Using Social Psychology to Motivate Contributions to Online Communities

Gerard Beenen[1], Kimberly Ling[1], Xiaoqing Wang[4], Klarissa Chang[1], Dan Frankowski[3], Paul Resnick[2], Robert E. Kraut[1]

CommunityLab*

[1]Carnegie Mellon University
5000 Forbes Ave
Pittsburgh PA 15221
(gbeenen, kling, changtt, kraut)
@andrew.cmu.edu

[2]University of Michigan
3210 SI North
Ann Arbor, MI 48109
presnick@umich.edu

[3]University of Minnesota
200 Union Street SE
Minneapolis, MN 55455
dfrankow@cs.umn.edu

[4]University of Pittsburgh
Pittsburgh, PA 15260
xiawang@katz.pitt.edu

ABSTRACT

Under-contribution is a problem for many online communities. Social psychology theories of social loafing and goal-setting can provide mid-level design principles to address this problem. We tested the design principles in two field experiments. In one, members of an online movie recommender community were reminded of the uniqueness of their contributions and the benefits that follow from them. In the second, they were given a range of individual or group goals for contribution. As predicted by theory, individuals contributed when they were reminded of their uniqueness and when they were given specific and challenging goals, but other predictions were not borne out. The paper ends with suggestions and challenges for mining social science theories as well as implications for design.

Categories and Subject Descriptors

J.4 Social and behavioral sciences, H.4.3 Communications Applications.

General Terms

Design, Experimentation, Human Factors

Keywords

Virtual Community / Community Computing, Social Computing and Social Navigation, Empirical Methods-Quantitative, Usability Testing and Evaluation, Social and Legal Issues, User Studies

1. MOTIVATING CONTRIBUTIONS IN ONLINE COMMUNITIES

Since at least 1979, when the first Usenet news sharing programs were created, online communities have co-evolved with the growth in computer networking. Today, 25 years later, people share news, information, jokes, music, discussion, pictures, and social support in hundreds of thousands of online communities. People benefit from the presence and activity of others in online communities—from the information and other resources they provide and the conversations they participate in.

Despite the vibrancy of online communities, large numbers of them fail. Participation is often sub-optimal, with only a small minority contributing. In many online groups, participation drops to zero. For example, Butler found that 50% of social, hobby, and work mailing lists had no traffic over a 4-month period [1]. Under-contribution is a problem even in communities that do survive. In a majority of active mailing lists, fewer then 50% of subscribers posted even a single message in a 4-month period [1]. Similarly, on the popular peer-to-peer music sharing service, Gnutella, two-thirds of users share no music files and ten percent provide 87% of all the music [2]. In open source development communities, four percent of members account for 50 percent of answers on a user-to-user help site [3], and four percent of developers contribute 88% of new code and 66% of code fixes [4]. Although not everyone needs to contribute for a group to be successful [5], groups with a large proportion of non-contributors have difficulty providing needed services to members. For example, in open source development environments, bugs are not fixed and enhancements are not delivered. In movie rating groups, obscure movies might not be evaluated. In medical support groups, important problems and treatments might not be discussed. We believe it is an important and difficult challenge to design technical features of online communities and seed their social practices in a way that generates ongoing contributions from a larger fraction of the participants.

In this paper, we attempt to tackle the problem of under-contribution in an online community called MovieLens[6]. MovieLens is a web-based movie recommender community where members can rate movies, write movie reviews, and receive recommendations for movies. More than 20% of the movies listed in the system have so few ratings that the recommender algorithms cannot make accurate predictions about whether subscribers will like them. Here, the contributions we hope to motivate are ratings of movies, especially rarely-rated movies.

Social science theories have helped CSCW designers and developers make sense of failures and focus attention on difficulties that will need to be overcome in system design. For example, the ideas of

* CommunityLab is a collaborative project of the University of Minnesota, University of Michigan, and Carnegie Mellon University. http://www.communitylab.org/

critical mass and network externalities help explain the difficulty in getting systems adopted [7], and individual utility analysis based on the distribution of costs and benefits helped designers understand why electronic calendaring systems went unused in many organizations [8].

Many CSCW researchers have also drawn design inspiration and design guidelines from social science findings (e.g., [9-12]). For example, research on the importance of unplanned conversation in the workplace inspired the design of awareness systems to give people information about the availability and context of potential conversational partners (e.g., [9, 10, 12]).

Our aim is to build an even stronger link between social science theories and CSCW design. First, we aim to draw on behavioral theories that explain why people do things rather than just on empirical regularities (or stylized facts). By drawing upon theory, designers can craft mechanisms to engage the causal mechanisms even in settings that on the surface appear quite different from settings where the stylized facts were observed. Second, we seek to implement alternative designs for which the theories predict different outcomes, so that results from field experiments can test the underlying theories or resolve questions on which the theories were silent.

Motivating contributions, especially contributions to the communal good, is a topic that has received substantial attention in many branches of the social sciences. Social psychologists have documented a robust phenomenon across many settings that they call *social loafing*. People exert less effort on a collective task than they do on a comparable individual task (see [13] for a review). Economists and political scientists have come to similar conclusions. Across a wide range of settings, people contribute less than the optimal amount of public goods and consume more than their fair share of common pool resources, though the anti-social behavior is considerably less than theories based on pure short-term self-interest would predict (see [14] for a review).

Ostrom [15] provides a theory of the conditions that allow groups to effectively govern common pool resources. Kollock and Smith [16] treat these as design principles in an analysis of the successes and failures of Usenet, but they do not test the impacts of alternative designs.

In this paper, we describe two attempts to mine theories from social psychology on the motivators of individual effort in collective effort situations and on the motivating impacts of specific, challenging goals. We apply these principles to the design of appeals for soliciting contributions to MovieLens.

In each of two experiments, we first identified the abstract mental states that the theories propose should lead to contribution, such as believing that one's contributions are unique in a group or that they will benefit the group. We then translated them into specific mental states that a participant in the MovieLens community might have, such as believing that he or she rates movies that few other users rate or that the ratings help others. Next, we designed persuasive messages that we hoped would induce these mental states. We conducted controlled experiments by delivering different versions of an email message inviting existing MovieLens subscribers to rate movies. We relied on email primarily because of ease of implementation. Much as low-fidelity prototypes provide a low-cost way to test interface designs, these email interventions provide a low-cost way to test the impacts of manipulating relevant mental states. These email messages probably have only a fleeting influence on the mental states. Once experiments reveal the appropriate manipulations, more costly changes could be made to the MovieLens interface.

We expect that permanent changes to the interface would have longer-lived influences on subscribers' mental states.

2. Study 1: Motivating contributions through framing uniqueness and benefit

Social loafing or free riding is the robust phenomenon that occurs when people work less hard to achieve some goal when they think they are working jointly with others than when they are working by themselves. Karau and Williams developed a collective effort model, a theory to explain why people often work harder individually than in groups [13]. According to this theory, people work hard when they think their effort will help them achieve outcomes that they value. Working in a group can influence how hard people work because it can change their perception of the importance of their contribution to achieving a specified level of performance, their likelihood of reaching the goal, and the value they place on the outcomes they gain by their efforts. (See Karau and Williams [13] and Figure 1 for a fuller description of the collective effort model.)

Figure 1: The collective effort model (adapted from [13])
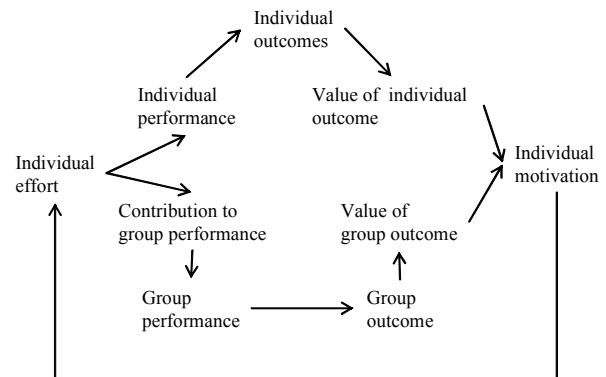
The collective effort model identifies conditions under which people will socially loaf less. These include (a) believing that their effort is important to the group's performance, (b) believing that their contributions to the group are identifiable, and (c) liking the group they are working with, among others. Social psychologists have tested the collective effort model by manipulating individuals' perceptions of the elements in Figure 1, both in laboratory settings (e.g., [17-19]) and online [20].

We attempted to apply the insights from the collective effort model to the problem of under-contribution in MovieLens. As described earlier, over 20% of the movies in MovieLens are rated by so few subscribers that the recommender system has insufficient data to provide recommendations for any user. We call these rarely-rated movies (RRM). This experiment sought to improve the quality of the MovieLens system by increasing subscribers' motivation to rate movies, both rarely-rated ones and other movies in the system.

Salience of uniqueness. The collective effort model posits that people will socially loaf less when they perceive that their contribution is important to the group [13]. If they believe that their contributions are redundant with what others in the group can offer, then their contribution is unlikely to influence the group's outcome. Conversely, if they think they are unique, they should be more motivated to contribute, because their contributions have a larger chance of influencing the outcome that they value.

In the case of MovieLens, individuals who rate rarely-rated movies make this type of important unique contribution. To motivate these members to action, one can make them aware of their uniqueness.

Hypothesis 1: MovieLens users will rate more movies when the uniqueness of their contributions is made salient.

Salience of benefit and the beneficiary. The collective effort model also posits that people are more motivated to contribute when they see the value that their contribution makes to an individual or group outcome [13]. MovieLens is a collaborative filtering system that uses other people's ratings to predict how much a subscriber will like a movie. In MovieLens, rating movies leads to useful movie predictions for both the rater and others. Raters personally gain benefit because predictions for them become more accurate, although with decreasing marginal returns as they rate more movies. Essentially, rating more movies allows the system to learn something about their preferences. Reminding subscribers of this individual benefit should increase their motivation to rate more movies.

> *Hypothesis 2a: MovieLens users will rate more movies when the personal benefit they receive from doing so is made salient.*

When individuals rate movies, they benefit the community as a whole by increasing the accuracy of recommendations that others receive. However, this benefit to the community may not be visible to members, because they do not have the data to see the correlation between their ratings and the accuracy of recommendations for others. Therefore, making explicit the benefit that the community receives from their ratings should increase their ratings. For example, Sheppard and Taylor found that participants increased their performance when they perceived a contingent tie between their individual performance and group outcomes [21].

> *Hypothesis 2b: MovieLens users will rate more movies when the benefit they provide to the community is made salient.*

Combining uniqueness and benefit. The collective effort model emphasizes that uniqueness leads to greater effort because the person believes that their effort has a positive impact on the group outcome [13]. Some MovieLens users, when informed of their uniqueness, will infer that they are uniquely helpful to the group, but others may not. Thus, emphasizing benefit to others should increase the power of the uniqueness manipulation.

> *Hypothesis 3: MovieLens users will rate more movies when the perception of both unique contribution and benefits to the community are made salient than when only unique contribution or benefits are made salient.*

2.1 Methods
Overview. We conducted our experiment on MovieLens.org, an online community administered by the University of Minnesota. MovieLens members rate movies and receive personalized recommendations provided by a collaborative recommender system. MovieLens has about 80,000 registered users, of whom about 7,000 were active in the six-month period before this research was conducted[1]. We recruited active MovieLens

---

[1] This research was approved by the Institutional Review Boards at both the University of Minnesota and Carnegie Mellon University.

members who had rated rare movies by sending them an electronic mail message inviting them to participate in a movie rating campaign. We used different messages in this invitation email to vary their perceptions of their uniqueness and who would benefit from their contributions. We then tracked their contribution behavior following the invitation.

Subjects. The subject population consisted of 904 active MovieLens subscribers who had rated rarely-rated movies. Members who logged on to the MovieLens website at least one time in 2003 were considered active. We sampled members who had rated at least 3 rarely-rated movies (i.e., those in the bottom 30% of all movies) or for whom rarely-rated movies comprised at least 15% of all movies they had rated. Of the 904 members that we contacted, 74 of the emails bounced, leaving us with 830 participants.

All subjects received an email message inviting them to participate in a campaign to rate more movies. The text of these email manipulated two variables, which we will call *uniqueness* and *benefit*.
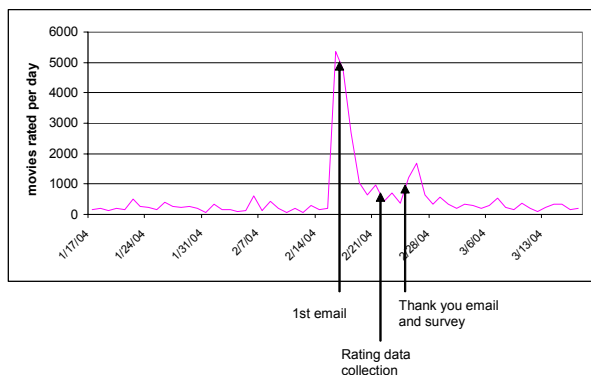
Uniqueness. Participants who received the uniqueness manipulation were sent a personalized email that told them they were selected for the campaign because they tended to rate movies that few other MovieLens users had rated. The message said, "We are contacting you because as someone with fairly unusual tastes, you have been an especially valuable user of MovieLens. In the past, you have rated movies that few others have rated, such as …" followed by titles of three rarely-rated movies they had previously rated. Participants who received the non-unique manipulation were told they were recruited because they had previously rated movies that many other MovieLens subscribers had rated. The message said, "We are contacting you because as someone with fairly typical tastes you have been an especially valuable user of MovieLens. In the past, you have rated movies that many others also rated, such as…" followed by titles of frequently rated movies.

Benefit. The benefit manipulation contained 4 conditions: no benefit, only benefit to self, only benefit to others, and benefit to both self and others. Participants who received the *self-benefit* manipulation received a message that said, "Rating more movies helps you! The more ratings you provide, the easier it is for MovieLens to identify people with similar taste to yours, and thus make accurate recommendations for you." Participants who received the *other-benefit* manipulation received a message that said, "Rating more movies helps the MovieLens community! The more ratings you provide, the more information we have about each movie and the easier it is to make accurate recommendations for other people. " Participants in the *both self and other benefit* condition received a combination of these messages, and those in the *no-benefit* condition received neither.

Measuring contribution. Figure 2 shows the number of ratings made by experimental participants surrounding the email invitation. Because ratings surged during the week of the invitation, and then rapidly fell to their pre-invitation level, we treated the week following the email invitation as the period when the experimental manipulation was active. We logged data from the participants, including their ID, movies rated, rating scores given, and time of rating, for one week following the invitation. Following data collection we sent a second personalized email reporting on each participant's individual rating behavior, summarizing the rating behavior of participants as a whole, and thanking them for their participation in the campaign.

Figure 2: Number of ratings surrounding an invitation to contribute

Table 1: Study 1. Mean Number of Movie Ratings

		No Benefit	Benefit to self	Benefit to others	Benefit to self and others
Non-unique	N	103	110	101	106
	P(logged in)	50.5	40.9	46.5	45.3
	# Ratings	25.16	13.19	13.22	19.21
Unique	N	99	101	103	107
	P(logged in)	59.6	46.5	42.7	51.4
	# Ratings	31.53	19.28	9.48	23.66
Total	N	202	211	204	203
	P(logged in)	55.0	43.6	44.6	50.7
	# Ratings	28.28	16.1	11.33	21.45

Note. N is the number of subjects to whom email was successfully delivered. P(logged in) is the percentage of participants who logged into MovieLens during the week of the experiment. # Rating is the mean number of ratings for the N subjects.

The dependent variable is the number of ratings that participants made during the week following their invitation. Because the number of ratings is skewed, with many respondents not logging in to the MovieLens site during the week of data collection and a small number generating many ratings, we used the log transform of the number of ratings as the dependent variable. We conducted the analysis in two ways. In the first, we used data from everyone who received an email invitation and therefore was presumably exposed to the experimental manipulations. In this analysis, because the logarithm of 0 is not defined, we consider those who did not log in as contributing one rating. We conducted some supplementary analyses only among those who logged in.

2.1.1 Data Analysis and Results

Of the 830 participants who received email, 397 (47.8%) members logged in and rated at least one movie. Descriptive analysis including all participants showed that they rated an average of 19.26 movies during the week, far higher than the 5.4 movies per week they had rated in the 6 months before the invitation. During the experiment, they rated an average of 1.56 rarely-rated movies. As Figure 2 illustrated, merely sending the email led to a dramatic increase in ratings. In contrast, only 3.2% of a matched control group logged in during the week of the experiment. They rated 9.1 movies per member logging in, compared to 39.7 movies per participant who received the email reminder and logged in.

We analyzed the data using a 2 (uniqueness) x 2 (self-benefit) x 2 (other benefit) analysis of variance. Binary logistic regression testing the likelihood of logging in at least once during the week showed that there was no difference between the experimental groups on this measure. Therefore we used all participants in our analyses, using the number of ratings (log transformed) and the number of ratings of rarely-rated movies (log transformed) as dependent variables. The results are qualitatively the same if we restrict analysis only to those who logged in at least once.

Uniqueness

Hypothesis 1 stated that making unique rating behavior salient would increase people's motivation to rate more movies in general and more rarely-rated movies in particular. The results confirmed this. As seen in Table 1, participants in the unique group rated 18% more movies than those who got the non-unique message (means = 20.92 vs. 17.65, p=.045). Moreover, the differential was even higher when we consider only rarely-rated movies, where the participants in the unique group rated 40% more movies than those who got the non-unique message (means

= 1.82 vs. 1.30, p=.019). Supplementary analyses using robust regression analysis to guard against outliers and the non-normal distribution yielded similar results.

Benefit

Hypothesis 2 stated that making salient personal and group benefits would increase people's motivation to rate more movies. The results were more complex. Mention of benefit depressed ratings rather than increased them. As Table 2 shows, participants receiving email mentioning benefits rated an average of 16.36 movies during the experiment, while on average the combined no-benefit groups rated 28.28 (p=.006). This finding disconfirms the prediction from the collective effort model.

The average difference in ratings between the benefit and no-benefit groups is qualified by the highly significant self-benefit by other benefit interaction (See Figure 3). People who received no benefit information and those who received both the self-benefit and the other-benefit information rated more movies than those who received only the self-benefit or the other-benefit information. Neither the main effect of the benefit manipulation nor the self-benefit by other-benefit interaction was significant for the ratings of rarely-rated movies.

Combining Uniqueness and Benefit

Hypothesis 3 stated that participants will rate more movies when the perception of both uniqueness of their contribution and benefits of increased rating are made salient than when just one of them is made salient, so that a combination of these two manipulations would have an effect beyond the sum of the individual effects. The results did not support our hypothesis. There was no interaction effect between uniqueness and benefit

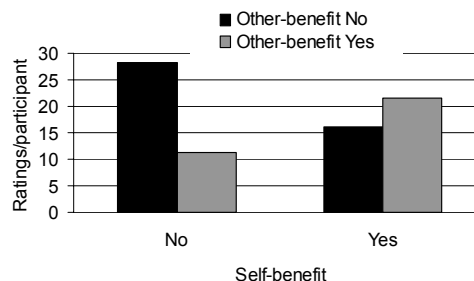**Figure 3: Effects of benefit manipulation on ratings**

Table 2: Study 1 Results Summary

Hypothesis	Results	
	All movies	**Rarely-rated movies**
H1:Unique group rates more than non-unique group	Supported	Supported
H2a: Benefit-to-self group rates more than no-benefit groups	Disconfirmed (significant in the opposite direction)	Not supported
H2b: Benefit-to-others group rates more than no-benefit group	Disconfirmed (significant in the opposite direction)	Not supported
H3: Interaction between Uniqueness and Benefit	Not supported	Not supported

on either overall ratings or rarely-rated movies ratings.

A summary of the hypotheses testing results is shown in Table 2.

2.2 Discussion

The results of this experiment confirm what telemarketers know: email messages can motivate people in an online community simply by reminding them of an opportunity to contribute. More interestingly, the content of the message made a difference, partially in line with the collective effort model. Making members of the community feel unique encouraged them to contribute more in general and especially to contribute in the domain where they were unique. In the case of MovieLens, reminding some members of their rarely-rated-movie history increased their willingness to rate these types of movies (and all movies) compared to members who did not receive this reminder.

Highlighting the benefits that participants or other members of MovieLens receive from ratings had a more complicated relationship to contributions. According to the collective effort model, uniqueness should have its effects on contributions by making people feel that their contributions were more valuable. Based on this model, we expected that the effects of uniqueness on increasing contributions would be greater for participants reminded of the utility of their contributions, but this prediction was not supported by the data.

The collective effort model also led us to expect that reminding people of the utility of their contributions would increase their motivation to contribute; however, the results were inconsistent with this expectation. They show that reminding participants of the benefits that either they or others receive from contributions depressed the number of ratings they made compared to participants who received no reminders of benefit. On the other hand, telling participants simultaneously about both benefits that they and others receive led to more effort than telling them about either one alone.

Surveys of the MovieLens membership suggest that they rate movies primarily to improve the accuracy of recommendations that they receive from the system, because the acts of remembering movies and rating them is intrinsically fun, and, to a lesser extent, to help other subscribers to the system. It is possible that reminding participants of the instrumental reasons for rating—more accurate predictions for themselves and others—may undermine their intrinsic motivation to rate [35]. In a similar vein, it is also possible that in choosing a population of users who had rated rarely-rated movies, we may have contacted a segment of the population that was already highly committed to MovieLens and well aware of the various benefits of their contributions. Sending these people a message highlighting only a single benefit of contribution might have paradoxically undermined their prior beliefs that contributions had broad utility. That is, the reminder of a single kind of benefit might have caused them to narrow their focus from many benefits of participation to only the one mentioned. This may have actually demotivated the members by not validating the other beliefs about benefit that the rare raters already held. In contrast, the groups receiving a reminder of both benefits may not have narrowed their focus as much.

It is also possible that the content of our emails did not sufficiently convey the types of benefit we were attempting to make salient. Users may already have had strong models of how MovieLens works and so were not be swayed by the short explanations in the emails. Moreover, the effort required to understand the messages about benefits may have drawn attention away from the main message of requesting users to rate movies. According to this interpretation, merely including an extra paragraph in the message was a demotivator, but once that paragraph was included, it was better to make it more complete.

Overall, it is clear that calling users' attention to their uniqueness and to the benefits they provide through their action has an impact on their rating effort. More research is needed to understand the puzzle of why subjects rated more movies when no mention was made of the benefits their ratings create for themselves or other users.

3. Study 2: Motivating contributions through goal-setting

The designers of online communities rarely provide participants specific descriptions of the type and amount of contribution that is expected of them. Open source software development communities are an exception, posting lists of bugs to be fixed, but even there, goals are rarely set for specific individuals. Although MovieLens provides users feedback information about the number of ratings they have made, it does not give an indication of the number of ratings they should make or which movies would be most helpful to rate. Our purpose here is to test both the benefits and limits of specific, high-challenge goals in online community.

Benefits of High-challenge Goals. Abundant research since the 1960s shows that providing people with specific, high-challenge goals stimulates higher task performance than easy or "do your best" goals (see [22] for a review). This phenomenon is among the most robust psychological findings on human motivation. The straightforward design recommendation from the goal-setting literature for online communities is that these communities should set specific and challenging contribution goals for their members.

We know of only one study that has investigated goal-setting in an online environment [23]. This study used an Internet search task to show that subjects with goals worked longer and produced more search results than those without goals, though with lower search accuracy. These findings however, do not necessarily generalize to contribution in online communities where task outputs are in the form of individual contributions useful to other

community members. We know of no research to examine the impact of goals on contributions to online communities.

High-challenge goals energize higher performance in at least three ways [22, 24]. First, self-efficacy [25], or a person's belief in his or her own abilities to successfully perform a task, positively moderates commitment to a goal. Persons high in self-efficacy are more likely to set or accept specific high-challenge goals, provided such goals are perceived as achievable and reasonable. Second, goals provide "normative information to the individual by suggesting what level of performance the individual could be expected to attain"[24]. This normative information operates in a feedback loop in which individuals regulate their task effort so that their performance does not deviate too much from the expected level. Third, achieving the goal leads to task satisfaction. This satisfaction enhances self-efficacy and future goal commitment, which in turn boost future task performance. This upward performance spiral leads Locke and Latham to refer to their goal-setting theory as the High Performance Cycle [22].

The goal-setting theory specifies the setting that stimulates high performance [22]. Goals must be relatively difficult, specific, context-appropriate, and immediate rather than long-term. Mediating variables include individual commitment, importance assigned to the goal, and self-efficacy. The goal effects are stronger when people have specific feedback concerning performance against the goal and weaker with more complex tasks.

Contributing rating of movies to MovieLens satisfies these conditions. Contributions are voluntary (implying high task commitment), low in task complexity, and available with immediate feedback (e.g., the user interface shows ratings immediately after they are entered).

> Hypothesis 4: In an online community, specific, numeric goals will motivate greater contributions than non-specific goals.

Group Goals. Most goal-setting research has focused on individual level goals. Generally, the same motivating effects of high-challenge goals have been shown at the group level (see [26] for a review). In some small group settings where individual and group interests are aligned, group goals have been associated with higher performance than individual goals [27, 28]. However, consistent with the social loafing effect, the performance benefits of group goals tend to disappear and even reverse as the size of the group increases beyond 6-8 members [28]. Given that group size in our experiment was set at 10 members, we should expect that members assigned group goals in general will contribute less than members with comparable individual goals.

> Hypothesis 5: Members assigned individual goals will provide more contributions than members assigned group goals.

Limits of high-challenge goals. Despite the robust findings of goal-setting theory, little research has looked at the limits of high-challenge goals. Locke and Latham assert that performance will increase monotonically with goal difficulty, until "subjects reach the limits of their ability at high goal difficulty levels; in such cases the function levels off" [24]. If this is true, then no goal can be too high, since performance will plateau but not drop as goals become increasingly difficult. If this is not true, goals that are too high may result in a performance drop, not just a plateau. There has been insufficient previous research to definitively resolve this question, although White et. al. found high-challenge goals above the 93rd percentile in performance resulted in lower performance than "do your best" goals when subjects did not believe they would be evaluated by the experimenter [29].

In online communities where performance (in terms of individual contribution level) varies greatly, it is important to know if difficult goals boost contribution and if overly difficult goals sub-optimize contribution. If online community members view difficult goals as unattainable, member contribution could decline for several reasons. First, overly difficult goals could undermine a member's commitment to the goal, diminishing or eliminating engagement with the task of contributing to the online community. Second, overly difficult goals could undermine self-efficacy by providing a performance standard viewed as beyond reach. Third, overly difficult goals could reduce the attractiveness of community membership if that community is perceived as making unreasonable demands. These factors lead to our third hypothesis.

> Hypothesis 6: In an online community, contribution will drop off when goals exceed some difficulty threshold.

3.1 Methods

Overview. We again conducted our experiment in MovieLens.org. Members were recruited by sending them an email message inviting them to participate in a movie rating campaign to enhance the ratings prediction in MovieLens. As in study 1, we used different messages in this invitation email to vary participants' perceptions of their group assignment and goal challenge. We then tracked the rating behavior of participants who were sent invitations.

Subjects. The subject population consisted of MovieLens members who had logged in at least once in the five month period between July and December 2003. On average, these subjects had rated 147 total movies (standard deviation=124). In the five months prior to the experiment, members had rated an average of 8 movies per week (standard deviation=2). Of the 900 members that we contacted, 66 emails bounced, leaving us with 834 subjects.

Experimental manipulations. We manipulated two variables: group assignment (2 conditions: group or individual), and specificity of goals (2 conditions: "do your best" or a specific numeric goal). Within the specificity of goal condition, we varied the difficulty of the goal (4-levels) as displayed in Table 3.

All subjects received an invitation email saying, "The quality of recommendations MovieLens makes depends on the number of ratings that members contribute. Currently, many of the movies in MovieLens have too few ratings to make accurate recommendations about them. That's why we're conducting a seven day campaign to increase movie ratings on MovieLens."

Group assignment. For the group assignment manipulation, some participants were told they belonged to a group of 10 active MovieLens members called the "Explorers." Group size was set at 10 members for two reasons. First, a group size of 10 minimized the amount of cognitive effort required for group condition subjects to comprehend their fair share of the group goal, since most people can mentally divide by 10. We did not want to explicitly state that "a group goal of 'x' ratings translates into 'y' ratings per member," because doing so would confound the group goal with the individual goal. Second, because past research suggested that group goals are less effective than individual goals above group sizes of 8, we wanted to determine if these findings would be replicated by setting group size above 8.

Table 3: Study 2. Mean Number of Movie Ratings

Goal	Group	Do	80	160	320	640	Total
	Individual	your best	8	16	32	64	
Group	N	86	85	84	83	85	423
	P(logged in)	26.7	35.3	35.7	30.1	29.4	31.4
	# Ratings	14.9	16	19	20	16	17.2
Individual	N	78	85	83	83	81	410
	P(logged in)	23.1	27.1	30.1	32.5	28.4	28.3
	# Ratings	5	6.7	12	16	12	10.4
Total	N	164	170	167	166	166	833
	P(logged in)	25.0	31.2	32.9	31.3	28.9	29.9
	# Ratings	10.3	12	16	18	14	13.9

Note. N is the number of subjects to whom email was successfully delivered. P(logged in) is the percentage of participants who logged into MovieLens during the week of the experiment. # Rating is the mean number of ratings for the N subjects.

Subjects in the individual task condition were not told about group membership.

Goal specificity. For the goal specificity manipulation, subjects in the non-specific goals condition were told to "do your best" to rate the movies. Specifically, they were told, "[Together, the ten Explorers]/[You] have a goal of doing [their]/[your] best to rate additional movies over the next seven days. You can begin rating now by clicking http://www.movielens.org." Subjects in the specific goals condition were given a specific number of movies to rate in a week. Group goals, for the 10-person groups, were set at ten times the individual goals. We asked participants in the individual conditions to either rate 8, 16, 32 or 64 movies in the next week. In the group condition, we told them that the explorer's task was to rate either 80, 160, 320 or 640 movies during the week.

We chose 8 ratings per week as the baseline "low challenge" goal, because it was approximately the mean weekly contribution in the past. We selected a one-week period to collect data because goal-setting theory specifies that goals are more effective if they are of shorter duration than longer duration.

Measuring contributions. For the week following the invitation email, we tracked user rating activities. After one week, we tabulated total ratings and then sent a thank-you email reporting the rating behavior of all participants and summarizing their individual and group (if applicable) rating behavior.

3.2 Analysis and Results

We excluded one outlier in the individual, do-your-best condition, who rated 601 movies (>7 standard deviations above mean ratings). This left 833 subjects, of whom 30% (249) logged in at least once. Table 3 summarizes the rating contributions for these 833 subjects.

As the distribution of ratings was skewed, the dependent variable in our analyses was the log of the number of ratings. Because the log of zero is undefined, we treated participants who rated no movies as if they had rated one. Hypothesis 4 tested whether people contribute more when given specific numeric goals versus non-specific do-your-best ones. This was confirmed. A multiple regression analysis comparing ratings of participants who received the do-your-best email to those who received a specific

numeric goal (all specific goal conditions combined) showed that those with specific goals made more contributions than those with do-your-best goals. (mean do-your-best = 10.3, mean specific=13.9 F(1,832) =5.95, p<.02).

Hypothesis 5, that participants assigned individual goals would contribute more than those with group goals, was disconfirmed. Participants with group goals rated more movies than those with comparable individual goals. (group mean=17.2, individual mean=10.4, F(1,831)=2.99, p=.08).

Figure 4: Effects of goals on number of ratings
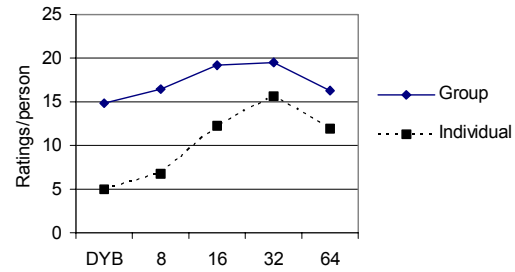
Hypothesis 6 was that the relationship between goal specificity and performance is curvilinear, with the highest challenge goals leading to declines in ratings. As shown in Figure 4, ratings appeared to drop with the highest goals, especially in the individual goal condition. To test for statistical significance, we excluded the do-your-best subjects. We created orthogonal linear and quadratic contrasts for the 4-levels of assigned goals (linear contrasts = -1.5, -.5, .5, 1.5; quadratic contrasts = 2.25, .25, .25. 2.25). We then ran regression models with the contrasts (representing the goals and goals squared) as predictors. Although a curvilinear model fit the data better than a linear one, neither the linear nor the quadratic coefficient was significant. We also reran the model including only participants who rated movies. Here the quadratic coefficient were significant (F(1,95))=4.16, p=.05). This is weak support for the hypothesis that extremely difficult goals would reduce contributions. A summary of the hypothesis testing results is shown in Table 4.

Table 4: Study 2. Results Summary

Hypothesis	Result
H4: Specific goals more effective than non-specific goals	Supported
H5: Individual goals more effective than group goals	Disconfirmed (reverse of hypothesis at p<.10 level)
H6: Difficulty of goal has convex effect on contributions	Weak support (only for participants in the individual goals condition who rated)

3.3 Discussion

The most robust result from this experiment was that specific goals led to higher contribution rates than non-specific ones. This is the first study to document that this finding from goal-setting theory applies to contributions to an online community and should encourage designers to be more specific about assigning goals or providing opportunities for individuals to declare contribution goals for themselves.

A finding inconsistent with social loafing and the collective effort model was that subjects in the group conditions contributed more than those in the individual conditions, although this difference was only marginally significant. There are several possible explanations for this reversal. First, some in the group condition may have misinterpreted group goals as an individual goal, despite careful wording to reduce this risk. For example, they may have believed that they individually rather than their group were being asked to rate 320 movies. However, the high contribution even in the group do-your-best condition makes this an unlikely explanation. Second, social loafing is more difficult to replicate among people committed to their groups than to among those less committed or in field than in lab research [13], and our study was a field experiment among volunteers. Although commitment and field conditions often diminish the social loafing effect, they should not reverse it, with individuals in a group setting working harder than those in an individual setting. Finally, it is possible that the invitation email, which identified subscribers by their email address, and the feedback they were promised, may have reversed the effects of social loafing. Overall, it appears that being assigned to a group provided greater impetus to contribute to the community than being a solo contributor, a puzzle that we will return to in the conclusion of the paper. This result is consistent with some research contrasting individual and group goals [26, 27].

This study also is among the first to suggest high performance goals have upper limits, and beyond those limits, performance may drop, not just plateau as previously supposed [24]. The experiment here was a weak test of the limits of goals, since even the highest level of individual challenge (64 movies in a week) was probably perceived as attainable, since 45% of the subjects had rated more than 64 movies in a single day at least once previously. We suspect that a less attainable goal might have more clearly demonstrated a reduced motivational effect.

This study provides a number of insights that contribute to solving the problem of under-contribution in online community. First, specific, challenging goals have been shown to be powerful motivators of online contributions, particularly when contributors are not part of a group. The fact that this was shown using a simple email manipulation, without interface modifications, suggests specific challenging goals can have a strong effect on increasing individual contribution in an online community. Second, assignment to a group condition in the context of a large, anonymous online community seemed to raise contribution levels, even though it was a group in name only, with members neither knowing the identities of other members nor interacting with them. Integrating both these findings with usability design principles should provide an even greater performance boost. For example, providing an interface that facilitates elements of the Collective Effort Model and Goal-setting Theory such as real time feedback, member identifiability, group cohesion, and so forth, could enhance the effects observed in this field experiment. Some of these elements may already be included in interfaces; integrating them in the broader context of these findings should serve to increase their effectiveness. Finally, this study suggests that goals that are overly difficult may result in reduced contributions. This is more relevant for online community managers than designers, but future research could focus on algorithms that provide individuals and groups with goals that optimize contribution.

4. Conclusion

This paper attempted to use principles from social psychology theory to redesign an online community to increase contributions. We now reflect on the larger lessons about the theories available to mine in social psychology, why they are an under-utilized resource for design, and some of the difficulties we had in applying them.

4.1 The success in applying social science theory to design

Our attempt to drive design from theory was successful in the sense that the theories led to design innovations that are rarely seen in existing online communities. One key insight from the collective effort model is that people will be more likely to contribute to a group task if they think their contribution does not duplicate what others can provide and is thus needed for accomplishing the group's goal. Many online communities provide feedback on the number or assessed quality of their contributions, like the "top reviewer" designations given to some contributors on the www.epinions.com website. However, we know of no online community that provides feedback to contributors about the uniqueness of their contributions. Similarly, the key insight from Locke's theory of goal-setting is that people work hard to achieve specific, challenging goals, but online communities rarely provide potential contributors with specific, challenging goals to reach. On-air fundraising drives for public television and radio do ("We need $500 in the next hour to meet a donor's matching grant"), but this technique is rarely used in online communities, whether they are soliciting conversation or more quantifiable contributions.

Our attempt to drive design from theory was also successful in that applying some of these design principles led to increased contributions. A simple email message making salient the uniqueness of potential contributions caused recipients to rate more movies than a comparable message that emphasized commonality. Without the collective effort model, it would not be obvious whether emphasizing uniqueness or commonality would be more effective. A simple email message assigning recipients specific ratings goals led recipients to rate more movies than a comparable message just urging them to rate more. Without the prior research on goal-setting, it would not be obvious whether a specific goal would be helpful, since it might even discourage contributions above the goal.

4.2 Failures of implementation and theory

However, not all the design ideas derived from the theories led to increased contributions. Results from Experiment 2 were inconsistent with a fundamental prediction from the collective effort model, that people would exert less effort when they believed their output would be pooled rather than being individually identified. Although the collective effort model stresses that people are more motivated to contribute when they believe their contributions will have benefit, in Experiment 1 making salient *either* the benefit that the contributors themselves would receive from their ratings *or* the benefit that others would receive depressed their contributions. On the other hand, reminding them of both their own and other benefits together was better than mentioning either one alone.

Why did the design choices inspired by social psychology theories sometime fail to increase contribution? Here we consider three classes of explanation.

Failures of implementation. At one extreme, we may have started with correct and applicable theory, but our implementation failed

to capture the design principles appropriately. In the present experiments, the design principles extracted from theory were implemented as short, single electronic mail messages. They may have been poorly worded. For example, in Experiment the email manipulations combined a rationale along with the assertion that ratings have benefit. One possible follow-on experiment would test the impact of pure assertions ("Rating more movies helps the MovieLens community!") versus explanations ("The more ratings you provide, the more information we have about each movie and the easier it is to make accurate recommendations for other people.") Alternately, the email manipulations may simply have been too weak. In MovieLens it would be possible to build a reminder of uniqueness or benefit directly into the system, for example by calculating the number of people who would be influenced by a potential rating whenever a particular movie is presented. By presenting an index of benefit simultaneously with the opportunity to make a contribution and tailoring the index to a particular contribution, this implementation may more powerfully influence contribution. By testing with a lower-fidelity intervention (i.e., email messages), we may have underestimated the effects that would come with stronger manipulations.

Mismatches between engineering and scientific disciplines. The failures, however, may reflect a deeper mismatch of goals and values of HCI and CSCW research with those of social psychology. HCI and CSCW are primarily engineering disciplines, where the primary goal is problem-solving. In contrast, social psychology views itself as a behavioral science, whose goal is to uniquely determine the causes for social phenomena.

We treat only one consequence of this difference in orientation here. The collective effort model asserts that people will contribute less when they think their contributions are pooled with those of a group, but that the nature of the group and the individual's relationship to it make a difference. A behavior scientist would want to distinguish the effects of commitment to a group from the effects of pooling contribution. To do so, the behavioral scientist would carefully contrast groups that differ on only a single factor. In contrast, a designer would be more interested in constructing the groups to maximize contribution. To do so, the designer would typically incorporate many factors to shape the group. For example, in the second experiment, when assigning subjects to the group condition, wearing our designers' hat, we gave the group a name with a positive valence ("Explorers") rather than a neutral name "Group A" or no name at all. But the name may have induced some sense of group commitment that overpowered the social loafing effect and thus made it difficult to test the hypotheses involving main effects or interaction effects predicted by social loafing theory.

Incomplete theories. In some cases, social science theories may simply not be up to the task when multiple features vary simultaneously, as they do in tests of real designs. Although designers know that the importance of design elements depends upon context, e.g., whether a system is used by young or older adults, the norm in much social psychological research is to abstract these contextual details away. The goal is to have a theory that is as general as possible.

Social psychology theories may therefore be incomplete because they identify isolated main effects but not interaction effects or the effects of context. Similarly, they typically do not provide any way to say which effect will predominate when two main effects have opposite predictions. For example, assigning people to the "Explorers" group may generate a sense of group identity, which

makes people care more about the collective outcome and thus contribute more. But it also reduces the importance of any individual's contribution, reducing motivation. The theory is silent about what the net effect will be in a particular situation.

This lack of detail forces the designer to improvise when attempting to apply social psychological knowledge to solve design problems. While social science theory can inspire design, by suggesting options, it does not eliminate the need for it.

4.3 The Way Forward
Despite the problems identified above, we believe that mining social science theory as a source of principles for design innovation is a useful general strategy for the design of CSCW systems (see [30] for a fuller discussion). Although we focused our effects in this paper on applying two social psychological theories to the problems of under-contribution to online communities, the approach is far more general. Since the turn of the 20th century [31] and especially since World War II, the field of social psychology has developed a rich theoretical base for understanding and predicting group behavior. However, unlike theories in cognitive psychology, this theoretical base has been inadequately mined in the HCI and CSCW literatures.

Social psychological theories relevant to encouraging contributions to groups include, for example, those on group cohesion, group identity, interpersonal attraction, and altruism, among others. Recent handbooks (e.g., [32-34]) and graduate-level textbooks provide useful resources.

The approach we advocate here will not always be easy to follow. But if we hope to create scientifically informed processes and guidelines for CSCW designers to follow, more work is needed that maps behavioral theories to hypotheses that can be tested in lab and field experiments.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Butler, B., When is a group not a group: An empirical examination of metaphors for online social structure, in Social and Decision Sciences. 1999, Carnegie Mellon University: Pittsburgh, PA.

[2] Adar, E. and B.A. Huberman, Free riding on Gnutella. First Monday, 2000. 5(10): p. NP.

[3] Lakhani, K.R. and E.V. Hippel, How open source software works: "Free" user to user assistance. Research Policy, 2003. 32: p. 923-943.

[4] Mockus, A., R.T. Fielding, and H. Andersen, Two case studies of open source software development: Apache and Mozilla. ACM Transactions on Software Engineering and Methodology, 2002. 11(3): p. 309-346.

[5] Nonnecke, B. and J. Preece, Lurker demographics: Counting the silent, in Proceedings of CHI'2000: Human Factors in Computing Systems. 2000: Hague, The Netherlands. p. 73-80.

[6] Cosley, D., et al., Is Seeing Believing? How Recommender Systems Influence Users' Opinions, in Proceedings of CHI

2003: Human Factors in Computing Systems. 2003: Fort Lauderdale, FL. p. 585-592.

[7] Markus, L., Towards a "critical mass" theory of interactive media: Universal access, interdependence, and diffusion. Communication Research, 1987. 14: p. 491-511.

[8] Grudin, J., Why groupware applications fail: Problems in design and evaluation. Office: Technology and People, 1989. 4(3): p. 245-264.

[9] Dourish, P. and S. Bly, Portholes: Supporting awareness in a distributed work group, in Proceedings of CHI92: Human Factors in Computing Systems. 1992, ACM: New York. p. 541-547.

[10] Erickson, T. and W. Kellog, Social translucence: An approach to designing systems that support social processes. Transactions on Computer-Human Interaction, 2000. 7(1): p. 59-83.

[11] Preece, J., Online Communities. 2000, New York, NY: John Wiley & Sons.

[12] Viegas, F.B. and J.S.Donath, Chat circles, in Proceedings of CHI'99: Conference on Human Factors in Computing Systems. 1999, ACM Press: New York. p. 9-16.

[13] Karau, S. and K. Williams, Social loafing: A meta-analytic review and theoretical integration. Journal of Personality and Social Psychology, 1993. 65(4): p. 681-706.

[14] Ledyard, J., Public goods: A survey of experimental research, in The Handbook of Experimental Economics, J.H. Kagel and A. Roth, Editors. 1995, Princeton University Press: Princeton, NJ. p. pp. 111-194.

[15] Ostrom, E., Governing the Commons: The Evolution of Institutions for Collective Action. 1990, Cambridge: Cambridge University Press.

[16] Kollock, P. and M. Smith, Managing the virtual commons: Cooperation and conflict in computer communities, in Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives, S.C. Herring, Editor. 1996, John Benjamin: Amsterdam. p. pp. 109-128.

[17] Harkins, S.G. and R.E. Petty, Effects of task difficulty and task uniqueness on social loafing. Journal of Personality and Social Psychology, 1982. 43: p. 1214-1229.

[18] Kerr, N.L., Motivation losses in small groups: A social dilemma analysis. Journal of Personality and Social Psychology, 1983. 45: p. 819-828.

[19] Kerr, N.L. and S. Bruun, The dispensability of member effort and group motivation losses: Free rider effects. Journal of Personality and Social Psychology, 1983. 44: p. 78-94.

[20] Markey, P.M., Bystander intervention in computer mediated communication. Computers in Human Behavior, 2000. 16(2): p. 183-188.

[21] Shepperd, J.A. and K.M. Taylor, Social loafing and expectancy-value theory. Personality and Social Psychology Bulletin, 1999. 25(9): p. 1147-1158.

[22] Locke, E.A. and G.P. Latham, Building a practically useful theory of goal setting and task motivation: A 35 year odyssey. American Psychologist, 2002. 57(9): p. 705-717.

[23] Thompson, L.F., J.P. Meriac, and J. Cope, Motivating online performance: the influences of goal setting and Internet self-efficacy. Social Science Computer Review, 2002. 20(2): p. 149-160.

[24] Locke, E.A. and G.P. Latham, A Theory of Goal Setting and Task Performance. 1990, Englewood Cliffs, NJ: Prentice-Hall.

[25] Bandura, A., Perceived self-efficacy in cognitive development and functioning. Educational Psychologist, 1993. 28(2): p. 117-148.

[26] Weldon, E. and L.R. Weingard, Group goals and group performance. British Journal of Social Psychology, 1993. 32: p. 307-334.

[27] Matsui, T., T. Kakuyama, and M. Onglatco, Effects of goals and feedback on performance in groups. Journal of Applied Psychology, 1987. 72(3): p. 407-415.

[28] Streit, M.C., The effects of group size on goal-setting behavior, task performance, goal commitment, and evaluation potential, in Department of Psychology. 1996, Hofstra University: Long Island, NY.

[29] White, P.H., M.M. Kjelgaard, and G. Harkins, Testing the contribution of self-evaluation to goal setting effects. Journal of Personality and Social Psychology, 1995. 69(1): p. 69-79.

[30] Kraut, R., Applying social psychological theory to the problems of group work, in HCI Models, Theories and Frameworks: Toward A Multidisciplinary Science, J.M. Carroll, Editor. 2003, Morgan Kaufman: New York. p. 325-356.

[31] Ross, E.A., Social Psychology: An Outline and Source Book. 1908, New York: The Macmillan Company.

[32] Brown, R. and S.L. Gaertner, Blackwell Handbook of Social Psychology: Intergroup Processes. 2001, Oxford, UK: Blackwell.

[33] Gilbert, D.T., S.T. Fiske, and G. Lindzey, The Handbook of Social Psychology. 1998, Boston: McGraw-Hill.

[34] Higgins, E.T. and A.W. Kruglanski, Social psychology: Handbook of Basic Principles. 1996, New York: Guilford.

[35] Deci, E.L., R. Koestner, and R.M. Ryan, A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin, 1999. 125(6): p. 627-66