

A stylized illustration of a person in a white lab coat holding a document, surrounded by large, colorful eyes. The person is standing on a red carpet, and the background is composed of various colored geometric shapes. The eyes are large and stylized, with different colors like blue, brown, and green. The overall style is graphic and modern.

Twitter Research Ethics: Beyond the Review Board

Nicholas Proferes
Arizona State University
@moduloone

Talk Outline

- I. “Ethics” and the use of publicly available data.
- II. Twitter: Gap between what researchers do and what users expect.
- III. Values-tensions example.

Ethics as Contemplation

Ethics: systems of principles that we use to guide us in making moral evaluations.

- Utilitarianism
- Deontology
- Virtue Ethics, etc.

Using our capacities for **reason** and **judgment** to critically examine our actions and our character.

“Ethics” as Compliance

Process of ensuring conformity with all relevant laws, policies, and guidelines.

These laws/guidelines/policies often draw on principles, such as:

- Respecting human dignity and autonomy
- Maximizing benefit/minimizing harm
- Justice and beneficence

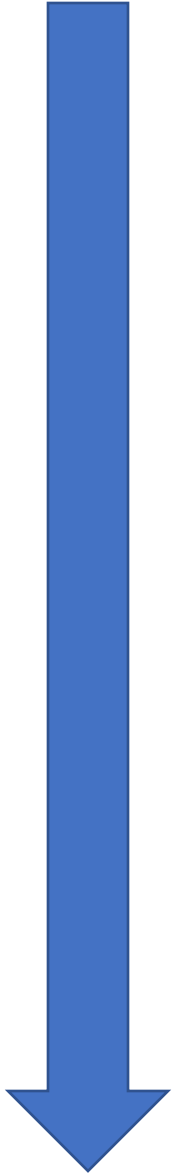
Ensuring researchers do not violate certain baseline conditions for the treatment of others.

Compliance / Contemplation

Rules and policies developed for the biomedical setting.

“Public” data not considered (in the U.S. at least) as “research involving human subjects.”

But, projects using publicly available data from social media sites do not lack ethical dimensions.



Data Collection

- Hacked Data, Deleted Data, How “Public”
- Studying Marginalized Groups
- Aggregation of Data Points
- Removing Data

Data Use

- Ends
 - Phrenology 2.0
- Author’s Rights
- Protecting Researchers

Data Sharing

- Representing Data-Subjects
- Sharing w/ Community
- Replicability

There is no singular “ethics portion” of a research project.

We must continuously weigh the tension between values and duties.

Disclosive ethics ([Brey, 2000](#)) can help us.

- A process of ethical consideration that involves both a *descriptive* and *normative* component.
- Asks us to think about gaps.

Part II: Twitter

- Twitter a major data source for academics.
 - +2k research papers in the past 3 years using Twitter data.
 - Projects regularly include billions of Tweets.
- Why?
 - Textual
 - Easy(ish) APIs
 - “Public”
- Few projects report going through ethics review. ([Zimmer & Proferes, 2014](#))



“Do users know their content is being used for academic study and how do they feel about it?”

Fiesler, C., & Proferes, N. (2018). [“Participant” Perceptions of Twitter Research Ethics](#). *Social Media+ Society*, 4(1), 2056305118763366.

Surveyed
Twitter
users,
asking:

Whether users think researchers are “allowed” to use their content without re-consent.

Levels of comfort with their tweets being used.

Contextual factors (permission asked; study content; size of dataset; who is doing analysis; kinds of content; status of content; quoted in study).



61.2% of respondents thought researchers were **forbidden** from using public Tweets without having to ask user permission.



64.9% thought researchers shouldn't be able to use tweets **without permission**.



Many users are *somewhat comfortable* with the idea of their content being used, **if asked**. But, this is **extremely** contextually driven.

	Very uncomfortable	Somewhat uncomfortable	Neither uncomfortable nor comfortable	Somewhat comfortable	Very comfortable
... you were not informed at all?	35.1%	31.7%	16.4%	13.4%	3.4%
... you were informed about the use after the fact?	21.3%	29.1%	20.5%	22.0%	7.1%
... it was analyzed along with millions of other tweets?	2.6%	18.7%	25.5%	30.0%	23.2%
... it was analyzed along with only a few dozen tweets?	16.5%	30.3%	24.0%	20.2%	9.0%
... it was from your “protected” account?	54.9%	20.5%	13.8%	6.0%	4.9%
... it was a public tweet you had later deleted?	31.3%	32.5%	20.5%	10.4%	5.2%
... no human researchers read it, but it was analyzed by a computer program?	2.6%	14.3%	30.5%	32.3%	20.3%
... the human researchers read your tweet to analyze it?	9.7%	27.6%	25.0%	25.4%	12.3%
... the researchers also analyzed your public profile information, such as location and username?	32.2%	23.2%	21.0%	13.9%	9.7%
... the researchers did not have any of your additional profile information?	4.9%	15.4%	25.1%	34.1%	20.6%
... your tweet was quoted in a published research paper, attributed to your Twitter handle?	34.3%	21.6%	21.6%	13.1%	9.3%
... your tweet was quoted in a published research paper, attributed anonymously?	9.0%	16.8%	26.5%	28.4%	19.4%

User understandings
of data-uses are
limited.

Contextually-driven
levels of acceptance.

Important to understand
and study phenomena.

Notifying/consenting
users could create
anxiety.



Part III: Exploring Tensions



Describing the tensions...

Compliance:

- Twitter's ToS forbids full JSON data-sharing, but allows for Tweet IDs to be shared.
- Our local laws may/may not allow re-sharing of data.
- We may be asked by our funders to share data.

Contemplation:

- Users may not want to be in a labeled data-set.
 - Privacy
- Even distribution of risk?
- Data-loss from users deleting content could threaten validity.

Addressing tensions...

- Documenting the Now's "[Hydrator](#)" lets researchers easily "rehydrate" JSON data from Tweet IDs.
- Find ways to anonymize user-data when possible, particularly if possibly sensitive content or labeling.
- Re-sharing by request only, [with your own user agreement](#).
- Document decision-making and discuss it!

- Association of Internet Researchers.
- Just shy of two decades of tackling ethics questions involving Internet Research.
- Has published two ethical decision-making guides + flowchart: <https://aoir.org/ethics/>
 - 2019 Ethical Guidelines (v 3.0).



Thanks!