

Using Social Psychology to Motivate Contributions to Online Communities

Kimberly Ling¹, Gerard Beenen¹, Pamela J. Ludford³, Xiaoqing Wang⁴, Klarissa Chang¹,
Xin Li², Dan Cosley³, Dan Frankowski³, Loren Terveen³
Al Mamunur Rashid³, Paul Resnick², Robert E. Kraut¹
CommunityLab*

¹Carnegie Mellon University
5000 Forbes Ave
Pittsburgh PA 15221
{kling, gbeenen, changtt, kraut}
@andrew.cmu.edu

²University of Michigan
3210 SI North
Ann Arbor, MI 48109
{xinl, presnick}
@umich.edu

³University of Minnesota
200 Union Street SE
Minneapolis, MN 55455
{ludford, arashid, dfrankow,
terveen}@cs.umn.edu

⁴University 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 lead to mid-level design goals to address this problem. We tested design principles derived from these theories in four field experiments. In experiment 1, members of an online movie recommender community participated in an online discussion group either with others with similar movie tastes or not and were told of conversational topics where they held unique information or not. In experiment 2, members of the movie recommender community were reminded of the uniqueness of their contributions and the benefits that follow from them. In experiment 3, they were reminded of different types of benefits that they or others would gain from rating movies. In experiment 4, 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. However, other predictions were disconfirmed. For example, in Experiment 4, participants given group goals contributed more than those given individual goals. The paper ends with suggestions and challenges for mining design implications from social science theories.

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, 26 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. 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 (Butler, 1999). For example, on the popular peer-to-peer music sharing service, Gnutella, ten percent provide 87% of users of all the music (Adar & Huberman, 2000) and in open-source development communities, four percent of members account for 50% of answers on a user-to-user help site (Lakhani & Hippel, 2003), and four percent of developers contribute 88% of new code and 66% of code fixes (Mockus, Fielding, & Andersen, 2002). Although not everyone needs to contribute for a group to be successful (Nonnecke & Preece, 2000), groups with a large proportion of non-contributors have difficulty providing needed services to members. 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 address the problem of under-contribution in an online community called MovieLens (Cosley, Lam, Albert, Konstan, & Riedl, 2003). MovieLens is a web-based movie recommender community where members 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. We describe interventions designed to boost contributions both to conversations about movies and ratings of them.

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 (e.g., Grudin, 1989; Markus, 1987). Other CSCW researchers have drawn design inspiration and design guidelines from social science findings (e.g., Dourish & Bly, 1992; Erickson & Kellog, 2000; Preece, 2000).

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. Economists and political scientists have observed that 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 Ledyard, 1995 for a review). Social psychologists have identified an analogous phenomenon called *social loafing*. People exert less effort on a collective task than they do on a comparable individual task (see Karau & Williams, 1993 for a review). Karau and Williams developed the collective-effort model to explain social loafing (Karau & Williams, 2001). In this paper, we describe several attempts to mine the collective effort model and other theories from social psychology on the motivators of individual effort. We apply these principles to the design of appeals for soliciting contributions to MovieLens.

In each of four 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 others rate or that their ratings help others. Next, we designed persuasive messages or other manipulations that we hoped would induce these mental states. We then conducted controlled randomized experiments by exposing existing MovieLens subscribers to these different manipulations. Three of the four experiments have been reported previously (Beenen et al., 2004; Ludford, Cosley, Frankowski, & Terveen, 2004). This paper analyzes more comprehensively and rigorously the impact of the different manipulations on contribution behavior, especially movie ratings.

2. Experiment 1: Motivating conversational contributions through group homogeneity and individual uniqueness

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 think they are working by themselves. Karau and Williams' collective-effort model, is a type of utility theory that claims that people work hard when they think their effort will help them achieve outcomes they value (Karau & Williams, 1993). 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 (Harkins & Petty, 1982; Kerr, 1983; Kerr & Bruun, 1983). (See Karau and Williams (2001) and Figure 1 for a fuller description of the collective effort model.)

The collective effort model identifies conditions under which people will socially loaf less. These include, among others, (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. Social psychologists have tested the collective effort model by manipulating individuals' perceptions of the elements in

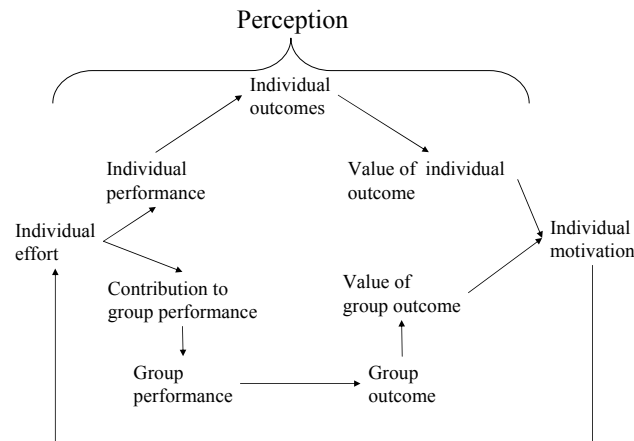
Figure 1: The collective-effort model (adapted from Karau and Williams, 1993)

Figure 1, both in laboratory settings (e.g., Harkins & Petty, 1982; Kerr, 1983; Kerr & Bruun, 1983) and online (Markey, 2000).

We attempted to apply the insights from the collective effort model to the problem of under-contribution in MovieLens. MovieLens is a movie recommender system that recommends to subscribers movies that they would enjoy, based on movie evaluations from other subscribers with similar tastes. In Experiment 1, we expanded MovieLens' functionality by adding new online discussion groups. The goal of the experiment was to identify ways of organizing the discussion groups so that people would offer conversational posts in the group and subsequently rate more movies in MovieLens. In experiment 1 we tested predictions from the collective effort model that people will contribute more to a group when they think their contributions are likely to be unique and when they like the group more. More details about this experiment are available in Ludford et al. (2004).

Uniqueness of contribution. The collective effort model posits that people will socially loaf less and contribute to a group more, the more they see their contribution as important to the group (Karau & Williams, 1993). If they believe that their contributions are redundant with those that others in the group can provide, then there is little reason to contribute because their contributions have little likelihood of influencing the group. Conversely, if they think they are unique, they should be more motivated to contribute, because their contributions are more likely to influence the group. In Experiment 1, we manipulated subjects' perceived uniqueness by reminding them of movies they had seen but that others in the group had not.

Hypothesis 1: People will contribute more to online communities when given personalized information showing their contributions would be unique.

Similarity/homogeneity of the group. The collective effort model posits that people will socially loaf less and contribute more to a group the more they like it (Karau & Williams,

1993). By doing so, they increase their own utility by benefiting the group. In contrast, they do not receive the same benefit if they contribute to groups they dislike. Social psychologists have identified many bases for members' attachment to a group, including their liking for individual members. People tend to like others who are similar to themselves (Byrne & Griffith, 1973; Byrne, 1997) and to dislike groups composed of dissimilar members (Williams & O'Reilly, 1998). In Experiment 1, we manipulated subjects' liking for their discussion group by populating the group with others who had either similar or dissimilar tastes in movies.

Hypothesis 2: People will contribute more to online communities when they believe they are similar rather than dissimilar to others in the group.

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. We recruited subjects by email, inviting MovieLens users who had rated at least 50 movies to participate in conversations about movies with other MovieLens members. Of the nearly 8500 invitations sent, approximately 2800 bounced. Two hundred and forty-five people volunteered to participate.

Subjects participated in online movie discussions, implemented as asynchronous message boards. To jump-start the conversations, we provided an incentive, entering subjects who posted at least one message during four of the five weeks of the study for a drawing for one of five gift certificates.

The experiment ran for five weeks, from the end of July to the start of September 2003. Subjects could talk about whatever they wanted in the forums. In addition, the research team posted one discussion topic each week. These questions kept the forums from being empty when the experiment began and assured new content appeared regularly. The five weekly topics were:

1. What is a little-known film you have seen that you'd recommend to other people?
2. What acting performance was worthy of an Oscar, but did not win?
3. Was acting better in the 1950s (and earlier) than it is today?
4. Take a look at the films MovieLens recommends for you. There's probably at least one on the list you're unfamiliar with or are not sure you want to see. Discuss a recommendation you're unsure about with the group.
5. In your opinion, what makes a love story 'click' in a movie?

Subjects were randomly assigned to one of 8 experimental groups, arranged in a two (*Uniqueness*) by two (*Similarity*) factorial design.

Uniqueness. Subjects in the four uniqueness groups received a weekly message telling them how their MovieLens ratings differed from others in their group vis-à-vis the discussion topic for the week. For example, for the discussion topic for week one, on a little-known movie, we identified movies that each participant had rated favorably but that few others in the subject's group had rated. To insure that these movies were less well-known, we selected ones that fewer than 1000 MovieLens users had rated. Under these conditions, subjects' information and opinions about the movie were likely to be unique, not duplicated by other group members. We did not explicitly tell subjects to mention their uniqueness information when they posted. Instead, we simply explained that they might find the information relevant to the discussion topic; many of them did.

Subjects in the four *non-unique* groups were merely sent an email identifying the topic of the week, but did not get the name of a movie that only they had seen.

Similarity. While the uniqueness condition was implemented via weekly email, the similarity manipulation was implemented via group composition, which was fixed at its creation. We constructed the four *similar* groups so that on average pairs of people had seen many movies in common and agreed in their evaluations of them, while in the four *dissimilar* groups subjects saw many movies in common and disagreed in their evaluations. Pairs of participants agreed in their evaluation of a movie if their evaluations were on the same side of a threshold of 6 on a 10-point Likert scale of liking. They disagreed if one rated the movie less than six and the other rated it six or greater. The co-agreement similarity score between pairs of users is the number of movies on which they agreed. Our algorithm builds similar (dissimilar) groups by starting with the most (least) similar pair of users, then adding the user that resulted in the highest (lowest) average pair-wise similarity among group members. The algorithm adds users to a group until it has the desired number of members. After enough groups are formed, it improves the results by swapping users between groups as long as the total difference between similar and dissimilar groups increases. We imposed one additional constraint, forcing the algorithm to assign subjects to groups so that the distribution of number of ratings among members in each group was roughly the same.

2.2 Results

Preliminary results. On average, pairs of participants rated approximately the same number of movies in common in the similar and dissimilar groups, but agreed upon them twice as often in the similar groups than in the dissimilar one. As a manipulation check for the similarity manipulation, we had each subject rate every other subject who posted at least once in their group on 4-point Likert scales to indicate how similar to themselves they judged the other and how much they disagreed with the other. We analyzed the data with a mixed model regression, with the rater as a random factor. Subjects in the similar group perceived themselves as having views more similar to others in the group than did those in the dissimilar group ($\bar{X}_{similar} = 2.72$, $\bar{X}_{dissimilar} = 2.47$, $t(1,853) = 4.97$, $p < .001$).

Surprisingly, subjects in the uniqueness condition also perceived themselves as having views more similar to others in the group than did those in the non-unique group

($\bar{X}_{similar} = 2.72$, $\bar{X}_{dissimilar} = 2.40$, $t(1,853) = 2.37$, $p < .02$).

The 230 subjects in 8 forums posted a total of 1473 messages over the course of the study. 163 subjects posted at least one message. Posting followed an inverse power law, with 9% of subjects accounting for 50% of all posts. Because of the skewed distribution, before analysis we transformed the data, taking the log to the base 10 of the number of posts, adding one because the log of zero is undefined.

We tested our hypotheses concerning participation by analyzing the quantity of posts in the various conditions and the results of a post-experiment survey, using the gllamm module in Stata to compute a mixed model regression, with group treated as a random effect. Table 1 shows the major results.

Table 1: Experiment 1. Mean Number of Movie Ratings

		Dissimilar	Similar	Total
Non-unique	N	58	57	115
	Mean posts	2.93	1.99	2.43
	Mean ratings	5.61	3.64	4.54
Unique	N	58	57	115
	Mean posts	5.53	3.18	4.24
	Mean ratings	11.74	8.98	10.29
Total	N	116	114	230
	Mean posts	4.06	2.53	3.24
	Mean ratings	8.17	5.80	6.91

Note. *N* is the number of subjects in each condition. Statistics for Posts and Ratings were calculated in the log scale and then transformed back for ease of understanding.

Consistent with Hypothesis 1, subjects posted more messages in the uniqueness condition, when they were given personalized information about how their knowledge of movies differed from others (See table 1: $z=2.88$, $p<.004$). However, Hypothesis 2 was disconfirmed. Subjects posted fewer messages when conversing in groups constructed so that members had similar tastes in movies than in groups with heterogeneous members ($z=-2.45$, $p<.05$). There was no similarity by uniqueness interaction ($z=-.59$, $p > .50$).

In addition to posting more messages in the unique condition than in the non-unique condition, they also rated more movies (See Table 1). In order to test for statistical significance of the differences in rating between the different experimental conditions, we must account for the differences in probability of logging in to the rating site when computing average levels of rating. Ordinary least squares regression using only the subjects who contributed (or using all the subjects) would yield biased estimates of the number of movies rated per participant who did log in. Instead, we used a Heckman regression model (Heckman, 1979; StataCorp, 2003) to predict numbers of movies rated and number of less well-known movies rated, while controlling for biases associated with missing data.

The Heckman method involves estimating two equations: (1) a maximum likelihood estimate (MLE) probit selection equation that models whether participants logged in, and (2) an ordinary least squares (OLS) regression equation that models the level of contribution in terms of the number of movie ratings, while controlling for whether participants logged in. Although we do not observe the rating behavior of participants who do not log in during the experiment, we do have information about their rating history prior to the experiment which we can use in the first step of Heckman's method to model the likelihood of participation. This procedure successfully accounts for unobservable factors and is widely adopted by economists and other researchers in the social sciences (Heckman et al., 1996). Because subjects within a single group are not independent of each other, we computed the Heckman analysis, clustering by group.

To model whether participants logged in, we used the number of weeks in the past year where they logged in and made at least one rating and the number of weeks since they last logged in as control variables, along with the experimental variables (dummy variables for uniqueness, similarity, and their interaction). The second stage of the Heckman model predicts numbers of ratings from the experimental manipulations for those who logged in at least once during the experimental period, using ordinary least squares regression but including a covariate that is calculated from the first stage. An additional covariate not included in the first-stage selection model was the mean number of ratings per week for the subject over the last year, counting only weeks in which the subject rated at least one movie. Because the number of ratings is skewed, with a few subjects generating many ratings, we used the log transform of the number of ratings as the dependent variable.

The first stage of the Heckman model showed that participants were more likely to log in the more frequently they had logged in previously ($z=2.19$, $p<.03$), and the fewer weeks had elapsed since their last login ($z=-5.91$, $p>.001$). Participants in the unique condition were more likely to log in than those in the non-unique condition. ($z=3.70$, $p>.001$). Participants with homogeneous groups were less likely to log in than those in the non-similar condition. ($z=-3.17$, $p>.002$).

The second-stage model predicts the number of ratings that participants made, controlling for their likelihood of logging in. Participants made more ratings in the unique condition than in the non-unique condition ($z=2.32$, $p<.02$). However, neither the similarity manipulation ($z=-.4$, $p>.50$) nor the similarity by uniqueness interaction ($z=-.88$, $p > .30$) were significant.

2.3 Discussion.

Both posting and rating data show that people contributed more when they were made to see themselves as having unique information to contribute. In retrospect, the finding that subjects posted more to the conversation forum when they were least similar to those they were talking to may also reflect the influence of uniqueness. Subjects in groups with similar others may have run out of topics of conversation, while those in the heterogeneous groups could have lively disagreements. Both quantitative data, in which subjects in the dissimilar groups rated disagreeing with other participants more highly

than did those in the similar groups, and examination of the transcripts are consistent with this interpretation. For example, a typical comment from a subject in the dissimilar group comparing acting in the 1950s and today was, “I’m not sure what you are thinking of here because I can hardly agree with you. I would take the exact opposite view on most of your points and will explain why....” In contrast, comments in the similar group were “I have to agree with the general consensus that today’s acting is no worse than that in the 50s...” and often included less follow-up. Alternatively, subjects in groups composed of people with similar tastes may never have occasion to discover the issues on which they differ. People tend to talk about the topics on which they hold similar positions when left to their own devices (Stasser & Titus, 1985).

3. Experiment 2: Motivating contributions through framing uniqueness and benefit

One problem with Experiment 1 is that the uniqueness manipulation was confounded with the specificity of the message participants received that reminded them to post to their group. In the unique condition, the reminder message provided a very specific suggestion for a topic of communication, while in the non-unique condition it merely reminded them to post. Although this confound is unlikely to influence the number of ratings participants made, it could directly influence their motivation to post because it lowers the effort associated with writing a message. We manipulated uniqueness in Experiment 2 without this confound, thus more cleanly testing the prediction from the collective effort model that people will contribute more to a group if they think their contributions are unique.

Experiment 1 also had the problem of a very small sample of groups, each with its idiosyncratic history. Although our statistical procedures took into account the non-independence of subjects in each group, the small number of groups raises doubts about the generalizability of the findings. Abandoning the conversational setting for the research, we conducted Experiment 2 so that subjects were truly independent. Experiment 2 is reported more fully in Beenen et al. (2004).

In addition to fixing problems in Experiment 1, this experiment tests the utility underpinnings of the collective effort model: People will contribute more to a group if they perceive their contributions as benefiting themselves or the group as a whole.

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 to any user. We call these rarely-rated movies. 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. As noted in Hypothesis 1, the collective effort model posits that people will socially loaf less when they perceive that their contribution is important to the group (Karau & Williams, 1993). In the case of MovieLens, making individuals who rate rarely-rated movies aware of their unique contribution should motivate them.

Salience of benefit and the beneficiary. The collective effort model also posits that people are more motivated to contribute when they perceive the value that their contribution makes to an individual or group outcome (Karau & Williams, 1993). MovieLens is a collaborative filtering system that uses other people's ratings to predict how much a subscriber will like a movie. As participants rate more movies, the system learns about their preferences improving the accuracy of recommendations for them, although with decreasing marginal returns. Reminding subscribers of this individual benefit should increase their motivation to rate movies.

Hypothesis 3a: 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.

Hypothesis 3b: MovieLens users will rate more movies when the benefit they provide to the community from doing so is made salient.

3.1 Methods

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 once 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.

All subjects received an email message inviting them to participate in a ratings campaign. Of the 904 members that we contacted, 74 of the emails bounced, leaving us with 830 participants who presumably received the email invitation. The text of these emails 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 they had previously rated.

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, but those in the *no-benefit* condition received neither.

Measuring contribution. Because participants’ ratings surged during the week after the invitation containing the experimental manipulation, and then rapidly fell to the pre-invitation level, we logged data from this week.

3.1.1 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 830 participants showed that they rated an average of 19.26 movies during the week following the invitation, far higher than the 5.4 movies per week they had rated in the previous 6 months. Participants who logged in during the experiment rated on average 39.7 movies, far higher than the 9.1 ratings made by individuals from a matched control group who logged in during the week

Table 2: Experiment 2. Number of Ratings by Condition

		No Benefit	Benefit to self	Benefit to others	Benefit to self and others	Total
Non-unique	N	103	110	101	106	420
	N logged in	52	45	47	48	192
	P(logged in)	50.5	40.9	46.5	45.3	45.7
	# Ratings	25.16	13.19	13.22	19.21	17.65
Unique	N	99	101	103	107	410
	N logged in	59	47	44	55	205
	P(logged in)	59.6	46.5	42.7	51.4	50.0
	# Ratings	31.53	19.28	9.48	23.66	20.92
Total	N	202	211	204	203	830
	N logged in	111	92	91	103	397
	P(logged in)	55.0	43.6	44.6	50.7	47.8
	# Ratings	28.28	16.1	11.33	21.45	19.26

Note. *N* is the number of participants to whom email was successfully sent. *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.

of the experiment. Table 2 summarizes the participation rates for the different treatment conditions.

We again used the Heckman procedure in the statistical analysis to account for the differences in probability of participation when computing average participation levels. The first stage of the Heckman model showed that participants were more likely to log in during the week of data collection the more frequently they had logged in previously ($z=6.2$, $p<.001$), and the fewer weeks had elapsed since their last login ($z=-3.0$, $p>.001$). Participants who were given the uniqueness message did not differ significantly from those who were given the non-uniqueness message ($z=.92$, $p>.35$). Compared to participants in the no-benefit control condition, participants who received the self-benefit email were less likely to log in ($z=2.47$, $p<.05$). So too were those who received the other-benefit email ($t(821)=2.01$, $p<.05$). Participants who received both emails with both benefit messages did not differ significantly from the no-benefit control condition ($z=1.41$, $p>.20$).

The second-stage model predicts the number of ratings that participants made, controlling for their likelihood of logging in. Consistent with Hypothesis 1, participants who were reminded of the uniqueness of their movie tastes rated more movies than those who got the non-unique message ($z=1.92$, $p<.05$). Moreover, the differential was even higher when we consider only rarely-rated movies as the dependent variable: the participants in the unique groups rated 40% more movies, and it was statistically significant ($z=2.30$, $p<.01$).

However, contrary to Hypothesis 3a, participants who received the self-benefit message reduced their number of ratings ($z= -2.52$, $p<.01$). In addition, contrary to Hypothesis 3b, participants who received the other-benefit message also reduced their number of ratings ($z=-2.46$, $p<.01$). Participants who received both self-and other-benefit messages increased their number of ratings almost to the level of the control condition ($z=1.41$, $p>.20$).

3.2 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. These results are consistent with those from Experiment 1, where mention of unique information caused participating to post more messages in a conversational forum.

Highlighting the benefits received from ratings had a more complicated relationship to contributions. Based on the collective effort model, Hypothesis 3 predicted that reminding people of the utility of their contributions would increase their motivation to contribute. However, the results were inconsistent. Instead, reminding participants of the benefits that either they or others would 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 benefits that both they and others would receive led to more effort than telling them about either one alone. In the follow-up experiment described below, we test two explanations for the disconfirmation of predictions from the collective effort model.

4. Experiment 3: Following up motivating contributions through benefits

Surveys of MovieLens members suggest that they rate movies for multiple reasons (Harper, Li, Chen, & Konstan, In press). They rate primarily to improve the accuracy of recommendations that they receive from the system and because the acts of remembering movies and rating them are intrinsically fun, and to a lesser extent, to help other subscribers to the system. It is possible that in Experiment 2, highlighting only the instrumental, extrinsic benefits may have undermined participants' intrinsic interest in rating. Previous research has shown that when people are intrinsically motivated to perform some behavior, the promise of extrinsic rewards such as money or grades reduces their intrinsic interest in it (Thompson, Meriac, & Cope, 2002). As a result, they are less likely to perform the behavior in the absence of the reward, compared to those who were never offered a reward. Deci and his colleagues proposed that extrinsic rewards may decrease intrinsic motivation by “thwart[ing the] satisfaction of the need for autonomy, lead to a more external perceived locus of causality” (Deci et al., 1999). If this explanation is correct, one would expect to boost contribution by reminding participants of the intrinsic motives for contributing. Reminding them of the intrinsic motivation may even reverse the effect of making extrinsic rewards salient.

Hypothesis 4: Members who receive messages that increase salience of intrinsic motivation will rate more movies than those who receive messages that do not increase salience of intrinsic motivation.

An especially perplexing finding from Experiment 2 is that mentioning either self-benefit or other-benefit reduced ratings from a control condition, but mentioning both together did not. Because each MovieLens subscriber may have multiple motives for rating movies – e.g., fun, improved accuracy for themselves, help to other people—it is possible that highlighting only a single benefit narrows the otherwise broad utility they associate with rating. If mentioning a single benefit narrows focus, then mentioning more should reduce this narrowing. In Experiment 2 we saw that mentioning two benefits led to more ratings than mentioning one. If the number of benefits mentioned is important, then mentioning three should in turn lead to more ratings than mentioning two.

Hypothesis 5: Members who receive who are reminded about the multiple benefits that contribution provides will rate more movies than those who are reminded of only a single benefit.

4.1 Methods

Subjects. The subject population consisted of 900 active MovieLens subscribers who had logged in at least twice during the last 8 months. Unlike the earlier experiment, we did not select raters of rarely-rated movies. Of the 900 members that we contacted, 94

messages bounced, leaving us with 806 participants. As in the previous experiment, all subjects received an email message inviting them to participate in a campaign to rate more movies. The persuasive message in the invitation email contained the experiment manipulation, with different text mentioning intrinsic and extrinsic motivations and with different numbers of reasons to participate. Subjects were randomly assigned to conditions in a way that balanced the number of ratings they had contributed in the past eight months.

Intrinsic motivation. Invitations in the intrinsic-benefit condition contained the text, “Most MovieLens subscribers tell us they rate movies because it’s fun! We hope you think so too.” Invitations in the no-intrinsic-benefit condition did not contain this text.

Benefit conditions. The extrinsic benefit manipulation was similar to the “benefit” condition in the Experiment 2. It contained 4 conditions: no extrinsic benefit, only benefit to self, only benefit to others, and benefit to both self and others. Invitations in the *self-benefit* condition contained the text, “In addition, rating more movies helps you! The more ratings you provide, the easier it is for MovieLens to assess your taste. This information allows MovieLens to make accurate recommendations for you.” Invitations in the other-benefit condition contained the text, “Rating more movies helps the MovieLens community! The more ratings you provide, the more information MovieLens has about each movie. This information allows MovieLens to make accurate recommendations for other MovieLens subscribers.” Invitations in the self-and-other-benefit condition contained both of these texts, while those in the no-extrinsic-benefit condition contained neither.

Measuring contribution. The variables used for measuring contribution were number of ratings transformed using the logarithm of the ratings, as in the previous experiment.

4.2 Results

Table 3 shows the results from this experiment. As in the earlier experiments, we analyzed the data using the Heckman model, separately predicting likelihood of logging in and number of ratings. Again participants were more likely to log in the more frequently they had logged in in the past and the more recently they had logged in. There were no significant differences among the experimental conditions, however, in likelihood of logging in.

Intrinsic Motivation. Hypothesis 4 predicted that participants who were reminded of their intrinsic motivation for ratings (“It’s fun”) would rate more than participants that did not receive this message. Although participants who received the intrinsic motivation message rated more movies than those who did not (means = 29.83 versus 27.42), this effect did not approach statistical significance ($z=.68$, $p>.50$).

Benefits. Hypotheses 2a and 2b predicted that participants would rate more movies when their self-benefit and benefit to others were made salient. As in Experiment 2, messages mentioning either one alone were associated with declines in the number of ratings ($z=-.10$, $p>.90$ for self-benefit and $z=-1.14$, $p>.20$ for other-benefit) and the message

reminding participants of self-benefit and other-benefit simultaneously had a higher mean than either condition alone (for the interaction, $z=1.09$, $p>.20$). None of the differences in this study, however, were statistically significant.

Broad utility. Hypothesis 5, that participants would rate more movies if they receive either no mention of benefits or description of no benefits or more than one benefit, was not supported ($z=.43$, $p > .60$).

Table 3: Experiment 3: Mean Number of Ratings by Condition

		No Benefit	Benefit to Self	Benefit to Other	Benefit to Self and Other	Assertion Only
Email alone	N	100	na	na	na	na
	N logged in (%)	33	na	na	na	na
	Average # ratings	35	na	na	na	na
Intrinsic Motivation	N	100	100	100	100	na
	N logged in (%)	32	38	29	32	na
	Average # ratings	30.52	24.61	26.69	25.7	na
No Intrinsic Motivation	N	na	100	100	100	100
	N logged in (%)	na	40	30	36	33
	Average # ratings	na	24.63	30.57	34.53	20.86

4.3 Discussion

In Experiment 2, the collective effort model (Karau & Williams, 1993) could not account for all of the results. Ratings declined when the invitation letter mentioned either self-benefit or other-benefit, but not when it mentioned both. The pattern of means for the current experiment were similar, although none of the experimental manipulations had significant effects when this experiment is considered by itself. To draw conclusions across the two experiments, we combined comparable results from Experiments 2 and 3 using meta-analysis. We calculated the combined z-values from both experiments, using

techniques described in Rosenthal & Rosnow (1991; $Z_{combined} = \frac{\sum z}{\sqrt{2}}$). Positive z-values

indicate that an experimental manipulation increased ratings. According to this meta-analysis, reminding participants of the benefit they would receive from rating had no reliable effect on the number of ratings they made ($z=1.21$, $p>.20$). However, reminding them that their contributions would help others depressed ratings ($z=2.5$, $p=.01$), and mentioning both benefits increased ratings ($z=1.77$, $p=.08$).

Although Experiment 3 was designed to examine explanations for the anomalous results in Experiment 2, none of our alternative explanations were clearly supported by the experiment, and the mystery remains. It is possible that we have simply run up against a limited ability to manipulate the relevant psychological states using just email messages. The majority of MovieLens users report intrinsic benefit as their greatest motivation when asked; however, email messages saying that it is fun to rate may not be an effective way to evoke intrinsic benefit. The nature of this type of motivation requires that the individual becomes personally motivated. Increasing the salience of “fun” from an external source, the research team at MovieLens, may even have depressed rather than stimulated the effect.

5. Experiment 4: Motivating contributions through goal-setting

Online communities rarely specify the type and amount of contribution expected of members. Open source development communities are an exception, displaying bug fix lists, though goals are rarely assigned to members. MovieLens also displays user feedback on the number of ratings a member has made, yet does not assign ratings goals. Goal setting theory, a robust theory of motivation in social psychology, has shown that assigning people challenging specific goals causes them to achieve more (Locke & Latham, 2002; 1990). Study 4 tests both the benefits and limits of the theory in an online community.

Benefits of High-challenge Goals. Hundreds of studies with over 40,000 subjects have shown that specific, challenging goals stimulate higher achievement than easy or “do your best” goals (Locke & Latham, 1990). High-challenge assigned goals energize performance in three ways (Bandura, 1993). First, they lead people to set higher personal goals, in turn increasing their effort. Second, assigned goals enhance self-efficacy, or belief in one’s own ability to complete a task (Bandura, 1993). Third, achieving an assigned goal leads to task satisfaction, which enhances both self-efficacy and commitment to future goals, resulting in an upward performance spiral.

The theory claims that difficult, specific, context-appropriate, and immediate goals rather than long-term goals motivate people most, and that they do so especially in tasks that are simple and non-interdependent and give them feedback on their performance against the goal. Providing MovieLens members with challenging numeric goals about the number of movies to rate satisfies these conditions. The straightforward design implication for online communities is that members should contribute more if they are assigned challenging contribution goals.

Hypothesis 6: Members who are assigned challenging, specific numeric goals will rate more than members assigned non-specific do-your-best goals.

Group Goals. Although most research on assigned goals has assigned them only to individuals, assigning goals to groups shows the same motivating effects (see Weldon & Weingart, 1993 for an overview). The collective effort model (Karau & Williams, 1993) predicts that individual goals and feedback will be more motivating than group goals, because in a group setting people can believe that their contribution is partially

redundant and that if they shirk, others can take up the slack. Although some studies have found that group goals are more motivating than individual goals, these findings are reversed as group size increases beyond 3-5 members (Streit, 1996). With group size in our experiment set at 10 members, members assigned group goals should contribute less than members with comparable individual goals.

Hypothesis 7: Members assigned individual goals will rate more than members assigned group goals.

Limits of high-challenge goals. According to Locke and Latham (1990), increases in goal difficulty are associated with steady performance increases until “subjects reach the limits of their ability at high-goal difficulty levels; in such cases the function levels off.” If this is true, assigned goals can never be too high. For extremely hard goals, performance should plateau, but not decline. Limited research, however, suggests otherwise. High-challenge goals above the 93rd percentile in performance have been associated with lower performance than “do your best” goals after subjects no longer believed they would be evaluated (White et al., 1995). Also, goals viewed as threatening rather than challenging resulted in lower performance than easy goals (Drach-Zahavy & Miriam, 2002).

In the MovieLens community, assigned goals that members view as overly difficult could reduce contributions in three ways. First, such goals could reduce goal commitment, lowering both personal goals and ratings. Second, overly difficult goals could impair self-efficacy by being viewed as an unattainable performance standard. Third, community membership may seem less desirable if a community makes unreasonable demands of members. This suggests exceedingly difficult goals should reduce contributions.

Hypothesis 8: Members assigned exceedingly difficult specific goals will rate less than members assigned difficult specific goals.

5.1 Methods

Overview. The subject population included active MovieLens members who logged in at least once in the period between July and December 2003. Of 900 members we contacted, 66 emails bounced, leaving us with 834 subjects. The invitation emails contained different text to manipulate members’ perceptions of whether they were part of a group or not and of the rating goals they were assigned. We tracked ratings of those who received invitations. More detail about the methods and preliminary results are available in Beenen et al (2004).

Group assignment. Participants were randomly assigned to a group-goal or an individual-goal condition. The invitation in the group-goal condition said, “We've enrolled you in the Explorers, a group of ten active MovieLens members...” and a rating goal for Explorers as a whole was given. We set group size at 10 for two reasons. First, 10 minimized the cognitive effort required for subjects to translate their fair share of the group goal, since most people can mentally divide by 10. Second, since past research

suggests that group goals are less effective than individual goals above groups of 3-5, we wanted to see if these findings would hold for larger groups. The invitation email in the individual-goal condition did not mention group membership. Participants were simply assigned a personal goal.

Goal specificity. Non-specific goal condition subjects were told to “do your best” to rate movies. Their message said, “[You/The Explorers] have a goal of doing [your/their] best to rate additional movies over the next seven days.” In the specific-goal condition, subjects were assigned a specific number of movies to rate. We asked individual-goal-condition subjects to either rate 8, 16, 32 or 64 movies in a week, and subjects in the 10-member “Explorers” group to rate either 80, 160, 320 or 640 movies in a week. We set 8 ratings per week as a baseline goal based on subjects’ mean weekly contribution in the past.

Measuring contributions. We tabulated user ratings for one week after sending the invitation email. We then sent a thank-you email summarizing their individual and group (if applicable) rating behavior.

5.2 Analysis and Results

Preliminary analysis showed one outlier (>7 std dev above mean) changed the size of coefficients, but not their direction and significance levels. We therefore excluded the outlier, leaving 833 subjects of whom 30% (249) logged in at least once. Table 4 summarizes subjects’ ratings.

We used the Heckman method for analysis as in the previous experiments, using the experimental condition variables (i.e., non-specific vs. specific goals; individual vs. group assignment, and their interaction term) to predict number of the log of the number of ratings subjects made. As in Experiments 2 and 3, participants were more likely to log in the more frequently they had logged in in the past. Subjects who received specific goals were marginally more likely to log in than those who received do-your-best goals ($z=1.73$, $p=.08$).

Hypothesis 6, which predicted that members given specific numeric goals would rate more than those given do-your-best goals, was supported. Subjects rated 27% more movies when given one of the specific goals than the non-specific do-your-best goals ($z=.2.87$, $p<.01$). Moreover, the marginally significant group goal*specific goal interaction ($z=1.67$, $p<.10$) indicates specific goals had a larger effect in the individual-goal condition than in the group condition. (See Figure 2 to examine the shape of the interaction.)

Table 4: Study 4. Mean Number of Movie Ratings

		Goal Specificity					Total
Group Assignment	Group	DYB	80	160	320	640	
	Individual		8	16	32	64	
Group	N	86	85	84	83	85	423
	N logged in	23	30	30	25	25	133
	% (logged in)	26.7	35.3	35.7	30.1	29.4	31.4
	# Ratings	14.9	16.4	19.2	19.5	16.3	17.2
Individual	N	78	85	83	83	81	410
	N logged in	18	23	25	27	23	116
	% (logged in)	23.1	27.1	30.1	32.5	28.4	28.3
	# Ratings	5	6.7	12.3	15.6	12	10.4
Total	N	164	170	167	166	166	833
	N logged in	41	53	55	52	48	249
	% (logged in)	25.0	31.2	32.9	31.3	28.9	29.9
	# Ratings	10.3	11.5	15.8	17.5	14.2	13.9

*Note: DYB = "Do Your Best" non-specific goal
 N = number of subjects to whom email was successfully delivered.
 % (logged in) = percentage of participants who logged into MovieLens during the week of the experiment.
 # Rating = mean number of ratings for the N subjects*

Hypothesis 7, which predicted members given individual goals would rate more movies than those with group goals, was disconfirmed. Subjects in the individual-goal condition rated 42% of the movies they rated in group-goal condition ($z=-2.43$, $p<.02$).

Hypothesis 8, predicting the highest challenge goal would lead to a decline in ratings, was weakly supported. To test this, we created orthogonal linear and quadratic contrasts for the 4-levels of assigned goals (linear contrasts = -1.5, -.5, .55, 1.5; quadratic contrasts = -1, 1, 1, -1) and conducted a Heckman analysis among participants in the specific-goals condition. The coefficient for the linear contrast was not significant ($z=-.36$, $p>.70$), but the coefficient for the quadratic contrast was marginally significant ($z=1.85$, $p=.06$). As seen in Figure 2, participants made the most ratings when they received intermediate goals, but made fewer when they were given the unchallenging goal of 8 ratings or the most challenging goal of 64 ratings.

5.3 Discussion

Consistent with hypothesis 6, specific goals predicted higher contribution rates than do-your-best goals, and this effect was stronger in the individual condition. The inverse U-

shaped relationship between the size of the goal and the contribution rate suggests that goals have upper limits, and that beyond those limits, the goals may demotivate members of online communities rather than motivate them. Our experiment was a weak test of these limits, since the highest goal (64 movie ratings in a week) was not be a stretch for a sizeable proportion of the MovieLens community. In the past, 45% of subjects had rated over 64 movies in a single day at least once.

Although, based on the collective effort model, we had expected that individual goals would be more effective than group goals in stimulating contribution, we found the reverse. Two alternatives may explain this reversal. First, naming the group “Explorers” may have motivated greater effort by making members identify with this in-group, to the exclusion of unnamed out-groups (Kane et al., 2005; Tajfel & Turner, 1986). Second, the presence of both individual and group level feedback in the group condition may have resulted in social facilitation (Zajonc, 1965) rather than social loafing. Social facilitation is an effect whereby the real or imagined presence of evaluative others results in greater effort on a group task. Some have classified social loafing and social facilitation as related effects (Paulus, 1983; Jackson & Williams, 1985), distinguished by the presence (facilitation) or absence (loafing) of normative performance feedback (Sanna, 1992; Harkins, 1987). Social loafing experiments therefore do not include group and individual feedback, since it can eliminate loafing (Karau & Williams, 1993). We included feedback in our experiment as specified by goal-setting theory, by telling participants that they would receive an accounting at the end of the campaign of the number of movies they and their group had rated.

6. Conclusion

This paper attempted to use principles from social psychology theory to redesign an online community to increase contributions, and tested these principles in four field experiments using the MovieLens online community. Table 5 summarizes the empirical results.

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. Although we applied only two theories—Karau and William’s collective effort model and Locke and Latham’s Goal-setting Theory—we think the benefits we obtained from using theory and problems we observed are likely to characterize other attempts to systematically apply social science theory to the design of online communities.

Table 5: Summary of empirical results

Hypothesis	Support
H1: People will contribute more to online communities when they believe that their contributions will be unique.	Supported, Experiments 1, 2, & 3
H2: People will contribute more to online communities when they believe they are similar rather than dissimilar to others in the group.	Disconfirmed, Experiment 1
H3a: MovieLens users will rate more movies when the personal benefit they receive from doing so is made salient.	Not supported, Experiments 2 & 3
H3b: MovieLens users will rate more movies when the benefit they provide to the community from doing so is made salient.	Disconfirmed, Experiments 2 & 3
H4: Members who receive messages that increase salience of intrinsic motivation will rate more movies than those that receive messages that do not increase salience of intrinsic motivation.	Not supported, Experiment 3
H5: Members who are reminded about the multiple benefits that contribution provides will rate more movies than those who are reminded of only a single benefit.	Not supported, Experiment 3
H6: Members who are assigned challenging, specific numeric goals will rate more than members assigned non-specific do-your-best goals.	Partially supported, Experiment 4, only for individual-level goals
H7: Members assigned individual goals will rate more than members assigned group goals.	Disconfirmed, Experiment 4
H8: Members assigned exceedingly difficult specific goals will rate less than members assigned difficult specific goals.	Marginally supported, Experiment 4

6.1 The success in applying social science theory to design

We identify two criteria for applying social psychology theory to design. First, does the theory generate design goals that lead to design options that are not obvious from current design practice? Second, does applying the design options generated from theory lead to the desired communal outcomes? For shorthand, we can label these criteria as *inspiration* and *prediction*, respectively.

By the inspiration criterion, 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 will 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. 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.

By the prediction criterion, our attempt to drive design from theory was only partially successful. Applying some of the design principles inspired by theory led to increased contributions in MovieLens. Making salient the uniqueness of potential contributions caused recipients to rate more movies than they would have otherwise. This result occurred whether the sense of uniqueness was reinforced by conversations or whether it was simply asserted in a simple email message. Without the collective effort model, it would not be obvious whether emphasizing uniqueness or commonality would have been more effective. Assigning specific ratings goals to members of the community led them 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 could plausibly discourage contributions above the goal.

For three of experiments conducted here, we implemented the theory-inspired designs superficially, by crafting the text of email messages we sent to subscribers, although in Experiment 1, theory was also implemented by group composition and conversation. Even though we recognized that the email text provided a weak manipulation, we selected this technique because of its ease of implementation. Having evidence that providing subscribers with specific goals and making them aware of their uniqueness increases their rating behavior, we could integrate goals and indicators of uniqueness more deeply into the user interface of MovieLens. For example, based on historical ratings, we can predict for each subscriber the movies that a subscriber is likely to see early in their run in theatres, but that few others will see at all, or will see only late in

their run. The system can request that the subscriber rate these movies, with an indicator that the subscriber is making a unique contribution by doing so.

6.2 Failures of prediction

However, not all the design ideas derived from the theories led to increased contributions. Results from Experiment 1, in which participants posted fewer messages when they were most similar to other group members, is inconsistent with a corollary of the collective effort model, that people contribute most to groups they like. Results from Experiment 4 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 for themselves and others, in Experiments 2 and 3 making salient the benefit that others would receive from their ratings depressed their contributions. In Experiment 2, making salient the benefit that the contributors themselves would receive also depressed contributions. On the other hand, in both experiments, reminding contributors of their own and others' benefits together was better than mentioning either one alone.

Why did the design choices inspired by social psychology theories sometimes fail to increase contribution, and even decrease it? The inconsistencies with theory may have been partly a failure of prediction and partly a failure of implementation. Here we consider both classes of explanation.

Poor implementation. It is unlikely that the social psychology theories we used as the basis for design were fundamentally wrong, in the sense that they incorrectly explain people's motivation in the domains in which they were developed. Both goal-setting theory and the collective effort model are robust theories of human motivation, consistent with and able to explain a wide variety of field and experimental data, as evidenced by recent literature reviews (Locke & Latham, 2002; Karau & Williams, 1993). Similarly, one of the most robust findings in the literature on interpersonal attraction is that people like those who are similar to themselves.

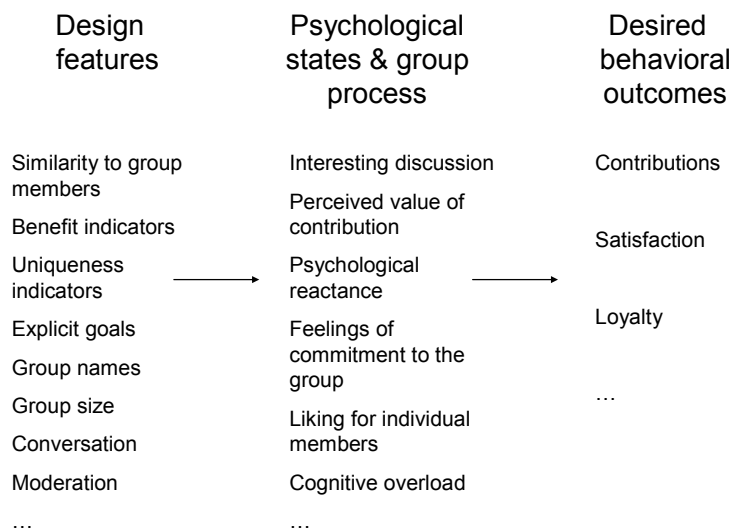
It is possible, however, that the email manipulations of group connection, uniqueness, and benefit and the group composition manipulation of attraction may simply have been too weak to effectively motivate members of MovieLens. Weak manipulations, however, cannot explain reliable reversals of predictions derived from the theories. Although the collective effort model predicts that people will contribute less when they think their contributions would be aggregated with those of others than when they think their contributions stand apart, Experiment 4 appears to find the opposite. In particular, subscribers contributed more ratings to MovieLens when they were given group goals that made them think their ratings would be combined with others in their nominal group than when they were given individual goals and others were not mentioned. Similarly, a basic premise of the expectancy-value model, underlying both the collective effort model and goal-setting theory, is that people contribute to a group when they perceive they will personally benefit, either directly or indirectly, because their contributions are helping a group that they value. Yet results from Experiments 2 and 3 seem to show the opposite;

participants contributed less when either benefit to the participant or to the group as a whole was made salient. Finally, the collective effort model predicts that people will contribute more to groups they like, a prediction at odds with the findings from Experiment 1.

Although we may have started with correct and applicable theory, our implementations could have failed to capture the design principles appropriately. In Experiments 2, 3, and 4, the design principles extracted from theory were implemented as short, single electronic mail messages. They may have been poorly worded. For example, the messages used to manipulate self- and other-benefit in Experiments 2 and 3 may have been too weak to manipulate perceptions of benefit, but instead may have undermined intrinsic motivation. Results from Experiment 3, however, do not support the explanation that mention of benefits undermined intrinsic motivation. Alternatively, these attempts to make benefit salient may have stimulated feelings of psychological reactance (Brehm, 1966), which would lead people to try to assert their autonomy. In Experiment 1, attraction to the group was manipulated by constituting groups so that members had either similar or dissimilar tastes in movies. However, it is possible that because all the participants were movie buffs, this basic similarity may have overwhelmed the influence of specific differences in taste, and the rich conversations inspired by discussing movies with someone with different tastes may have led to liking for the group.

Incomplete theories. In some cases, social science theories may simply not be up to the task of guiding design when multiple features of an online community must be determined simultaneously, as they do in tests of real designs. Figure 3 illustrates one problem in attempting to use social psychological theory to drive design. Figure 3 is a variant of the familiar input-process-output model often used to analyze group behavior (McGrath, 1984; Hackman, 1987). The designer desires multiple outcomes from an

Figure 3: Multi-determined Outcomes



online community. In the case of MovieLens, for example, its designers want the site to provide useful predictions for its users and therefore for subscribers to contribute ratings. They also want subscribers to have a satisfying experience on each visit and to return often. Typically, these outcomes of groups— known generically as production, member support, and group maintenance (McGrath, 1984) – are only weakly correlated. To achieve these outcomes, the designer can modify features of the group composition or technology in an attempt to influence subscribers' psychological states and group process. In the experiments described here, for example, email messages providing goals and emphasizing the uniqueness and benefits of contributions were designed to influence the recipients' perceptions of the norms and benefits for contribution, and selection criteria were used to influence members' liking for their group.

The problem is that each design feature can potentially influence multiple intervening psychological states and group processes. Each of these intervening states and process, in turn, can have multiple determinants. Finally, each intervening psychological state and group process can have multiple and potentially inconsistent effects on the outcomes the designer wants to influence. Concretely, for example, similarity among group members, manipulated in Experiment 1, may have had the desired outcome of getting people to like their group more. However, it may have had a side effect of making conversation in these groups less engaging, causing participants in these groups to return to the discussion less frequently. In Experiments 2 and 3, it is possible that participants perceived messages highlighting benefits as information about the consequences of their contributions but also as attempts to manipulate their behavior. If so, this design feature could have led them to refrain from contributing in order to reassert their autonomy. Finally, assigning participants a group goal may have made them think they were redundant, thus reducing their perceptions of the value of their contribution. Simultaneously, though, the group assignment may have increased participants' attachment to MovieLens overall, thereby increasing their motivation to help their group.

In summary, design features have multiple consequences and intervening states, and behavioral outcomes are multiply determined. These complexities are general phenomena. They apply to the experimental manipulations deployed in the four experiments reported here, but also more generally whenever one tries to leverage social science theory as a basis for design. For example, there is abundant evidence that interpersonal communication is one mechanism for getting people to like each other (Berscheid & Reis, 1998; Bossard, 1932; Duck, 1998, 1991) and to develop attachment to the community as a whole (Sassenberg, 2002). However, conversation often frequently leads to cognitive overload, which in turn can drive people from the group (Jones, Ravid, & Rafaeli, 2004; Kraut, 2003).

The failure of social psychology to produce theory that is sufficiently complete for the purposes of design stems from fundamental differences in the goals and values of social psychologists and of HCI and CSCW researchers. Social psychology is primarily a behavioral science, whose goal is to unambiguously determine the causes for social phenomena and explain them. The conventional analytic approach is to examine the influence of a small number of variables through measurement or experimental manipulation, while holding other variables constant through statistical techniques or

experimental control. In addition, a value in the social science tradition is to produce theory that is as general as possible. While social psychologists recognize the importance of context, they attempt to create theories that abstract across “irrelevant” differences in context. For example, if they were interested in the degree to which group members’ liking for each other influenced their willingness to contribute to the group, they would prefer a theory of contribution that allowed them to generalize across different bases for liking—e.g., similarity, familiarity, attractiveness, or reciprocity— unless data said these bases made a difference for the amount or type of contribution. A consequence of this scientific orientation is that many social psychological theories are sparse, describing the relationships among a small set of variables.

In contrast, HCI and CSCW are primarily engineering disciplines, where the primary goal is problem-solving. In trying to build an online community whose members like each other, for example, the bases of liking irrelevant to a psychological theory of contribution would be very relevant to the design of the online community. The designer would want to know which source of liking is most powerful, what sources are compatible with each other, and how they hold up over time. The designer also wants to know about the interaction among variables implicated in different theories. For example, does uniqueness, which the collective effort model suggests can be a basis of contribution, conflict with perceived similarity to other group members, which can lead to liking for the group, which is another basis of contribution (cf. Experiment 1)? Because psychologists have not set themselves the task of designing the optimum configuration of features to achieve a desired outcome, they provide little guidance about which effect will predominate when a single design feature can have opposite effects.

The net result is that designers get insufficient guidance from the social science theory, because the theory isn’t general and complete enough (Landauer, 1991). 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.

6.3 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 generally useful strategy for the design of CSCW systems (see Kraut, 2003 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 (Ross, 1908) 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 the problem of under-contribution in online communities include, for example, those on group cohesion, group identity, interpersonal attraction, and altruism, among others. Recent handbooks (e.g., Brown & Gaertner, 2001; Gilbert et al., 1998; Higgins & Kruglanski, 1996) and graduate-level textbooks provide

useful resources. However, as we have suggested above, this accumulated knowledge is insufficiently detailed to provide concrete guidance for designers.

We believe these theories would be more useful if they could be expressed in language that is more adequate to the task of representing the relationships among scores or hundreds of variables simultaneously, rather than the handful that characterize most theory in social psychology. It would also require theories to predict the magnitude of effects rather than just their direction. Recent innovations in simulating groups and organizations (Epstein & Axtell, 1996; Axelrod, 2003; Gilbert & Troitzsch, 1996; Prietula et al., 1998) may make this approach possible. As Axelrod (2003) notes, simulation approaches to expressing theory allow one to take complicated input and generate their consequences as predictions, allowing researchers and designers to reason about processes that would otherwise be too complex. The simulation approach has led to major advances in cognitive science (Anderson et al., 2004; Simon, 1995), including the application of theories of psychology to inform the technological design (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Anderson, Douglass, & Qin, 2005). Perhaps it is time to use these techniques to synthesize knowledge in the social sciences and package the knowledge for design.

An alternative approach would be to try to modify designs so that the predictions of multiple theories were aligned rather than conflicting. With this approach, a designer would start, as we did, with a single theory that suggests a design goal such as highlighting uniqueness in order to enhance the perceived value of contributions. The designer would then, as we did, proceed to generate a specific implementation. At that point, however, the designer would examine the specific implementation to see if any other theories suggested other psychological states or group processes that would be impacted. If any indicated likely behavioral outcomes are in the opposite direction from those desired, the designer would try to modify the implementation so that it still manipulated the originally identified psychological state but also manipulated the other states and group processes in a direction that would predict desirable outcomes.

It is not clear how often design alternatives can be crafted so that the predictions of all the relevant theories are aligned. Nor is it clear whether relevant theories can be parameterized sufficiently to guide designers about which effects will predominate when theories are not aligned in their predictions. Thus, the approach we advocate here, mining social psychology theories for design guidance, 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.

ACKNOWLEDGMENTS

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

This material is based upon work supported by the National Science Foundation under Grant No. 0325837. Anupriya Ankolekar, John Graham, Janice Golenbock, Dan Frankowski, John Riedl and Steve Karau provided technical assistance or advice in implementing this research.

REFERENCES

- Adar, E., & Huberman, B. A. (2000). Free riding on gnutella. *First Monday*, 5(10), NP.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167-207.
- Anderson, J. R., Douglass, S., & Qin, Y. (2005). How should a theory of learning and cognition inform instruction? In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 47-58): Washington, DC, USical Association, 2005, xxii, 265.
- Axelrod, R. (2003). Advancing the Art of Simulation in the Social Sciences. *Japanese Journal for Management Information System*, 12(3).
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148.
- Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., et al. (2004). Using social psychology to motivate contributions to online communities. In *CSCW '04: Proceedings of the ACM Conference On Computer Supported Cooperative Work*. New York: ACM Press.
- Berscheid, E., & Reis, H. T. (1998). Attraction and close relationships. In D. T. Gilbert, S. T. Fiske et al. (Eds.), *The handbook of social psychology, Vol 2* (4th ed., pp. 193-281). New York: McGraw-Hill.
- Bossard, J. (1932). Residential propinquity as a factor in marriage selection. *American Journal of Sociology*, 38, 219-224.
- Brehm, J. (1966). *A theory of psychological reactance*. New York: Academic Press.
- Brown, R. & Gaertner, S.L. (2001) *Blackwell handbook of social psychology: Intergroup processes*. Oxford, UK: Blackwell.
- Butler, B. (1999). *When is a group not a group: An empirical examination of metaphors for online social structure*. Carnegie Mellon University, Pittsburgh, PA.
- Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social & Personal Relationships* 14(3), 417-431.
- Byrne, D. & W. Griffith (1973). Interpersonal attraction. *Annual Review of Psychology*, 317-336.

Cosley, D., Lam, S. K., Albert, I., Konstan, J., & Riedl, J. (2003). Is seeing believing? How recommender systems influence users' opinions. In *Proceedings of CHI 2003: Human factors in computing systems* (pp. 585-592). New York: ACM Press.

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

Dourish, P., & Bly, S. (1992). Portholes: Supporting awareness in a distributed work group. In *Proceedings of CHI 1992: Human factors in computing systems* (pp. 541-547). New York: ACM.

Drach-Zahavy, A. & Miriam, E. (2002). Challenge versus threat effects on the goal-performance relationship. *Organizational Behavior and Human Decision Processes*, 88(2), 667-682.

Duck, S. (1998). *Human relationships* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Duck, S., Rutt, D., Hurst, M., & Strejc, H. (1991). Some evident truths about conversations in everyday relationships: All communications are not created equal. *Human Communication Research*, 18(2), 228-267.

Epstein, J. & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Washington, DC: Brookings Institution and Cambridge, MA: MIT Press.

Erickson, T., & Kellog, W. (2000). Social translucence: An approach to designing systems that support social processes. *Transactions on Computer-Human Interaction (TOCHI)*, 7(1), 59-83.

Gilbert, D.T., Fiske, S.T., & Lindzey, G. (1998). *The handbook of social psychology*. Boston: McGraw-Hill.

Gilbert, N. & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Berkshire, UK: Open University Press.

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

Hackman, R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of Organizational Behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Harkins, S. (1987). Social loafing and social facilitation. *Journal of Experimental Social Psychology*, 23(1), 1-18. Englewood Cliffs, NJ: Prentice-Hall.

Harkins, S. G., & Petty, R. E. (1982). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology*, 43, 1214-1229.

- Harper, M., Li, X., Chen, Y., & Konstan, J. (In press). *An economic model of user rating in an online recommender system*. Paper presented at the Proceeding, UM'05 - 10th International Conference on User Modeling, Edinboro, Scotland.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-162.
- Heckman, J., Ichimura, H., Smith, J. & Todd, P. (1996). Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. *Proceedings of the National Academy of Sciences USA*, 93, 13416-13420.
- Higgins, E.T. & Kruglanski, A.W. (1996). *Social psychology: Handbook of basic principles*. New York: Guilford.
- Jackson, J. & Williams, K. (1985). Social loafing on difficult tasks: Working collectively can improve performance. *Journal of Personality and Social Psychology*, 49, 937-942.
- Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces. *Information Systems Research*, 15(2), 194–210.
- Kane, A., Argote, L. & Levine, J. (In Press). Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes*.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4), 681-706.
- Karau, S. J., & Williams, K. D. (2001). Understanding individual motivation in groups: The collective effort model. In M. E. Turner (Ed.), *Groups at work: Theory and research* (pp. 113-141). Mahwah, NJ :Lawrence Erlbaum Associates.
- Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, 45, 819-828.
- Kerr, N. L., & Bruun, S. (1983). The dispensability of member effort and group motivation losses: Free rider effects. *Journal of Personality and Social Psychology*, 44, 78-94.
- Lakhani, K. R., & Hippel, E. V. (2003). How open source software works: "free" user to user assistance. *Research Policy (Special issue)*, 32, 923-943.
- Landauer, T. K. (1991). Let's get real: A position paper on the role of cognitive psychology in the design of humanly useful and usable systems. In J. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 60-73). New York: Cambridge University Press.
- Ledyard, J. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. Roth (Eds.), *The handbook of experimental economics* (pp. 111-194.). Princeton, NJ: Princeton University Press.

- Locke, E.A. & Latham, G.P. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice-Hall.
- Locke, E.A. & Latham, G.P. (2002). Building a practically useful theory of goal setting and task motivation: A 35 year odyssey. *American Psychologist*, 57(9), 705-717.
- Ludford, P. J., Cosley, D., Frankowski, D., & Terveen, L. (2004). Think different: Increasing online community participation using uniqueness and group dissimilarity. In *Proceedings of Human Factors in Computing Systems, CHI 2004* (pp. 631-638.). NY: ACM Press.
- Markey, P.M. (2000). Bystander intervention in computer mediated communication. *Computers in Human Behavior*, 16(2): p. 183-188.
- Markus, L. (1987). Towards a "critical mass" theory of interactive media: Universal access, interdependence, and diffusion. *Communication Research*, 14, 491-511.
- McGrath, J. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.
- Mockus, A., Fielding, R. T., & Andersen, H. (2002). Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*, 11(3), 309-346.
- Nonnecke, B., & Preece, J. (2000). Lurker demographics: Counting the silent. *CHI 2000, ACM Conference on Human Factors in Computing Systems, CHI Letters*, 4(1), 73-80.
- Paulus, P. (1983). Group influence on individual task performance. In P.B. Paulus (Ed.), *Basic group processes* (97-120). New York: Springer-Verlag.
- Preece, J. (2000). *Online communities*. New York, NY: John Wiley & Sons.
- Prietula, M. J., Carley, K. M. & Gasser, L. (Eds.). (1998). *Simulating organizations: Computational models of institutions and groups*. Cambridge, MA: MIT Press.
- Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw Hill.
- Ross, E.A. (1908). *Social psychology: An outline and source book*. New York: The Macmillan Company.
- Sanna, L. (1992). Self-efficacy theory: Implications for social facilitation and social loafing. *Journal of Personality and Social Psychology*, 62(5), 774-786.
- Sassenberg, K. (2002). Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats. *Group Dynamics*, 6(1), 27-37.
- Simon, H. A. (1995). The information-processing theory of mind. *American Psychologist*, 50(7), 507-508.

- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making. *Journal of Personality and Social Psychology*, 48, 1467-1478.
- StataCorp. (2003). Stata statistical software: Release 8.0. College Station, TX: Stata Corporation.
- Streit, M. C. (1996). *The effects of group size on goal-setting behavior, task performance, goal commitment, and evaluation potential*. Hofstra University, Long Island, NY.
- Tajfel, H. & Turner, J. C. (1986). An integrative theory of intergroup conflict. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 2-24). Chicago: Nelson-Hall.
- Thompson, L. F., Meriac, J. P., & Cope, J. (2002). Motivating online performance: The influences of goal setting and Internet self-efficacy. *Social Science Computer Review*, 20(2), 149-160.
- Weldon, E. & Weingard, L.R. (1993). Group goals and group performance. *British Journal of Social Psychology*, 32, 307-334.
- Williams, K. & O'Reilly, C. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, 20, 77-140.
- Zajonc, R. (1965). Social facilitation. *Science*, 149, 269-274.