# Online Field Experiments: A Selective Survey of Methods*

Yan Chen Joseph Konstan

February 9, 2015

**Abstract**

The Internet presents unprecedented opportunities to conduct field experiments. Based on online field experiments in Economics and Computer Science, we present an analysis of the design choices, with particular attention to the underlying technologies, for conducting online field experiments and report on lessons learned.

**JEL Classifications**: C93, H41
Keywords: online field experiment

*We would like to thank Eytan Bakshy, Sara Kiesler, Robert Kraut, David Reiley, Paul Resnick, John Riedl and Loren Terveen, for many helpful conversations on the topic. The financial support from the National Science Foundation through grants no. IIS-0325837 is gratefully acknowledged. Chen: School of Information, University of Michigan, 105 State Street, Ann Arbor, MI 48109-2112. Email: yanchen@umich.edu. Konstan: Department of Computer Science and Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455. Email: konstan@cs.umn.edu.

# 1 Introduction

Field experiments attempt to bring together the control and rigor of laboratory experiments with some of the ecological validity of field studies. They have a long tradition of use in medicine (Lohr, Brook, Kamberg, Goldberg, Leibowitz, Keesey, Reboussin and Newhouse 1986), economics (Harrison and List 2004), and social psychology (Lerner, Gonzalez, Small and Fischhoff 2003). One of the challenges of field experiments, however, is the substantial cost of conducting them, particularly at a sufficient scale to study high-variance social phenomena. In the realm of online communities, however, it becomes significantly more practical to conduct field experiments. Given sufficient access to a community of users and the software substrate for their community, researchers can conduct wide-ranging manipulations and study their effects, both short-term and long-term.

In this paper, we present an analysis of the design choices for online field experiments using representative studies from both Economics and Computer Science. Within Computer Science, we focus on two subfields, i.e., Human-Computer Interactions (HCI), and Computer-Supported Collaborative Work (CSCW). In what follows, we summarize current methods for conducting online field experiments, with particular emphasis on the underlying technologies, and offer some insights and advice for social scientists interested in conducting such studies.

With the recent flurry of online field experiments, it is not possible to conduct an exhaustive survey within the page limit. Instead, we choose a number of studies which use a variety of technologies and cover a broad spectrum of sites, including social networking sites (Facebook, LinkedIn), sites with user-generated content (Wikipedia, MovieLens), e-commerce (eBay, Yahoo!), online game (World of Warcraft), crowdfunding (Kiva), and crowdsourcing sites (Google Answers, TopCoder, oDesk, Taskcn). We do not include experiments conducted on Amazon's Mechanical Turk, as they have been covered in a separate survey (Horton, Rand and Zeckhauser 2011). We also note that a large number of commercial online field experiments are conducted everyday to help improve product or interface design. A vast majority of these experiments are not published. Therefore, we focus on a selective number of experiments by academic researchers.

# 2 Technologies for Intervention

In this section, we present three basic experimental technologies for intervention, from email, modified web interface to bots and add-ons. For each technology, we use a couple of experiments to demonstrate how the underlying technology can be used for intervention.

## 2.1 Email

Email is among the most common but also most blunt intervention technologies. Compared to modified web interface, email is more likely to get participant attention. In what follows, we report two case studies, each using email as its tool for intervention.

Ling et al. (2005) reports on a set of four field experiments conducted with members of the MovieLens online movie recommender community (`http://www.movielens.org`). In three of these experiments, selected users of the system received email messages asking them to rate more movies (i.e., to contribute effort and knowledge to the community). In all, over 2,400 users were sent an email invitation. These messages were crafted to test certain hypotheses based on the Collective Effort Model from social psychology (Karau and Williams 1993). Specifically, the experiments found that highlighting the member's uniqueness by pointing out that they had rated movies rarely rated by others increased rating behavior. Setting specific rating goals (either for individuals or for a group) also increased rating behavior. Surprisingly, highlighting the benefit of rating, either to the member or to others, did not increase ratings.

This experiment also demonstrated the importance of proper controls. Rating activity peaked after the mailings, but also after the later thank-you email and survey. Indeed, any reminder about the site seems to promote more visits.

The fourth experiment reported on in the paper shows the potential of somewhat longer-term studies. This $2 \times 2$ field experiment organized users into discussion groups to explore the effects of group homogeneity vs. heterogeneity (as measured by similarity in movie ratings) and of individual awareness of uniqueness. The experiment showed that uniqueness did indeed lead to greater participation (as hypothesized) but found that intra-group similarity did not. This experiment shows the potential for creating longer-lived field experiments to test the permanence of an effect.

In a recent field experiment on Kiva (`http://www.kiva.org`), the first microlending website to match lenders with entrepreneurs in developing countries, Chen, Chen, Liu and Mei (2014) run a large-scale randomized field experiment ($n = 22,233$) by posting team forum messages. The experimenters investigate two factors which might influence team activity levels, coordination through a reduction of search cost vs. competition through goal setting. Compared to the control, they find that lenders make significantly more loans when exposed to a goal-setting and coordination message, whereas goal-setting alone significantly increases lending activities of previously inactive teams. While Ling et al. (2005) collaborate with MovieLens to send their emails to users, Chen et al. (2014) use a built-in feature in Kiva that, each day, all forum messages are summarized into one email and sent to each team member's inbox. Thus, their experimental intervention is incorporated into the normal flow of emails that lenders receive.

In preparation for an online field experiment, it is often useful to download and analyze archival data from the site. Many websites have a public application programming interface (API), which enables researchers to download data the site collects about its users. For example, Chen et al. (2014) contains an empirical section where they analyze naturally occurring data on Kiva, obtained from its public API (`http://build.kiva.org/`). Through the empirical analysis, they are able to assess the role of teams on lending activities, which provides guidance for the design of their subsequent field experiment.

## 2.2 Modified Web Interface

Randomized experiments through modified web interface is often used in the technology industry, sometimes called A/B tests, to evaluate the effects of changes in user interface design. Software packages, such as PlanOut,[1] have been developed to facilitate such experimentation (Bakshy, Eckles and Bernstein 2014).

In a large-scale field experiment on Yahoo!, Reiley, Li and Lewis (2010) investigate whether the competing sponsored advertisements placed at the top of a webpage ("north" ads) exert externalities on each other. They run a field experiment by randomizing the number of north ads from zero to four for a representative sample of search queries on Yahoo! The experiment was conducted on 2% of all Yahoo! Search users, for whom 10% of searches were randomly selected. Later, a second experiment was conducted using an 18% larger sample than the first experiment. Overall, the number of observations per treatment was around $100,000$. Surprisingly, the researchers find that rival north ads impose a positive, rather than a negative externality on existing north listings. That is, the top north listing receives more clicks when additional north listings appear below it. The researchers offer several plausible behavioral explanations for this finding.

A new form of advertising is social advertising, where producers use information about consumer peers from a social network. Bakshy, Eckles, Yan and Rosenn (2012) investigate the effect of social cues on consumer responses to ads on Facebook in two large-scale experiments. In the first experiment ($n = 23,350,087$), the researchers randomize the number of social cues present in word-of-mouth advertising from one to three, and measure how responses increase as a function of the number of cues. The second experiment ($n = 5,735,040$) examines the effect of augmenting ad units with a minimal social cue about a single peer. On average, this cue causes significant increases in consumer clicks and connections with the advertised entity. Using a measurement of tie strength based on the total amount of communication between subjects and their peers, they find that these influence effects are greatest for strong ties.

[1] PlanOut is an open source software package developed by Facebook researchers. For detailed information, see `https://facebook.github.io/planout/`.

In a recent field experiment on LinkedIn, Gee (2014) varies the amount of information seen by two million job seekers when viewing 100,000 job postings (`https://www.linkedin.com/job/`). While users randomized into the treatment group see the true number of people who previously started an application, those in the control group would see no such information for any of the postings during the 16 days of the experiment. The additional information in the treatment increases the likelihood a person will start and finish an application by 2-5%. Furthermore, the treatment also changes the makeup of the applicant pool by increasing the number of women who apply, which has important implications for the advertising firms which are predominantly in the high tech and finance industry, where women are under-represented.

As a tool for intervention, modified web interface can also be used in combination with emails. For example, Chen, Harper, Konstan and Li (2010a) design a field experiment on MovieLens to explore the use of social comparison to increase contributions to an online community. Overall, 398 users were sent a personalized email newsletter, each containing social or personal information according to the experimental treatments. All of the newsletters contained the same five links: (1) rate popular movies, (2) rate rare movies, (3) invite a buddy to use MovieLens, (4) help us update the MovieLens database, and (5) just visit the MovieLens home page. To track user behavior over time, the researchers modified user interface accordingly. Participants who visited MovieLens during the month after receiving the newsletter were given a slightly modified interface with the four links from the email newsletter included in the "shortcuts" pane of the main MovieLens interface visible from each page in the system.

The authors find that, after receiving behavioral information about the median user's total number of movie ratings, users below the median demonstrate a 530% increase in the number of monthly movie ratings, while those above the median do not necessarily decrease their ratings. When given outcome information about the average user's net benefit score, above-average users mainly engage in activities that help others. These findings suggest that effective personalized social information can increase the level of public goods provision. One key benefit of this study has been the ability to follow user behavior for an extended period of time to determine how durable an effect such a prompt has.

## 2.3 Bots

Many online communities provide ways in which an experimenter may add features that can be differentially deployed to different users. One such community is Wikipedia, which allows bots to be deployed on site. Bots are programs or scripts that make automated edits or suggestions. Before deploying a bot on Wikipedia, the experimenter must get the approval of a group of designated Wikipedia users who have the technical skills and wiki-experience to oversee and make decisions

on bot activity.[2] Bots on Wikipedia are governed by the following policy, "The burden of proof is on the bot-maker to demonstrate that the bot is harmless, is useful, is not a server hog, and has been approved" by the Bot Approvals Group.

Cosley, Frankowski, Terveen and Riedl (2007) deployed an intelligent task-routing agent, called SuggestBot, in the Wikipedia community to help study how workload distribution interfaces can affect the amount of work members of a community undertake and complete. SuggestBot pre-processed a dump of Wikipedia to build a learning model of what articles a user might be interested in editing based on their past editing behavior and then recommended needed work to users through their talk pages. A six month study in which over 1,200 people received recommendations found that personalized recommendations led to nearly four times as many actual edits as random suggestions.

The SuggestBot study adds two important factors. First, it demonstrates how one might use bots for experimental interventions. One challenge to deploy bots on a third-party website is the level of detail of observation available (e.g., we could measure edits, but not reading behavior), but this is all determined by the nature of the extension interface. Second, it shows that field experiments can be used not only to address traditional social science research questions, but also to address technical design questions motivated by social science.

# 3 Design Choices

Based on existing literature, we present a set of design choices in online field experiments, including (1) the access and degree of control the experimenter exerts over the online venue, (2) issues of recruiting and informed consent, (3) identifiability and authentication of subjects, and (4) the nature of the control group. These dimensions exclude the three core design features of any experiment - the hypotheses, the experimental conditions, and the presentation of the experimental manipulation - because those vary substantially with each individual study. We also skip power analysis which can be found in multiple statistical textbooks.

## 3.1 Access and Degree of Control

When one has the flexibility to choose among several different sites to conduct the study, the degree of experimenter control is a crucial factor in the final decision.

[2]Wikipedia's bot policy can be found at `https://en.wikipedia.org/wiki/Wikipedia:Bot_policy`. The approval procedure can be found at `https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval`.

1. **Experimenter-as-user** involves minimal or no collaboration of the site owners. In most online sites, it is feasible for experimenters to establish identities as users for the purposes of both gathering data and introducing interventions. Naturally, both the types of manipulation possible and the data that can be gathered are limited by the system. Furthermore, some online communities have usage agreements or codes of conduct that prohibit such research uses. In the absence of such restrictions, however, experimenter-as-user can be not only an effective way to gather field data, but also to inject manipulations.

   Experimenter-as-seller in auctions has been used since the first economic field experiment conducted over the Internet by Lucking-Reiley (1999), where he auctioned off pairs of identical collectible Magic: the Gathering trading cards using different auction formats to test the revenue equivalence theorem. The auctions were conducted in an Internet newsgroup, which was exclusively devoted to the trading of cards and where a variety of auction mechanisms were already used by other traders before the experiment was conducted, and where there was substantial trading volume. He found that the Dutch auction produced 30-percent higher revenues than the first-price auction, which violates the theoretical prediction and what had been observed in laboratory studies.

   On eBay, Resnick, Zeckhauser, Swanson and Lockwood (2006) conducted a field experiment to study Internet reputation systems. A high-reputation, established eBay seller sold matched pairs of vintage postcards under his regular identity and under seven new seller identities (also operated by him). They were able to measure the difference in buyers' willingness-to-pay, and put a price on good reputation. Since eBay was not involved in the experiment, data were collected directly from the eBay webpage using a Web spider.

   Similarly, experimenter-as-employer has become increasingly common for experiments on crowdsourcing sites, testing theories of social preference on the now defunct Google Answers (Harper, Raban, Rafaeli and Konstan 2008, Chen, Ho and Kim 2010b), labor market sorting on TopCoder (Boudreau and Lakhani 2012), and contest theory on Taskcn (Liu, Yang, Adamic and Chen forthcoming).

   In one such experimenter-as-employer study, Pallais (forthcoming) evaluates the effects of employment and feedback on subsequent employment outcomes on oDesk (`https://www.odesk.com/`), an online labor market for freelance workers. The experimenter hired 952 randomly-selected workers for data entry jobs. Afterwards, she gave workers either detailed or coarse public evaluations. Using oDesk administrative data, she finds that both hiring workers and providing detailed evaluations substantially improved their subsequent employment rates, earnings and reservation wages. The experiment has important public policy implications for markets for inexperienced workers and reputation building.

2. **A site with a public interface** might allow substantial experimenter control. Social networking sites, such as Facebook and LinkedIn, use the OpenSocial API standards, which facilitate the integration of third party applications. Wikipedia is another example of a site that encourages such research. Cosley et al. (2007) used the site data dumps made available by Wikipedia to build a model of users (based on editing behavior) and articles to identify the articles a user might be be interested in editing, then deployed SuggestBot which recommended editing work to users through their talk pages. This use of Wikipedia is a success story, but it also illustrates the challenges of working through an open interface; there was no way to build profiles of interest based on reading behavior, and as a result, the profiles could not be generated for people without editing experience already.

   Some other systems encourage the construction of extensions and make participant data available. For example, World of Warcraft, a massively multiplayer online game, supports a variety of such data gathering through the interfaces made available to player bots (interfaces which provide statistics on currently active players). Williams et al. (2006) used this method in their study of the social life of WoW guilds.

3. **A collaborative relationship with the site owner** can also give you a fair amount of data and control. Keep in mind, however, that most site owners are primarily concerned about the user experience rather than your experiments, which imposes limits on the types of experiments you can do. Chen, Li and MacKie-Mason (2006) worked with the Internet Public Library to test the effectiveness of various fund-raising mechanisms proposed in the literature. These were implemented through a variety of interfaces through which a solicitation could be delivered (e.g., pop-up messages, pop-under messages, and in-window links). The authors find that, although the gift size is not significantly different across mechanisms, the Seed Money and Matching mechanisms each generate significantly higher user click-through response rate than the Premium mechanism. Having the collaboration of IPL staff gives us the opportunity to collect micro-behavioral data, such as user click-streams. Such collaborative relationships can be extremely effective, but tend to develop slowly as the site owner gains trust in the collaborating researcher. We find they are best thought of as a substantial investment in research infrastructure rather than as a quick target for a single study.

4. Lastly, **owning your own site** gives the experimenter the most control and flexibility in the experimental design and data collection. There are great advantages to having your own community. We created MovieLens more than a decade ago, and it has given us the ability to control and measure every aspect of the system and of user interaction with it. Many interesting experiments involve presenting different experiences to different users and watching their behavior as time progresses. We can modify the interface, implement varying inter-

faces for different experimental groups, and use usage data to assign users into experimental groups. We can also email users (except for those who have requested otherwise) to invite them into experiments, broadening our reach in experiments).

For example, in Chen et al. (2010a), we were interested in the effects of social information on contribution behavior. We chose to influence behavior through personalized email newsletters that presented varying information about social comparisons and that included directly links to different activities. Because we controlled the site, we were able to use user history data (e.g., number of movies rated, frequency of login, and other usage data) to assign subjects to groups and to personalize their newsletters. We also were able to track their activity in the month following the newletter mailing (and beyond) to determine the effect of the manipulation on their interaction with the site as a whole – not just the link selected in the newsletter. Finally, we were able to present these users with a modified web interface to the MovieLens site that would present the email newsletter links within the site itself. This level of control and observation would be difficult without direct control over the site.

## 3.2 Recruiting and Informed Consent

Online field experiments use two types of subject recruiting. The first type is natural selection. In the eBay field experiments discussed above, the experimental tasks are natural tasks that participants undertake as bidders. Whoever is interested in buying the auctioned items might end up in the experiment. In such experiments, participants do not know that they are in an experiment. These are natural field experiments (Harrison and List 2004). In nearly all cases, no informed consent is presented to the participants as the informed consent process itself would disturb the natural interaction of the buyer or answerer.

The second type of recruiting method is sampling. When the experimenter has access to a database of site users (either by owning the site, through collaboration with the site owner, or because that data is made available through an accessible interface), it is possible to generate a pool of potential subjects and to in some way recruit them into the study. The pool of potential subjects may be all subjects, or may be limited to a set that meet some criteria, such as Cosley et al.'s (2007) limitation to subjects with editing history. From the pool, the experimenter may invite a random sample, may create a stratified or other systematic sample, or may simply implement the experimental interface across the entire pool.

We should be clear that "recruiting" is a broad term here. Some experiments, such as Chen et al. (2010a), involve explicit recruitment by email. This study invited a random subset of users who met certain activity thresholds, but only those who replied became subjects. Other experiments, such as the email studies shown in Ling et al. (2005) , simply made a random selection of users

and assigned those users into groups. Being sent the email was the experimental treatment. Sen et al's (2006) tagging experiments presented the interface to the entire community. For experiments which accept convenience samples of those users who volunteer, who visit the site, or who otherwise happen to stumble across the manipulation, it is not uncommon to have significantly unrepresentative samples that are highly biased towards more frequent users of the site. Even studies that do not require explicit consent, such as Cosley et al. (2007) or Sen et al. (2006) , will have samples biased towards those users who frequent the site (and in the case of Wikipedia, towards those who actually look at their personal talk pages).

The recruitment strategy is closely related to the question of informed consent. Compared with laboratory experiments, it is much more common for field experiments to request a waiver of informed consent so as to avoid changing the behavior of the subject.

## 3.3 Identification and Authentication

While some field experiments explore only one-shot interactions, most studies benefit from the ability to identify users over a period of time. Methods of identification and authentication on a site are important as they determine how accurately you can track individual activities.

Identification requires that a user offer a unique identifier, such as a registered login name. Authentication is a process that verifies the proffered identity, to increase the confidence that the user proffering the identity is actually the owner of that identity. An identification and authentication system may also be concerned with ensuring that a real-world user has only a single identity in the online community (Friedman and Resnick 2001). Sites that provide personalization or reputation systems typically require login with an ID and password. E-commerce sites may require login, but often do not do so until a purchase is being made. Many information services, from CNN.com and ESPN.com to the Internet Public Library, do not require users to identify or authenticate themselves. For these sites, creating an identification system that requires users to create accounts and login for an experiment might adversely affect usage and public satisfaction with the service, and would therefore likely to be discouraged by the site owners.

Three methods that are commonly used for tracking users on sites without logins are session tracking, IP addresses, and cookies. Each has strengths and weaknesses. Session tracking in a web server can help identify a sequence of user actions within a session, but does not track users across sessions. IP addresses - the internet address of the computer from which the user is accessing a system - can be used to track a user across multiple sessions originating from the same computer, but do not follow a user from computer to computer; also, in many cases they are impermanent, being reissued to a new computer while the original computer gets a new address. Cookies are small files that a website can ask a user's web browser to store on the user's computer and deliver

at a later time. This can can identify a user even if her IP address changes, but don't help if user moves to a different computer (or use a different browser). Users may also set their browser to reject cookies.

In Chen et al. (2006), we used cookies as one of the methods for ensuring that a user remained in the same experimental group throughout the experiment. If the user had our cookie stored, it identified that user and we would assign her to the same campaign message. If not, we would try to write a cookie, and if successful, would create an ID for the user in our experiment database so we could track her current and future interactions. Of course, this did not protect against users returning from multiple machines, but it was a practical solution. We should note that people who reject cookies may well be more technologically savvy than the average user, which raises sample bias questions for some studies. Also, we should note that there is no perfect method of identification and authentication online. It is a trade-off.

### 3.4 Control Group

Designing appropriate control conditions for online field experiments can be challenging. In many cases, it is necessary to have at least two different control groups. One group receives as carefully matched a stimulus as possible, with the exception of the hypothesized active ingredient. For example, if studying personalization, the control group receives an unpersonalized version of the interface; if studying varying content, the control group receives the same media, but different content; if studying the medium, the control group receives the same content, but with a different medium. This type of control is similar to the placebo in medical experiments. Frequently a second control is introduced - a control in which users are not given any experimental treatment. This second control can be used to help estimate the extent of any Hawthorne effects. To be effective, this control needs to be selected from the group of recruits, volunteers, or other eligible subjects. Simply following those who volunteer is inadequate, as that group does not match the subjects.

To illustrate the value in having this second control, consider the first email study in Ling et al. (2005). At the end of the data collection period, the experimenters sent out a thank-you email to all participants, which did not contain any intentional stimuli or experimental manipulation. That thank-you email created a second peak of contributions to movie-rating.

## 4 Conclusion

The number of field experiments in economics has grown tremendously in recent years, spanning fields as diverse as public finance, labor economics, industrial organization and development economics. With the expansion of the Web and e-commerce, we expect more field experiments to be

conducted over the Internet in the future. Working at the intersection of economics and computer science, we summarize the main technologies for conducting such experiments and a small number studies to highlight the potential for online field experiments.

# References


**Bakshy, Eytan, Dean Eckles, and Michael S. Bernstein**, "Designing and Deploying Online Field Experiments," in "Proceedings of the 23rd International Conference on World Wide Web" WWW '14 ACM New York, NY, USA 2014, pp. 283–292.

**__ , __ , Rong Yan, and Itamar Rosenn**, "Social Influence in Social Advertising: Evidence from Field Experiments," in "Proceedings of the 13th ACM Conference on Electronic Commerce" EC '12 ACM New York, NY, USA 2012, pp. 146–161.

**Boudreau, Kevin J. and Karim R. Lakhani**, "High Incentives, Sorting on Skills-or Just a Taste for Competition? Field Experimental Evidence from an Algorithm Design Contest," 2012.

**Chen, Roy, Yan Chen, Yang Liu, and Qiaozhu Mei**, "Does Team Competition Increase Pro-Social Lending? Evidence from Online Microfinance," 2014. University of Michigan Manuscript.

**Chen, Yan, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li**, "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens," *American Economic Review*, September 2010, *100* (4), 1358–1398.

**__ , Teck-Hua Ho, and Yong-Mi Kim**, "Knowledge Market Design: A Field Experiment at Google Answers," *Journal of Public Economic Theory*, 2010, *12* (4), 641–664.

**__ , Xin Li, and Jeffrey MacKie-Mason**, "Online Fund-raising Mechanisms: A Field Experiment," *Contributions to Economic Analysis & Policy, Berkeley Electronic Press*, March 2006, *5* (2), Article 4.

**Cosley, Dan, Dan Frankowski, Loren Terveen, and John Riedl**, "SuggestBot: using intelligent task routing to help people find work in wikipedia," in "Proceedings of the 12th international conference on Intelligent user interfaces" 2007, pp. 32 – 41.

**Friedman, Eric J. and Paul Resnick**, "The Social Cost of Cheap Pseudonyms," *Journal of Economics & Management Strategy*, June 2001, *10* (2), 173–199.

**Gee, Laura K.**, "The More You Know: Information Effects in Job Application Rates by Gender In A Large Field Experiment," 2014. Tufts University Manuscript.

**Harper, F. Maxwell, Daphne Raban, Sheizaf Rafaeli, and Joseph A. Konstan**, "Predictors of answer quality in online Q&A sites," in "CHI '08: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems" ACM New York, NY 2008, pp. 865–874.

**Harrison, Glenn W. and John A. List**, "Field Experiments," *Journal of Economic Literature*, December 2004, *42* (4), 1009–1055.

**Horton, John J., David G. Rand, and Richard J. Zeckhauser**, "The online laboratory: conducting experiments in a real labor market," *Experimental Economics*, 2011, *14* (3),

399–425.

**Karau, Steven J. and Kipling D. Williams**, "Social loafing: A meta-analytic review and theoretical integration," *Journal of Personality and Social Psychology*, 1993, *65*, 681–706.

**Lerner, Jennifer S, Roxana M Gonzalez, Deborah A Small, and Baruch Fischhoff**, "Effects of Fear and Anger on Perceived Risks of Terrorism: A National Field Experiment," *Psychological science*, 2003, *14* (2), 144–150.

**Ling, Kimberly, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid et al.**, "Using social psychology to motivate contributions to online communities," *Journal of Computer-Mediated Communication*, 2005, *10* (4), 00–00.

**Liu, Tracy X., Jiang Yang, Lada A. Adamic, and Yan Chen**, "Crowdsourcing with All-pay Auctions: a Field Experiment on Taskcn," *Management Science*, forthcoming.

**Lohr, Kathleen N, Robert H Brook, Caren J Kamberg, George A Goldberg, Arleen Leibowitz, Joan Keesey, David Reboussin, and Joseph P Newhouse**, "Use of medical care in the RAND Health Insurance Experiment: diagnosis-and service-specific analyses in a randomized controlled trial," *Medical care*, 1986, pp. S1–S87.

**Lucking-Reiley, David**, "Using Field Experiments to Test Equivalence between Auction Formats: Magic on the Internet," *American Economic Review*, 1999, *89* (5), pp. 1063–1080.

**Pallais, Amanda**, "Inefficient Hiring in Entry-Level Labor Markets," *American Economic Review*, forthcoming.

**Reiley, David H., Sai-Ming Li, and Randall A. Lewis**, "Northern Exposure: A Field Experiment Measuring Externalities Between Search Advertisements," in "Proceedings of the 11th ACM Conference on Electronic Commerce" EC'10 ACM New York, NY, USA 2010, pp. 297–304.

**Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood**, "The Value of Reputation on eBay: A Controlled Experiment," *Experimental Economics*, June 2006, *9* (2), 79 – 101.

**Sen, Shilad, Shyong K Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F Maxwell Harper, and John Riedl**, "Tagging, communities, vocabulary, evolution," in "Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work" ACM 2006, pp. 181–190.

**Williams, Dmitri, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell**, "From tree house to barracks the social life of guilds in world of warcraft," *Games and culture*, 2006, *1* (4), 338–361.