics into the introductory psychology course curriculum," *Teaching of Psychology*, vol. 24, no. 3, pp. 172-175, 1997.
- [55] B. J. Brummel, C. K. Gunsalus, K. L. Anderson, and M. C. Loui, "Development of Role-Play Scenarios for Teaching Responsible Conduct of Research," *Science and Engineering Ethics*, vol. 16, no. 3, pp. 573-589, 2010.
- [56] C. E. Harris, M. Davis, M. S. Pritchard, and M. J. Rabins, "Engineering Ethics: What? Why? How? And When?," *Journal of Engineering Education*, vol. 85, no. 2, pp. 93-96, 1996.
- [57] "Protection of Human Subjects," 45 C.F.R. §46.101(b), 2009.
- [58] Statista. "Which of the following activities has your child ever carried out on a mobile device such as a smartphone or tablet?" <https://www.statista.com/statistics/204510/use-of-portable-media-devices-among-americas-children/> (accessed September 1, 2020).
- [59] Statista. "Percentage of children aged 8 and 14 years old having a Snapchat account in France in 2018." <https://www.statista.com/statistics/1011320/sha-re-children-subscribers-snapchat-france/> (accessed September 1, 2020).
- [60] R. Bapna, J. Ramaprasad, G. Shmueli, and A. Umyarov, "One-Way Mirrors in Online Dating: A Randomized Field Experiment," *Management Science*, vol. 62, no. 11, pp. 3100-3122, 2016.
- [61] M. L. Gray and S. Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston, MA: Houghton Mifflin Harcourt, 2019.
- [62] Z. M. Schrag, *Ethical imperialism: Institutional review boards and the social sciences, 1965–2009*. Baltimore, MD: The Johns Hopkins University Press, 2010.
- [63] M. Israel, *Research Ethics and Integrity for Social Scientists: Beyond Regulatory Compliance*, 2 ed. Thousand Oaks, CA: SAGE Publications, 2015.
- [64] S.-M. Huang, "Developing Institutional Monitoring Protocols for Humanities and Social Science Research in Taiwan," in *Integrity In The Global Research Arena*, N. H. Steneck, M. S. Anderson, S. Kleinert, and T. Mayer Eds. Singapore: World Scientific, 2015, pp. 203-208.
- [65] C.-J. Guo and M.-H. Chiu, "Research Projects on Science Education Funded by the National Science Council in Taiwan from 1982 to 2012: A Historical Review," in *Science Education Research and Practices in Taiwan: Challenges and Opportunities*, M.-H. Chiu Ed. Singapore: Springer, 2016, pp. 11-41.
- [66] Digitimes. "MOST announces AI R&D guidelines." <https://www.digitimes.com/news/a20190923PD209.html> (accessed September 1, 2020).