Evaluating *Irrelevant Search*: A Case Study for Internet Ethics Education

Anne Bowser
Microsoft Research
One Microsoft Way
Redmond, WA 98052
anne.bowser@gmail.com

Janice Tsai
Microsoft Research
One Microsoft Way
Redmond, WA 98052
janice.tsai@microsoft.com

Every submission will be assigned their own unique DOI string to be included here.

Abstract
Existing standards for ethical human subjects research may not always fit the needs of Internet researchers. We explore how ethics education can be designed for an industry research lab. Specifically, we present a case study on search engine experimentation— "Irrelevant Search"— and explore how the ethics of this case study are evaluated by two different populations, graduate students and industry researchers.

Author Keywords
Research ethics; ethics education; case studies; Irrelevant Search

ACM Classification Keywords
K.4.1. Computer and Society: Public Policy Issues—ethics

Introduction and Background
July 2014 was an important month for Internet research ethics. First, Facebook's study on emotional contagion called into question the roles that consent, evaluation of harm, and IRB approval play in online experiments. Then, OKCupid founder Christian Rudder published a blog post (provocatively titled: "We Experiment on Human Beings!") defending Facebook's research and outlining recent OKCupid experiments [6].

We hope that the outcry surrounding this research ensures that the topic of Internet research ethics remains a key social issue. Regardless, Internet researchers themselves have been discussing the ethics of their work for decades [1]. These conversations support individual researchers, and increasingly inspire institutions seeking to codify ethical practices as business standards. Facebook, for example, outlined changes to their research standards including enhanced training and a new public research repository [9].

The authors are dedicated to supporting ethical research at a second institution, Microsoft Research (MSR). Our current efforts [3], are: the implementation of an internal research review system, and the selection of ethics training materials. The research we conducted to support our ethics training produced the case study, "Irrelevant Search," described below.

As many consider case studies ideal for ethics education [1, 2, 4, 9], one contribution of this paper is a vetted pedegological tool. We also explore how two audiences—graduate students and industry researchers—evaluate the ethics of Irrelevant Search. These assessments shed light on how ethics are understood by both groups, with implications for designing teaching materials and for understanding the ethical challenges of Internet research. After presenting key background information, we describe the methodology for designing our training materials, and present our case study Irrelevant Search. We then report on quantitative and qualitative evaluations of this case study via a survey with 21 graduate students and 21 industry researchers, and interpret our results.

Background

Our ethics training will support industry researchers engaged in Internet research. While we draw on existing resources when possible, we recognize that our population is somewhat unique. For example, [8] found differences in how students and professionals evaluate ethical issues such as copyright, while [5] observed that identities as researcher and practitioner are often in conflict, complicating ethical valuations.

Classroom-based ethics education often begins with foundational theory, drawing on perspectives as diverse as Utilitarianism and Ubuntu [2]. While foundational theory supports agile evaluation, it may be most powerful linked with real disciplinary practices [9]. This can be achieved through teaching with case studies, an approach designed "to assess precisely the kind of difficult, borderline questions facing Internet researchers" [4, p.23]. Case studies may be implemented in generalized education, such as the CITI training offered by many universities (http://www.citiprogram.org). Programs like CITI benefit from being modular and online. However, while some modules (e.g., "History and Ethical Principles") may help our population, others (e.g., "Research with Children") are more narrowly applicable. Furthermore, generalized training neither directly prepares researchers for the specific challenges of Internet research, nor encourages learners to conduct their own ethical valuations. For these reasons, we have designed our ethics training as a hybrid between both classroom-based, in-person and modular, online approaches.

We evaluated four potential ethics training modules. "Motivating History" establishes the relevance of ethics to computer science and Internet research. "Guiding

Ethical Principles" articulates the three principles advanced by the Belmont Report (respect for persons; beneficence; justice) and three applications (informed consent; assessment of risks and benefits; selection of subjects). Together, these modules are designed to represent foundational ethical theory. "Conflict Zones" offer a short list of situations where norms regarding human subjects research are challenged by the questions and methods that new technologies support.

Our fourth and final model contains seven case studies. Two are modified from the case studies presented by [4]; the remaining five are our own. For each case study, participants read the study carefully and then evaluate the research proposed. We discuss Irrelevant Search for two reasons. First, 70% of our survey respondents rated this case study as "extremely helpful" or "helpful" for inspiring critical contemplation regarding research ethics. Yet, students and industry researchers differed on their specific valuations of this case study, like by disagreeing on the necessity of securing informed consent.

Irrelevant Search

"An intern at a major search engine wishes to study whether the relevance of search results is related to future use of a search engine. He designs an online A/B test where some users are deliberately exposed to an algorithm that returns both relevant, up-to-date information and old, potentially out-of-date information over a two month period. Users neither provide explicit consent, nor are debriefed. The intern hopes to publish the results of his research at a major conference."

After reading the scenario, participants were first asked, "Noting that ethics are highly subjective…what ethical questions does this research raise?" Answers were presented as check boxes, encouraging participants to select more than one concern: *Potentially harming research participants*; *Failing to secure informed consent from search engine users*; *Publishing the results of this research*. Following this was a button, "See what an ethicist says," providing the following information when clicked:

- *Potentially harming research participants:* Displaying outdated search results could be a minor inconvenience, or cause significant harm. Researchers should implement this carefully, if at all.
- *Failing to secure informed consent from search engine users*: The Internet research community grapples with issue of conducting research where consent may be extremely difficult to obtain.

Participants were then asked two yes or no questions- "Do the benefits of this research (to individuals and/or society) outweigh the risks?" and, "Should this research be allowed to proceed as proposed?" Following these, all were asked to, "Please evaluate the ethics of this research in 2-3 sentences." Finally, participants were presented with the option to "See what others have to say." This button provided information similar to "See what an ethicist says," but in a more conversational tone. Both options are implemented as a digital alterative to the dialogic nature of classroom-based case studies [2, 4].

Evaluation

Participants were 21 graduate students and 21 industry researchers recruited via snowball sample. When asked to identify the ethical considerations of Irrelevant

Search, 19 graduate students and 15 industry researchers checked, "potentially harming research participants." Additionally, 15 graduate students and eight Industry researchers checked "Failing to secure informed consent from search engine users." Eight graduate students and four industry researchers checked, "publishing the results of this research." Notably, five respondents—all industry researchers—did not check any ethical violations. Nine industry researchers and three students believed "the benefits of this research …outweigh the risks." Only seven industry researchers and two students indicated the research should proceed "as proposed." All valuations are graphed in Figure 1.

Chi-square tests of independence determined that the following differences were significant:

- Graduate students were more likely than industry researchers to check "failing to secure consent from search engine users" as an ethical violation, X^2 (1, 22) = 4.709, $p < .05$
- Industry researchers were more likely to report that the benefits of this research outweigh the risks, X^2 (1, 22) = 4.200, $p < .05$
- Industry researchers were more likely to report that the research should be allowed to proceed as proposed, X^2 (1, 22) = 3.535, $p < .10$

Responses to "Please evaluate the ethics of this research" elucidate these findings. Some respondents elaborated on the ethical violations suggested above. Many offered suggestions for correcting these violations. Respondents drew heavily on social and research norms in their evaluations.

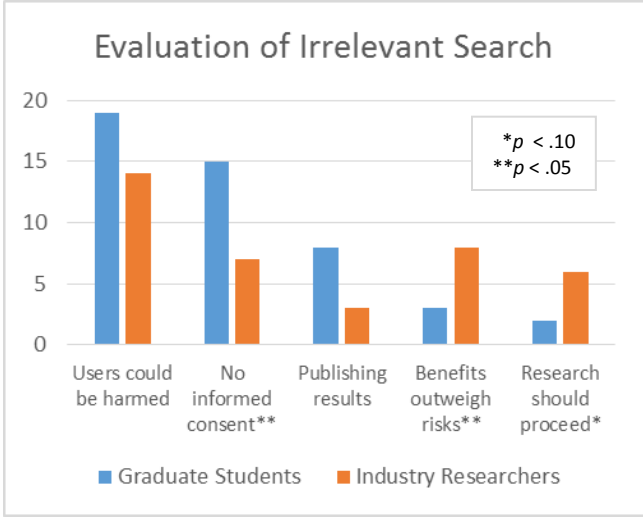

Figure 1. Graduate Student and Industry Researcher evaluation of Irrelevant Search.

Potential for harm

Students and industry researchers alike believed the potential for harm was determined by the query—"*outdated sports information may be irritating … dated medical information could have very harmful effects*." Respondents from both groups recommended restricting this research to a *"whitelist"* of benign queries. Industry researchers offered additional solutions. One recommended a disclaimer *"suggesting results should be examined carefully for accuracy."* A second noted that researchers *"should try to mitigate the fact that a user clicks on a bad result by almost immediately bringing their attention to the latest most up to date result."* A third suggested a lab experiment user study as a more suitable design, for reasons of research validity as well as ethics.

Respondents also identified potential harm to the company. For students, harm comes through reputational damage—"*if this study is approved as is…the search engine company will set themselves up for a Facebook like situation*." In contrast, industry researchers percieved damage in decreased use of the search engine and related loss in revenue.

In line with the Belmont Principles, respondents often contrasted the risks and benefits of participation. While a number of students saw no benefits, a few researchers wrote that *"customers benefit when A/B tests lead to better products."* A second industry researcher suggested that risks would be mitigated if participants were "*compensated for the trouble*."

Failure to gain informed consent
Many identified consent as *"a sticky issue."* Some students characterized the lack of consent as a blanket violation; industry researchers were more likely to suggest consent "*for this study*" due to specific aspects of the research design. One explained, *"Many sites use A/B tests to test improved versions prior to shipping. In this case, in contrast, the users are exposed to a degraded version of the service, and I would expect using a group of participants that are briefed."*

Both students and industry researchers noted, *"While gaining [Informed Consent] is impractical, debriefing may not be."* Some students suggested that users should be permitted to *"opt in"* or "*opt out*" of research as an alternative that does not "*compromise the integrity of the study*." Making a slightly different point, one industry researcher wrote, *"if users are dissatisfied with the results of their search, they can always opt-out by using another search engine."*

Publication of research results
The majority of students who identified an ethical issue with publication were concerned with protecting the anonymity of participants, namely by categorizing queries rather than quoting them "*verbatim*." One student also suggested that publication exacerbates other ethical grievances- *"the publishing of results gained unethically is an unethical behavior in itself."* Interestingly, one student believes debriefing is only necessary if the research is published. While the underlying rationale is unclear, it should be noted that this assertion is consistent with the definition of research as actions to produce general (i.e., public) knowledge, as opposed to business operations.

Social and Research norms
Many considered social norms as part of their evaluations. Noting the potential for harm, one industry researcher explained, *"We have as a society learned to trust search engine results."* However, another believes, *"users have learned that not all search results are good, and have learned how to evaluate [a] page credibly."* Respondents form both groups noted that *"similar research is common and largely not known by normal internet users,"* and viewed *"educating people about the issue"* as a solution to some ethical concerns.

A few participants compared the scenario described in Irrelevant Search with hypothetical research projects. Two students compared Irrelevant Search to Facebook's Emotional Contagion study, while a third student asserted, *"if the intern was showing biased results instead of just old ones, I would have a bigger problem."* And, as noted earlier, industry researchers often compared the specific A/B test described in Irrelevant Search to the broader practice of A/B testing.

This finding is consistent with the understanding of how case studies encourage ethical contemplation through the comparison of relatively straightforward paradigm cases with less clear situations [4].

Discussion

Our research validates adopting a case study approach, and the specific case study Irrelevant Search. We find respondents able to elaborate on the suggestions offered by three potential ethical issues and the optional "show me" sections, articulating their underlying reasoning and connecting their rationale to social and research norms. And, 70% of respondents report that Irrelevant Search is useful for understanding Internet research ethics.

Consistent with [8], we observed a number of differences between our two target populations. Industry Researchers were more lenient in determining that the benefits of our scenario outweigh the risks, and in suggesting that the research may proceed as proposed. The former finding, may be explained by the observation that industry researcher propose more as well as more varied solutions to the ethical (and methodological) issues with Irrelevant Search as graduate students do. The later finding may be related to articulated research norms, and to the understanding that *"customers benefit when A/B tests lead to better products."*

We also find that Students are more likely to identify "failure to secure informed consent" as an ethical violation. Based on the open-ended responses, this may be because students consider consent less negotiable than industry researchers, who emphasize the unique aspects of a research proposal, do.

Analyzing the remaining case studies will further clarify these findings, and shed additional light on the ethical valuations of those who engage in Internet research.

References

[1] Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee. http://www.aoir.org/reports/ethics.html

[2] Fleischmann, K., Robbins, R. and Wallace, W. Collaborative learning of ethical decision-making via situated cases. *Proc. iConference 2011,* ACM Press (2011), 319-326.

[3] Goel, V. As data overflows online, researchers grapple with ethics. http://www.nytimes.com/2014/08/13/technology/the-boon-of-online-data-puts-social-science-in-a-quandary.html?ref=technology&_r=1.

[4] McKee H. and Porter, J. *The Ethics of Internet Research: A Rhetorical, Case-Based Process*. Peter Lang Publishing, New York, NY, USA, 2009.

[5] McGinn, M. and Bosacki, S. Research ethics and practitioners: Concerns and strategies for novice researchers engaged in graduate education. *Forum: Qualitative Social Research, 5,* 2 (2004), 1-14

[6] Rudder, C. We Experiment on Human Beings! http://blog.okcupid.com/index.php/we-experiment-on-human-beings/

[7] Schroepfer, M. Research at Facebook. http://newsroom.fb.com/news/2014/10/research-at-facebook/

[8] Sumner, M. and Werner, K. On-line ethics: A comparison of the attitudes of freshmen, MIS majors, and practitioners. *Proc. SIGCPR '97,* ACM Press (1997).

[9] Quinn, M. Case-based analysis: A practical tool for teaching computer ethics. *Proc. SIGCSE '06,* ACM Press (2006), 520-524.