EDITOR'S COMMENTS

Opportunities and Challenges for Different Types of Online Experiments

By: **Elena Karahanna**
University of Georgia
ekarah@uga.edu

Izak Benbasat
University of British Columbia
izak.benbasat@sauder.ubc.ca

Ravi Bapna
University of Minnesota
rbapna@umn.edu

Arun Rai
Editor-in-Chief, *MIS Quarterly*
Georgia State University
arunrai@gsu.edu

Motivation for and Objectives of the Editorial

Experimental research has been an important research method in the Information Systems (IS) discipline. Recently, we have seen an expansion in the types of experiments conducted beyond traditional laboratory (lab) and field experiments. These new types of experiments, which leverage the online environment, provide new opportunities as well as new challenges for IS researchers.

This diversity also creates the need for authors and reviewers to understand the respective strengths and limitations of various types of experimental research, and not mechanically apply the lens of their favorite type of experiment.

The purpose of this editorial is to highlight the reasons that have propelled new types of experiments, categorize these along a set of dimensions, discuss their strengths and weaknesses, and highlight some new issues that emerge with these new opportunities for research. Our objective is not to be exhaustive in terms of the various types of experiments but to highlight opportunities and challenges that emerge for online variants that are more prominently seen in IS research. We, therefore, constrain our focus to lab, field, and natural experiments and their online variants.[1]

Changing Landscape of Experiments in IS Research

Experiments have been a major research method in IS research since the origins of the field. We have recently seen a stronger interest in experiments, especially those occurring online. This can be attributed to the Internet providing two sets of opportunities: (1) a field setting for experimentation as a prominent locus of economic transactions and social interactions (for field and natural experiments), and (b) opportunities to recruit larger subject pools more efficiently and reach more diverse samples with reduced administrative and financial costs (for lab experiments) (Hergueux and Jacquemet 2015).

Online transactions and interactions have created both the need and the opportunity for *online field experiments* to understand the various types of social and economic activities in which people engage online. The availability of persistent trace data for these online transactions and inter-

[1]Harrison and List (2004) use six criteria to define the "field" context of an experiment: "the nature of the subject pool, the nature of the information that the subjects bring to the task, the nature of the commodity, the nature of the task or trading rules applied, the nature of the stakes, and the nature of the environment that the subject operates in" (p. 1012). Based on these characteristics, they classify experiments into four categories: a traditional lab experiment, an artefactual field experiment (lab experiment but with subjects that are representative of the population), a framed field experiment, and a natural field experiment. Their first two categories correspond to lab experiments, whereas the third and fourth categories correspond to field and natural experiments, respectively.

actions also facilitates leveraging *natural experiments* for research by providing the ability to capture before and after measures of behavior for a treatment that occurs naturally.

Further, online labor markets (e.g., Amazon Turks) and the ability to deliver treatments remotely via the Internet facilitate online *lab experiments* by expanding the pool of subjects available beyond local convenience samples to more representative samples to increase external validity, if done correctly. In the same vein, IS researchers have also adopted the lab-in-the-field approach by conducting lab experiments that leverage the subjects' real-world settings for the context to increase external validity and realism.

Criteria for Evaluation of Experiments

Research methods are typically assessed on three criteria (McGrath et al. 1982):

- precision in control and measurement of variables related to the behavior of interest (i.e., internal validity)
- generalizability (i.e., external validity)
- realism of the context in which behaviors are observed

McGrath et al. call these three criteria the three-horn dilemma because any given research method cannot be strong on all three. "The very choices and operations by which one can seek to maximize any one of these will reduce the other two; and the choices that would 'optimize' on any two will minimize the third" (p. 74).

Lab experiments are associated with precision in control since experimental and control groups, random assignment to conditions, and control over extraneous factors enable causal inferences. However, they have traditionally been considered as low on realism and generalizability. Field and natural experiments maximize realism of context, are moderate with respect to "precision with regard to measurement, manipulation, and control of behavior variables" but trade off "generalizability with regard to populations" (*ibid.*, p. 75).

However, as we alluded to, the move to online experiments may lead to different strengths and tradeoffs for the different types of experiments. Further, new threats to validity may arise that need to be mitigated. Below we discuss these issues for different types of experiments (see Table 1).

Types of IS Experiments and Their Online Variants

Lab Experiments

A lab experiment is an experimental study conducted in a setting designed or created by the researcher for the purpose of answering or investigating particular research questions (for an extended discussion of this topic as related to IS studies, see Benbasat 1990). The first reaction to the mention of a lab experiment is typically its lack of realism and its low external validity (generalizability). While this is mostly correct, there are many advantages unique to lab experiments that have made them a viable research method.

From a methodological perspective, the major benefit of a lab experiment is its very high *internal validity* (demonstration of causality). This is because in a lab setting the researcher has the ability to create (multiple) treatment and control conditions and effect *random* assignment of participants to these conditions (balancing out potential nuisance variables that might affect outcomes). The researcher first administers the treatment, and then measures the consequent outcome variables, which generates the time precedence relationship between cause and effect needed to show causality. The researcher can also define the stimulus or treatment in a precise fashion as specified by the constructs in the theory being tested (which is the main reason for conducting most lab experiments) effecting high construct validity. The researcher's ability to create clearly defined treatments is much stronger in a lab experiment as compared to a field experiment (see Zmud et al. 1990). Another main advantage is that in the lab, the participants being removed from real-world settings, the nuisance variables that might influence the relationship between cause and effect, over and above those of the treatments, are controlled for. Furthermore, well designed lab experiments can be more easily replicated compared to studies conducted by other methods. Such replications, combined with systematic changes in the manipulations of some treatment or control variables, and in the characteristics of the participant populations and tasks (which enhance generalizability or external validity to some extent), can be testbeds for programmatic and cumulative experimental studies, allowing the researcher to study a particular research topic exhaustively.

The artificiality of the lab setting combined with the researchers' ingenuity can afford the study of IT artifacts that do not currently exist in the real world. The best examples of these are the so-called "Wizard of Oz" (WOZ) experiments (wikipedia.org/wiki/Wizard_of_Oz_experiment), first conducted in the field of human–computer interaction where the participants believed that they were interacting with a natural

Table 1. Issues and Best Practices for Online Experiments			
	Strengths and Limitations of Traditional Versions	**Advantages of Online Variants**	**Issues and Best Practice for Online Variants**
Lab Experiments	• High internal validity (control) • Low realism • Low generalizability • Ease of replicability	• Ability to recruit a broader range of subjects increases external validity • Ability to accommodate large numbers of subjects who can participate in an online lab experiment simultaneously removes time and space constraints and increases efficiency of conducting the studies • Ability to leverage participants' own settings for lab-in-the field experiments to increase realism	• Loss of control over subjects and environment can threaten internal validity. • Loss of ability to collect verbal protocols, but can leverage online observational trace data to rule out alternative causal mechanisms. • Use diversity in subject pool to detect heterogeneous treatment effects. • Use participants' settings in lab-in-field experiments to evaluate how differences in field settings influence treatment effects.
Randomized Field Experiments	• High internal validity (but less control than lab experiments) • High Realism • Generalizability depends on how representative the field setting is	• Availability of large samples can enable identification of heterogeneous treatment effects • Ability to leveraging trace data that captures users' online behaviors can help identify underlying causal mechanisms that are based on behaviors	• Ensure that treatment and control groups are equivalent by collecting pre-treatment data for treatment and control groups (A/A testing). • Potential selection bias in cases where subjects randomly assigned to a treatment can opt out of the treatment, which can threaten causal inference on downstream variables (other than the immediate outcome), needs to be statistically addressed. • Combine machine learning with causal inference methods to adjust for differences between treated and control groups in high-dimensional settings and detect and estimate heterogeneous treatment effects. • Underlying causal mechanisms involving perceptions or beliefs may be difficult to tap into; computational approaches applied to trace data can be used to estimate perceptions and beliefs in some cases. Supplemental lab experiments can also be used to tap into these. • Obtaining informed consent can be challenging; even when this is the case, researchers should assess risks and benefits and debrief participants after the study.
Natural Experiments	• Causal inference can be tenuous • High realism • Generalizability depends on how representative the field setting is	• Persistent online trace data enable researchers to leverage, *a posteriori*, changes in policies or design for causal inference	• Lack of randomization creates challenges in making causal inferences. • Collection of pre-treatment data is critical to making causal inference. • Matching with subjects that did not receive the treatment can be used to estimate the treatment effect; providing evidence of the quality of such matching is a key consideration.

*Unless otherwise stated, online variants share the same advantages and limitations as their traditional counterparts language understanding interface, or software, when such systems did not yet exist in practice, whereas in reality they were being responded to by a person "behind the curtain." One could imagine today such WOZ experiments taking place to gauge people's trust in artificial intelligence (AI)-based advice that is "simulated" in a lab setting by using human experts instead of AI software that might not currently exist. Finally, being in an artificial setting, the researcher has "license" for certain types of interventions that are not usually allowed in real settings, such as the ability to collect verbal protocols, asking questions prior to the experimental treatment to gauge initial conditions in outcome variables (pre/post designs) or assigning participants to treatments based on personality variables, and collecting data about perceptions and beliefs to test particular theories which are difficult to do in natural experiments. On the other hand, artificiality creates the challenge of making sure that participants in the lab experiment are motivated to work on the experimental task as though it is really important to them. This can be accomplished by developing task contexts that are of high interest to the participants (such as a challenging management game) and by providing monetary and other rewards to incentivize them to work conscientiously on the given experimental tasks.

In the last two decades, with the Internet being accessible to most people, lab experiments in IS have been moving out of labs, which are usually located in the confines of a university setting, to so-called online experiments. There are several companies that have access to a pool of subjects whom researchers can utilize for their experiments for a fixed fee per person, or more recently the MTurk platform, which has provided access to a much larger pool of potential participants who are willing to participate in online experiments for a much lower fee. The online experiment can be administered through the use of a particular web site that allows the participants who volunteer to be part of a study. The obvious advantage of online experiments is access to very large numbers of participants who can participate simultaneously and, as a consequence, completion of studies in a much shorter time span. This also allows for replications to be conducted more efficiently and frequently. Companies that provide access to participants will also stratify their pool according to various demographic characteristics, and sometimes according to certain skills and knowledge, thus allowing for higher external validity based on diverse populations alleviating the "student subjects" complaints for which lab experiments have been criticized.

In addition to recruiting subjects and delivering treatments over the Internet, IS researchers can leverage the online context to conduct "lab-in-the-field" experiments. These experiments leverage the subjects' real world online context as the setting for the experiment. For example, Bapna et al. (2017) designed a Facebook application (app) and recruited Facebook users for their treatment. The experiment leveraged these users' existing Facebook social networks as the setting rather than creating artificial social networks in the lab. The advantage of this approach is an increase in realism of the context.

However, online experiments still have lower realism relative to a natural setting since the participants are aware of being part of a study. More importantly, once the participants are "out of the lab" the researcher has lost the very important control afforded by the lab. As a consequence, the researcher does not know what is taking place within the context in which the participant is working on the experiment's task (e.g., there was evidence of collusion in Bapna et al. 2017), the environmental conditions that might affect the results, whether or not the subjects are multitasking or starting/stopping while working on the task, and who else might be assisting them in the completion of the task (for a discussion of methods to mitigate some of these problems, see Reips 2002). Therefore, internal validity, the ability to make causal inferences, which is the hallmark of a lab study, has been weakened somewhat. In addition, it is not uncommon to find or infer that a significant proportion of the participants in online experiments are not conscientiously filling out post-experimental questionnaires to measure key dependent variables.

In conclusion, the lab experiment is the method that provides the highest degree of internal validity via the controls it affords to the researcher. Without internal validity, the importance of realism and generalizability, which might exist, become secondary, when the epistemology is to discover causal mechanisms underlying a phenomenon. There is a second type of control in lab studies: the ability of the researcher to study the questions he/she wants to study (a top- down approach). This is best afforded by a lab experiment since the researcher can control and "create" conditions in the lab to study particular questions, but admittedly lab settings are not good for creating complex environments that exist over a long period of time (although lab-in-the-field experiments can provide this option, albeit with the loss of some control). On the other hand, they enable one to create environments and conditions that do not currently exist in the field and, thus, to create insights into the viability and pitfalls of forward-looking technology designs.

Randomized Field Experiments

A randomized field experiment is one where subjects are randomly assigned to treatments and the treatments are delivered in the subjects' naturally occurring environment. The field experiment methodology typically involves deep collaborations of researchers and partner companies, and championing by the senior leadership of collaborating companies. Extensive discussions have to be undertaken to identify mutually beneficial manipulations of interest, to identify the relevant subject pool, and resources have to be allocated to deploy the treatment, and to collect the pre-, during, and post-treatment data in a de-identified manner.

The field setting provides field experiments with a high level of realism that, in some cases, is not achievable via lab experiments. Randomized field experiments also have high internal validity; nonetheless, the validity is not to the extent of lab experiments where experimental conditions can be controlled with more precision and certainty. The generalizability of the findings will depend on the representativeness of the field setting and of the subjects, implying that researchers need to make the case as to why and how their results generalize beyond the study's specific setting. Therefore, depending on the setting, one can conceivably achieve reasonably high levels on all three dimensions of the McGrath et al. criteria of investigational design.

The emergence of wide-scale digital transformation of society and business has seen a concomitant growth in IS researchers deploying, often large-scale, *online* randomized field experiments in partnership with companies. The availability of trace data and the large samples offer distinct advantages for online experiments. First, the extensive availability of micro-level trace data on digital platforms is often used to dive deeper into the underlying mechanisms leading to the treatment effect. This trace data captures users' online behaviors including their usage of system features and their online social and business interactions. Therefore, trace data is well-suited for capturing underlying mechanisms that are based on users' online behaviors. However, trace data does not typically lend itself to tapping into underlying mechanisms that may involve perceptions and beliefs, although these can sometimes be inferred based on trace data. As such, additional lab experiments, offline or on platforms such as Amazon Turks, are sometimes carried out to identify such mechanisms. Second, the large samples that are potentially available in large-scale randomized field experiments hold promise for a new branch of analytics that combines machine learning based predictive modeling with causal inference from experiments, especially as it pertains to going beyond the average treatment effect to heterogeneous, conditional on users' characteristics, treatment effects.

Researchers engaging in online field experiments should be sensitive to three issues. First, the practice of A/A testing, the idea of having a pre-experimental period where the treatment groups and control groups are left assigned but unmanipulated and observed, is critically important. This is to make sure that the design and implementation are valid and that there are no preexisting differences between treatment and control that can invalidate causal inference. Second, while in a lab experiment all the subjects assigned to a treatment are ensured to receive the treatment, in certain types of field experiments, subjects randomly assigned to a treatment may choose not to receive the treatment, making this an additional consideration for causal inference on downstream variables.[2] Third, as we discuss later, issues of informed consent and assessment of risk and benefits need to be evaluated and addressed.

Randomized online field experiments have been deployed to study a variety of research problems in different research domains, such as social influence and engagement in online social networks (e.g., Bapna and Umyarov 2015), social proof for online word of mouth (Burtch et al. 2017), content contribution on social media prediction markets (Qiu and Kumar 2017), behavioral impact of wage cuts in online labor markets (Chen and Horton 2016), the effect of privacy enhancing features in different types of crowdfunding platforms (Burtch et al. 2015), and the complementarity of signals in equity crowdfunding (Bapna 2017), to name just a few.

Natural Experiments

A natural experiment is a special type of field experiment in which subjects are exposed to experimental and control conditions that are determined by nature or by factors outside the control of the researchers. These can be a change in government or organizational policy, a change in site design and features, or when there is exogenous variation in policy/laws or in the introduction of technology across geographical regions, that then "instrument" a change in the phenomenon of interest.

Natural experiments are similar to randomized field experiments in terms of realism and generalizability. However, because natural experiments do not always entail random assignment of subjects to treatment and control groups, causal inference can be tenuous. Random assignment by experimenters ensures that membership in the treatment and control conditions is exogenous and independent of the subjects' own decisions and motivations. As such, the treatment and control groups are assumed to be probabilistically equivalent prior to the intervention and any observed differences between the two can be causally attributed to the treatment (Cook and Campbell 1979; Murnane and Willett 2011). In the absence of random assignment, selection into treatment and control groups may be endogenous and the groups may differ systematically pre-treatment. These differences (and their interactions with history, maturation, and instrumentation) pose an alternative explanation for the outcomes (for a discussion, see Cook and Campbell 1979).

Therefore, for natural experiments, researchers have to convincingly argue for the pre-treatment equivalence of experimental and control groups and provide convincing empirical evidence for the fact (Murnane and Willett 2011). Collecting pre-treatment observational data for both treatment and control groups and knowledge of the assignment process so that any confounding due to selective exposure can be addressed become critical steps in assessing pre-treatment equivalence. Furthermore, using observational data *post hoc* to construct control groups that match experimental groups (e.g., using propensity score matching, Mahalanobis distance matching, Euclidean distance matching, or coarsened exact matching) is one common strategy that is used to create "equivalent" groups. Given that such matching methods may have their limita-

[2]For example, in an online randomized field experiment, Sun et al. (2017) provided two incentives to induce adoption of a mobile app. Some subjects in each treatment adopted the app. Given that the subjects who adopted in each treatment could be systematically different than those who did not, to causally examine the downstream effect of induced adoption on purchases, the authors used the local average treatment effect (LATE) method, where they used random assignment to a test group as an instrument for app adoptions.

tions, a careful assessment of alternative techniques is required. Furthermore, analysis methods such as difference-in-differences (DiD), interrupted time series, regression discontinuity, falsification tests, instrumental variables, and synthetic controls are frequently used to strengthen causal inference (for a discussion of the relative merits of different approaches, see Craig et al. 2017).

There are several examples of natural experiments in the IS literature. For example, some studies leveraged differences in the timing of introduction of specific technologies or platforms across geographical regions to assess their impact (e.g., the impact of the introduction of Craigslist on HIV transmission [Chan and Ghose 2014], the impact of the introduction of ride sharing services on alcohol-related fatalities [Greenwood and Wattal 2017]) or differences in privacy laws across states to examine the effect of IT integration on health outcomes (Miller and Tucker 2011).

Online natural experiments typically involve leveraging observational trace data where researchers capitalize on feature or policy changes by the online platform or policy changes by states or governments that affect users on the platform differentially. The persistence of trace data that enables capturing behaviors before and after the natural experiment makes such studies possible. For example, Zhang and Wang (2012) leveraged panel data collected from the Chinese language version of Wikipedia and a policy change that blocked it in mainland China but not in other regions of the world to examine the effect of an editor's position in a collaboration network on his or her contribution behavior. Shun et al. (2014) leveraged a policy change by eBay that removed sellers' ability to retaliate against buyers to examine the effect of reputation systems on sellers' behavior, and Huang et al. (2017) leveraged the integration of Yelp.com and Tripadvisor.com with Facebook to examine the effect of social network integration on user content generation.

Informed Consent and Risk-Benefit Assessment for Online Field Experiments

Online field experiments provide new opportunities to create both scientific value in terms of knowledge accretion as well as practical value in providing a better user experience and user outcomes, in improving organizational profitability, and in improving allocation of capital and labor. With new opportunities, new challenges often arise.

A series of revelations from Facebook and OkCupid that they experimented with their users[3] sparked a debate around the ethics of experimenting with millions of users and bystanders without their consent (see Grimmelmann 2015; Puschmann and Bozdag 2014).

Informed consent as well as assessment of risk and benefits are important considerations in academic research involving human subjects (Belmont Report 1979).

Informed consent "generally requires that subjects be given sufficient information about the research, that they comprehend the information they are given, and that their agreement to participate be free of undue influence" (Grimmelmann 2015, p. 226). A number of these large-scale online field experiments involve partnerships of academics with companies. While companies incorporate clauses in their user agreement that gives them permission to experiment with users for quality improvement (Puschmann and Bozdag 2014), a counterargument is that such user agreements, from an *academic research* perspective, are too broad to meet the information, comprehension, and voluntariness criteria for informed consent set forth in the Belmont Report. Specifically, they (1) do not inform participants of the risks involved in the research; (2) do not provide sufficient information on any specific research and do not ascertain users' comprehension that would make users' consent "informed"; and (3) do not give users the opportunity to opt-out of a specific experiment (Grimmelmann 2015).

Informed consent can be waived or altered based on three criteria (Belmont Report, p. 7; Grimmelmann 2015):

- *the minimal risk criterion*, that is, there are no undisclosed risks to subjects that are more than minimal
- *the impracticability criterion*, that is, incomplete disclosure is truly necessary to accomplish the goals of the research (e.g., if getting informed consent may impair the outcome of the study in that users' responses may change if they know they are participating in research or that informed consent may lead to biased samples [e.g., Bernstein 2014])
- *the debriefing criterion*, that is, there is an adequate plan for debriefing subjects, when appropriate, and for dissemination of results to them

While the large scale of online field experiments may impose challenges on opting-in, making the case that informed consent is "impractical" is not sufficient reason to waive or alter informed consent. The option to waive or alter informed consent for a study hinges on the study meeting the minimal risk criterion (which many of these studies do). Furthermore, although "opt-in procedures may be unrealistic for online experiments because of their ubiquity" (Puschmann and Bozdag 2014, p. 5), debriefing by retroactively informing subjects of the experiment

[3]For example, in its emotional contagion experiments, Facebook manipulated whether users saw positive or negative postings from their connections to examine the effect on the valence of the users' subsequent postings. OkCupid, a dating website, reversed compatibility scores (i.e., people who were compatible were shown as incompatible and vice versa) to examine the effects of compatibility scores on messaging behavior.

and its results is a viable and desirable alternative (e.g., Desposato 2014; Grimmelmann 2015).[4]

Given that field experiments have an impact on outcomes in the real world, a careful assessment of risks and benefits both to the subjects and on a broader scale is required (see Zmud et al. 1990). Treatments which are low-risk and normatively good (e.g., aimed at reducing secondary hospital infections, increasing literacy) would not be met with objections (Desposato 2014). However, treatments may involve risks to participants, may provide benefits to some at the expense of others (especially when the outcomes are zero-sum games), or may have consequences on broader outcomes beyond those of interest in the research.[5] Furthermore, interventions may spread to bystanders (e.g., others on online social networks or offline contacts) beyond those involved in the research (Desposato 2014). All these (risks and benefits to participants, bystanders, and broader outcomes) need to be carefully assessed to meet the standard "of systematic, nonarbitrary analysis of risks and benefits" stipulated in the Belmont Report and to adhere to its ethical principle of beneficence.

There have been several suggestions by scholars on how to minimize risks for large-scale experiments including proportionally scaling the sample size rather than running a large-scale experiment at the start of a study and using the minimum number of subjects required to meet the research objectives (e.g., hypothesis testing, assessment of heterogeneous treatment effects) to minimize the number of users impacted (Desposato 2014; Grimmelmann 2015). The approach taken to mitigate risk and adhere to informed consent standards will, of course, depend on the specifics of a study.

Finally, researchers may find it attractive to engage in large-scale online randomized experiments in those regions of the world where it is easier to partner with companies and/or where there is lower scrutiny by institutional research boards. While engaging in such experiments provides opportunities to advance knowledge, it is also important that standards for informed consent and assessment of risks and benefits are carefully considered.

Concluding Remarks

The diversity in the type of experiments provides IS scholars with the opportunity to make meaningful choices given their research objectives. These choices require consideration of precision in control and measurement, external generalizability, and realism. Whereas in traditional lab experiments the subjects, stimulus, and response are all within controlled lab environments, their online variants have some or all of these provided online, raising issues in terms of control.

The benefits of online experiments include (1) increasing external validity by recruiting larger and more diverse samples and increasing efficiency, (2) leveraging online trace data to understand mechanisms for treatment effects; (3) leveraging large sample sizes to uncover heterogeneous treatment effects; and (4) being able to capitalize on naturally occurring events (e.g., site design changes or site policy changes) *a posteriori* because of the availability of trace data.

To address challenges of control for causal inference, advances in the collection and analysis of fine-grained behavioral trace data during the pre- and post-treatment phases provide opportunities to assess equivalence of groups, match treatment and control groups, and evaluate the precision in matching. Further, given that online field experiments can have substantial economic and societal impacts on subjects and others at broad scale, researchers need to carefully consider issues pertaining to informed consent and assessment of risks and benefits.

Finally, while all types of experiments have commonalities, it is important to understand the strengths and limitations inherent in a specific type of experiment and assess its merits using criteria appropriate to its variant.

References

Bapna, R., Gupta, A., Rice, S., and Sundararajan. R. 2017. "Trust and the Strength of Ties in Online Social Networks: An Exploratory Field Experiment," *MIS Quarterly* (41:1), pp. 115-130.

Bapna, R., and Umyarov, A. 2015. "Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network," *Management Science* (61:8), pp. 1902-1920.

---

[4]For example, Jeffrey Hancock, one of the co-authors of the Facebook emotional contagion study, responding to criticism about their study, argued that while opt-in may be unrealistic, he was in favor of a notify-after approach. Specifically, he "argued in favour of retroactively informing users after an experiment has taken place, including more information about the study, and contact information for the researchers or an ombudsman" (LaFrance 2014, as cited in Puschmann and Bozdag 2014, p. 5).

[5]For example, manipulating political information to see how far it diffuses during an election may impact actual election outcomes. For instance, results of the "I Voted" experiment on Facebook increased turnout by about 340,000 votes (Bond et al. 2012).

Bapna, S. 2017. "Complementarity of Signals in Early Stage Equity Investment Decisions: Evidence from a Randomized Field Experiment," *Management Science* (forthcoming).

Belmont Report. 1979. "Ethical Principles and Guidelines for the Protection of Human Subjects of Research," National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, U.S. Department of Health & Human Services (available at http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html).

Benbasat, I. 1990. "Laboratory Experiments in Information System Studies with a Focus on Individuals: A Critical Appraisal," in *The Information Systems Research Challenge: Experimental Research Methods*, I. Benbasat (ed.), Boston: Harvard Business School Press, pp. 33-47.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature* (489:7415), pp. 295-298.

Burtch, G., Ghose, A., and Wattal, S. 2015. "The Hidden Cost of Accommodating Crowdfunder Privacy Preferences: A Randomized Field Experiment," *Management Science* (61:5), pp. 949-962.

Burtch, G., Hong, Y., Bapna, R., and Griskevicius, V. 2017. "Stimulating Online Reviews by Combining Financial Incentives and Social Norms," *Management Science* (64:5), pp. 2065-2082.

Chan, J., and Ghose, A. 2014. "Internet's Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission," *MIS Quarterly* (38:4), pp. 955-976.

Chen, D. L., and Horton. J. J. 2016. "Are Online Labor Markets Spot Markets for Tasks? A Field Experiment on the Behavioral Response to Wage Cuts," *Information Systems Research* (27:2), pp. 403-423.

Cook T. D., and Campbell D. T. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Boston: Houghton Mifflin Company.

Craig, P., Katikireddi, S. V., Leyland, A., and Popham, F. 2017 "Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research," *Annual Review of Public Health* (38) pp. 39-56.

Desposato, S. 2014. "Ethical Challenges and Some Solutions for Field Experiments," unpublished paper (http://desposato.org/ethicsfieldexperiments.pdf; accessed September 5, 2018).

Greenwood, B. N., and Wattal, S. 2017. "Show Me the Way to Go Home: An Empirical Investigation of Ride Sharing and Alcohol Related Motor Vehicle Homicide," *MIS Quarterly* (41:1), pp. 163-187.

Grimmelmann, J. 2015. "The Law and Ethics of Experiments on Social Media Users," *Colorado Technology Law Journal* (13), pp. 221-271.

Harrison, G. W., and List, J. A. 2014. "Field Experiments," *Journal of Economic Literature* (43:4), pp. 1009-1055.

Hergueux, J., and Jacquemet, N. 2015. "Social Preferences in the Online Laboratory: A Randomized Experiment," *Experimental Economics* (18:2), pp. 251-283.

Huang, N., Hong, Y., and Burtch, G. 2017. "Social Network Integration and User Content Generation: Evidence from Natural Experiments," *MIS Quarterly* (41:4), pp. 1035-1058.

LaFrance, A. 2014. "How Much Should You Know About the Way Facebook Works?," *The Atlantic*, August 20 (http://www.theatlantic.com/technology/archive/2014/08/how-much-should-you-know-about-how-facebook-works/378812).

McGrath, J. E, Martin, J, and Kulka R. A. 1982. *Judgment Calls in Research: An Unorthodox View of the Research Process,* Newbury Park, CA: Sage Publications.

Miller, A. R., and Tucker, C. E. 2011. "Can Health Care Information Technology Save Babies?," *Journal of Political Economy* (119:2), pp. 289-324.

Murnane, R. J., and Willett, J. B. 2011. *Methods Matter: Improving Causal Inference in Educational and Social Science Research*, New York: Oxford University Press.

Puschmann, C., and Bozdag, E. 2014. "Staking Out the Unclear Ethical Terrain of Online Social Experiments," *Internet Policy Review* (3:4), pp. 1-15.

Qiu, L., and Kumar, S. 2017. "Understanding Voluntary Knowledge Provision and Content Contribution Through a Social-Media-Based Prediction Market: A Field Experiment," *Information Systems Research* (28:3), pp, 529-546.

Reips, U. 2002. "Standards for Internet-Based Experimenting," *Experimental Psychology* (49:4), pp. 243-256.

Shun, Y., Gao, G., and Viswanathan, S. 2014. "Strategic Behavior in Online Reputation Systems: Evidence from eBay," *MIS Quarterly* (38:4), pp. 1033-1056.

Sun, T., Shi, L., Viswanathan, S., and Zheleva, E. 2017. "Motivating Effective Mobile App Adoptions: Evidence from a Large-Scale Randomized Field Experiment" (available at SSRN: http://dx.doi.org/10.2139/ssrn.2984506).

Zhang, X., and Wang, C. 2012. "Network Positions and Contributions to Online Public Goods: The Case of Chinese Wikipedia," *Journal of Management Information Systems* (29:2), pp. 11-40.

Zmud, R. W., Olson, M. H., and Hauser, R., 1990. "Field Experimentation in MIS Research," in *The Information Systems Research Challenge: Experimental Research Methods*, I. Benbasat (ed.), Boston: Harvard Business School, pp. 97-111.