“Ethical Issues in the Employment of User-generated Content as Experimental Stimulus: Defining the Interests of Creators”

Ben Merriman
April 1, 2015

Department of Sociology, University of Chicago. merriman@uchicago.edu


Abstract

Social experimental research commonly employs media to elicit responses from research subjects. This use of media is broadly protected under fair use exemptions to copyright, and creators of content used in experiments are generally not afforded any formal consideration or protections in existing research ethics frameworks. Online social networking sites are an emerging and important setting for social experiments, and in this context, the material used to elicit responses is often content produced by other users. This article argues that users may have a reasonable interest in controlling the use of their content in experiments conducted in online social networks. Matters of risk and autonomy in research ethics are explored by analogy to active debates in law over adhesion contracts, moral rights, and the right to be forgotten. The article concludes by considering practical difficulties in identifying and protecting the interests of creators.

Introduction

Human behavior in networks is a topic of significant and rapidly expanding interest in the social sciences and other disciplines. Extensive bodies of evidence show that networks have a considerable effect upon a wide range of social and individual outcomes (Burt 2000, Smith and Christakis 2008), and research is increasingly directed toward the use of social networks to alter behavior (Valente 2012). The advent of online social networking sites has been a particular spur to such research. Social networking sites may provide unusually large, rich, and orderly sources of data. They are also of undoubted social significance in their own right: in the United States, 74% of online adults use social networking sites (Pew Internet Project 2014). Many researchers have begun to use experimental
methods in studies of online social networks. Such work has substantiated the effect that online networks—and the manipulation of networks and networked content—can have on important outcomes such as health behaviors (Centola 2010, 2011), voting (Bond et al. 2012), and mood (Kramer, Guillory, and Hancock 2014).

The publication of Kramer, Guillory, and Hancock (2014) inspired considerable public controversy, and has also prompted researchers to give further thought to questions of risk and consent in networked experimental studies of this kind, though these ethical concerns are certainly not new (see for instance Ess and AoIR 2002). Experimental studies on social networking sites have aroused concern, in part, because they expose a major gap in the regulatory coverage of Institutional Review Boards (IRBs) in the United States (Salganik 2014) and Research Ethics Committees (RECs) in the United Kingdom (Harriman and Patel 2014). This is unsurprising, given that these frameworks were created to regulate very different forms of research in a different context (see Stark 2012). This study also provoked concern because participants did not consent to be the subjects of experiments (Tufekci 2014). The responses to the Kramer, Guillory, and Hancock are too numerous and diverse to discuss in depth here. However, it is worthwhile to call attention to points where there is significant agreement: as a matter of principle, informed consent and ethical oversight are both desirable (Watts 2014), and as a matter of practice, these goals are difficult to achieve in a research context that is rapidly evolving and spans many different institutions and institutional forms (Gray 2014).

In experimental studies on social networking and social media sites, individual subjects are selectively exposed to materials such as images, texts, film, or other media in a way that may have some measurable or patterned effect on their behavior. This exposure is commonly referred to as a “stimulus” or “treatment.” The material employed to induce a behavior in research subjects is often content generated by other users. The interests of content creators have, to date, received far less attention than the interests of persons who are the direct subjects of experiments. In addition, because content creators are not considered to be research subjects, no formal rules exist to govern the use of user-generated content in experiments, though this use of media is analogous to research uses for which guidelines do exist. When content creators and experimental subjects are embedded in the same network or social networking service, these interests may meaningfully coincide.

This article considers ethical issues in the employment of user-generated content as stimulus in networked experiments. It begins by describing the current regulatory framework, fair use doctrine, and shows that this approach, while appropriate for traditional media, poses problems for content produced and shared in experiments on social networking sites, the experiment often takes a form very similar to the routine practice of A/B testing, where different users are shown different content or different versions of the website. The distinction between research and the ordinary operation of a service is often unclear, as the modification of algorithms is an ubiquitous feature of life online (Gillespie 2014). It may be added that this ambiguity is of strategic value to providers of services (Gillespie 2010).
in social networks. The article then describes several open problems in privacy and intellectual property law and their relevance to matters of research ethics. The core argument of the article is that content creators have interests in the use of their creations that are very similar to the interests of research subjects, and may have a fair expectation of similar protections. Some of the practical challenges arising from this claim are developed in the discussion section.

**Absence of Formal Protections for Content Creators**

As noted in the introduction, experiments in online social networks commonly induce an effect by selectively exposing research subjects to text, images, or other media produced by other users of a service. Such media are generally entitled to copyright protection: ordinary creations such as status updates, tweets, blog postings, photographs, and webcam recordings are protected by copyright to the same extent as a best-selling novel or a Hollywood blockbuster. In the United States, many scholarly uses of such media are protected under the fair use exemptions set forth in the Copyright Law Act of 1976 (17 U.S. Code § 107). Of particular relevance for this discussion is the exemption to use “copyrighted material to stimulate response, discussion, and other reactions during research” (International Communication Association 2010: 11). This fair use exemption allows researchers to use copyrighted media in experiments without prior clearance or permission from copyright holders. This legal exemption is certainly crucial, as it allows research to proceed without high transaction costs or needless obstruction.

Although fair use provides researchers with a legal basis to employ user-generated content in experiments, it does not adequately address ethical issues specific to the context of social networking and social media. In online communities or networks, communicative acts are embedded in large, reciprocal chains of interaction and groups of peers (see Van Dijk 2009). In such contexts, the distinction between person and product is not clear. Media that are, for purposes of law, fixed records of performances produced by a person are, in practice, much closer to everyday activity (Cohen 2012) or persons themselves (Floridi 2005: 194-197). For this reason, the lack of a clear distinction between data and person as object of study is understood to be a major problem for the ethics of internet research in general (Markham and Buchanan 2012: 7). The issue is particularly thorny in experiments, where material is deployed for a different end within the same context: the relational aspects of “content” are complex and socially significant (Lomborg 2012; for a further treatment of relationality see Silver and Lee 2012). In short, the presumptions underlying the fair use exemption for the employment of media in experiments do not accurately represent the production and dissemination of media within social networks. The phenomenon is much more akin to social interaction than broadcast.

For this reason, it is not clear that social experiments within networks can
cleanly disentangle research subjects from the users whose content serves as the
stimulus to subjects. Indeed, experiments are motivated by the presumption
that network effects exist, breaking with the traditional assumption that the
experimental treatment of one individual does not affect the treatment of an-
other (see Cox 1958, Manski 2013). Network experiments examine, in part, the
reasons why effects are distributed unevenly: some individuals, for reasons re-
lated to the structure of a network, will be highly susceptible to the experiment,
while others will not. This is itself a key object of study (Aral and Walker 2012,
2014; for a classic treatment see Katz and Lazarsfeld 1955). This poses practical
problems for research ethics. First, the effects produced by experiments may
spread from the experimental subjects to others in their network, including in-
dividuals who generated the content for the experiment, though in many cases
content creators will also be research subjects. In addition, individuals creating
content in and for networked peer communities are generally less powerful than
traditional media producers. They are more vulnerable to reputational injuries
that may arise from manipulated or misleading representations of the content
they produce, particularly if this content is presented experimentally to much
or all of a user’s peer network.

The preceding serves to make a brief case that experiments in online social
network and social media sites may raise concerns related to content creators
as well as experimental subjects. The matter, in light of this, is to define
what ethical consideration researchers can or should extend to individuals who
provide content for experiments. This is not a question of legality, but law may
nevertheless provide useful ideas. The following section explores these issues by
briefly considering three current topics in law: adhesion contracts; the moral
rights of creators; and the right to be forgotten. Each of these issues in law
speaks to matters of fairness, autonomy, and privacy that have direct analogues
in research ethics.

Three Legal Issues and Their Research Ethical
Correlates

Computing technologies and the internet have unsettled many structures of
civil law just as they have created new problems for research ethics. For this
discussion, it is especially notable that these technologies have raised significant
concerns about fairness in contracts between those who provide and use services,
as well as the privacy and creative rights of individuals. This section offers
short characterizations of legal issues, and then considers what ethical questions
each issue may raise for the employment of user-generated content in online
experimental research.

Adhesion Contracts and Uninformed Assent

Adhesion contracts are “standard form contracts presented on a take-it-or-leave-
it basis” (Rakoff 1983: 1174). Most of us assent to such agreements regularly:
adhesion contracts are the masses of fine print that consumers generally do not read, and are particularly common in computing. Some come in the form of shrink-wrap or click-wrap licenses that users accept when making use of software or other services (see Ryan 1988). Others come in the form of “browse-wrap,” allowing users to give tacit assent to contractual terms without viewing them at all (Rambarran and Hunt 2007). By definition, these contracts are not open to negotiation; users agree to the terms or do not use the service. These contracts often impose highly disadvantageous provisions, many of which are likely to be unintelligible to a non-specialist, and some of which are so skewed that courts decline to recognize them (Korobkin 2003). Power imbalance, lack of consideration of the contract, and extreme unfairness of certain provisions all pose major problems for courts. The general legal enforceability of browse-wrap contracts remains uncertain (for a current overview see Laven 2014).

In Bond et al. (2012), Kramer, Guillory, and Hancock (2014), and other studies, a contract of this kind was taken to provide implied consent to be subject to an experiment. Notably, these terms also provided the researchers the use of the material used to elicit the responses from subjects. Facebook users, like users of social networking and media sites more generally, grant the service provider non-exclusive but very broad rights to employ user-created content and metadata. In such cases, users would not know when or how content they produced was used in experiments, nor would they be able to indicate a wish that their content not be used in this way, either in general or in specific studies.

For several reasons, this kind of contract would probably not pass muster as an agreement to be involved in research if such studies were subject to formal oversight. First, there are strong reasons to believe that users are broadly uninformed when they enter into agreements that make their creations available for experimental purposes; aside from the fact that users may assent to agreements without reading them, the text of the agreements is likely to be opaque to a person without legal training. Although comprehension is not required for a contract to be legally enforceable, understanding is a core value of the consent process in research. In addition, the take-it-or-leave-it provisions, as well as extensive retention of user-created content, do not provide an easy means for an individual to deliberate or withdraw, as might be expected in an agreement to participate in research. The extent to which this is a concern has much to do with the possible consequences of a given study; some experiments may indeed have a plausible claim of posing minimal risk to subjects, while others do not. For this reason, it unreasonable to assume that exemption or blanket, one-time implied consent can be sufficient to manage all possible risks.

Moral Rights and Creative Autonomy

An evolving issue in copyright law is the extent to which a creator’s moral rights are recognized. Moral rights are inalienable interests that individuals retain in work they create, commonly understood to include rights to “attribution, integrity, disclosure, withdrawal, and resale royalties” (Liemer 1998: 46). Creators retain these rights even if they have sold their work or assigned certain rights
to others. The idea of moral rights originated in civil law jurisdictions, and historically they have not enjoyed broad recognition in common law countries like the United States. However, scholars have argued that several principles in common law afford a basis to recognize moral rights (Damich 1988, Lee 2001, Kwall 2010), and moral rights have growing acceptance in many common law jurisdictions (Rigamonti 2006). The common law right of publicity may also have some points of substantive overlap with moral rights (for an overview see Bartholomew 2011). The legal notion of moral rights rests on a view of individuals as autonomous and creative: an injury to the individual’s ability to create freely and with integrity is also an injury to the individual as such. As is the case with copyright, moral rights attach to ordinary expressive products; a modest effort shared with a limited public of peers is still meaningful for the creator, and entitled to protection.

This view has strong affinities with the conception of individual autonomy undergirding regulations in human subjects research (Berry 2004). It is also concordant with major accounts of social practice online. Cohen (2012), for instance, understands creative everyday practice to be the central feature of individuals’ networked lives. Two dimensions of moral rights may be of particular relevance in deliberation over ethical research conduct. First, the right of attribution: a recognized interest in being accurately identified as the creator of a given work poses a challenge to experiments that manipulate the apparent source of a given creation or make false attributions of authorship (for instance, Canini, Suh, and Pirolli 2011). Second, researchers should consider what the integrity of a work might mean—some experimental contexts may use creations in ways that run strongly contrary to the intention of a work. Overly broad concession of these rights would probably make many kinds of experimental research impossible. Fairness in attribution and integrity is better understood as a question of degree; it may be worthwhile for researchers to consider when an experimental use of a creative product differs so much from the original use that it violates individuals’ legitimate interest in the integrity of their creative expressions. Although these considerations are partially distinct from questions of privacy, Nissenbaum’s conception of privacy as contextual integrity (2009), and the elaboration of this view by internet researchers (Ess 2014: 58-75), may provide a useful basis for applying these concerns in practice.

The Right to Be Forgotten and Individual Privacy

Moral rights are founded in an implicit belief in the value of creative effort. A recent development in privacy law, the right to be forgotten, has similar practical implications for researchers, but derives from distinct premises about individuals and their interests. The right to be forgotten, presently being implemented in the European Union, entitles individuals to have personal information deleted by other parties who hold such data (Rosen 2012). There are certain points of significant friction between this right and features of US law (Bennett 2012), and many practical obstacles that may limit successful implementation (Ausloos 2012). However, the right is broadly concordant with commonly accepted
notions about privacy.

The right to be forgotten has been constructed from basic principles. However, the perceived need for this right arises from serious practical difficulties in managing one’s disclosures online: this is likely to be prohibitively time consuming (Martin 2013), and apt to fail because information is embedded in networks partially or wholly controlled by other actors (Garg et al. 2013). In other words, a categorical right of self-deletion has emerged because it is impracticable to manage personal content through series of private agreements between users and service providers (see Elkin-Koren 2005). This is particularly so when the terms of a private agreement have been dictated entirely by the service provider. The availability of user-generated content, then, may often be inadvertent or the result of regulatory failure, and for purposes of research, it should not always be assumed that publicity is intended (Nissenbaum 2011; on “private publicity” see Lange 2007).

Discussion

Fair use exemptions to copyright law allow scholars to employ the creative works of others in experimental contexts without formal permission. The creators of content employed in human subjects research are also not afforded specific protections under existing frameworks for research ethics. In themselves, neither of these facts are a cause for concern. However, the advent of online social networking has created a novel context for research, and experiments in this context commonly employ user-generated content to induce effects. This article has argued that content produced within online social networks is different from traditional media: such content is embedded within patterns of everyday social interaction and is intrinsic to free, expressive individual practice. The preceding review of legal and ethical scholarship has shown that both of these differences may be taken as reasons to afford some formal protections to individuals who produce content employed in online, networked experiments. The two issues correspond, respectively, to ethical concerns about beneficence and risk, and concerns about autonomy and respect for persons.

The employment of user-generated content for an alternative, experimental purpose within the same socially embedded context may pose risks, either from the diffusion of network effects induced by the experiment, or by affecting the way an individual is perceived by others in ways not directly intended by the purposes of the experiment. This claim is not meant to be alarmist: to date, no clear claim of harm has arisen. However, the issue merits consideration because the state of scientific knowledge about networks clearly suggests that such harm is possible. In light of the structural properties of networks, the small aggregate effects of most experiments should not be taken to mean that the possible effect on any given individual is also small.

The matter of autonomy is less speculative. Individuals have a legitimate interest in exercising some control over how their creations are used. This control is particularly important when individuals produce content intended for
an audience of networked peers. However, the preceding section has shown that there are serious legal and technological obstacles to full management of one’s creations online. The fact that content is technically accessible does not mean that creators intend for it to be available or understand it to be public or published.

There are a number of practical difficulties in protecting the interests of content creators. The first is simply recognizing that an ethical concern exists. Work in digital media ethics has identified many issues that are specific to online research; the blurred line between data and person and the highly contextual character of privacy are particularly salient here. Many questions of ethical practice in online research do not admit of a simple answer, but deliberation in itself may encourage researchers to conduct their work with greater care. Though such problems may be familiar to researchers with a background in internet research, online experiments are conducted by researchers in a number of disciplines. Many researchers may field experiments online simply because they are convenient, and may be unfamiliar with concerns specific to online research. Similarly, disciplines have their own conceptions of ethics and touchstones for ethical (and unethical) research. These can obstruct thinking about the issues posed by novel research contexts (Tolich 2014).

There is, of course, a very direct means of addressing the concerns raised in this article: researchers can ask users for permission to employ their content in experiments. This could be sought as part of the consent process for research participants in a given experiment, or could be solicited separately. Garnering such permission in studies that involve millions of individuals would be no simple matter, though Hong (2014) discusses prospects for massively scalable approaches to consent. To the extent that experiments of this kind almost necessarily include both academic and private researchers, a system of oversight that does not rely solely on universities is also desirable (boyd 2014). These considerations point to another difficulty: providers of social network services have a stated legal claim on the content produced by users, and an economic interest in understanding how their services are used. The purpose of this article has been to defend the interests of creators, not the interests of platforms, but the operators of platforms have legitimate interests. Some pragmatic means of balancing these different sets of interests is needed, particularly because private companies could readily exempt themselves from most regulations (and scrutiny) by excluding academic researchers and declining to publish the results of experiments run internally. Such an outcome would not be in the scientific or public interest.

This article has argued that users of social networking services have a plausible interest in how their content is used in experimental research on grounds of risk and autonomy. However, at present they have no control over this use of content, and are not afforded formal protections under existing frameworks for research ethics. The extent to which protections can or should be afforded to content creators is a complex practical problem. This article has pointed to relevant questions but offers no definite solution. Online experimental research is an emerging research paradigm that examines a rapidly evolving social con-
text. This necessarily creates new ethical questions. As this practice continues to develop, it is incumbent upon researchers to consider anew what it means to act with respect, beneficence, and fairness in their work.

References


