Online Field Experiments: Lessons from CommunityLab

Joseph A. Konstan[1] and Yan Chen[2] …

[1]CommunityLab[1], Dept. of Computer Science and Eng., University of Minnesota

[2]CommunityLab[*], School of Information, University of Michigan

Email address of corresponding author: konstan@cs.umn.edu

Abstract. We report briefly on a set of online field experiments conducted as part of the CommunityLab collaborative research project. Based on these projects, and the published research literature, we present an analysis of the design choices for online field experiments and report on lessons learned.

Introduction

Field experiments attempt to bring together the control and rigor of laboratory experiments with the ecological validity of field studies. They have a long tradition of use in medicine (Lohr et al., 1986), economics (Harrison and List, 2004), and social psychology (Lerner et al., 2003). One of the substantial 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 community, 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.

CommunityLab was established specifically to study the challenge of eliciting participation in online communities. We are an interdisciplinary team, including social psychologists (Sara Kiesler and Bob Kraut from CMU), an economist (Yan Chen from Michigan), computer scientists specializing in HCI and personalization (Joe Konstan, John Riedl, and Loren Terveen from Minnesota), and a specialist in socio-technical capital and information systems (Paul Resnick from Michigan). Together, we have conducted a variety of online field experiments using both online communities under our control and ones with public interfaces where we built add-ons to support our experimentation. In this paper we report on our experiences with this research methodology and offer some insights and advice for e-social scientists interested in conducting such studies.

---

[1] CommunityLab is a collaborative project of the University of Minnesota, the University of Michigan, and Carnegie Mellon University, funded by the National Science Foundation Information Technology Research program under grant IIS 03-25837. For further information, see http://www.communitylab.org.

The Anatomy of an Online Field Experiment

To achieve both the ecological validity needed to establish relevance to the real world and the level of control needed to establish causality, online field experiments should pay careful attention to the following design elements:

- Embedding the experiment in a real online activity or community, and among a set of users who naturally would be part of that activity or community. The realism is a key factor that distinguishes field experiments from laboratory studies. In some cases, the embedding may involve the creation of an online community that users see as serving a purpose apart from research studies (e.g., the MovieLens community which is primarily viewed as a place to rate and receive recommendations for movies). In other cases, researchers may embed their experiment in an existing online community such as Wikipedia (Cosley et al., 2007) or eBay (Resnick et al., 2006).

- Appropriate assignment of subjects to conditions. This may be purely random assignment, but in many cases is a stratified assignment or, in the case of social environments involving groups, a more complex assignment that intentionally either preserves groups within conditions or severs them across conditions. One of the key questions to address is whether to assign subjects from across the entire membership pool vs. eliciting volunteers.

- Appropriate controls for measuring the effect of the conditions. One of the clear lessons from our studies is that there are both significant effects associated with being in such a field experiment--whether the "Hawthorne Effect" commonly associated with being observed or the more mundane effect of simply being contacted and therefore reminded about the online venue. Accordingly, it is often useful to separately measure the effect of the manipulation against both a control manipulation (balanced for contact) and against a null control (no contact).

Communitylab Case Studies

Motivating Contributions through E-mail. Ling et al. (2005) reports on a set of four field experiments conducted with members of the MovieLens online movie recommender community. In three of these experiments, selected users of the system received e-mail messages asking them to rate more movies (i.e., to contribute effort and knowledge to the community). In all, over 2400 users were sent an e-mail invitation. These messages were crafted to test certain hypotheses based on Karau and Williams' Collective Effort Model (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 e-mail 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 2x2 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 where subjects are kept unaware of the nature of the manipulation. In this case, the discussions were created specifically for the experiment. Later, however, we decided to add them as a permanent feature of the site, making possible future experiments where participation in discussions is even more of a natural part of site usage for subjects.

Economic Modeling of User Rating. Harper et al. (2005) conducted a series of studies targeted at building an economic model of user rating behavior and then using that model to test theories of motivation. In this work, we solicited a set of 356 volunteers (and in this case ended up with a disproportionately high-usage pool of users) and used a combination of their behavioral data plus survey data to model their rating behavior. After calibrating a model of their behavior, we were able to explain 34% of the between-user variation in rating behavior. The work has continued into a study of inequality aversion and conformity based on a set of variant newsletters sent to volunteer members that compared their rating behavior (conformity) or economic net benefit (inequality aversion) to other users in a peer group. 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.

SuggestBot and Wikipedia. Cosley et al. (2007) deployed an intelligent task-routing agent (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 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 1200 people received recommendations found that personalized recommendations led to nearly 4 times as many actual edits as random suggestions. Other promising data includes dozens of positive comments from users.

The SuggestBot case adds two important factors. First, it shows that it is not necessary to "own" the community in order to conduct online field experiments there. Many online communities provide ways in which an experimenter may add features that can be differentially deployed to different users. One challenge 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 design questions motivated by social science.

Tagging. Sen et al. (2006) studied the evolution of tagging vocabulary in a community by deploying a tagging feature in MovieLens and then dividing the user base into four groups. One group saw only their own tags. Three other groups saw different version of shared tags, including randomly selected, most popular, and those based on a recommendation algorithm. Through this experiment, it became clear that a user's view of the community's tagging has a significant effect on which tags a user chooses to apply himself or herself.

Analysis and Discussion

Field experiments conducted over the Internet have several distinct characteristics compared to those conducted in the brick and mortar world. First, the nature of the task is often unique to the online world, even though the underlying behavioral principles might be the same. Second, typically the experimenter can obtain more information than traditionally available for experiments conducted in other media. For example, some software can track the detailed activities of each user, including a user's click stream and a time stamp for each activity. From these data, the designer can infer

important underlying user preferences and the time cost of each activity. Such information has been used to target customers in e-commerce, as in Amazon.com's book recommendations. Finally, for online communities with an authentication process, one can track past and future behavior.

Based on the case studies presented above, and on other field experiments in the research literature, we present a taxonomy to help researchers understand the design alternatives in online field experiments. This taxonomy looks at (1) the degree of control the experimenter exerts over the online venue where the experiment is being carried out, (2) issues of recruiting and informed consent, (3) identifiability and authentication of subjects, (4) the nature of the control group, and (5) participation incentives. 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.

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:

1. 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 about 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 interfaces 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., 2007) 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.

2a. A site with a public interface that allows substantial experimenter control can be a useful alternative, especially when you don't already have a user community available. Unfortunately, few sites give that level of control to outsiders, but there are some examples that do, at least in certain domains. Wikipedia is one example of a site that encourages such research Cosley et al. (2007) used the site 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. Over a six month study, over 1200 people received recommendations; their subsequent editing was monitored (which is made easy since editing is both public and displayed on the users' page), leading to the discovery

that personalized recommendations led to nearly 4 times the number of edits as random ones. 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.

2b. A close, 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 et al. (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.e., 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.

3. Experimenters can pose as ordinary users which involves minimal or no collaboration of the site owners. In most online sites, it is feasible for experimenters to establish identities as ordinary users for the purposes of both gathering data and introducing some stimulus. 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, posing as ordinary users can be not only an effective way to gather field data, but also to inject manipulations.[2]

Posing as an ordinary user for purposes of conducting an experimental manipulation differs from creating an account for purposes of participant-observation or conducting in-context interviews (for an example of those, see Nardi and Harris (2006)). It is often the case that researchers set up multiple accounts for the purpose of experimenting with different conditions in parallel, and they may take additional steps to avoid detection.

Lucking-Reiley (1999) presents the first economic field experiment conducted over the Internet, 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

---

[2] Posing as users raises important ethical considerations. Our general view is that such posing is acceptable when it fulfills three criteria: (a) it does not disrupt the normal operation of the site, (b) it does not deceive users in a way that would cause them any harm, and (c) it represents a small enough percentage of the usual activity on the site so as to prevent establishing some new social expectation that will be disturbed when the experiment ends. Of course, depending on the details of the experiment, the planned manipulation may need to be reviewed by a human subjects review committee.

higher revenues than the first-price auction, which violates the theoretical prediction and what had been observed in laboratory studies.

More recently, Resnick et al. (2006) conducted a field experiment on eBay 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 spider.

The posing technique is also effective for non-economic studies. Burke et al. (2007) studied the effect of different rhetorical strategies for first-time posters in Usenet news groups. They reposted old messages (from new identities) altered to randomly assign the message to different categories of self-introduction and request. They then analyzed the responses (which are publicly posted) to find that introductions referencing group history increase reply counts, though requests do not.

4. Scrape data after a natural shock: sometimes there are natural shocks to a system or online community which lead to interesting comparisons and results even though the shocks are not designed by an experimentalist. One can then scrape data from the site, before and after the shock, to study the effect of the particular variable which changed.

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* as defined by 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. (2007), 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. Also, we should make it clear that few field experiments go to the lengths of traditional survey research to obtain statistically valid samples. Most experiments accept convenience samples of those users who volunteer, who visit the site, or who otherwise happen to stumble across the manipulation. As a result, it is not uncommon to have significantly unrepresentative samples that are highly biased towards more frequent users of

the site. Harper et al. (2005) was highly skewed towards power users; they were more likely to respond to the invitation email and complete a survey. 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. Conceptually, this is no different online than offline; a researcher wondering to what extent lower prices would result in higher purchases of healthy snacks from vending machines (French et al., 2001) cannot first intercept prospective buyers for informed consent.

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 (even one-shot studies may want to ensure that the same individual doesn't participate multiple times). Methods of identification and authentication on a site are important as they determine how far and 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 move to a different computer (or use a different browser). Users may also set their browser to reject cookies.

In Chen, Li and Mason (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 didn't 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.

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 doesn't 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.

Participation Incentives

Many online field experiments rely on the natural incentives in the tasks that participants perform in their natural environment, rather than explicit incentives provided by the experimenter. For example, experiments conducted on eBay, bidders do not need to be induced to bid; indeed, the experiment would be invalid if they were paid to bid.[3]

Certain less natural activities may require additional compensation. This compensation may be in-band (such as providing special access or status in the system, or system-usable currency), or may be out-of-band (such as cash payments or souvenir merchandise). In some cases such incentives are delivered to each participant, and in others entry into a prize drawing is used. Such participation incentives are particularly common when surveys or interviews are coupled with the experiment. For example, in Chen et al. (2007) we awarded entries into a prize drawing (for cash prizes) for filling out surveys and for participating in the study. $300 in prizes (one $100 prize, two $50 prizes, and five $20 prizes) yielded us a post-experiment survey completion rate of 78%.

Lessons Learned

We have learned a great deal from carrying out these online field experiments ourselves and from from reading the experiences of others. Some key lessons include:

- An interdisciplinary team is of great value. We have found that the collaboration between technologists (from Computer Science) and social scientists has been a critical part of our success. Technologists not only can set up the experimental apparatus, but they can envision

[3] Indeed, Lucking-Reiley (1999) notes that, rather than spending money, he made a 25% profit for his online field experiment.

new ways of manipulating and measuring the community. Social scientists ground the research in fundamental questions and can provide design insight grounded in theory.

- The experimenter's relationship with the research site gives them various degrees of cotnrol. There are great advantages to having your own community. We have had MovieLens for about a decade, and it has given us the ability to control and measure every aspect of the system. Many interesting experiments involve presenting different experiences to different users and watching their behavior as time progresses. Working with the site owner or a site with a public interface that allows substantial experimenter control gives the experimenter a fair amount of data and control as well. Lastly, the experimenter posing as a user can be effective in a variety of studies.

- Carefully consider the user reaction. In most cases, we've found users to be extremely supportive of serving as subjects in our experiments. In a few cases, however, users have become very upset at what they see as social engineering or simply making the system they enjoy worse. Remember that users (and their happiness) is a resource you need to manage.

- Exploit the power of large numbers. When you need it. It is often possible to track, electronically, the behavior of hundreds or even thousands of individuals. An important corollary is to be very careful in selecting the measurements and metrics you use.

Finally, we have come to recognize the tremendous value of stable online communities that are open to (and open for) research. For this field to prosper, we believe that many more such communities are needed, and would advocate making research infrastructure funding (which all-too often is limited to equipment) explicitly available to fund community development and/or community management efforts designed to broaded research access.

Acknowledgments

We would like to thank the entire CommunityLab team. This work was supported by the National Science Foundation under grant: IIS 03-24851.

References

Burke, M., Joyce, E., Kim, T., Anand, V., and Kraut, R. (2007): 'Introductions and Requests: Rhetorical Strategies That Elicit Response in Online Communities', *Third International Conference on Communities & Technologies*, 2007.

Chen, Y., Harper, M., Konstan, J. and Li, X. (2007): `Social Comparisons and Contributions to Online Communities: A Field Experiment at MovieLens', Manuscript.

Chen, Y., Li, X. and MacKie-Mason, J (2006): `Online Fund-raising Mechanisms: A Field Experiment.' *The B.E. Journal of Economic Analysis & Policy* vol. 5, issue. 2 (Contributions), article 4.

Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. (2007): 'SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia', In *Proceedings of IUI 2007*.

French, S.A., Jeffery, R.W., Story, M., Breitlow, K.K., Baxter, J.S., Hannan, P. and Snyder M.P. (2001): `Pricing and promotion effects on low-fat vending snack purchases: the CHIPS Study', *American Journal of Public Health*, vol 91, issue 1: pp. 112-117.

Friedman, E. and Resnick, P. (2001): `The Social Cost of Cheap Pseudonyms', *Journal of Economics and Management Strategy*, vol. 10, no. 2: pp.173-199.

Harper, F., Li, X., Chen, Y. and Konstan, J. (2005): `An Economic Model of User Rating in an Online Recommender System', In *Proceedings of The 10th International Conference on User Modeling*, Edinburgh, UK, 2005.

Harrison, G.W. and List, J.A. (2004): `Field experiments.' *Journal of Economic Literature*, vol. 42, no. 4: pp.1009–1055.

Karau, S. and Williams, K. (1993): `Social loafing: A meta-analytic review and theoretical integration', *Journal of Personality and Social Psychology*, vol. 65, no. 4: pp. 681-706.

Lerner, J.S., Gonzalez, R.M., Small, D.A. and Fischhoff, B (2003): `Effects of fear and anger on perceived risks of terrorism: A national field experiment',*Psychological Science*, vol. 14, no. 2: pp. 144–150.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., and Kraut, R. (2005): `Using social psychology to motivate contributions to online communities', *Journal of Computer-Mediated Communication*, vol. 10, no. 4, article 10.

Lohr, K.N., Brook, R.H., Kamberg, C.J., Goldberg, G.A., Leibowitz, A., Keesey, J., Reboussin, D. and Newhouse, J.P. (1986): `Use of Medical Care in the Rand Health Insurance Experiment: Diagnosis- and Service-Specific Analyses in a Randomized Controlled Trial', *Medical Care*, vol. 24, no. 9: pp. S1-S87.

Lucking-Reiley, D. (1999): `Using field experiments to test equivalence between auction formats: Magic on the internet', *American Economic Review*, *89*, 5: 1063–1080.

Nardi, B. and Harris, J. (2006). 'Strangers and friends: collaborative play in world of warcraft' In *Proceedings of CSCW 2006*, 149-158.

Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K (2006): `The Value of Reputation on eBay: A Controlled Experiment.' *Experimental Economics*. 9, 2: 79-101.

Sen, S., Shyong K., Lam, D., Rashid, A., Frankowski, D., Harper, F., Osterhouse, J., and Riedl, J. 'tagging, community, vocabulary, evolution', In *Proceedings of CSCW 2006*.

Williams, D., Ducheneaut, N, Xiong, L., Zhang, Y., Yee, N., and Nickell, E (2006): `From Tree House to Barracks: The Social Life of Guilds in World of Warcraft,' *Games and Culture*, vol. 1, no. 4: pp. 338-361