Responsible Data Science

Applied Ethics in Data Science

Prof. George Wood

Center for Data Science
New York University

data RESPONSIBLY

NYU

Applied Ethics in Data Science

1. Benefits, costs, and externalities
2. Ethical frameworks and principles
3. Case studies

Motivation: Costs, benefits, and risks in DS

- Barebones DS pipeline:

Aims ⟶ Task ⟶ Data

Crude cost-benefit analysis

$$\text{Benefit}_{\text{aims}} > \text{Cost}_{\text{task}} + \text{Cost}_{\text{data}} \rightsquigarrow \text{Do it}$$

Decomposing the cost, benefits, risks

- *Aim*: YouTube wants to optimize views
- *Task*: Experiment with recommendation engine
- *Data*: User profiles, page views, time spent watching, etc

Decomposing the cost, benefits, risks

Stakeholders

- *Aim*: YouTube wants to optimize views
- *Task*: Experiment with recommendation engine
- *Data*: User profiles, page views, time spent watching, etc

Are there any other stakeholders?

Decomposing the cost, benefits, risks

The New York Times

Can YouTube Quiet Its Conspiracy Theorists?

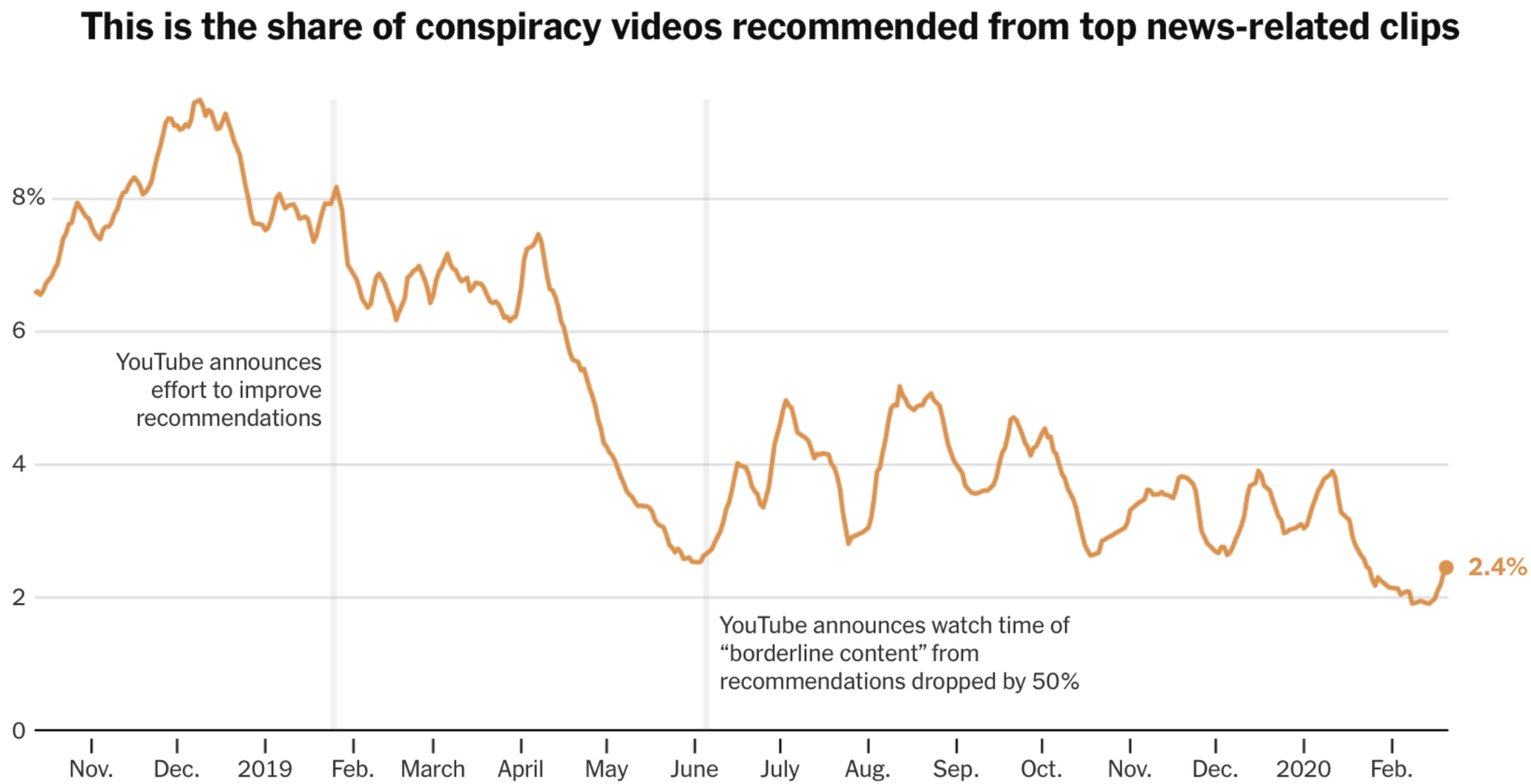
A new study examines YouTube's efforts to limit the spread of conspiracy theories on its site, from videos claiming the end times are near to those questioning climate change.

By Jack Nicas

Produced by Rumsey Taylor, Alana Celii and Dave Horn

March 2, 2020

Decomposing the cost, benefits, risks



Note: Recommendations were collected daily from the "Up next" column alongside videos posted by more than 1,000 of the top news and information channels. The figures include only videos that ranked 0.5 or higher on the zero-to-one scale of conspiracy likelihood developed by the researchers. ▪ Source: Hany Farid and Marc Faddoul at University of California, Berkeley, and Guillaume Chaslot

Decomposing the cost, benefits, risks

Sensitive data

- *Aim*: YouTube wants to optimize views
- *Task*: Experiment with recommendation engine
- *Data*: User profiles, page views, time spent watching, etc

Decomposing the cost, benefits, risks

Potential for repurposing

- *Aim*: YouTube wants to optimize views
- *Task*: Experiment with recommendation engine
- *Data*: User profiles, page views, time spent watching, etc

Decomposing the cost, benefits, risks

Potential for repurposing

TITLE	CITED BY	YEAR
You only look once: Unified, real-time object detection J Redmon, S Divvala, R Girshick, A Farhadi Proceedings of the IEEE conference on computer vision and pattern …	14989	2016
YOLO9000: Better, Faster, Stronger. J Redmon, A Farhadi Proceedings of the IEEE conference on computer vision and pattern recognition	7576	2017
Yolov3: An incremental improvement J Redmon, A Farhadi arXiv preprint arXiv:1804.02767	6805	2018

Joseph Redmon
@pjreddie

Replying to @pjreddie

I stopped doing CV research because I saw the impact my work was having. I loved the work but the military applications and privacy concerns eventually became impossible to ignore.

> **Roger Grosse** @RogerGrosse · Feb 20, 2020
> Replying to @skoularidou
> What's an example of a situation where you think someone should decide not to submit their paper due to Broader Impacts reasons?

11:09 AM · Feb 20, 2020 · Twitter Web App

Externalities in DS

Internal

Data Scientist

↓

Product

↓

Users

External

→ Production externalities
(external costs and benefits)

→ Consumption externalities

Externalities in DS

YouTube recommendation engine

Internal

Data Scientist

↓

Product

↓

Users

External

→ Production externalities
gambling company uses engine

→ Consumption externalities
non-users exposed to anti-maskers

What are the incentives for YouTube to capture these externalities?

How clear are the benefits?

Barebones DS pipeline:

- Abstraction from the thing(s) we want to know to the things we can practically study
- How tightly can we model the complexity of the thing(s) we want to know?
- To what extent does the data measure the phenomena?
- …

Downside risks

- Getting the task specification wrong:

Officer characteristics and racial disparities in fatal officer-involved shootings

David J. Johnson[a,b,1], Trevor Tress[b], Nicole Burkel[b], Carley Taylor[b], and Joseph Cesario[b]

- *Aim*: Investigate the degree to which Black civilians are more likely to be fatally shot than White civilians, and whether this varies by officer race

- *Task*: Estimate whether a person fatally shot was more likely to be Black (or Hispanic) than White

- *Data*: Fatal shootings, civilian race, officer race

- *Claim*: “White officers are not more likely to shoot minority civilians than non-White officers” (p. 15877)

Downside risks

Officer characteristics and racial disparities in fatal officer-involved shootings

David J. Johnson[a,b,1], Trevor Tress[b], Nicole Burkel[b], Carley Taylor[b], and Joseph Cesario[b]

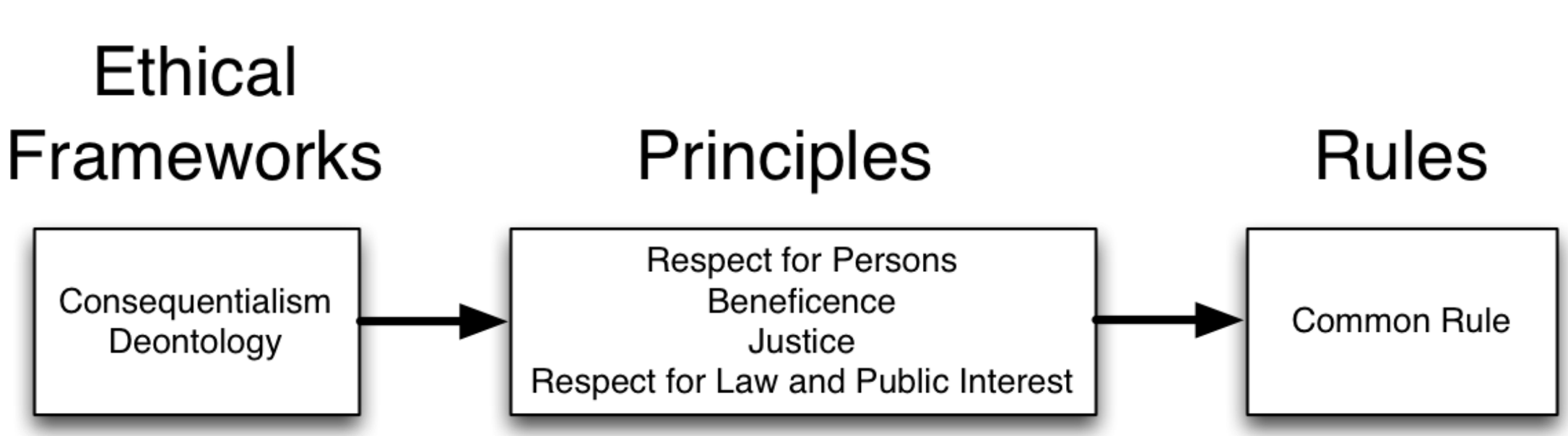
- The task they in fact carried out:
  - $Pr(\text{shot} \mid \text{civilian race}, \text{officer race}, X)$
  - i.e. whether a person fatally shot was more likely to be black or hispanic than white (see Knox and Mummolo, 2020)
  - Unobserved data problem; did not observe how many interactions happened between officers and civilians by race, nor the context of these interactions
  - Unfortunately, this paper was later cited in congress and prominently in a WSJ op-ed as evidence against systematic racism in policing

How should we evaluate our work?

- To do nothing is also an ethical judgement (*status quo bias*) and in many cases not a practical option
- How should we evaluate our work and whether to share data in cases of ethical uncertainty?

Principles-based approach to ethics

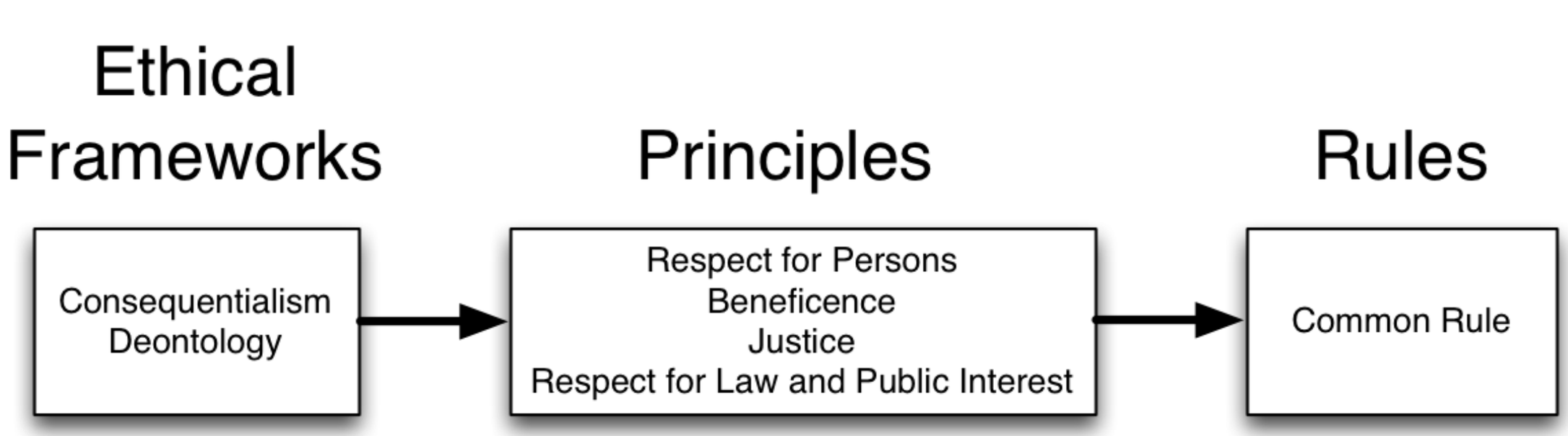
Bit by Bit, Chapter 6: Ethics



"The rules governing research are derived from principles that in turn are derived from ethical frameworks. A main argument of this chapter is that researchers should evaluate their research through existing rules—which I will take as a given and assume should be followed—and through more general ethical principles."

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Bit by Bit, Chapter 6: Ethics



The **Common Rule** is the set of regulations currently governing most federally funded research in the United States... The **four principles** come from two blue-ribbon panels that were created to provide ethical guidance to researchers: the Belmont Report and the Menlo Report.

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Principles and rules

- In simple terms, rules represent a codification of principles
- This codification can be useful, e.g. the Common Rule, prevents *ad hoc justification, ethical slippage*
- But codification has limits; the writing of rules requires all sorts of decisions, e.g.:
  - *Inclusion, exclusion, interpretation, changing contexts*

Bit by Bit, Chapter 6: Ethics

"Neither of these approaches—the rules-based approach of social scientists or the ad hoc approach of data scientists—is well suited for social research in the digital age. Instead, I believe that we, as a community, will make progress if we adopt a principles-based approach.

[…]

This principles-based approach helps researchers make reasonable decisions for cases where rules have not yet been written, and it helps researchers communicate their reasoning to the public."

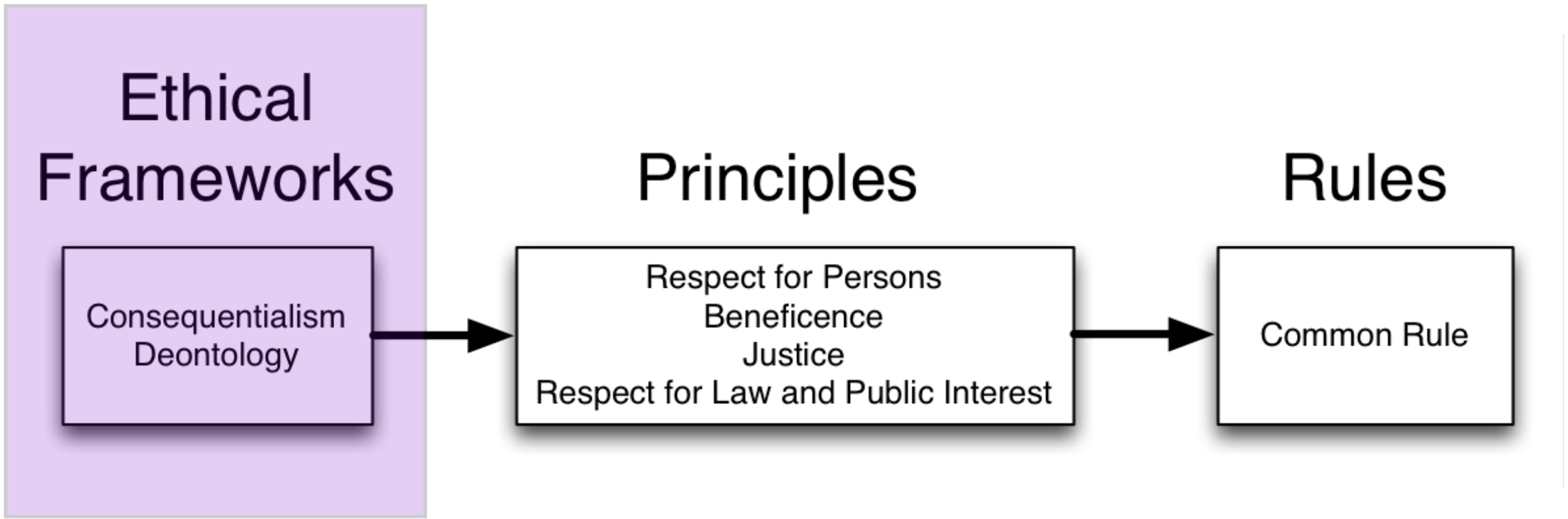
Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Bit by Bit, Chapter 6: Ethics

"In some cases the principles-based approach leads to clear, actionable solutions. And, when it does not lead to such solutions, it clarifies the trade-offs involved, which is critical for striking an appropriate balance. Further, the principles-based approach is sufficiently general that it will be helpful no matter where you work."

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Before we get to the principles…



Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Ethical frameworks

- Underlying moral theories:

Consequentialist

- The moral value of an act is determined by the value of performing that act
- External to the act; what happens as a result of doing it
- *John Stuart Mill*, *Jeremy Bentham*, utilitarianism

Non-consequentialist (deontology)

- The consequences of performing an act do not by themselves determine its moral value
- Value is something internal to the act itself
- *Immanuel Kant*: actions performed in accordance with moral obligations are good
- *Rawls*' Theory of Justice; veil of ignorance

Based on slide by Kent Baldner

Application: Informed Consent

- Individuals should be given the opportunity to consent (or not consent) to taking part in a study. They should be provided with sufficient information to understand the purpose, risks, and consequences of the study and how their data will be used and stored.
- Consequentialism and deontology support informed consent, but for different reasons.

Application: Informed Consent

A consequentialist argument:

> Informed consent helps prevent harm to participants by prohibiting research that does not properly balance risk and anticipated benefit. In other words, consequentialist thinking would support informed consent because it helps prevent bad outcomes for participants.

A deontological argument:

> Researcher has a duty to respect the autonomy of participants and obtain informed consent.

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Application: Informed Consent

"Given these arguments, a pure consequentialist might be willing to waive the requirement for informed consent in a setting where there is minimal risk, whereas a pure deontologist would not."

Salganik (2017), *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Kant's self-legislation:

> "The agents who are subject to moral requirements must be regarded as their legislators."

Reath (2006), *Agency and Autonomy in Kant's Moral Theory*

Four principles

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Tuskegee Syphilis Study

- In 1932, researchers from the US Public Health Service enrolled 399 black men from Tuskegee, Alabama, with syphilis
- Study was non therapeutic; designed to document the history of the disease
- Participants were deceived; told it was a study of “bad blood” and offered false and ineffective treatment
- As study progressed and treatment options were developed, researchers actively intervened to prevent participants from receiving treatment from elsewhere

Tuskegee Syphilis Study

Table 6.4: Partial Time Line of the Tuskegee Syphilis Study, adapted from Jones (2011)

Date	Event
1932	Approximately 400 men with syphilis are enrolled in the study; they are not informed of the nature of the research
1937-38	The PHS sends mobile treatment units to the area, but treatment is withheld for the men in the study
1942-43	In order to prevent the men in the study from receiving treatment, PHS intervenes to prevent them from being drafted for WWII
1950s	Penicillin becomes a widely available and effective treatment for syphilis; the men in the study are still not treated (Brandt 1978)
1969	The PHS convenes an ethical review of the study; the panel recommends that the study continue
1972	Peter Buxtun, a former PHS employee, tells a reporter about the study, and the press breaks the story
1972	The US Senate holds hearings on human experimentation, including Tuskegee Study
1973	The government officially ends the study and authorizes treatment for survivors
1997	US President Bill Clinton publicly and officially apologizes for the Tuskegee Study

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

The Belmont Report (1978)

- Boundaries between research and practice
- Ethical principles
  - Respect for Persons
  - Beneficence
  - Justice
- Applications

Boundaries between practice and research

For the most part, the term “practice” refers to interventions that are designed solely to enhance the wellbeing of an individual patient or client and that have a reasonable expectation of success. […] By contrast, the term “research” designates an activity designed to test a hypothesis, permit conclusions to be drawn, and thereby develop or contribute to generalizable knowledge

- Research seeks generalizable knowledge, practice includes everyday treatment and activities
- The general rule is that if there is any element of research in an activity, that activity should undergo review for the protection of human subjects

Respect for Persons

- Individuals should be treated as autonomous agents

To respect autonomy is to give weight to autonomous persons' considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others. To show lack of respect for an autonomous agent is to repudiate that person's considered judgments, to deny an individual the freedom to act on these considered judgements, or to withhold information necessary to make a considered judgement, when there are no compelling reasons to do so.

Respect for Persons

- Value of self-determination; benefits in terms of helping individuals protect themselves from harm
- Kant's self-legislation offers a different rationale: "The agents who are subject to moral requirements must be regarded as their legislators."
- Argument:

  Policy will no longer be based on how far it goes in the direction of offering people opportunities for personal deliberation. Instead, it will be rated by how well it protects people against deception and coercion.

  Kristinsson (2019), The Belmont Report's Misleading Conception of Autonomy

Respect for Persons

- People with diminished authority are entitled to protection

In some situations, however, application of the principle is not obvious. The involvement of prisoners as subjects of research provides an instructive example. On the one hand, it would seem that the principle of respect for persons requires that prisoners not be deprived of the opportunity to volunteer for research. On the other hand, under prison conditions, they ,ay be subtly coerced or unduly influenced to engage in research activities for which the would not otherwise volunteer.

Respect for Persons

- People with diminished authority are entitled to protection

Respect for persons would then dictate that prisoners be protected. Whether to allow prisoners to “volunteer” or to “protect” them presents a dilemma. Respecting persons, in most hard cases, is often a matter of balancing competing claims urged by the principle of respect itself.

Beneficence

Persons are treated in an ethical manner not only by respecting their decisions and protecting the from harm, but also by making efforts to secure their well-being. Such treatment falls under the principle of beneficence.

- Do no harm
- Maximize possible benefits and minimize possible harms

Beneficence

The Hippocratic maxim “do no harm” has long been a fundamental principle of medical ethics. Claude Bernard extended it to the realm of research, saying that one should not injure one person regardless of benefits that might come to others. However, even avoiding harm requires learning what is harmful; and, in the process of obtaining this information, persons may be exposed to the risk of harm.

Beneficence

Learning what will in fact benefit may require exposing persons to risk. The problem posed by these imperatives is to decide when it is justifiable to seek certain benefits despite the risks involved, and when the benefits should be foregone because of the risks.

- How might this be assessed under a consequentialist versus deontological framework?

Justice

- Who ought to receive the benefits of research and bear its burdens?

  Questions of justice have long been associated with social practices such as punishment, taxation, and political representation. Until recently these questions have not generally been associated with scientific research. However, they are foreshadowed even in the earliest reflections on the ethics of research involving human subjects. For example, during the 19th and early 20th centuries the burdens of serving as research subjects fell largely upon poor ward patients, while the benefits of improved medical care flowed primarily to private patients.

Justice

- Who ought to receive the benefits of research and bear its burdens?

  Questions of justice have long been associated with social practices such as punishment, taxation, and political representation. Until recently these questions have not generally been associated with scientific research. However, they are foreshadowed even in the earliest reflections on the ethics of research involving human subjects. For example, during the 19th and early 20th centuries the burdens of serving as research subjects fell largely upon poor ward patients, while the benefits of improved medical care flowed primarily to private patients.

Justice

- Who ought to receive the benefits of research and bear its burdens?

  Subsequently, the exploitation of unwilling prisoners as research subjects in Nazi concentration camps was condemned as a particularly flagrant injustice. In this country, in the 1940's, the Tuskegee syphilis study used disadvantaged, Black rural men to study the untreated course of a disease that is by no means confined to that population. These subjects were deprived of demonstrably effective treatment in order not to interrupt the project, long after such treatment became generally available.

  Against this backdrop, it can be seen how conceptions of justice are relevant to research involving human subjects.

Justice

- Rawls, “Justice as Fairness”
- *Negative thesis*: people are not entitled to more benefits simply because of morally arbitrary conditions (e.g. born into rich or poor family)
- *Positive thesis*: benefits should be distributed equally, unless an unequal distribution would be to everyone’s advantage

The Menlo Report (2012)

- Respect for Law and Public Interest

The Menlo Report

The Menlo Report calls on researchers to move beyond the narrow definition of “research involving human subjects” from the Belmont Report to a more general notion of “research with human-harming potential”

A principles-based approach means that researchers should not hide behind a narrow, legal definition of research involving human subjects, even if IRBs allow it. Rather, they should adopt a more general notion of “research with human-harming potential”

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

The Menlo Report

Principle	Application
Respect for Persons	Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
Beneficence	Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit.
Justice	Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects.
Respect for Law and Public Interest	*Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions.*

Table 1: Proposed guidelines for ethical assessment of ICT Research.

The Menlo Report

Transparency and accountability serve vital roles in many ICTR contexts where it is challenging or impossible to identify stakeholders (e.g. attribution of sources and intermediaries of information), to understand interactions between highly dynamic and globally distributed systems and technologies, and consequently to balance associated harms and benefits.

A lack of transparency and accountability risks undermining the credibility of, trust and confidence in, and ultimately support for, ICT research.

* ICTR — Information and Communication Technology Research

The Menlo Report

Accountability demands that research methodology, ethics evaluations, data collected, and results generated should be documented and made available responsibly in accordance with balancing risk and benefits. Data should be available for legitimate research, policy-making, or public knowledge, subject to appropriate collection, sue ,and disclosure controls informed by the Beneficence principle.

Keeping up with the capabilities

- The Belmont Report and the Common Rule were developed before modern ICT, digital age
- Menlo Report responds to ICT
- However, given pace of new capabilities, rules are typically reactive and seldom proactive
- Principles are generalizable and “extendable,” but also flexible (openness to interpretation)
- Rules arguably less generalizable, but also less flexible; trade-offs

Case Studies

Racial bias in resume screening

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

Marianne Bertrand

Sendhil Mullainathan

AMERICAN ECONOMIC REVIEW
VOL. 94, NO. 4, SEPTEMBER 2004
(pp. 991-1013)

We perform a field experiment to measure racial discrimination in the labor market. We respond with fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perception of race, each resume is assigned either a very African American sounding name or a very White sounding name. The results show significant discrimination against African-American names: White names receive 50 percent more callbacks for interviews. We also find that race affects the benefits of a better resume. For White names, a higher quality resume elicits 30 percent more callbacks whereas for African Americans, it elicits a far smaller increase. Applicants living in better neighborhoods receive more callbacks but, interestingly, this effect does not differ by race. The amount of discrimination is uniform across occupations and industries. Federal contractors and employers who list Equal Opportunity Employer' in their ad discriminate as much as other employers. We find little evidence that our results are driven by employers inferring something other than race, such as social class, from the names. These results suggest that racial discrimination is still a prominent feature of the labor market.

Racial bias in resume screening

Aims → Task → Data

- Aim: Do employers unlawfully discriminate against applicants based on race?
- Task: Audit study; respond to help-wanted ads with fictitious resumes
- Data: Features of resumes, employer attributes
- Stakeholders, sensitive data, potential for repurposing?

Racial bias in resume screening

Consent process can be analyzed as containing three elements (*Belmont Report*):

- Information
- Comprehension
- Voluntariness

Employers didn’t provide consent. In fact, they were actively deceived!

Racial bias in resume screening

Field-experiments to study discrimination are legally permissible if:

1. The harm to employers is limited, and

2. There is great social benefit to having a reliable measure of discrimination, and

3. Other methods of measuring discrimination are weak; and

4. Deception does not violate the norms of that setting

Salganik (2017) *Bit by Bit:* https://www.bitbybitbook.com/en/1st-ed/ethics/

Emotional contagion

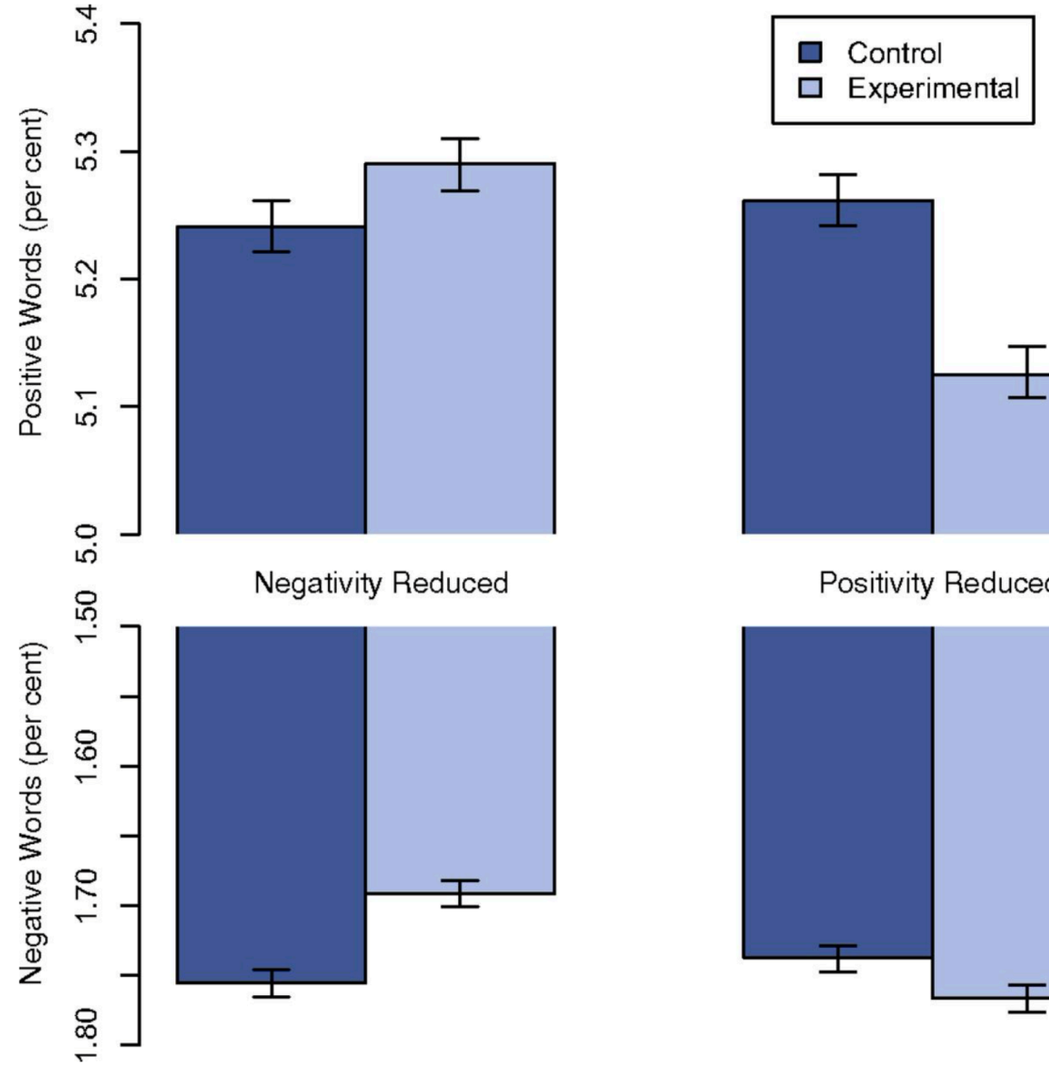
PNAS Proceedings of the National Academy of Sciences of the United States of America

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

Significance

We show, via a massive (*N* = 689,003) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.



Emotional contagion

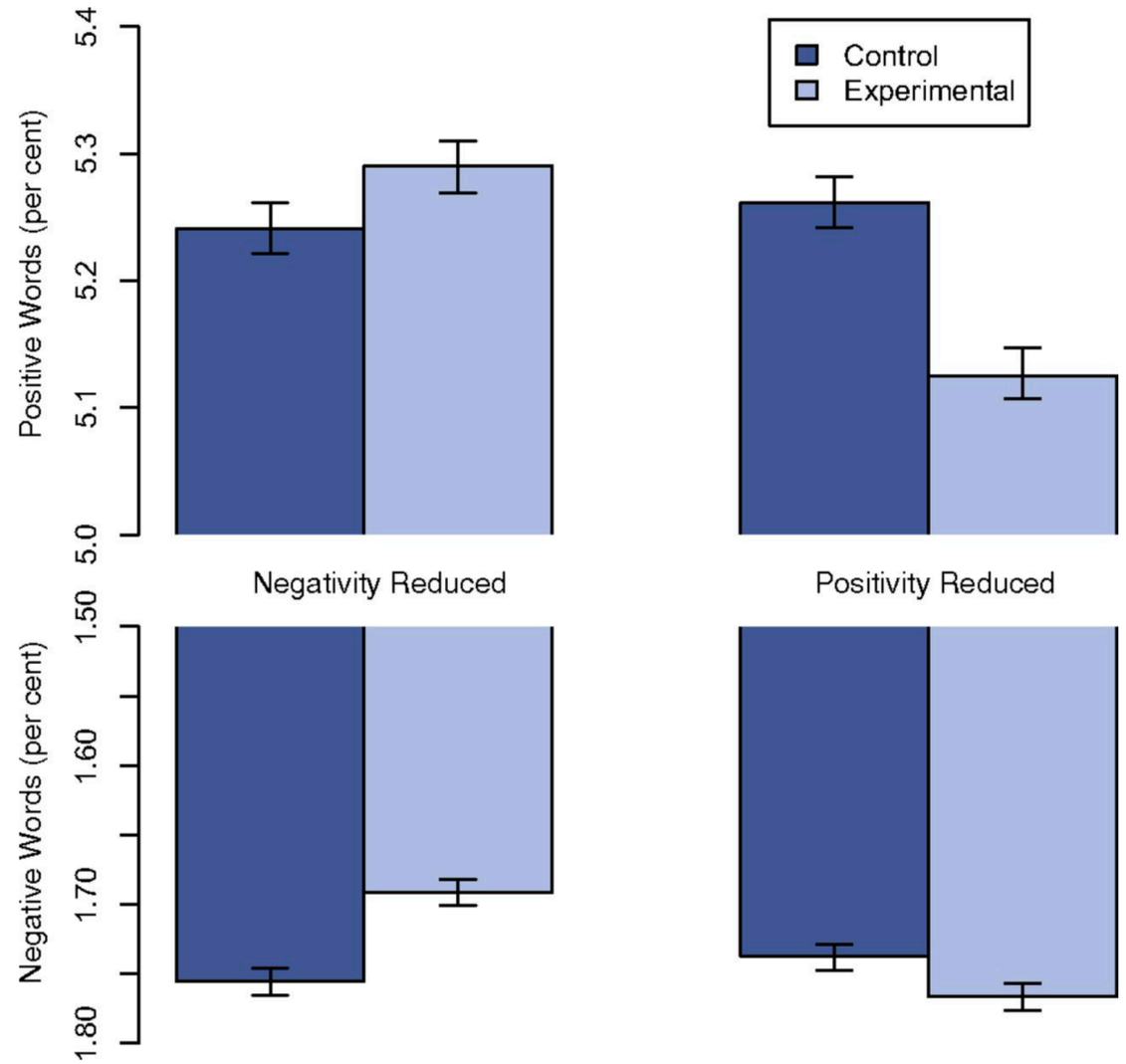
Participants assigned to one of four conditions:

A. Negativity-reduced (e.g. “sad” blocked)

B. Negativity-reduced control

C. Positivity-reduced (e.g. “happy” blocked)

D. Positivity-reduced control



Emotional contagion

Criticism from researchers and press:

1. Participants did not provide any consent (only standard Facebook terms of service)

   - “[The work] was consistent with Facebook’s Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research.”

2. Study had not undergone meaningful third-party ethical review

   - “Because this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB [Institutional Review Board] determined that the project did not fall under Cornell's Human Research Protection Program.”

Emotional contagion

Criticism from researchers and press:

1. Participants did not provide any consent (only standard Facebook terms of service)

   - “[The work] was consistent with Facebook’s Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research.”

2. Study had not undergone meaningful third-party ethical review

   - “Because this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB [Institutional Review Board] determined that the project did not fall under Cornell's Human Research Protection Program.”

Verma (2014), Editorial Expression of Concern, PNAS

Emotional contagion

Aftermath:

- PNAS placed a disclaimer on the article
- Facebook instituted an internal ethics review board

The potential for this type of experimentation on digital platforms is massive, consider:

social media, streaming services, news media