Evaluating Compliance-Without-Pressure Techniques for Increasing Participation in Online Communities

Mikhil Masli and Loren Terveen
GroupLens Research
Department of Computer Science and Engineering
University of Minnesota, Minneapolis, MN, USA
{masli, terveen}@cs.umn.edu

ABSTRACT
Social psychology offers several theories of potential use for designing techniques to increase user contributions to online communities. Some of these techniques follow the "compliance without pressure" approach, where users are led to comply with a request without being subjected to any obvious external pressure. We evaluated two such techniques – *foot-in-the-door* and *low-ball* – in the context of Cyclopath, a geographic wiki. We found that while both techniques succeeded, low-ball elicited more work than foot-in-the-door. We discuss design and research implications of applying these (and other such techniques) in online communities.

Author Keywords
Increasing participation; social production; online communities; compliance; foot-in-the-door; low-ball

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation: Group and Organizational Interfaces

General Terms
Human Factors, Design

INTRODUCTION
A large body of literature from social psychology suggests techniques that may be used to motivate people to contribute to these communities. One set of these techniques is known as *compliance without pressure* [12]: these techniques are designed to increase compliance to requests in the absence of any obvious sources of pressure. Two of the most popular techniques are:

- *Foot-in-the-door* (FITD): Once a person performs a small request, he/she is more likely to perform a subsequent, larger demand [12].
- *Low-ball* (LB): Once a person commits to a request, he/she is more likely to perform it at a later stage, even if its cost is increased [8].

While these techniques have been evaluated extensively in several offline contexts, there has been very little evaluation in the world of computer-mediated communication [16, 17], and none in online social production communities.

We evaluated FITD and LB techniques in the context of Cyclopath[1], an online geographic wiki that offers route-finding services for cyclists in the metropolitan area of Minneapolis-St. Paul, Minnesota, USA. We carried out a work campaign in which we requested Cyclopath users to contribute information to the map in the form of tags with the goal of improving the map for the next cycling season. We conducted a field study and a follow-up survey and found that LB succeeded in eliciting more contributions than the control condition, while FITD had limited success.

This paper is organized as follows. We first survey relevant related work and explain the motivation to carry out empirical evaluation in the online context. Second, we describe our field experiment, its results and their interpretation using theories from social psychology. Third, we describe our follow-up survey, and conclude with implications for design and future research.

RELATED WORK

Social Production Communities
Social production communities [2, 14] let loosely-connected users work together to produce information and artifacts of value [10]. Collaborative filtering systems like MovieLens[2] and Amazon leverage users' ratings of items (movies, consumer products, etc.) to enable personalized recommendations. Q&A sites like Yahoo! Answers and Stack Overflow form knowledge economies, where users spend points to ask or boost the priority of questions and earn them for answering. Wikis take user-provided content to its logical end: anyone can add, edit, or delete anything. Wikipedia is among the top 10 most popular sites on the Web, and smaller wikis are ubiquitous. Because of the significant value produced by online communities, ensuring that they receive adequate member participation is an important research topic.

[1] cyclopath.org
[2] www.movielens.org

Compliance Without Pressure

Compliance without pressure techniques are designed to lead people to comply with requests without any obvious source of external pressure. Researchers from social psychology and marketing have developed several techniques that fall under this category, e.g. FITD, LB, *door-in-the-face* [7], *that's-not-all* [3] and *bait-and-switch*. FITD is one of the oldest and the most popular of them with well over a hundred studies devoted to it over the last 50 years [4]. LB has also been found to be one of the most effective of these techniques [21, 22]. Because of their effectiveness, FITD and LB are interesting candidate solutions to the problem of under-contribution in online communities, and we thus chose to investigate them.

Several theories have tried to explain why and how FITD and LB work (or not). The most popular explanation for the success of the FITD technique is Bem's self-perception theory [1], which postulates that when a person is induced to comply with a smaller request, he/she is more likely to comply with a subsequent, larger request because of perceiving himself/herself as the type of person that does such tasks. Alternative theories including those citing psychological reactance, conformity to existing social norms including reciprocity and cognitive dissonance [11] have also been used to explain both successes and failures of the FITD technique. The LB effect has been generally explained using the theory of commitment [8], which says that people generally tend to stick to their commitments when acting in public view. We elaborate on these techniques in a later section in this paper while interpreting the results of our field experiment.

While compliance-without-pressure techniques may have potential to increase participation in online communities, if users perceive these techniques as manipulative, their use could negatively impact long-term member satisfaction and commitment. Accordingly, in this work, along with evaluating how successful FITD and LB are in an online social production context, we also evaluate the extent to which they are harmful and discuss the pragmatic issues designers of online communities might face when employing some of them.

NEED FOR EMPIRICAL EVALUATION

This research focuses on empirically evaluating the techniques of FITD and LB in the context of an online social production community. Is this necessary? Could we not merely assume findings from the offline settings of social psychology? Online communities have several key differences that makes them unique, and justify the need for an empirical evaluation of proven offline techniques:

1. **Anonymity on the Internet**. Psychological phenomena like cognitive dissonance, considered one of the factors behind the FITD and LB effects, as well as the LB effect itself have been shown to be more effective in public situations [5, 32], where visibility is higher (and consequently, anonymity is lower). However, since obtaining an identity on the Internet is cheap [13], online communications have a higher degree of anonymity than offline situations. Hence, it is not clear if online social production communities are public (contributions are shared) or private spaces (people can work from the confines of their homes), thereby raising questions over the effectiveness of the FITD and LB techniques.
2. **Nature of online communication**. Email and similar modes of communication are used most often by online social production communities to interact with their members, as it is the norm on the Internet. These modes are fundamentally not as personal as face-to-face or telephone communication – modes used by most studies of compliance in offline situations. Further, due to problems of email overload, it easy to ignore incoming messages. These reasons affect any interventions that use email as the underlying mode of communication.

Although online communities are different from their offline counterparts in some respects, they are similar in many others. Indeed, the social sciences have provided design inspiration and specific techniques that Computer Supported Cooperative Work (CSCW) designers and researchers have used to create more effective online communities [28]. Researchers have successfully applied theories and models like the Collective Effort Model [23, 25], goal-setting [25] and social comparison [20] to address the problem of under-contribution. Closer to this work, the FITD technique has been shown to be effective over email [16, 17].

Drawing inspiration from these successes, our research extends prior work in significant ways:

- We empirically evaluate the FITD and LB techniques in the context of an online social production community using a live field experiment and a follow-up survey.
- We demonstrate that the LB technique is largely successful in eliciting a higher quantity of contributions, while the FITD technique receives mixed results.
- Based on what we learned from our studies, we discuss the advantages and limitations of employing these (and similar) techniques in real-world online social production communities.

RESEARCH PLATFORM

Cyclopath went live in the summer of 2008. As of May 17, 2011, Cyclopath had 3,232 registered users and visits from 69,754 different IP addresses. Cyclopath has served as an excellent research platform for prior research on user behavior in social production communities, including studying efficient ways of eliciting contributions [29], user lifecycles [27] and specialization of contributions [26].

Like any wiki, users contribute information to Cyclopath in sets of edits called revisions. When a user makes the desired edits in the Cyclopath interface (e.g. add a point) and clicks "Save Changes", they are saved as a single revision associated with the username of the user (if the user is logged in), IP address and a time-stamp (see [30] for a more in-depth description). All revisions are publicly visible and revertible (via the Recent Changes list). So far, the Cyclopath map has seen 12,949 revisions (11,258 by 699 registered users and 1,691 anonymously).

Each revision may contain edits to points, blocks, regions, notes and tags. Blocks (roads and trails like the *Kennilworth Trail*) are useful for route-finding, points help define route endpoints (e.g., from *CS Building* to *Lagoon Theater*), tags enable more customized route requests (e.g., a route that is more *scenic*, but less *bumpy*), and notes supplement blocks included in a route with subjective information useful for evaluatint the route (e.g., "icy during the winter").

Prior work in Cyclopath suggests that many notes contain potential tags [33]. If extracted, these tags could provide more options for users to customize their routes. Further, this task could be performed by any person, regardless of his/her familiarity with the note or the area on the map where the note was applied (just like the task of fixing intersections [29]). Hence we chose to ask Cyclopath's users to do the task of extracing tags from notes. Prior to the start of our experiment, users had added 6,373 tags (applied to 4,312 blocks and points) and 8,405 notes (applied to 6,084 blocks).

EXPERIMENT DESIGN

Hypotheses and Variables

We conducted a field study in Cyclopath during winter 2010-11. Like bike-riding (at least, in colder climates), Cyclopath use is seasonal – it sees over 150 route requests per day and over 100 revisions per week during the summer and only about a third of those numbers during the winter. This made winter a good time for our field study – any effects we observe are primarily due to our manipulation and not the general motivations that drive contributions during summer. As a part of our field study, we requested Cyclopath users to contribute tags to the resource.

Our experiment design was guided by two hypotheses:

H1. The FITD technique will result in higher compliance for the contribution request than the request being presented alone.

H2. The LB technique will result in higher compliance for the contribution request than the request being presented alone.

We measured compliance using two outcome (dependent) variables:

- **Response:** Whether the user responded at all and the number of responses per user (we allowed users to respond multiple times, as described later) to the target request.
- **Quantity of Work:** The amount of work done by the user in response to our request, measured in terms of number of tag-applications to blocks and points, and number of new tags introduced into the tag vocabulary.

Method

Partitioning Users

We partitioned users into three groups – FITD (953 subjects), LB (953 subjects), and a control group (951 subjects). However, since participation on Cyclopath is highly unequal (like most voluntary online activities), we took care to ensure that user experience levels were more-or-less evenly distributed among the three groups, i.e. not all highly experienced users (measured by the number of prior revisions made) were assigned to the same group. Specifically, we ordered users in descending order by the number of revisions they saved and stepped through this list assigning users to control, FITD and LB groups in a cyclic fashion.

Procedure

General structure. All communication with users was conducted via email. We presented both experimental groups (FITD and LB) with their *initial contacts* on December 9, 2010 at 5 pm. We then followed this up with the *target request* (the real request for which we wanted compliance) to all groups, via email 6 days later[3], on December 15, 2010 at 5 pm.

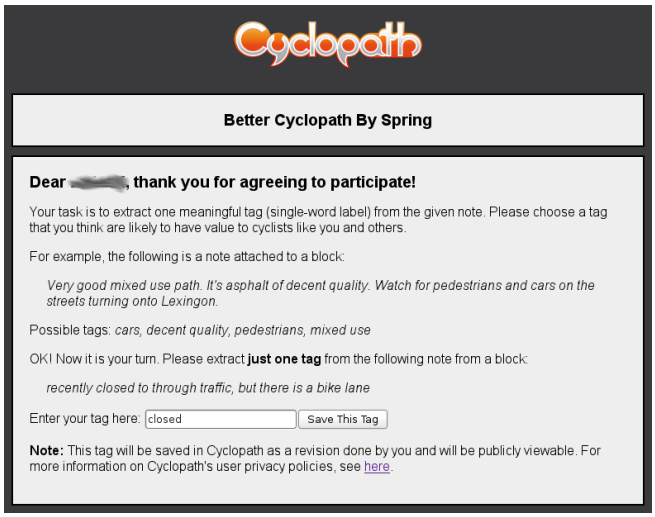
It is important to note here that essentially, we are comparing 3 techniques of eliciting work: "simply asking users to do work" (control) and two "enhanced ways of asking users to do work" (FITD, LB). We could have evaluated other, intermediate techniques like Freedman and Fraser's *familiarity-only* (related to FITD) [12] and *commitment-but-no-increase-in-cost* (related to LB) as well. However, in this paper, we have explicitly chosen to evaluate FITD and LB in their entirety and not part-by-part.

The following paragraphs detail the initial contact and the target request for the various groups (the control group received no initial contact).

Initial contact – FITD. We sent email to the subjects in the FITD group asking them to do the initial (small) task of extracting just one tag from a single note, with the subject line "A tiny favor: Cyclopath needs your help!". The key passage was: *Would you spare a minute to help this campaign? All we are asking you to do is to read the following note and extract only one meaningful tag (one or two word label) from it.* The email also contained the note itself (we thought that once the subject eyeballed it, the task of extracting one tag would seem trivial), a sample note-tag pair, and a link to an interface to submit the tag. This link launched a simple HTML-based interface where the subject could re-read the note and enter the tag he/she extracted (see Figure 1(a)). For example, a subject extracted the tags *"Dirt path"* from the note *"dirt path connects to Arcade beware of curb"*. Since the initial FITD contact had to be short and easy, we limited subjects to just one attempt at it: if a subject clicked the link again, the system would display a message saying he/she has already completed the task. After the subject submitted the tag from the HTML interface, he/she was thanked.

Initial contact – LB. We asked the subjects in the LB group (with the same subject line) to agree to extract one tag from a single note, explicitly asserting that this task should take only a minute to complete. The key passages were: *Would you spare a minute to help this campaign? All we are asking you to do is to read a note and extract only one tag (single word label) from it (this should just take you a minute)* and

[3]Prior research suggests that a short (near-zero) delay between the initial and the target requests when both originate from the same requester can tend to counteract the FITD effect [4, 19].



(a) FITD Interface

(b) LB Interface

Figure 1. Initial contact. We designed simple HTML-based interfaces for the FITD subjects to perform their small task, and for the LB subjects to express their commitment.

If you agree now, we'll contact you with specific instructions in near future, after we have heard from more people. The email also contained a sample note-tag pair and a link to an interface to express commitment. This link launched a simple HTML-based interface (similar to the FITD interface) where the subject was thanked for agreeing (see Figure 1(b)).

Target request – all groups. Our target request to all three groups was to extract as many tags as possible from a given set of notes. To avoid any biases resulting from subjects' self-selection, this email request was sent to every subject in the FITD and LB groups, regardless of their response to the initial contact. Crucially, for the FITD and LB conditions, we needed to explain why we were asking for more work than in the original request. Our explanation was that Cyclopath had miscalculated the amount of work required of each user and that more work per user was needed in order to analyze all notes by spring time[4]. The key passage in this email was: *You will be presented with a area on the map. Read any notes you find on blocks in this area and extract around 2-3 potentially useful tags (one or two word labels) from each note. This should take you 15-20 minutes.* The link in the email

[4]This argument was also included in the email sent to the control group for uniformity, though not framed as an "excuse".

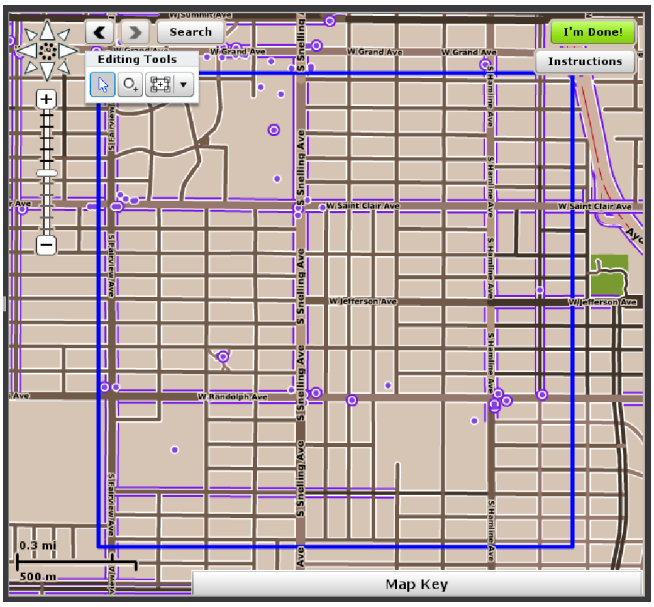


Figure 2. Target request. Each subject could perform as many experimental trials as he/she wishes. The blue box showing the area the subject was asked to work in, with the "I'm Done" button to end the trial at the top right.

launched an experimental trial (the procedure as well as the code for conducting trials was adapted from [29]):

1. *Begin trial.* The subject begins the trial by clicking on the link provided in the email.

2. *Select viewport.* The system chooses a viewport (100 km^2 area) randomly on the map, ensuring that there are at least 5 notes available to extract tags from.[5]

3. *Display viewport.* The map pans and zooms to the selected viewport, and highlights it with a blue box. An instruction dialog box pops up that details the task to be done, with brief instructions on how to do it and links to help (see Figure 2). All blocks and points with notes were highlighted with a purple halo.

4. *Subject does work.* The subject is now free to use the system. Although we explicitly asked the subject to work in the assigned area, the system did not restrict him/her in any way – the subject was free to make edits anywhere on the map.

5. *End trial.* The subject clicks a button labeled "I'm Done", which results in the completion of one trial. The subject can now either leave Cyclopath, use it for other purposes, or start another trial.

We recorded when the subjects started and finished the trials, the number of trials requested, and everything they did as a part of those trials.

[5]This was to ensure that there was *some* work available to do. Further, at the outset of our analyses, we also verified that there were no statistically significant differences in the number of notes available per user and per trial among the three groups.

Group	Response
FITD	11.4% (109)
LB	12.7% (121)

Table 1. Response to the initial contact. Response rates are low, but typical of contributory behavior in social production communities.

Group	Response	Trials per subject
Control	6.14% (58)	0.22 (205)
FITD	5.14% (49)	0.10 (92)
LB	7.35% (70)	0.29 (272)

Table 2. Response to the target request. The numbers in parantheses are the respective absolute counts. There are no statistically significant differences between the three groups of subjects.

Results

We took a snapshot of the database on January 16, 2011. By this date, 190 users completed 597 trials. We removed 28 trials from our analyses – 15 were unusually long (over 2 hours[6]) and 13 were done by 6 users who were not in any of our groups (these users joined Cyclopath after the experiment started). This left us with 569 trials by 177 subjects. We present our results organized by our outcome variables.

Response

Initial contact response. Table 1 shows the response to the initial contact. Out of the 953 subjects in the FITD condition to whom we sent the initial contact emails, 109 (11.4%) completed the short task. Out of the 953 subjects in the LB condition to whom we sent the initial contact emails, 121 (12.7%) responded.

Target request response. Table 2 shows how subjects responded to the target request. We see no significant differences among groups in the response rates, or the number of trials per subject. However, we found interesting differences in the quantity of work done, as described in the following sub-section.

Quantity of Work

Metrics. When a tag that does not already exist in the system is applied to a new map feature, a new tag is introduced into the tagging vocabulary. We call such an action "creating a tag." On the other hand, if an existing tag is applied, it is merely re-used, i.e., a new connection is added between the existing tag and the map feature to which it is applied. We call such an action "applying a tag." Naturally, when a tag is created, it is also applied to some map feature (e.g., a block), resulting in a tag application. We counted the number of tag applications (split into tag applications on blocks and points) and the number of new tags added, per user and per trial.

Work done per user. LB subjects applied more tags on average (about 3 times) than FITD or control subjects. These results were statistically significant. We performed statistical tests (for p-value computations) on $log(1 + x)$-transformed variables to account for the non-normality (long-tail nature) of the data and compensate for counts of zero. We first did a one-way analysis of variance (ANOVA) on the number of tag applications ($F(2, 2854) = 2.78, p = .06$), then did post-hoc pairwise comparisons using Tukey's HSD test (the difference between LB and control groups was marginally significant, $p = .06$). On an average, LB subjects also created nearly thrice as many tags as control subjects (ANOVA: $F(2, 2854) = 2.70, p = .07$, then Tukey's HSD: LB vs. control: $p = .08$). Table 3 summarizes the number of tag applications and new tags added per user by subjects from the three groups.

[6]The median length of a trial was 1.5 minute. We hypothesize that the unusually long trials represent cases where the subject forgot to click the "I'm done" button, remembering only much later. Consequently, we could not idenfity, with reasonable confidence, the parts of the trial that were done as a response to the target request.

(a) Tags created per user.

Group	Total	Mean	SD	p
Control	70	0.07	0.65	
FITD	99	0.10	1.05	
LB	193	0.20	1.64	$*$

(b) Tags applied per user.

Group	Total	Mean	SD	p
Control	542	0.57	5.07	
FITD	571	0.60	6.05	
LB	1702	1.79	18.07	$*$

Table 3. Tagging work done per user in response to the target request. $*$ indicates statistical significance at the 0.1 level when compared with control.

(a) Tags created per trial.

Group	Total	Mean	SD
Control	70	0.34	1.08
FITD	99	1.08	2.43
LB	193	0.71	2.54

(b) Tags applied per trial.

Group	Total	Mean	SD
Control	542	2.66	7.20
FITD	571	6.23	13.11
LB	1702	6.25	13.79

Table 4. Tagging work done per trial in response to the target request.

Work done per trial. We see evidence of the success of the LB technique at the per-trial level too. The average LB trial produced more than twice as many tags (6.25) as the average control trial (2.66). Table 4 summarizes these comparisons. Looking back at Table 2, we see that the average FITD subject did fewer trials than the average control subject. This explains why there is a significant difference between the FITD and control groups in terms of work done per trial, despite there being none in terms of work done per user.

Summary of results. Our results suggest that the LB technique significantly increased the quantity of work subjects put in (hypothesis H2). The FITD technique, on the other hand, is not as successful – an average FITD trial produced more work than a control trial, but on an average, FITD subjects did no more work than control subjects ($0.57 \approx 0.60$) (hypothesis H1). There also was evidence that the LB technique fared better than the FITD technique.

Discussion – Why these results?

We have just described an empirical investigation of the FITD and LB techniques of compliance-without-pressure in the context of an online social production community, Cyclopath. Why did we obtain the results we saw? What aspects of our procedure caused these results, and what aspects produced opposing forces? Answering these questions is important for crafting recommendations concerning the use of these techniques for designers of online communities. We now refer back to relevant social psychological theories to make sense of our results.

Why did LB succeed?

The LB effect is theoretically simpler: it is generally explained by the theory of commitment. The seminal work on the LB effect [8] attributed its success to the development of commitment to the task in the minds of the person as a consequence of the initial contact. Further research [6] has shown that commitment to the requester, not the task, is actually responsible for the increased likelihood of compliance to the target request. In our case, either is possible: task commitment, since we explicitly asked LB subjects to agree to the task of extracting a tag from a note, or requester committment, as Cyclopath is a community resource, and people tend to care for the communities they are a part of.

Why did FITD meet with limited success?

There have been several attempts at explaining the FITD effect (or the lack of it) using a variety of psychological processes and theories [4]. Based on these theories, we speculate on the following reasons behind our results.[7]

Self-perception not strong enough. The self-perception theory [1] is the most popular explanation for the success of the FITD effect. The reasoning is that, when a person is induced to comply with the smaller request, he/she infers from this behavior that he/she is the kind of person that does such tasks and supports the cause behind the request; this self-perception later causes him/her to comply with the larger request. However, if a weak FITD effect is observed (as our results indicate), then we can hypothesize that the inferential process of self-perception triggered by the smaller request did not take place or was not strong enough.

Reactance as a result of excessive requesting. Psychological reactance may also undermine the FITD effect – the subject may simply feel badgered and pressured by the repeated requests and refuse the second, larger request. Our results showed a significant difference in the work done per trial, but no difference in the work done per user between the FITD and control groups. This is because although the number of subjects from the FITD group who responded to the target request is about the same as that from the control group, the FITD group completed fewer trials than the control group (see Table 2). A possible hypothesis for this behavior is that some FITD subjects might have either felt badgered by the repeated requesting or felt that they have "already done their part" when they complied with the smaller request.

Reciprocity: Cyclopath did not return the favors asked. Expectation of reciprocity may actually work against the processes driving the FITD effect. If the subject, after complying with the smaller request, expects the requester to give something in return, then the fact that he/she instead comes back with a larger request has the potential of inviting a refusal on the part of the subject. On the other hand, had the subject turned down the initial contact, then the norm of reciprocity would drive him/her to comply with the target request even if it is larger. Indeed, 21 subjects who did not respond to our initial request did respond to the target request. Whether reciprocity was the reason for this behavior is worthy of investigation.

Conformity to norms: Was the target request not to be done? Another factor that might affect compliance is the knowledge of existing social norms. For example, if the subject learned that the norm is that the larger request is not complied with, then he/she also might refuse. However, in this study, Cyclopath saw nearly 500 revisions to the map after the target request was sent out: so, if any norm was present, it was the norm of contributing.

Consistency needs. The cognitive dissonance theory [11] stipulates that people have a need to appear consistent and resolve any dissonances in their selves as soon as possible, even if it means making irrational decisions. In the FITD context, when a subject is presented with the larger request, he/she behaves consistently with his/her prior behavior in a similar context, i.e., the compliance to the initial smaller request. Although social psychologists have developed a scale to measure the preference for consistency [15], any conclusion we draw here would be about the people who use Cyclopath, and not the design of the system itself.

We conducted a follow-up study to investigate which of these processes may have affected our results.

FOLLOW-UP STUDY

Research Questions

We designed a follow-up study to address the following research questions that we constructed from analyzing the results of our field study:

RQ-FITD. Why did the FITD technique meet with limited success? Possible reasons include:

- FITD subjects did not perceive themselves to be the type of people who respond to requests from community sites like Cyclopath (self-perception).
- FITD subjects did not feel motivated enough to comply with the target request because they felt that they had already done their part when they responded to the initial contact (reactance).
- FITD subjects felt that the repeating requesting behavior from Cyclopath was badgering them (reactance).

[7] Burger's review [6] of about 100 FITD studies clearly points out that the degree to which each of the psychological processes known to enhance/reduce the FITD effect operate depends on the specific procedures used in the intervention. Hence, one must exercise caution while generalizing any speculations we make regarding how strongly they operated in our case.

FITD group statements:

About whether the subject would respond to initial request:
- *"I would do the task because it was requested by Cyclopath."*
- *"I would do the task because it is small and easy."*
- *"I would do the task regardless of who requested it."*

About how the subject would feel after responding to the initial request were presented:
- *"I have done my part towards helping Cyclopath."*
- *"I am the type of person who typically responds to requests for help from community sites."*

About the target request:
- *"Asking me to do something else after only a week has gone by is too often."*
- *"I've already done my part when I responded to the earlier request."*
- *"(If I hadn't done the task from the earlier request) I will do this one since I haven't done the task from the earlier email."*
- *"I'm the type of person that helps Cyclopath whenever called upon."*

LB group statements:

About whether the subject would respond to initial request:
- *"I would commit to such a request because it was requested by Cyclopath."*
- *"I would commit to such a request because it is small and easy."*
- *"I would commit to such a request regardless of who requested it."*

About the target request:
- *"This is unfair, Cyclopath just increased the amount of work it asked me to do."*
- *"I'm a person of my word: I agreed to help, so I'll do it now."*
- *"I'm the type of person that helps Cyclopath whenever called upon, so I would do this one too."*

Control group statements:

About the target request (in addition to statements for the FITD group):
- *"It is too much of a favor to ask on the part of Cyclopath."*
- *"I typically do not respond to such requests."*

General statements:

About perceived norms:
- *"If Cyclopath sent the above email to all registered users, how many people do you think (other than you) would respond to this request?"*
- *"It would be bother me if few people did the requested task."*

Figure 3. Follow-up survey items. Each experimental group was presented with a different set of scenarios and items. Subjects rated each item (except the first of the general statements) on a 7-point Likert scale.

- FITD subjects who turned down the initial contact felt obligated to comply with the target request on the grounds of reciprocity (reciprocity).

RQ-LB. Why did the LB technique succeed in eliciting more contributions? Possible reasons include:

- LB subjects complied with the target request because they felt a sense of commitment towards the task of extracting tags from notes (task commitment).
- LB subjects complied with the target request because they felt a sense of commitment towards Cyclopath (requester commitment).

Method

We sent out a survey on April 13, 2011 to subjects from all the groups. In the survey, subjects were asked to imagine scenarios similar to the experimental procedure of our field study; e.g., the FITD group was presented with the scenario: *Imagine this: You receive an email from Cyclopath asking you to do a task. The task is an easy one: you're asked to read a 20-word note, and create a single meaningful tag for it (a 1-2 minute task). What would your response be?* We took this approach (of asking people to imagine) instead of relying on subjects' memories of their experimental experience for two reasons: First, the interval between the start of the experiment and the survey was about 4 months, and it was unlikely that subjects would accurately remember their experiences over such a large interval. Second, this approach made it possible for us to administer the survey even to those who had not participated in the experiment.

Note that for subjects who participatedin the field study, we assigned them a survey consistent with their experience in the field study, i.e. FITD-relevant items be presented only to the FITD group, and so on. Users who registered after the start of the experiment (and hence were not assigned to any group) were randomly assigned to one of the three groups for the purpose of the survey.

Subjects were asked to rate several statements on a 7-point Likert scale (1-completely disagree to 7-completely agree) with regards to each scenario. Figure 3 lists some of the items.

Results and Discussion

We received 148 responses to our survey by April 20, 2011, a week after it was launched. Out of these, 43 (29%) of the users were among those who had participated in the previously described field study, whereas 105 (71%) were new. There were no statistically significant differences in the responses of these two sets of users. Table 5 presents the most important results. Our survey results showed some condition-specific patterns and some general population-wide patterns.

We observed general ideological patterns across all groups: respondents said they would do the requests because they came from Cyclopath ($M \geq 5.13, 1.23 \leq SD \leq 1.38$ in all cases) and would not do so in general ($M \leq 3.13, 1.58 \leq SD \leq 1.73$ in all cases). Moreover, respondents from the control group disagreed with the statement that the task requested of them (the target request) was too much of a favor ($M = 2.64, SD = 1.15$)! This shows users' general commitment to the Cyclopath community and can be a powerful motivation to contribute whenever requested.

FITD. Respondents from the FITD condition agreed with the statement that they would perform the task requested in the initial contact because it was small and easy ($M = 5.42, SD = 1.54$) – a verification of the FITD design. Further, we saw that FITD respondents agreed with the fact that they felt that they had *done their part towards helping Cyclopath* ($M = 5.47, SD = 1.16$) and that asking them to do something more after just a week had passed by was too often ($M = 4.62, SD = 1.32$). This is an indication that a likely reason for the lower-than-expected contributions to the FITD target request was that people felt that they were badgered.

The results also seem to indicate that this reason outweighs the self-perception effect: contributions from the FITD group in the field study were lower despite FITD respondents in the

Group	Survey Item	Mean (M)	Std. Dev. (SD)
FITD	Initial contact: How would you respond?		
	I would do the task because it was requested by Cyclopath.	5.53	1.23
	I would do the task because it is small and easy.	5.42	1.54
	I would do the task regardless of who requested it.	3.13	1.73
FITD	Initial contact: How would you feel after you did it?		
	I have done my part towards helping Cyclopath.	5.47	1.16
	I am the type of person who typically responds to requests from community sites.	4.70	1.65
FITD	Target request: How would you respond?		
	Asking me to do something else after only a week has gone by is too often.	4.62	1.32
	I'm the type of person that helps Cyclopath whenever called upon.	4.43	1.39
LB	Initial contact: How would you respond?		
	I would commit to such a request because it was requested by Cyclopath.	5.57	1.35
	I would commit to such a request regardless of who requested it.	2.65	1.58
	I would commit to such a request because it is small and easy.	4.98	1.50
LB	Target request: How would you respond?		
	This is unfair, Cyclopath just increased the amount of work it asked me to do.	4.47	1.50
	I'm a person of my word: I agreed to help, so I'll do it now.	4.63	1.50
Control	Target request: How would you respond?		
	I would do the task because it was requested by Cyclopath.	5.13	1.39
	I would do the task regardless of who requested it.	3.00	1.63
	It is too much of a favor to ask on the part of Cyclopath.	2.64	1.15

Table 5. Results of our follow-up survey.

survey agreeing that after doing the task requested during the initial contact, they would feel like they are the type of people that would respond to requests for help from community sites ($M = 4.70, SD = 1.65$).

We did not find any statistically significant evidence for reciprocity; however, one respondent said, *"yeah if you help us back, this ain't a one-way street buddy"* as a response to how he/she would feel after responding to the initial contact.

LB. Respondents from the LB group verified the LB design by agreeing to the statement that they would commit to the request in the initial contact because it was small and easy ($M = 4.98, SD = 1.50$). Further, they also agreed to the statement that they would contribute to the target request because they were *persons of their word* ($M = 4.98, SD = 1.50$); in other words, committed to doing the task requested.

IMPLICATIONS AND SUMMARY

We believe this work is useful to both researchers and practitioners designing interventions to increase community participation. The success of the LB technique showed that an intervention as subtle as requesting prior commitment produced large, significant increases in participation. This low-cost-high-returns mechanism is potentially of great use to large scale campaigns such as WikiProjects and crowd-sourcing initiatives. However, care should be taken while applying theories from social psychology to online communities: there are subtleties and points of caution that must be considered.

Compliance-without-pressure in the online world. Earlier in the paper, we outlined how features such as anonymity and the loose, detached nature of online communication (as opposed to the more involved face-to-face) motivated the need for an empirical evaluation of the FITD and LB techniques. Looking at our results in the light of these premises, we see

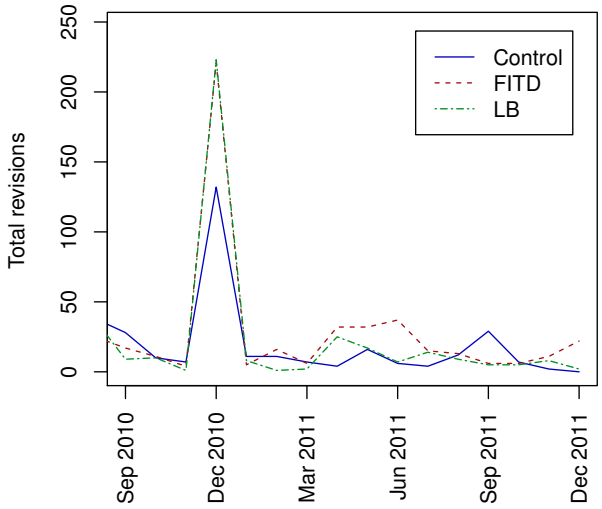


Figure 4. One-shot Effect. Our intervention saw a burst of activity just after the experimental emails were sent out (Dec, 2010), which dropped off steeply soon after. Note that the activity of Cyclopath's most prolific contributor has been excluded from this figure.

that the LB technique succeeded due to a heightened feeling of commitment despite the relative anonymity associated with the online community. However, FITD only produced limited success because it was very easy for subjects to simply close their browsers if they felt badgered by the repeated requesting. Thus, differences between online and offline interactions must be carefully considered before planning any intervention using a compliance-without-pressure technique.

Short versus long-term effects. We have seen in the past that interventions that try to boost member participation succeed in doing so only while they are active, and that participation

drops to pre-intervention levels once it is withdrawn [29]. Our FITD and LB interventions fell within the same category. We saw a burst of activity (measured as number revisions saved) immediately following the initial contact and the target request, which dropped off steeply soon after (Figure 4). The self-perception effect, one of the explanations for the success of the FITD technique, is one of the most likely processes that could have caused long-term effects, whereas psychological reactance would work against any. In our case, we did not see any evidence for self-perception, which is consistent with our understanding of why our work campaign produced a one-shot effect.

While the failure to cause a long-term increase in participation levels is a limitation of the techniques we evaluated, we believe that they have utility in one-shot interventions. Online social production communities do require one-shot bursts of member participation at times – e.g, if a WikiProject in Wikipedia wishes to categorize several hundred articles into a hierarchical structure, or if a geographic open content community like FixMyStreet[8] needs to identify all potholes in a locality as a part of a neighborhood campaign.

That said, research efforts should try to focus on producing long-term increases in participation. In other words, researchers must try to devise techniques that enable the community to support itself. One way of tackling this issue is to employ the powerful and ubiqutous norm of reciprocity, an approach we are currently exploring.

Member satisfaction and retention. From past research, we know that although users can be manipulated into behaving in a particular manner, they tend to detect the manipulation and become less satisfied with the system [9]. This reduces the value of perhaps the greatest asset of an online community: the trust and commitment of its members.

Similarly, in the context of our study, we must consider the possible effects on member satisfaction of implementating the type of compliance-without-pressure techniques we studied. Our users certainly felt that the repeated emailing was excessive (in case of FITD), and that increasing the cost of the requested task was unfair (LB) (Table 5). From an online community designer's perspective, users might be viewed as *workers* – in which case they can be manipulated to achieve maximum throughput – or as *community members* – in which case they must be carefully preserved and nurtured. Recognizing the strengths of the community (e.g., existing levels of user commitment) is critical: if apparently effective in the short term, techniques like FITD and LB might do more long term damage to users' perceptions of and commitment to a community. After all, the tangible product of a group activity is not the only outcome that matters; in the long run, users will continue to participate only if their needs are met [18, 24]. Finally, since there are so many online communities, users who are not satisfied with one are likely to leave it in favor of an alternative, more attractive one that serves the same purpose.

Limitations and future work. An alternative explanation for the LB effect is the role of time: people tend to discount how busy they are or how much effort is needed when they make a commitment in the future [31]. LB studies that vary the amount of time between the initial contact and the target request are required to test this explanation.

Our study used a simple survey to conduct the qualitative follow-up to the field experiment. A limitation of post-hoc surveys are that they are presented out-of-context. Although we tried to mitigate this problem by making the survey as standalone as possible, better qualititative information could be collected if the survey was interwoven into the field experiment itself.

Summary. We described an evaluation of the compliance-without-pressure techniques foot-in-the-door (FITD) and low-ball (LB) in the context of an online social production community, using a field experiment and a follow-up survey. We found that despite the subtlety in its manipulation, the LB technique elicited nearly 3 times more participation. However, while both techniques managed to produce an increase in contribution, this was short-lived. These findings are applicable to both researchers and designers who create interventions to boost participation in online communities. However, more research is needed to look into other techniques (like the norm of reciprocity) that may yield lasting effects.

[8]www.fixmystreet.com

ACKNOWLEDGEMENTS

We thank the members of GroupLens Research, the Cyclopath team, the Cyclopath user community, and our reviewers. We also thank Dr. Reid Priedhorsky, Katie Panciera, Jason Weaver, Dr. Donald Ross Jr., Dev Oliver, and Ginny Sims for providing valuable feedback and guidance on research and writing. This work was supported in part by the National Science Foundation grant IIS 08-08692.

REFERENCES

1. Bem, D. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological review 74*, 3 (1967), 183–200.
2. Benkler, Y. *The wealth of networks: How social production transforms markets and freedom*. Yale Univ Pr, 2006.
3. Burger, J. Increasing compliance by improving the deal: The that's-not-all technique. *Journal of Personality and Social Psychology 51*, 2 (1986), 277.
4. Burger, J. The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and Social Psychology Review 3*, 4 (1999), 303.
5. Burger, J., and Cornelius, T. Raising the Price of Agreement: Public Commitment and the Lowball Compliance Procedure. *Journal of Applied Social Psychology 33*, 5 (2003), 923–934.
6. Burger, J., and Petty, R. The low-ball compliance technique: Task or person commitment. *Journal of Personality and Social Psychology 40*, 3 (1981), 492–500.
7. Cialdini, R., et al. Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique.

Journal of Personality and Social Psychology 31, 2 (1975), 206.

8. Cialdini, R., et al. Low-ball procedure for producing compliance: Commitment then cost. *Journal of Personality and Social Psychology 36*, 5 (1978), 463–476.
9. Cosley, D., et al. Is seeing believing?: How recommender system interfaces affect users' opinions. In *Proc. CHI*, ACM (2003), 585–592.
10. Cosley, D., et al. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proc. CHI* (2006).
11. Festinger, L. *A theory of cognitive dissonance*. Stanford University Press, 1957.
12. Freedman, J., and Fraser, S. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology 4*, 2 (1966), 195–202.
13. Friedman, E., and Resnick, P. The social cost of cheap pseudonyms. *Ann Arbor 1001* (1999), 48109–1092.
14. Gilbert, E., and Karahalios, K. Using social visualization to motivate social production. *IEEE Transactions on Multimedia 11*, 3 (2009), 413.
15. Guadagno, R., et al. When saying yes leads to saying no: Preference for consistency and the reverse foot-in-the-door effect. *Personality and Social Psychology Bulletin 27*, 7 (2001), 859.
16. Guéguen, N. Foot-in-the-door technique and computer-mediated communication. *Computers in Human Behavior 18*, 1 (2002), 11–15.
17. Guéguen, N., and Jacob, C. Fund-raising on the web: the effect of an electronic foot-in-the-door on donation. *CyberPsychology & Behavior 4*, 6 (2001), 705–709.
18. Hackman, J. The design of work teams. *Handbook of organizational behavior 315* (1987), 342.
19. Harari, H., et al. Faculty Helpfulness to Students. *Personality and Social Psychology Bulletin 6*, 3 (1980), 373.
20. Harper, F., et al. Social comparisons to motivate contributions to an online community. *Persuasive Technology* (2007), 148–159.
21. Hornik, J., et al. Comparison of three inducement techniques to improve compliance in a health survey conducted by telephone. *Public Health Reports 105*, 5 (1990), 524.
22. Joule, R. Tabacco deprivation: The foot-in-the-door technique versus the low-ball technique. *European Journal of Social Psychology 17*, 3 (1987), 361–365.
23. Karau, S., and Williams, K. Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology 65*, 4 (1993), 681.
24. Kraut, R. Applying social psychological theory to the problems of group work. *HCI models, theories and frameworks: Toward a multidisciplinary science* (2003), 325–356.
25. Ling, K., et al. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication 10*, 4 (2005), 00–00.
26. Masli, M., et al. Task specialization in social production communities: The case of geographic volunteer work. In *Proc. ICWSM* (2011).
27. Panciera, K., et al. Lurking? cyclopaths? A quantitative lifecycle analysis of user behavior in a geowiki. In *Proc. CHI*, ACM (2010), 1917–1926.
28. Preece, J. *Online communities*. Wiley & Sons, 2001.
29. Priedhorsky, R., et al. Eliciting and focusing geographic volunteer work. In *Proc. CSCW*, ACM (2010), 61–70.
30. Priedhorsky, R., and Terveen, L. Wiki Grows Up: Arbitrary Data Models, Access Control, and Beyond. In *Proc. WikiSym*, ACM (2011).
31. Roy, M., et al. Underestimating the duration of future events: Memory incorrectly used or memory bias?. *Psychological Bulletin 131*, 5 (2005), 738.
32. Tedeschi, J., et al. Cognitive dissonance: Private ratiocination or public spectacle?. *American Psychologist 26*, 8 (1971), 685.
33. Torre, F., et al. bumpy, caution with merging: an exploration of tagging in a geowiki. In *Proc. GROUP 2010*, ACM (2010), 155–164.