

The ethics of online research with unsuspecting users: From A/B testing to C/D experimentation

Research Ethics
2017, Vol. 13(3-4) 200–218
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1747016116680664
journals.sagepub.com/home/rea


Raquel Benbunan-Fich

Baruch College, CUNY, USA

Abstract

This article analyzes recent cases of company-sponsored online experiments with unsuspecting users and discusses the ethical aspects of such experimentation. These cases illustrate a new type of online research where companies modify their algorithms to intentionally misinform or mislead users. Unlike typical forms of A/B testing, where two versions of the same website are presented to different users to evaluate interface changes, algorithm modification is a deeper form of testing where changes in program code induce user *deception*. Thus, we propose to call this new approach C/D experimentation to distinguish it from the surface-level website evaluation associated with A/B testing. Three aspects raise ethical concerns regarding C/D experimentation: the absence of user consent to participate in research, the presence of intentional deception, and the complete lack of protection for human subjects who partake in privately funded behavioral research. Three recommendations are proposed to address these issues: (i) to develop an ethical code of conduct for subject protection shared by online companies, (ii) to include special provisions for C/D experiments in social networking platforms, and (iii) to create an independent user advocacy board to protect the rights of users who partake in online research conducted in the private sector.

Keywords

deception, ethics, human subject protection, online research, social networks

Corresponding author:

Raquel Benbunan-Fich, Baruch College, CUNY, One Bernard Baruch Way, Box B11-220, New York NY 10010, USA.

Email: rbfich@baruch.cuny.edu

Introduction

In the summer of 2014, a prominent academic journal published the results of a study about emotional contagion conducted by a Facebook researcher in collaboration with two faculty members (Kramer et al., 2014). The premise of the study was to test whether emotional states can spread through social networks. The research consisted of two parallel experiments (reduced-positive and reduced-negative), each with a treatment and a control condition. In the reduced-positive experiment, when friends' positive posts were algorithmically removed from the News Feed, users posted slightly fewer positive words and more negative words in their subsequent status updates. The opposite happened in the reduced-negative experiment. When friends' negative posts were eliminated from the News Feed, users' words reflected more positive emotional content. These results show *emotional contagion* in both treatment conditions. Although the effect was very small (one-tenth of a percent of the observed change or less), its statistical significance provided evidence that emotional states can spread through social networks.

The Facebook emotion experiment is just one example of a new wave of company-sponsored online experiments with unsuspecting users. These experiments are novel because online companies modify the inner workings of their algorithms to change or curate information about the users themselves or about their friends (i.e. social network connections). As such, this is a deeper and more advanced form of behavioral experimentation than the traditional surface level or interface usability testing aimed at improving the design of websites, known as A/B testing. When programming code is altered to manipulate results about some users or about their connections, without forewarning, intentional deception takes place. To highlight the fact that code is altered and deception takes place and to avoid confusion with A/B testing, we propose a new name – C/D experimentation – for this type of research.

This article analyzes several cases of company-sponsored online experiments with unsuspecting users, distinguishes A/B testing from C/D experimentation, and discusses the need to establish ethical guidelines for this form of experimentation. Although company-sponsored research provides valuable data, three aspects raise ethical concerns regarding C/D experimentation: the absence of user consent to participate in research, the presence of intentional deception, and the complete lack of protection for human subjects who partake in privately funded behavioral research. To address these issues, three recommendations are proposed: (i) to develop an ethical code of conduct for subject protection shared by online companies, (ii) to include special provisions for C/D experiments in social networking platforms, and (iii) to create an independent user advocacy board to protect the rights of users who partake in online research conducted in the private sector.

Company-sponsored online experiments

Online companies such as Facebook, Google, and others are in a unique position to conduct large-scale research with their users, and obtain valuable data about human behavior. Some advocates welcome corporate-led experimentation because of its potential to identify unsafe or ineffective organizational practices (Meyer and Chabris, 2015). The benefits of company-sponsored research are not limited to the improvement of the firm's internal operations. Given the prevalence of online activities nowadays, this research has the potential to shed light on modern human behavior. Moreover, when the results are disseminated via journal or newspaper articles, they advance the collective knowledge about how people behave online, how ideas spread through social networks, and how users think, feel, and act (Manjoo, 2014).

The upside of deploying experiments via social networks is the immediate access to a large sample of online users, the convenience of running simultaneous conditions, and the option to study online and offline human behavior from a new perspective (Manjoo, 2014). Only the companies providing these online platforms can conduct this kind of large-scale research. The downside of this approach is the danger of overlooking fundamental safeguards for the protection of human subjects who unknowingly become participants in these experiments. We argue that the risks are more severe when company-sponsored research includes some form of deception. This point will be illustrated by reviewing two specific cases: Facebook's emotional contagion study and OKCupid's mismatching experiment.

Facebook's emotional contagion study

For the emotional contagion study, Facebook researchers developed a special program to automatically determine the valence of the emotional content in the users' News Feed and adjust it according to the experimental condition. The News Feed is a feature that displays comments, videos, pictures, and links posted by other people in the user's social network, based on an algorithm that automatically determines what to show for each user. The News Feed results are personalized for each user and updated continuously depending on the users' connections and activity on Facebook (Facebook Help Center¹).

As is the case with many other software features, the algorithm behind News Feed has been subject to evolutionary changes since its inception. Facebook describes its News Feed as a 'digest' of relevant news updates built with each user's preferences and with information about their connections. Whereas the content shown by the News Feed algorithm is tailored to each user, the process to generate the results is common to all users; it shows updates from friends and pages with which users interact frequently.

The results produced by the News Feed algorithm were altered to implement two parallel experiments (reduced-positive and reduced-negative) conducted within the emotional contagion study. Both conditions reduced user exposure to certain kinds of emotional content posted by friends and displayed via the News Feed. In the reduced-positive condition, upbeat words were suppressed, whereas in the reduced-negative condition, downbeat words were eliminated from the News Feed of selected users. Each experiment had its own control condition, in which a similar proportion of posts in the News Feed were omitted regardless of their emotional content.

Data for the study were gathered for 1 week in January of 2012 from 689,003 randomly selected individuals using the English version of Facebook. This version is used by about half of the 1.3 billion active Facebook users. The results indicate that users in the reduced-positive condition had a larger percentage of negative words in their own status updates and a smaller percentage of positive words. The opposite pattern was observed in the reduced-negative condition. Although the effect size for the manipulations was very small ($d = 0.001$), the findings provided statistically significant evidence of emotional contagion via social networks (Kramer et al., 2014).

Some critics caution that given the large sample size (more than 600 thousand users), the results may be ‘statistically significant but substantively trivial’ (Morin, 2014). In larger samples like this one, even small differences pass tests of statistical significance. For example, in the reduced-positive condition, the number of negative words used in status updates increased by 0.04% on average (i.e. only about four more negative words for every 10,000 written by these participants), and the number of positive words decreased by only 0.1%. Likewise, in the reduced-negative condition, seven fewer negative words were posted per 10,000 and the number of positive words increased by about six per 10,000. These results would not have been detected in a smaller sample size.

By implementing the emotional contagion manipulation in the results of the News Feed algorithm for randomly selected users, without forewarning, deception occurred. Facebook users understand that the News Feed algorithm works in the same way for all, though the results are different for each user depending on their connections. Thus, the assumption that the News Feed uses a similar algorithm for all users was violated. In fact, the addition of the extra program that analyzed the nature of emotional content and automatically decided which content to suppress depending on the condition introduced an additional layer of ‘editorial curation’ that was never disclosed to users.

News Feed manipulation and influence on offline behavior

A prior Facebook experiment about political turnout also manipulated the content displayed in the News Feed for research purposes, albeit with a different type of

curation and more startling results. In an experiment conducted during Election Day in 2010, Facebook researchers randomly divided 61 million American users into three conditions: social message ($n = 60$ million), informational message ($n = 611,000$), and a control group ($n = 613,000$) (Bond et al., 2012a). Each of the two message-conditions (social message and informational message) was shown a different, but non-partisan, get-out-to-vote statement.

The *social message group* was shown a statement at the top of their News Feed to encourage the user to vote, along with a link to local polling places, a clickable button with the words ‘I Voted,’ a counter indicating how many other Facebook users reported voting, and up to six small randomly selected profile pictures of the user’s Facebook friends who had already clicked the ‘I Voted’ button. The *informational message group* was shown the message, poll information, counter, and button, but none of the friends’ pictures. The *control group* did not receive any message at the top of their News Feed. The results indicate that those who received the social message were more likely to vote than users who received the informational message and those in the control group. There was no difference in turnout between the informational and control condition.

The effects of the News Feed’s manipulation on actual voting were validated ‘through examination of public voting records’ (Bond et al., 2012a: 295). According to the Supplementary Information of this study (Bond et al., 2012b), several states provided Facebook with their publicly available voting records for research purposes,² allowing Facebook to compile a list of over 6.3 million matched subjects with their corresponding online user account and offline voter information. This matched sample was used to perform a statistical analysis on the relationship between online treatment conditions and real-world voting behavior.

Compared to the emotional contagion project, the political turnout experiment generated much less publicity and very little backlash. Both experiments crossed the line between online and offline behavior by analyzing how social influence and contagion can change people’s voting behavior or emotional states. In the turnout experiment, Facebook’s evidence indicates that social messages showing the faces of friends who voted had a direct effect on increased turnout directly by approximately 60,000 voters and an indirect effect through social contagion. These findings indicate that ‘seeing faces of friends significantly contributed to the overall effect of the message on real-world voting’ (Bond et al., 2012a: 296).

This political turnout experiment provided evidence of the connection between online information displayed via social networks and actual behavior, and demonstrated how companies can effectively match online users with records of offline activity. Although this raises a number of privacy concerns,³ Facebook researchers argue that they never ‘see’ individual records because the matching is done with computerized programs. A similar argument was used by the researchers in the emotional contagion study when they claimed that the modification of the News

Feed was done with a custom-developed program. The issue is not whether the manipulation was performed manually or automatically, but the fact that it was implemented without notification to alter the results of a random sample of unsuspecting users. Although non-disclosure of News Feed curation could be considered a more subtle form of misinformation compared to deliberate lying, both are forms of deception.

In both studies, Facebook manipulated the News Feed algorithm to either suppress emotional information or enhance get-out-to-vote messages displayed in the automatically-generated and user-personalized flow of updates. Thus, deception occurred through the modification of the News Feed algorithm *results for some users* without previous notice. Deception through information manipulation takes place when a transmitting party intentionally introduces some form of misrepresentation to influence the behavior of the receiver (Johnson et al., 2001). Misrepresentation occurs either through *distortion* (disclosing half-truths created with concealment and/or embellishment) or through *falsification* (communicating false information as if it was true) (Ekman, 1988). The next section describes a case of falsification.

OKCupid experiments with matching algorithm

Other online companies have also pushed the boundaries of online experimentation. At around the same time as the publication of the Facebook emotion experiment, the online dating website OKCupid admitted that their researchers also conducted experiments with its users. OKCupid is a free dating site that matches users through mathematical algorithms based on answers to questions about their preferences and tastes. At the time of the study, the size of its user base was 12 million people, according to estimates reported in the business press (Suddath, 2014).

The company blog described three *separate* experiments (Rudder, 2014). In the first and second, OKCupid tested different aspects of their website interface. However, in the third experiment, OKCupid altered the compatibility percentage automatically provided by their matching algorithm to suggest that people were a much better or worse match than their actual match score (Wood, 2014).

The OKCupid mismatching experiment, dubbed ‘the power of suggestion,’ took pairs of users deemed as bad matches by the matching algorithm (30% compatibility level) and changed the level to display a 90% match, suggesting that they were a good fit for each other. As expected, misled users sent more first messages to their potential matches when they thought they were compatible, and engaged in conversations (or exchanged four messages or more) with their partners. At the conclusion of this experiment, affected users were notified of the correct match percentage. This is considered a form of debriefing because researchers informed participants about the manipulation after the deception occurred.

Upon examining the number of conversations that took place between mismatched partners, OKCupid was concerned that people interact because of the power of suggestion (induced by the fictitious compatibility level) and that their matching algorithm had limited power to predict real compatibility. In order to rule out this possibility, OKCupid tested additional combinations where compatibility matches of 30, 60 and 90 were either accurately displayed or changed for one of the other two percentages. The ideal combination, measured by the odds of an initial message turning into a conversation, was found in the 90%–90% condition, where people were a good match according to the algorithm, and they could see the true compatibility percentage displayed for them. By deliberately mismatching people, OKCupid tested its matching algorithm with selected users to examine their behavior.

The other two experiments conducted by OKCupid consisted of testing the effectiveness of the pictures and text displayed in the interface. In the first experiment, the company removed the pictures from all profiles in ‘love is blind’ day (15 January 2013). During the picture-blackout period, more communication and information exchange took place among ‘blind’ users, but those conversations stopped when the pictures were restored at 4 pm that day. In the second experiment, OKCupid changed the interface to show profile pictures with or without profile text and replaced the rating scales for personality and looks with only one scale measuring how ‘cool’ the person in the profile was perceived. The results show that the coolness ratings are entirely driven by the profile picture, and the profile text has no significant influence on the ratings given by other users.

These two experiments conducted by OKCupid (picture-blackout and pictures with/without text) are instances of what is commonly known as ‘A/B testing.’ In this type of testing, a fraction of users are diverted to a different version of a web page to compare their behavior with others who are using the standard site. In these cases, the changes to the interface are obvious and/or announced and do not involve deception. In contrast, in the mismatching experiment, the results of an algorithm were altered in a concealed way, thereby deceiving users. This is not a website design change; it is an instance of misinformation via falsification.

C/D experimentation: A new form of online research

A distinction must be drawn between traditional A/B testing and an alternative form of experimentation where algorithm results are modified for a fraction of users for research purposes. In A/B testing, interface design characteristics – such as arrangement of buttons, layout, or explanatory text – are blocked or rearranged to test their effects. Many online companies routinely perform A/B testing with their users to assess the impact of website design changes. However, a new form of experimentation emerges when the programming *code* of a website’s algorithm

is altered to induce *deception* with manipulated results. This is a deep form of testing, which we call code/deception or *C/D experimentation* to distinguish it from the surface level testing associated with A/B testing. C/D experimentation should be distinguished from the ongoing efforts of online companies aimed at improving their algorithms for operational purposes. Such cases of optimization do not involve deception because the objective is to produce better (more accurate) results *for all the users*. In contrast, in C/D experimentation the *results* of the algorithm are altered (i.e. distorted or falsified) for *some users for research purposes*.

The Facebook emotion and the OKCupid mismatching experiments involved instances of deception either through the *distortion* of program results (by reducing emotional content in News Feed) or through *falsification* (by communicating outright lies about compatibility percentages resulting from the matching algorithm). What makes C/D experimentation different from A/B testing is that misinformation about the user or about people in the user's social network is produced by a program that operates in the background but is presented in its usual format. In contrast, the manipulation in A/B testing is about the design of site (buttons, explanations, descriptive text, colors, etc.), and it is noticeable because the website itself has changed. Through the introduction of deception, C/D experimentation engenders additional risks for users, quite different from A/B testing.

The failure to recognize the distinction between these forms of online research may lead to the erroneous belief on the part of researchers and sponsors that this activity is covered by 'terms of use' agreements. For example, the researchers involved in Facebook's emotional contagion study argued that the acceptance of Facebook's Data Use Policy, which is a condition for establishing a user account in Facebook, provided consent for their study. However, it should be noted that the Facebook data use policy in effect at the time of the emotional contagion experiment in January of 2012 did not mention the possibility of using information collected by Facebook for 'research' purposes. To address this gap, a few months later, in May of 2012, Facebook's policy was amended to reflect several changes, including the addition of 'research' to the list of potential 'internal operation' uses (Hill, 2014). Whereas Facebook's terms of use did not include 'research' as a possible use for the information collected, the user agreement in effect at OKCupid did incorporate the possibility of using data for research and analysis purposes.

Implicit consent via terms of service agreements

Typical user agreements include some language to indicate that the company will use data for testing, troubleshooting, and service improvements. The argument is that online companies automatically acquire *implicit consent* for research when a user accepts the terms of service (TOS). Therefore, it can be argued that, unlike

Facebook, OKCupid did have *implicit* consent for the mismatching experiment by virtue of its TOS agreement.

User acceptance of the TOS agreement by clicking on a checkbox is one of the requirements for account creation in most social networking and other commercial websites. The binding action is exercised with a click-through instead of a signature, and for this reason these contracts are known as click-through agreements. These agreements are complex and difficult to read and thus raise doubts on the validity of ‘informed consent’ (Luger et al., 2013). Because of their length, most people fail to read the content of TOS agreements, and are unaware of their content. For example, the length of Facebook’s TOS at the time of the emotion study⁴ was about 6,700 words, and OKCupid’s current TOS are about 3,700 words. At a rate of 200 words per minute, it would have taken an average reader about 33.5 minutes to read Facebook’s TOS, and 18.5 minutes to read OKCupid’s. Yet, research shows that people spend an average of half a minute before clicking on the agreement box (Bakos et al., 2014).

Another study based on a content analysis of the TOS of 30 popular websites found that, owing to language complexity and the use of legal terminology, users may not understand which rights they are granting when they post their creative content on these sites, even if they take the time to read the terms (Fiesler and Bruckman, 2014). A software solution in the form of a browser extension has been developed to help individuals understand in plain language the main provisions of TOS.⁵

The lack of reading or understanding of TOS applies to all conditions and restrictions that users ‘accept’ when they click-through. Thus, users may not realize that they are implicitly consenting to participate in company-sponsored research without additional notice. Strictly speaking, because OKCupid did anticipate, and explicitly listed, research as one of the potential uses, it could claim that it had obtained implicit consent for experimentation. In contrast, because Facebook failed to include research purposes in the list of anticipated uses of information collected, it cannot assume that it had implicit consent for the emotional contagion study. In either case, *implicit* consent for research is not the same as *informed* consent for a specific study.

Informed consent and protection of human subjects

The main objective of informed consent is to make prospective participants aware of the research and give them the option to opt out of the study. The requirement to secure informed consent is the cornerstone of human subject protection. Regulations for the protection of human subjects emerged from unethical treatment of participants in the US Tuskegee study and abroad in the Nazi experiments. From the 1930s to the 1970s, the US Public Health services conducted a series of

experiments – called the Tuskegee Study – in which it withheld treatment and medical information from rural African-American men suffering from syphilis. In Europe, the Nazi experiments conducted in concentration camps during World War II resulted in the creation of the Nuremberg Code in 1949, which was followed in 1964 by the Declaration of Helsinki (Bulmer, 2001).

Similarly, in the USA, the public uproar caused by the Tuskegee experiments resulted in the passage of the National Research Act by the US Congress in 1974, and eventually led to a set of guidelines for the protection of human subjects used in research, known as the ‘Common Rule’ (US Department of Health and Human Services, 1991). According to Common Rule 45 CFR §46.102(d), *research* is defined as ‘a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.’⁶

With the requirement to obtain consent for research participation, prospective subjects are notified about the study. The goal of the consent procedure is twofold: (i) to *inform* prospective participants of the nature of the research study, and (ii) to give them the chance to participate in it if they so desire. Hence the term *informed* consent. Academic researchers often struggle with the amount of information to disclose in consent forms. On the one hand, explicit consent forms must include a description of the study, and outline the risks and benefits, so that participants can opt in with awareness of participation requirements and possible consequences. On the other hand, full disclosure of research details bears the risk of priming participants and altering their otherwise ‘natural’ behavior to conform to experimenters’ expectations (i.e. the experimenter–participant artifact problem). To balance the ethical duty to inform prospective participants with the need to leave their behavior unaffected, consent forms usually include only general descriptions of research objectives, risks, and benefits.

To ensure that all US federally-funded research complies with existing regulations, research protocols must be reviewed and approved by an institutional review board (IRB), before each research study is conducted. The review of research procedures is meant to ensure that risks to participants are minimized and commensurate with anticipated benefits. It must be noted that the Common Rule only applies to US federally funded behavioral and biomedical research conducted at academic and other institutions ‘for which a federal department or agency has specific responsibility for regulating as a research activity (45 CFR §46.102(e)).’⁷ Private companies like Facebook or OKCupid are not legally required to follow the regulations in the Common Rule. This legal loophole creates inequality in the protection of human subjects who participate in company-sponsored research. The absence of a legal regulatory framework creates opportunities for unethical research practices, with minimal adverse consequences.

The core ethical principles used as the guideline for the Common Rule were initially outlined in the Belmont report (Belmont Report, 1979). According to the

report, the three basic principles for ethical research conducted with human subjects at institutions that receive federal funds are: (i) respect for people, (ii) beneficence, and (iii) justice. *Respect for people* includes two related principles: individuals should be treated as autonomous agents, and individuals with diminished autonomy (children, prisoners, etc.) are entitled to protection. *Beneficence* refers to the obligation to treat individuals in an ethical manner by protecting them from harm, and making efforts to ensure their well-being. *Justice* is concerned with the equitable distribution of the benefits and burdens of research.

The C/D experiments described above did not comply with any of these three principles. First, because participants were uninformed and prevented from opting out, they were not treated as autonomous agents or afforded respect. Second, the manipulation of emotions or potential dating matches carried the risk of causing psychological harm to some users exposed to these practices, without any efforts to ensure their well-being, both of which are inconsistent with the principle of beneficence. Third, an undue burden was imposed on those who were randomly selected for experimental conditions that were clearly intended to benefit the sponsoring companies (by allowing them to test the power of emotional contagion or the power of suggestion), and therefore an injustice was committed against these unsuspecting users.⁸

On the basis of this analysis, it can be concluded that these experiments did not follow ethical guidelines for the treatment of research subjects. The unethical treatment of participants stems from the lack of specific consent and from the use of deception without safeguards. Kimmel et al. (2011) suggest an additional set of principles to follow when research involves deception. These principles highlight the need to forewarn and debrief subjects who consent to participate in research that involves deception (Kimmel et al., 2011).

Uninformed and deceived research participants

Intentional deception to stage experimental manipulations must be distinguished from the absence of full disclosure of research design information to prospective participants. Withholding the details of a study in the consent form is a common research practice and it is not considered an instance of deception because participants are informed that the study is taking place (Kimmel et al., 2011). In contrast, research practices that include ‘withholding of information to obtain participation, concealment and staged manipulations in field settings’ are instances of intentional deception (Kimmel et al., 2011: 226).

Both Facebook and OKCupid’s emotion experiments employed intentional deception with the potential to harm users who are emotionally vulnerable. Many people rely on Facebook for communication and support from their friends. Similarly, people who sign up for a dating service such as OKCupid are looking

for companionship and eventually long-term relationships. Manipulating the information users receive through their News Feed or through their compatibility matches may disproportionately harm those who are depressed, lonely, or in an emotionally fragile state.

In federally-sponsored and in academic research in the US, when a study involves deception, a partial or total waiver of informed consent may be obtained from the IRB. A waiver is granted when it is essential to carry out the research, when the study presents no more than minimal risk (typical risk encountered in everyday life), and when the waiver does not adversely affect subjects. Typically, when a partial or total waiver of consent is obtained, subjects are given additional pertinent information after their participation in a study, through a debriefing procedure. Both elements (consent and debriefing) are considered safeguard mechanisms for research studies using deception.

The nature of deception in OKCupid's mismatching and in Facebook's emotional contagion experiments is different. In OKCupid's experiment, deception occurred by giving users intentional and explicit misinformation about their compatibility matches (i.e. by lying about the results of the matching program). People sign up to OKCupid with the expectation that their matches will be accurately reported so that they can derive benefits from the site (dating, friendship, companionship, etc.), and were never told that the compatibility matches could be falsified for research purposes. Although OKCupid's TOS did include the possibility of conducting research, users were neither explicitly informed about the study, nor given the opportunity to opt out. There was, however, debriefing at the conclusion of the experiment.

In Facebook's emotional contagion experiment, deception occurred through the distortion of results of the News Feed algorithm to suit research purposes. This is an instance of violation of user expectations or assumptions that involves deception (Hertwig and Ortmann, 2008). In an interview, Ralph Hertwig of the Max Planck Institute for Human Development in Berlin specifically commented on Facebook's News Feed manipulation, '[A]n unannounced change to the digital code controlling what gets posted on Facebook users' News Feeds may be an 'implicit violation' of the site's contract with users who expect something else entirely.'⁹ Users, who view the News Feed as an unadulterated collection of random or best updates from their connections, as extracted by the News Feed program, were deceived by this additional emotion manipulation. Their results were distorted for research purposes.

Ethical analysis

Different ethical theories suggest alternative solutions for the dilemma of whether intentional deception in research is morally permissible. A fundamental distinction

in ethical theories is whether an act is evaluated in terms of its intrinsic adherence to moral rules or in terms of its consequences (Chatterjee et al., 2009; Mingers and Walsham, 2010). From the perspective of deontological theory, which maintains that the ethical decision is the one that reflects the adherence to moral rules or duties, deception in research is not an option. Doing so would conflict with the duty to be honest with participants. In the deontological view, consequences are not the most important consideration. In contrast, from the perspective of utilitarian theories, the morally right action is the one that maximizes the net expected utility for all parties affected by a decision or action. This theoretical perspective focuses on the consequences of an action. Therefore, deception in research is acceptable if it produces net positive value when comparing the benefits (i.e. advancement of science) to the costs (i.e. potential harm to participants). The normative prescription of this approach is difficult to implement because of the practical challenges of determining who is affected and who is favored, and the range of potential consequences (Kimmel et al., 2011).

The use of deception in research is typically justified from a utilitarian standpoint with the notion that society at large might benefit from the outcomes of this type of research. This benefit typically occurs when researchers disseminate their results for the advancement of science. This is not necessarily the objective that private companies like Facebook and OKCupid pursue with their internal research. In most instances, advancing knowledge about online user behavior is a circumstantial byproduct of research conducted for the company's own benefit.

Some argue that, even without public dissemination of results, society also indirectly benefits because these companies are able to provide better online services because of their internal research (Meyer and Chabris, 2015). In essence, however, those who stand to benefit the most are the companies pursuing the research, whereas those who bear all of the potential risks are the uninformed accidental participants.

In research, deception is a last-resort tool, only to be used when no other alternatives are available and more than minimal harm to subjects is not likely. Thus, research that is expected to cause psychological discomfort or severe emotional distress should not be conducted with deception because the inherent risk to cause disproportionate harm to participants would offset any potential benefits. At the very least, users should be given the choice to voluntarily partake in the study. When there is a power imbalance between those who conduct research for their own benefit and participants who bear all the risks, an independent entity is necessary to ensure that participants are not harmed and that the appropriate safeguards are in place.

Facebook's emotional contagion experiment and OKCupid's mismatching study manipulated emotions in different ways. With their respective manipulations came the potential to cause damage to emotionally vulnerable individuals, either

by contagion of negative emotions or by despair of finding unsuitable dating matches. In both cases, the risks to participants include negative feelings resulting from the study, which could disproportionately affect those who are in a fragile psychological state. The researchers of the Facebook study recognize the link between online experiences and offline emotions and behavior: ‘the well-documented connection between emotions and physical well-being suggests the importance of these findings for public health. Online messages influence our experience of emotions, which may affect a variety of offline behaviors’ (Bond et al. 2012a: 298). Given the possible carryover effects of online experiences to offline behavior, it is possible that psychological stress from online-induced negative feelings translates into instances of offline physical harm to oneself or others. In a classic experimental setting, these risks are mitigated by virtue of the IRB review process, via consent, debriefing, or modification of the research protocols. None of these safeguards is considered when the research takes place within corporations.

Recommendations

The attention garnered by the publication of Facebook’s emotional contagion experiment prompted a flurry of social media comments¹⁰ and also more formal letters¹¹ to the Federal Trade Commission (FTC) urging them to investigate Facebook’s research practices. The Electronic Privacy Information Center (EPIC) filed a complaint with the FTC on the basis that ‘the company purposefully messed with people’s minds.’¹² Although there was some history of complaints filed against Facebook’s privacy practices, this is the first instance in which Facebook’s research practices were called into question on non-privacy grounds. In October of 2014, Facebook announced revised guidelines concerning research, including internal review of research by a panel of senior employees from different departments (Schroepfer, 2014). Involving members of the company may not provide the independence and objectivity necessary to advocate for users’ rights. Much more can be done to protect human subjects who partake in online research conducted by corporations.

The possibility of unethical behavior in company-sponsored online research is not restricted to Facebook. Unfortunately, it took a voluntary publication in a highly visible academic journal (*PNAS*) and a deliberate self-disclosure by OKCupid in its company blog to find out the details of these studies. There could be many more similar cases that have not surfaced or have not garnered enough public attention. There is also the risk that after public debate and policy changes companies might stop their collaborations with academics to avoid the public dissemination of results of their internal research. This would be an unfortunate consequence, as there is social value in the research that these companies conduct. Nevertheless, users’ rights ought to be protected.

These conditions create momentum to make specific changes to guarantee the ethical treatment of online users. Previous calls for action issued by the same journal that published the emotion study, and by other outlets (Fiske and Hauser, 2014; Goel, 2014; Kahn et al., 2014) have prompted a much needed revision to the Common Rule (US Department of Health of Human Services, 2015). Aside from the proposed changes to the US federal guidelines, there are other steps to protect human subjects in non-federally funded online experiments. First, online companies should jointly create, or abide by, a common *code of ethics* to protect online users who partake in company-sponsored behavioral research. Second, this code should incorporate *special safeguards for deception* and for the mitigation of negative effects from online to offline behavior. Third, an important addition to the ongoing changes in the regulatory framework is to create a *separate entity to advocate for user rights* – a user review board – to regulate online behavioral research (particularly studies that belong in the C/D category of experimentation). Each one of these recommendations is elaborated below.

The first recommendation is to develop a code of ethics to ensure subject/user protection in online research. Given the large scale of social network platforms, any systematic investigation with users should be considered research because it has the potential to directly (through publications or press releases) or indirectly (through innovations or best practices) contribute to the advancement of knowledge on user behavior. As such, internal research conducted in social networking platforms in the private sector should adhere to an ethical code of conduct. This code could be developed along the lines of the British Psychological Society's (2013) ethics guidelines for internet-mediated research, and expanded to include provisions for online companies that experiment with their own online users by employing deceptive tactics. A practically applicable code of ethics, consistent with the new realities of online user experimentation, would articulate the users' rights and the companies' responsibilities in a more explicit way.

The second recommendation is to apply specific safeguard mechanisms through forewarning and debriefing when deception is used in privately-sponsored behavioral research. Our conceptual separation between A/B studies and C/D experimentation helps to raise awareness of deceptive research tactics in C/D experimentation, where algorithmic manipulation is not evident. According to our ethical analysis, deception in research should be used cautiously and judiciously, and when it is the only mechanism to create situations of interest that can be systematically studied. When deception is involved, some participants are exposed to more than the minimal risk they would face in everyday life (Fiske and Hauser, 2014). Therefore, researchers must take every precaution to minimize any risks related to discomfort and distress, both before and after the experiment. Before the experiment, to avoid revealing the nature of the study and thus biasing the results, we propose a mini-consent pop-up window that simply alerts users about the study and allows them to

choose (or decline) participation. In addition to alerting participants *a priori* and securing consent, we recommend the implementation of online debriefing procedures *a posteriori* to mitigate any negative effects caused by the experimental manipulations.

Third, because participants and researchers are at opposite ends of the deception-use dilemma, and there is a power difference between them, an objective third party should balance the potential benefits and costs of deception. Thus, an independent user advocacy board under the purview of the Federal Trade Commission (FTC) or another relevant entity should issue guidelines regarding experimentation through social networking platforms. This is akin to the idea of consumer subject review boards (CSRB) proposed by Prof. Ryan Calo. He argues that '[t]he accelerating asymmetries between firms and consumers must be domesticated, and the tools we have today feel ill-suited. We need to look at alternatives. No stone, particular one as old and solid as research ethics, should go unturned' (Calo, 2013). In this case, our proposed variation in the form of an independent user review board (URB) should be sensitive to the different types of online research (A/B testing vs C/D experimentation) conducted by non-federally funded entities.

One possibility is to model this new entity as the existing FTC's Bureau of Consumer Protection in the USA.¹³ Similarly, the proposed URB could be tasked with collecting complaints and conducting investigations, developing rules for the conduction of research by private companies, and educating users about their rights or responsibilities. Given the geographical location of the headquarters of the major social networking companies, this initiative could start in the USA but have international influence owing to the global reach of US-based companies such as Facebook. The function of this new URB would be regulatory by issuing guidelines, rather than supervisory by approving specific research studies conducted by social networking companies. In particular, the board should examine the circumstances that allow the use of deception and the risks of carryover effects between online experiences and offline behavior, and ensure that appropriate safeguards, via forewarning and debriefing, are in place.

Conclusion

The analysis presented in this article underscores the need to differentiate instances of company-sponsored online research and develop an ethical framework and an independent user advocacy board to regulate what we call C/D experimentation. With respect to the ethical framework, there is a need to develop a code of ethics for online research that involves human subjects regardless of the nature of the funding (federal vs private) or the type of institution that conducts such research. With respect to deception, there is a need to develop explicit rules to guide when and how this approach is acceptable in online research and what are the minimum

safeguards for participants. With respect to users' rights, an independent entity (or URB) must be in charge of regulating C/D research and balancing private interests with public benefits obtained from user participation. Users who participate in company-sponsored online research deserve an equal level of protection as human subjects who participate in federally funded or academic research.

Acknowledgements

An earlier version of this article, entitled 'The ethics of online experimentation with unsuspecting users' won the *2014 Abraham J Briloff Prize in Ethics* in the Faculty Category at Baruch College, on 26 March 2015. The current version has benefited from additional research, as well as feedback from various colleagues.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Notes

1. Facebook Help Center (n.d.) How News Feed works. Available at: <https://www.facebook.com/help/327131014036297/> (accessed 5 July 2015).
2. 'The cost of state records varied from \$0 to \$1500 per state ... The resulting list of states included Arkansas, California, Connecticut, Florida, Kansas, Kentucky, Missouri, Nevada, New Jersey, New York, Oklahoma, Pennsylvania, and Rhode Island. These states account for about 40% of all registered voters in the USA, and their records yielded 6,338,882 matched observations of voters and abstainers that we could use to compare to treatment categories from the experiment.' Massive turnout supplement (Bond et al., 2012b: 3–4).
3. This discussion is focused on unethical practices of online experimentation with social networking users, excluding instances of potential privacy breaches. Cases of privacy breaches in online experimentation deserve separate treatment, and as such they are outside the scope of this article.
4. Facebook User Agreement dated September 2011 is posted by the Coudrain Group at <http://thecoudrain.com/files/documents/Facebook-Data-Use-Policy.pdf> (accessed 5 July 2015).
5. The project is called 'Terms of Service; Didn't Read,' and it features a categorization of TOS based on user ratings: <http://tosdr.org> (accessed 5 July 2015).
6. <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102> (accessed 1 March 2016).
7. The focus of this article is on behavioral research. Biomedical research conducted by pharmaceutical companies and other contract research organizations must also comply with the Common Rule. These companies conduct research to file applications seeking approval of new drugs or medical devices from the US Food and Drug Administration

(FDA). As such, the research procedures should have been previously reviewed and approved to ensure compliance with the protection of human subjects.

8. The nature of this analysis is slightly different for the Political Turnout experiment because it does not involve the manipulation of individual emotions.
9. Quoted in <https://www.sciencenews.org/blog/scicurious/main-result-facebook-emotion-study-less-trust-facebook> (accessed 5 July 2015).
10. See hashtag #facebookemotionstudy results at: <https://twitter.com/hashtag/facebookemotionstudy?src=hash> (accessed 5 July 2015).
11. <http://james.grimmelmann.net/files/legal/facebook/FTC.pdf> (accessed 5 July 2015).
12. See the full text of the complaint at <https://epic.org/privacy/ftc/facebook/Facebook-Study-Complaint.pdf> (accessed 5 July 2015).
13. <https://www.ftc.gov/about-ftc/bureaus-offices/bureau-consumer-protection> (accessed 7 July 2016).

References

- Bakos Y, Marotta-Wurgler F and Trossen DR (2014) Does anyone read the fine print? Testing a law and economics approach to standard form contracts. *Journal of Legal Studies* 43(1): 1–35.
- Belmont Report (1979) *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Available at: <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html> (accessed 5 July 2015).
- Bond RM, Fariss CJ, Jones JJ, et al. (2012a) A 61-million-person experiment in social influence and political mobilization. *Nature* (489): 295–298.
- Bond RM, Fariss CJ, Jones JJ, et al. (2012b) *Supplementary information for: A 61-million-person experiment in social influence and political mobilization*. Available at: http://fowler.ucsd.edu/massive_turnout_supplement.pdf (accessed 5 July 2015).
- British Psychological Society (2013) *Ethics guidelines for internet-mediated research*. Available at: <http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf> (accessed 5 July 2016).
- Bulmer M (2001) The ethics of social research. In: Gilbert N (ed.) *Researching Social Life*. London: SAGE: 45–57.
- Calo R (2013) Consumer subject review boards: A thought experiment. *Stanford Law Review Online*, 3 September 2013. Available at: <http://www.stanfordlawreview.org/online/privacy-and-big-data/consumer-subject-reviewboards> (accessed 5 July 2015).
- Chatterjee S, Sarker S and Fuller M (2009) Ethical information systems development: A Baumanian postmodernist perspective. *Journal of the Association for Information Systems* 10(11): 787–815.
- Ekman P (1988) Self-deception and detection of misinformation. In: Lockard JS and Paulhus DL (eds) *Self-Deception: An Adaptive Mechanism*. Englewood Cliffs, NJ: Prentice Hall.
- Fiesler C and Bruckman A (2014) Copyright terms in online creative communities. *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 26 Apr–01 May 2014, Toronto, ON, Canada.
- Fiske ST and Hauser RM (2014) Protecting human research participants in the age of big data. *PNAS – Proceedings of the National Academy of Sciences* 111(38): 13675–13676.
- Goel V (2014) As data overflows online, researchers grapple with ethics. *New York Times*, 12 August. Available at: <http://www.nytimes.com/2014/08/13/technology/the-boon-of-online-data-puts-socialscience-in-a-quandary.html> (accessed 5 July 2015).

- Hertwig R and Ortmann A (2008) Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior* 18(1): 59–92.
- Hill K (2014) Facebook added ‘research’ to user agreement 4 months after emotion manipulation study. *Forbes*, 30 June 2014. Available at: <http://www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study> (accessed 5 July 2015).
- Johnson PE, Grazioli S, Jamal K and Berryman RG (2001) Detecting deception: Adversarial problem solving in a low base-rate world. *Cognitive Science* 25(3): 355–392.
- Kahn JP, Vayena E and Mastroianni AC (2014) Opinion: Learning as we go: Lessons from the publication of Facebook’s social-computing research. *PNAS – Proceedings of the National Academy of Sciences* 111(38): 13677–13679.
- Kimmel AJ, Smith NC and Klein JG (2011) Ethical decision making and research deception in the behavioral sciences: An application of social contract theory. *Ethics and Behavior* 21(3): 222–251.
- Kramer AD, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *PNAS – Proceedings of the National Academy of Sciences* 111(24): 8788–8790.
- Luger E, Moran S and Rodden T (2013) Consent for all: Revealing the hidden complexity of terms and conditions. *CHI ‘13 – Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 27 April–2 May 2013, Paris, France.
- Manjoo F (2014) A bright side to Facebook’s experiments with its users. *The New York Times*, 2 July 2014. Available at: http://www.nytimes.com/2014/07/03/technology/personaltech/the-bright-side-of-facebooks-social-experiments-on-users.html?_r=0 (accessed 5 July 2015).
- Meyer MN and Chabris CF (2015) Please, corporations, experiment on us. *New York Times*, 19 June 2015. Available at: <http://www.nytimes.com/2015/06/21/opinion/sunday/please-corporations-experiment-on-us.html> (accessed 5 July 2015).
- Mingers J and Walsham G (2010) Toward ethical information systems: The contribution of discourse ethics. *MIS Quarterly* 34(4): 833–854.
- Morin R (2014) Facebook’s experiment causes a lot of fuss for little result. *Pew Research Center*, 2 July. Available at: <http://www.pewresearch.org/fact-tank/2014/07/02/facebook-experiment-is-just-the-latest-to-manipulate-you-in-the-name-of-research> (accessed 5 July 2015).
- Rudder C (2014) We experiment on human beings. *OKCupid Company Blog*, 28 July 2014. Available at: <http://blog.okcupid.com/index.php/we-experiment-on-human-beings> (accessed 5 July 2015).
- Schroepfer M (2014) Research at Facebook. *Facebook Newsroom*, 12 October 2014. Available at: <http://newsroom.fb.com/news/2014/10/research-at-facebook/> (accessed 7 July 2016).
- Suddath C (2014) The science behind the perfect dating profile. *Bloomberg BusinessWeek*, 4 September 2014. Available at: <http://www.businessweek.com/articles/2014-09-04/mining-okcupids-data-reveals-how-we-date-now> (accessed 5 July 2015).
- US Department of Health and Human Services (1991) *Federal rules for the protection of human subjects (common rule)*. Available at: <http://www.hhs.gov/ohrp/humansubjects/commonrule/> (accessed 5 July 2015).
- US Department of Health and Human Services (2015) *Notice of Proposed Rulemaking – NPRM 2015 Summary*. Available at: <http://www.hhs.gov/ohrp/humansubjects/regulations/nprm2015summary.html> (accessed 1 March 2016).
- Wood M (2014) OKCupid plays with love in user experiments. *New York Times*, 28 July. Available at: <http://www.nytimes.com/2014/07/29/technology/okcupid-publishes-findings-of-user-experiments.html> (accessed 5 July 2015).