

In “Taste, ties, and time”¹, a group of researchers from Harvard University and the University of California, Los Angeles detail the collection and analysis of a new dataset, joining personal, social, and cultural information from Facebook profiles with university housing information for a large cohort of Harvard University undergraduate students. The ways in which this data was collected, managed, and published demonstrate several ethical miscalculations by the research team involved. In terms of the four principles of ethical research outlined by Salganik², researchers in the “Taste, ties, and time” study clearly violate the principles of *Respect for Persons*, *Beneficence*, and *Justice*. The ethical infractions demonstrated by this project were largely caused by the improper and unauthorized collection of information, a lack of adequate anonymization of data, and a failure to properly assess and remediate potential harm to participants. Ultimately, the researchers demonstrated a failed consequentialist approach to performing ethical research with personal, online data. Further research of this kind must reevaluate not only the potential consequences of such data usage, but also the deontological duties of researchers when such consequences may be unknown or unforeseeable.

By not obtaining explicit consent from research subjects³, and in certain cases, taking advantage of in-network research assistants’ access to private Facebook pages⁴, the researchers involved in “Taste, ties, and time” demonstrated a lack of *Respect for Persons*⁵.

The researchers, it seems, approached the ethical questions of their research with a consequentialist framework. After receiving initial criticism, one author, Jason Kaufman, responded “What might hackers want to do with this information, assuming they could crack the data and ‘see’ these people’s Facebook info?”⁶. While ultimately part of an attempt to assuage concerns about

¹ Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, Nicholas Christakis, Taste, ties, and time: A new social network dataset using Facebook.com, *Social Networks*, Volume 30, Issue 4, October 2008, Pages 330-342

² Salganik, Matthew J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Open review edition.

³ Lewis et al, *Loc cit*

⁴ Zimmer, M. *Ethics Inf Technol* (2010) 12: 313. doi:10.1007/s10676-010-9227-5

⁵ Salganik, *Bit by Bit*, 6.4.1 *Ethics: Respect for Persons*

⁶ Zimmer, *Loc cit*

deanonymization, the words from Kaufman provide insight into the research team's ethical considerations; even if someone could get this information, what consequence could that have on the subjects?

With such a mindset concerning the consequences of their research, the team chose to forego explicit participant consent. Furthermore, while most Facebook data obtained was said to be publicly available, the researchers admitted to accessing some information only available to them via the personal Facebook accounts of the undergraduate and graduate research assistants working on the project⁷. In other words, research assistants with privileged access to subject Facebook data (either through Facebook "friend" connections, or broader "network" connections") downloaded information that would have been otherwise unavailable. Not only is this likely a breach of Facebook's terms of service, it also demonstrates a further lack of *Respect for Persons*. By relying on personal connections research assistants had with research subjects, not only were subjects further denied the right to informed consent, but the consent they had given in confidence (via Facebook connections) was taken advantage of and violated.

While perhaps permissible when the risks to participants are assuredly low, it is not the case that access to this privileged dataset is entirely harmless, and thus the researchers violated the principle of *Beneficence*⁸.

The data from Facebook was combined with housing information from Harvard University⁹ and upon release the dataset was quickly identified as being collected at Harvard University¹⁰. Not only does this breach of anonymity put the privacy of a group (the undergraduate class) as a whole at risk, it also risks the privacy of individuals in that group. Many subgroups within the class were represented by a very small number of students, putting those particular students (often ethnic minorities) at a greater risk for identification. Likewise, network and cultural data that is comprehensive, as it often is on one's facebook account, provide a very unique source of identification. Individuals, especially members of minority groups, were at great risk of being identified in this dataset.

Again, the researchers saw little harm in the potential for subject identification. The greatest risk to subjects was the leaking of sensitive, personal information. For much of this information, it is

⁷ Zimmer, Loc cit

⁸ Salganik, Loc cit, 6.4.2 *Ethics: Beneficence*

⁹ Lewis et al, Loc cit

¹⁰ Zimmer, Loc cit

impossible to determine if its release would be harmful to an individual, because its significance to each individual is subjective and inaccessible to the researchers. In other words, no one can claim to know what information about another person is too sensitive to be released, or simply harmless. Especially in the case of marginalized and private identities, the release of certain information (religion, gender identity, sexual orientation, etc.) may be not only emotionally disturbing, but also physically endangering in certain contexts. Combined with the high possibility of identification, the leaking of such information posed a great risk to the study subjects.

As has been touched upon, the privacy and safety of minority groups was disproportionately placed at risk, violating the principle of *Justice*¹¹. While it is concerning that a small group of undergraduate students were placed at high risk, it is doubly concerning that these effects were only amplified for members of minority groups. With such a small sample size, individuals of minority groups (which naturally make up an even smaller subsample) were placed at an even greater risk of identification and the inherent risks as discussed above.

In addition to these specific ethical infractions, the research described in “Taste, ties, and time” demonstrates the limitations of a purely consequentialist ethical framework. The researchers’ ethical decisions were largely driven by the assumption that their observational study did not pose any risk to the research subjects. However, in order to respect the dignity and autonomy of research participants, large collections of personal, social, and cultural data must be constructed with careful consideration of the subjective significance of this data to each individual involved. As the ethical concerns of this study demonstrate, the consequences inherent in the collection and publication of such data are, at best difficult, if not impossible to foresee. Likewise, such collections of data may be more prone to deanonymization, especially in the face of rapidly developing computational capabilities, thereby increasing the likelihood of such consequences. While considering the consequences of research is obviously important, further research with similar data must be wary of using a strictly consequential framework, as this study demonstrates the difficulty of producing accurate predictions of what those consequences may be. Given the numerous and rather severe ethical concerns surrounding this data, I would refrain from and encourage others to refrain from using it.

¹¹ Salganik, Loc cit, 6.4.3 *Ethics: Justice*

Bibliography

- Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, Nicholas Christakis, Tastes, ties, and time: A new social network dataset using Facebook.com, *Social Networks*, Volume 30, Issue 4, October 2008, Pages 330-342, ISSN 0378-8733, <http://dx.doi.org/10.1016/j.socnet.2008.07.002>.
- Salganik, Matthew J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Open review edition.
- Zimmer, M. *Ethics Inf Technol* (2010) 12: 313. doi:10.1007/s10676-010-9227-5