The Effect of Performance Feedback on Social Media Sharing in Volunteer-Based Online Experiment Platforms

Bernd Huber
Harvard University
Cambridge, MA, USA
bhb@seas.harvard.edu

Katharina Reinecke
University of Washington
Seattle, WA, USA
reinecke@cs.washington.edu

Krzysztof Z. Gajos
Harvard University
Cambridge, MA, USA
kgajos@eecs.harvard.edu

ABSTRACT
As an alternative to online labor markets, several platforms recruit unpaid online volunteers to participate in behavioral experiments that provide personalized feedback. These platforms rely on word-of-mouth sharing by previous participants for recruitment of new participants. We analyzed the impact of performance feedback provided at the end of an experiment on 81,131 participants' sharing behavior. We show that higher performing participants share significantly more. We also show that self-verification has a moderating effect: people who expected to do poorly are not affected by a high score, but people who expected to do as well as others or better, are. In a second experiment, we evaluate three distinct social comparison designs for the presentation of the results. As expected, the design that most emphasized participants' relative success led to most sharing. Contrary to our expectations, people who expected to do poorly benefited from the most optimistic social comparison more than participants who expected to do better than others.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Social comparison; Volunteer-based Online Experiments; Self-Evaluation

INTRODUCTION
Volunteer-based online behavioral research platforms, such as TestMyBrain.org [9], Project Implicit [1], or LabintheWild.org [13], offer people feedback on their traits and skills in return for participation in experiments. Unlike conventional laboratory studies or those conducted on Amazon Mechanical Turk, where participants are frequently motivated by monetary compensation, volunteer-based platforms rely on participants' interest in receiving personalized results at the end of each study. To provide participants with engaging feedback, the personalized results are often presented using social comparison (e.g., "You scored better than the average participant"). The original social comparison theory explains how people compare themselves to others in order to evaluate their own abilities and opinions, which assumes that people have an inherent drive to gain accurate self-evaluation [8]. The feedback in online experiments is thought to motivate participants to exert themselves; previous work has found that participants provide reliable data and that experiments conducted on volunteer-based experiment platforms can accurately replicate laboratory study results [13, 9]. In addition, prior work found that this social comparison feedback is one reason why people share their behavioral experiment results on social networking sites such as twitter [13]. It therefore plays a key role in the recruitment of participants.

Despite the vital role of social comparison feedback in volunteer-based online experiments for data quality, motivating participation, and spreading the word, little is known how to present this feedback most effectively. Previous work demonstrates that social comparison interventions can increase participant contribution in online communities, in which contribution is a direct result of participant's effort [5, 12]. The effort of newcomers to online communities seems to be particularly receptive to social learning [4]. The goal of online experiment designers is not solely to increase participant's effort, but rather to increase engagement with the experiment results. Hence, social comparison may affect the way participants interpret them, ultimately influencing their sharing behavior. To find out why people share their results, we conducted two studies in the context of the Social Intelligence Test (rebranded Reading the Mind in the Eyes Test [3]). We implemented the test on LabintheWild.org, a popular volunteer-based online experiment platform. Participants taking this test have the option to share the experiment by clicking on social media buttons on the feedback page at the end of the test, which is the measure we used to estimate share rates (see Figure 1b).

In our first study, we analyzed sharing behavior when participant scores were communicated with the global average score as a reference point (see Figure 1b). We found that both participants' scores and the interaction between participants' self-conceptions and their actual scores affected their probability of sharing. Our results show that most people who take the test are attempting to make themselves feel good about themselves by receiving a positive evaluation of their social intelligence (a motivation that researchers have called "self-enhancement" [18]). The motivation for self-verification [17], i.e., people's desire to confirm their own judgements of their social intelligence, plays a moderating role.

CHI 2017, May 06 - 11, 2017, Denver, CO, USA.

ACM 978-1-4503-4655-9/17/05...$15.00 .
DOI: http://dx.doi.org/10.1145/3025453.3025553

In our second study, we tested three different designs to communicate the results to participants. We used social downward and upward comparison to emphasize that one's performance is better or worse than others, respectively. Both kinds of comparisons can motivate people, but not in all circumstances: For example, downward comparison can make people feel better because it might elevate the self-regard [10]. On the other hand, upward comparison may lead to the hope of self-improvement in some cases, but may be discouraging in others [16, 6]. Consistent with the results from Study 1, we found that more positive score communication led to more shares. These findings further suggest that participants in the *Social Intelligence Test* may be primarily driven by self-enhancement. Surprisingly, in this study people who expected to do poorly benefited from the most optimistic social comparison more than participants who expected to do better than others.

STUDY 1: WHAT MOTIVATES PARTICIPANTS?

To identify which of the three self-evaluation motives plays the greatest role, we studied how likely people were to share their results given a particular score, and how this is affected by their self-conceptions. The core assumption behind social comparison theory is that people have an inherent drive to gain an accurate self-evaluation [8]. The original theory explains how people compare themselves to others in order to evaluate their own abilities and opinions [8]. Subsequent research identified three distinct possible motivations for engaging in self-evaluation activities, such as completing skill or trait tests: (1) The *self-enhancement* motive to improve the positivity of one's self-concept [18], (2) the *self-assessment* motive to have an accurate and objective evaluation of the self [14], and (3) the *self-verification* motive to verify one's pre-existing self-conceptions [17]. These three motivations result in three distinct predictions about participants' likelihood of sharing their results depending on the feedback that they receive at the end of a study:

Self-enhancement predicts that the higher the score a participant receives, the more likely they will be to share.

Self-assessment predicts that the score will have no impact on the likelihood of sharing.

Self-verification predicts an interaction between the pre-existing self-conceptions with respect to the social intelligence, and the actual score: People whose score confirms pre-existing self-conceptions will be more likely to share than those who receive a score that contradicts their self-conceptions.

Participants

Our participants in this study were 75,120 online volunteers (aged 11–70, 42.7% male, 48.7% female, 2.3% other, the rest declining to answer). No financial compensation was given.

Task

Participants took the Reading the Mind in the Eyes test [3] (advertised as *Social Intelligence Test*) on LabintheWild.org [13], a volunteer-based online experiment platform. The test was originally developed in the context of autism research, but it has also been used to measure theory of mind (the ability to attribute cognitive and emotional states to others) in general

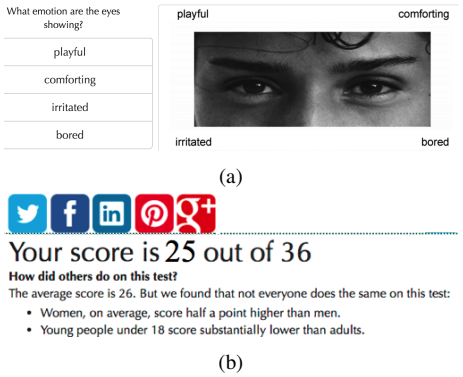


Figure 1: (a) During the test, participants choose emotions the person in the photograph is expressing. (b) At the end of the test, participants are presented with their own score, and information other participants' performed.

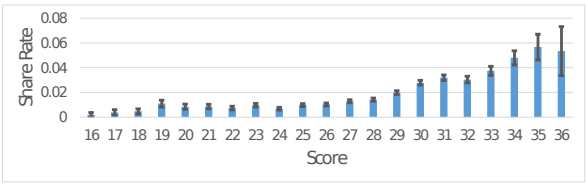
population and it has been demonstrated to be a strong predictor of success on team-based problem-solving tasks [19, 7]. This is why we decided that renaming it as a Social Intelligence Test was justified and better communicated its relevance to the general population. During the test, participants were presented with 36 images depicting the eyes of different people and were asked to tell what emotion the person in the image was expressing (see Figure 1a). The possible scores ranged between 0 and 36.

Procedure

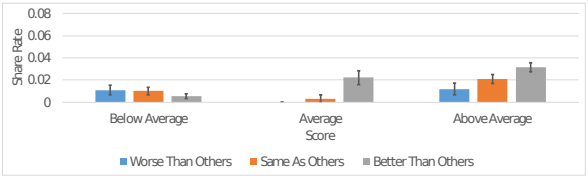
Participants arrived at the experiment site organically. Logs indicate that approximately 30% arrived at the test from social media sites or referrals. The landing page included test title, a tag line ("Test how well you can read emotions of others just by looking at their eyes") and a brief explanation of the test. Participants who clicked through were first presented with an informed consent statement, followed by a brief demographics questionnaire. The demographics questionnaire also included a question on participants' pre-existing self-conceptions with respect to social intelligence, phrased as "Compared to your family and friends, how good are you at reading people's emotions?". Participants were asked to answer this question on a 5-point Likert scale (1 = much worse, 5 = much better). All demographics questions were optional. Completing the actual test took 10 minutes on average. Before being presented with their results, participants were asked if they cheated in any way, experienced technical difficulties or had any other comments. The results page then showed the participant's score, the average score over all participants as a reference point, and background information on how to interpret the score (see Figure 1b). The results page also included links for sharing the test on social networking sites.

Design and Analysis

We conducted one analysis with 75,120 participants to analyze the effect of the score on the probability that a participant would share the test on a social networking site. We conducted a follow-up analysis with a subset of 7,068 additional participants to look for an interaction effect between pre-existing self-conceptions and the score. The self-conception question was added several months after the experiment was launched



(a)



(b)

Figure 2: Participants' scores on the *Social Intelligence Test* vs. share rates on social media. Error bars show standard errors. (a) There is a significant main effect of score on the probability that participants share. (b) There is a significant interaction between score and self-conception. Scores were binned for better visualization.

and answering that question was optional, which resulted in fewer participants in the second analysis of this study.

The share rate was calculated using the mean of the binomial variable share/not share across all participants per condition. Score was modeled as a continuous variable and pre-existing self-conception was treated as an ordinal variable with three levels (worse than others, same as others, better than others). To analyze the data, we used logistic regression (generalized linear model with binomial distribution). For better interpretation, we also include results of pairwise correlation analyses.

Prior to the analysis, we excluded participants who reported having taken the test before, having cheated, or having experienced technical difficulties. We also removed extreme outliers with respect to the score, that is, participants whose score was more than two standard deviations from the mean (i.e., participants with scores of 15 or lower). Self-conception responses were bucketed in *worse*, *equal* and *better* than others.

Results

We observed a significant main effect of score on the likelihood of participants sharing the test with their peers ($\chi^2_{1,N=75120} = 426.56, p < .0001$), with participants receiving a higher score being more likely to share (see Figure 2a and Table 1 for complete results).

We also observed a significant interaction between score and self-conception ($\chi^2_{2,N=7068} = 7.81, p < .05$). People who thought they were worse than others at reading emotions were not any more likely to share if they received a high score ($\chi^2_{2,N=1135} = 3.38, p = .18$; $r(19) = -.12, p = .62$), but people who thought they were better than others were significantly more likely to share the higher the score they received ($\chi^2_{2,N=3620} = 40.21, p < .0001$; $r(19) = .91, p < .0001$). These results indicate that self-enhancement plays a major role in participants' motivation, and that this effect is strongest for participants with high self-conception.

Predictor	β	SE β	Walds χ^2	df	p	e^{β}
Main Effect						
Intercept	-8.47	0.24	1815.08	1	<.0001	NA
Score	0.16	0.008	426.56	1	<.0001	1.17
Interaction Effect						
Intercept	6.68	0.74	90.14	1	<.0001	NA
Score*Self-Conception[worse]	0.08	0.05	3.13	2	.077	1.08
Score*Self-Conception[equal]	0.009	0.04	0.01	2	.90	1.009

Table 1: Logistic regression table for Study 1: The model predicts the likelihood for a participant sharing the experiment. e^{β} is the odds ratio, i.e., the odds of the variable having an effect.

STUDY 2: EFFECT OF RESULT PRESENTATION

Results of the first study indicated that people taking the Social Intelligence Test were primarily motivated by the desire to improve the positivity of their self-concept (the self-enhancement motivation) with self-verification playing a smaller role. Given that volunteer-based online experiments rely on participants' sharing behavior (and thus, recruiting others), the goal of our second study was to find out how those with lower scores could be encouraged to share if their scores were presented differently. The three designs were:

Upward Comparison "Your score is lower than XX% of all participants."

Downward Comparison "Your score is higher than YY% of all participants."

Downward/Equal Comparison "Your score is equal or higher than ZZ% of all participants."

The difference between how Downward and Downward/Equal comparisons are perceived may be large, as one more point may make up to 10% of participants. For example, the score of 25 is *higher than 35% of all participants*, but *equal or higher than 45% of all participants*.

Given that our participants are primarily motivated by self-enhancement, we hypothesized that presenting the results in the most positive light (the Downward/Equal Comparison) would result in more sharing than either Upward Comparison (negative framing of feedback) or Downward Comparison (lower apparent magnitude of the participant's achievement).

We additionally hypothesized an interaction effect between self-conception and comparison presentation: while participants with a more positive self-conception would prefer Downward/Equal Comparison, participants with more negative self-conceptions would prefer Upward Comparison.

Participants

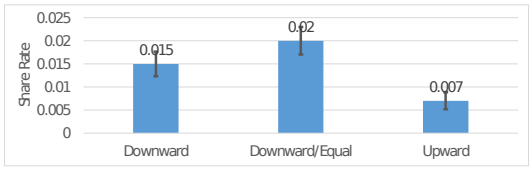
Participants were 6,011 online volunteers (aged 11–70, 45.7% male, 45.3% female, 1.6% other, the rest declining to answer). No financial compensation was given.

Experiment Design and Procedure

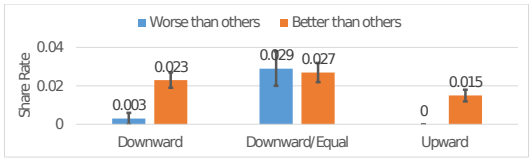
Task and procedure were the same as in Study 1 with the exception that for this study, we randomly assigned participants to one of the three result presentation conditions. Hence, we used a between-subjects design with one factor: presentation of the results.

Analysis

As in Study 1, we used logistic regression to analyze the data with presentation of the results and self-conception as the two



(a)



(b)

Figure 3: (a) Share rates for Downward, Downward/Equal and Upward comparison of 6,011 participants. Share rate is the fraction of participants that shared their results with others. (b) Share rates for Downward, Downward/Equal, and Upward, for different self-conceptions.

independent variables. The share rate was calculated the same way as in Study 1. In this study, we additionally controlled for score by including it as a covariate. We used the same exclusion criteria as in Study 1.

Results

We observed a significant main effect of the way the score was communicated on the likelihood of participants sharing the test with their peers ($\chi^2_{2,N=6011} = 14.00, p < .001$). Post hoc pairwise comparisons with Bonferroni correction showed that participants receiving Upward comparison (M=0.007) were significantly less likely to share than participants receiving Downward Comparison (M=0.015, $\chi^2_{1,N=4324} = 6.60, p < .05$) or participants receiving Downward/Equal Comparison (M=0.02, $\chi^2_{1,N=4397} = 14.22, p < .001$). There was no significant difference between participants receiving Downward/Equal or Downward Comparison ($\chi^2_{1,N=4407} = 1.44, p = .45$).

We also observed a significant interaction effect between self-conception and the way the score was communicated ($\chi^2_{4,N=6011} = 9.63, p < .05$). This effect was different from what we expected: People who expected to do poorly benefited more from the most optimistic presentation of the results than people who expected to do well.

DISCUSSION AND FUTURE WORK

In this work, we asked the question how to get people to share more often their results on online behavioral experiment platforms. To answer this question, we investigated whether people on volunteer-based online experiment platforms are driven by self-evaluation, self-enhancement or self-assessment. To do this, we analyzed participants' sharing behavior through the lens of social comparison theory in the context of the Social Intelligence Test. In Study 1, we observed a main effect of a participant's score on the probability that the participant would share the test on a social networking site. We also found an interaction effect between score and self-conception. These findings suggest that people taking the Social Intelligence Test are mostly driven by self-enhancement with some self-verification motives.

Predictor	β	SE β	Walds χ^2	df	p	e^{β}
Intercept	-9.71	1.41	58.69	1	<.0001	NA
Score	0.17	0.05	13.93	1	.0002	1.18
Presentation[Down]	-0.25	0.73	0	1	1	0.78
Presentation[Down/Equal]	1.60	0.59	13.94	1	.0002	4.95
Self-Conception[Low-Medium]	0.29	0.60	0.64	1	.42	1.34
Self-Conception[High-Medium]	0.77	0.27	9.97	1	.0016	2.16
Presentation[Down]*Self-Conc.[Low-Medium]	0.57	.79	0	1	1	1.77
Presentation[Down]*Self-Conc.[High-Medium]	-.13	.35	0.10	1	0.76	0.88
Presentation[Equal/Down]*Self-Conc.[Low-Medium]	-1.60	.67	9.07	1	.0026	0.20
Presentation[Equal/Down]*Self-Conc.[High-Medium]	0.35	.37	0.79	1	0.37	1.42

Table 2: Logistic regression table for study 2: The model predicts the likelihood for a participant sharing the experiment.

Given that volunteer-based online experiments rely on participants' sharing behavior (and thus, recruiting others), we conducted Study 2, in which we manipulated the presentation of participants' individual results. We found that the most positive social comparison condition resulted in the highest share rates, but this time we did not observe the expected moderating effect of pre-existing self-conception: Surprisingly, people with lower self-conception were more (rather than less) likely to share in the most positive social comparison condition. In contrast, the *self-verification* motive — partially supported by the results of the first study — would have predicted that lower self-conception might have led to higher sharing rates in the case of more negative score presentation [17]. Finding out why the self-conception had an opposite effect in our two studies will be an interesting further research direction.

Overall, these results show that high-performers are more likely to share their results, which could bias the participant pool if, for example, an online social network mainly consisted of other high-performers (there is ample evidence of homophily in social networks along many dimensions, including cognitive abilities [11]). Our results imply that we could manipulate the presentation of the results not just to increase the overall sharing rate, but also to reduce the difference in sharing rates between high and low performers. By personalizing participant feedback for self-conception and participant performance using social comparison, designers can significantly improve share rates and therefore might increase the overall impact and quality of their online experiment. For example, the feedback for participants with a low self-conception could emphasize the positivity of the results.

A limitation of our study is that we did not measure whether participants believed that their ability to read the emotions of others was a fixed trait or a skill that could be improved. Another limitation of our approach is that we did not measure the actual reasons for sharing. Measuring and understanding the role of the participant's need for self-affirmation, i.e. the need for projecting a certain image to others, might be a promising direction, as self-affirmation has been shown to be an important motive for sharing information with others [2]. These directions might affect the impact of social comparison [15] and should be considered in future work.

Acknowledgements. The authors would like to thank Ofra Amir for helpful comments and suggestions.

Online Appendix. The data used in this work can be found at http://iis.seas.harvard.edu/resources/.

REFERENCES

1. 2016. Implicit Association Test. https://implicit.harvard.edu/implicit/. (2016). Accessed: 2016-12-22.
2. Aliosha Alexandrov, Bryan Lilly, and Emin Babakus. 2013. The effects of social-and self-motives on the intentions to share positive and negative word of mouth. *Journal of the Academy of Marketing Science* 41, 5 (2013), 531–546.
3. Simon Baron-Cohen, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb. 2001. The Reading the Mind in the Eyes test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry* 42, 2 (2001), 241–251.
4. Moira Burke, Cameron Marlow, and Thomas Lento. 2009. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 945–954.
5. Yan Chen, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li. 2010. Social comparisons and contributions to online communities: A field experiment on movielens. *The American economic review* 100, 4 (2010), 1358–1398.
6. Rebecca L Collins. 1996. For better or worse: The impact of upward social comparison on self-evaluations. *Psychological bulletin* 119, 1 (1996), 51.
7. David Engel, Anita Williams Woolley, Lisa X. Jing, Christopher F. Chabris, and Thomas W. Malone. 2014. Reading the mind in the eyes or reading between the lines? Theory of mind. 12 (2014). DOI: http://dx.doi.org/10.1371/journal.pone.0115212
8. Leon Festinger. 1954. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.
9. Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review* 19, 5 (2012), 847–857.
10. Frederick X Gibbons. 1986. Social comparison and depression: Company's effect on misery. *Journal of Personality and Social Psychology* 51, 1 (1986), 140.
11. Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001), 415–444.
12. Petromil Petkov, Felix Köbler, Marcus Foth, and Helmut Krcmar. 2011. Motivating domestic energy conservation through comparative, community-based feedback in mobile and social media. In *Proceedings of the 5th International Conference on Communities and Technologies*. ACM, 21–30.
13. Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1364–1378.
14. Constantine Sedikides. 1993. Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of personality and social psychology* 65, 2 (1993), 317.
15. Diederik A Stapel and Willem Koomen. 2000. Distinctiveness of others, mutability of selves: Their impact on self-evaluations. *Journal of personality and social psychology* 79, 6 (2000), 1068.
16. Jerry Suls, Rene Martin, and Ladd Wheeler. 2002. Social comparison: Why, with whom, and with what effect? *Current directions in psychological science* 11, 5 (2002), 159–163.
17. William B Swarm Jr. 1983. Self-verification: Bringing social reality into harmony with the self. *Social psychological perspectives on the self* 2 (1983), 33–66.
18. Joanne V Wood. 1989. Theory and research concerning social comparisons of personal attributes. *Psychological bulletin* 106, 2 (1989), 231.
19. Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.