"You Can Always Do Better!" The Impact of Social Proof on Participant Response Bias

Aditya Vashistha[†] **Fabian Okeke**[§] **Richard Anderson**[†] **Nicola Dell**[§]
[†]University of Washington
{adityav,anderson}@cs.washington.edu
[§]The Jacobs Institute, Cornell Tech
{fno2,nixdell}@cornell.edu

ABSTRACT
Evaluations of technological artifacts in HCI4D contexts are known to suffer from high levels of participant response bias—where participants only provide positive feedback that they think will please the researcher. This paper describes a practical, low-cost intervention that uses the concept of social proof to influence participant response bias and successfully elicit critical feedback from study participants. We subtly exposed participants to feedback that they perceived to be provided by people 'like them', and experimentally controlled the tone and content of the feedback to provide either positive, negative, or no social proof. We then measured how participants' quantitative and qualitative evaluations of an HCI artifact changed based on the feedback to which they were exposed. We conducted two controlled experiments: an online experiment with 245 MTurk workers and a field experiment with 63 women in rural India. Our findings reveal significant differences between participants in the positive, negative, and no social proof conditions, both online and in the field. Participants in the negative condition provided lower ratings and a greater amount of critical feedback, while participants in the positive condition provided higher ratings and a greater amount of positive feedback. Taken together, our findings demonstrate that social proof is a practical and generalizable technique that could be used by HCI researchers to influence participant response bias in a wide range of contexts and domains.

Author Keywords
HCI4D; ICTD; response bias; social influence; social proof.

INTRODUCTION
HCI researchers and practitioners are increasingly interested in engaging with marginalized communities to design new technologies to have a positive impact on people's lives, including low-income [18, 51], low-literate [43, 52], rural [4, 55, 64], disabled [47, 56], and other communities [14, 26, 62]. One characteristic that these diverse contexts share is that there are frequently large differences between researchers and their participants, such as differences in background, social status, culture, language, education, and technical expertise. Unfortunately, these differences have been shown to substantially impact researchers' efforts to evaluate their new designs or interventions. In particular, usability studies and field evaluations frequently suffer from high levels of participant response bias [15], defined as the extent to which participants provide researchers with feedback or results that will please the researchers or help to achieve the research goals [22, 46]. As a result, many researchers have found it challenging to obtain critical or negative feedback from participants that could help them to improve their designs or interventions [2, 26]. Although participant response bias is present in *all* studies with human participants, its effects have been shown to be significantly amplified in studies involving marginalized communities [15]. Although a growing number of studies acknowledge the potential for participant response bias to impact their results (*e.g.*, [29, 38, 54]), little progress has been made on developing practical tools and techniques that could help HCI researchers to cope with response bias in their studies.

The goal of our research is to fill this gap by contributing a generalizable technique to influence response bias and encourage participants to provide constructive feedback, particularly critical feedback. We conducted a series of controlled experiments that systematically influence participant response bias using the concept of social proof (or informational social influence) from the field of social psychology [16, 53]. Social proof refers to the psychological phenomenon where people assume the actions of others in an attempt to reflect correct behavior in a given situation. In other words, when people are uncertain about what to do, they assume that the people around them, such as experts, celebrities, and friends, have more knowledge about what should be done.

We conducted two controlled experiments: an online experiment with 245 workers recruited through Amazon's Mechanical Turk (MTurk) platform, and a field experiment with 63 low-income, low-literate participants in rural India. Working within an existing HCI project, the *Projecting Health* project in India [36, 37, 58], we asked participants to evaluate a community-created video. In both experiments, participants were randomly assigned to one of the three conditions: positive social proof, negative social proof, and no social proof (*i.e.*, baseline). Prior to watching the video, participants in the positive and negative conditions received social proof through subtle exposure to three positive and negative 'video reviews', respectively, that they perceived to have been provided by other participants 'like them'. Participants in the baseline

CHI 2018, April 21–26, 2018, Montréal, QC, Canada.

http://dx.doi.org/10.1145/3173574.3174126

condition were not exposed to any reviews. We hypothesized that participants in the positive and negative conditions would provide feedback that conformed to the tone of the reviews they encountered. We structured each experiment to examine the effect of social proof on participants' quantitative ratings and qualitative feedback on the artifact being evaluated.

At a high level, our findings show that social proof had a profound effect on participants' evaluations of the artifact in both the online experiment and the field experiment. We found statistically significant differences between the three experimental conditions for *both* the quantitative ratings *and* the qualitative feedback provided by participants. In general, participants in the negative social proof condition gave the video lower ratings and provided a greater amount of critical feedback than participants in the baseline condition. On the other hand, participants in the positive social proof condition gave the video higher ratings and provided a greater amount of positive feedback than participants in the baseline condition. These findings confirm that social proof is an effective way to influence response bias and, in particular, that negative social proof is an effective way to elicit critical feedback from participants, both online and in the field.

Our intervention possesses several key benefits that make it practical for researchers and practitioners to implement. For example, the technique effectively elicits negative feedback even when participants are evaluating a single artifact that is known to be associated with the researcher [15]. It is also a low-cost intervention that does not require any additional equipment beyond the artifact being evaluated. Moreover, the procedure is relatively simple to understand for organizations working in the field and for participants. Finally, by conducting two experiments in different contexts—with MTurk workers online and with low-literate participants in the field—we demonstrate that our intervention could be applied by HCI researchers to a wide range of contexts and domains.

BACKGROUND AND RELATED WORK

There has been a growing concern within the HCI community about the effects of participant response bias in evaluations of new designs or technological artifacts. A number of studies have discussed the difficulty of eliciting critical or negative feedback from participants, particularly in HCI for Development (HCI4D), where there are often large social and cultural differences between researchers and participants [2, 26, 29, 38]. Brown et al. studied the challenges of conducting HCI trials in *"the wild"* and documented the effects of demand characteristics [46], in which participants adjust their behavior to match the expectations of the researchers. Dell et al. [15] conducted a study in India to quantify the effects of participant response bias, and found that participants were 2.5 times more likely to prefer a technological artifact that they believed to have been developed by the researcher, even when the alternative was identical. In addition, when the researcher was a foreigner who required a translator, the response bias with low-income Indian participants increased to five times. Trewin et al. [54] analyzed participants' subjective Likert-scale responses in accessibility studies, and found that participants in non-anonymous studies gave more positive ratings than those in other studies.

HCI researchers have suggested a variety of approaches to try and reduce participant response bias. Brown et al. [7] suggested postponing the evaluation of technologies altogether until the technologies can be better understood by users. Chavan [9] encouraged participants to submit critical feedback by situating user studies within dramatic storylines. Molapo et al. [45] recommended role playing and skits to motivate front-line workers to share their opinions. Other researchers have explored reducing response bias by dissociating themselves from designs or artifacts [48, 59], limiting direct contact with participants [23, 57], or spending more time with participants in the field in the hope that they would be comfortable enough to provide critical feedback [21]. However, for the most part, the impact of these approaches on reducing response bias has not been systematically quantified.

Our study uses the concept of social proof from the field of social psychology to influence response bias and encourage participants to provide constructive, critical feedback to researchers. Social proof [53] refers to the psychological phenomenon of assuming the actions of others in an attempt to reflect correct behavior. Also known as informational social influence, social proof occurs when people experience uncertainty about what decision they should make, assume that the people around them possess more (or better) information, and accept information gleaned from other people's behavior as evidence about reality [16, 17]. Examples of social influence include presuming that the food at a restaurant is good because the queue is long, endorsing a political candidate because everyone else approves of the person, or giving a product excellent reviews because an expert or celebrity positively reviewed the same product. The effects of social proof have also been shown to differ across countries and cultures [10]. For example, prior research has demonstrated that people living in collectivist cultures (such as India) tend to conform to social proof more often than those in individualist cultures [5].

There is a growing interest within the HCI community in understanding and applying the concept of social proof to a range of application domains, such as interpreting graphical information and visualizations [27], influencing user opinions in recommendation systems [11], prompting people to explore and adopt better security habits [12, 13], and affecting people's intention to adopt privacy behaviors [44]. Several scholars have also studied social proof, or the broader concept of social influence, in the context of online platforms. For example, Bond et al. [6] found that showing people that their Facebook friends have voted increased voter turnout. Burke et al. [8] showed that social learning played an important role in influencing how novice Facebook users interact with the platform. Kramer [34] found that people were more likely to share emotional content that matched the content shared by their friends. Malu et al. [41] used social influence to encourage people to contribute personal content to an online community. Finally, Wu and Huberman [63] examined social influence in the context of online opinions, news, and product reviews, and found that awareness of others' opinions leads to increasingly extreme views. Our paper extends this body of work by conducting controlled experiments that measure the impact of social proof in the evaluation of an HCI artifact.

To the best of our knowledge, ours is the first paper to apply the concept of social proof to influence response bias in HCI. We are also the first to study the effects of social proof with low-literate populations in resource-constrained settings.

INTERVENTION DESIGN

We situated our study in the context of *Projecting Health*, an existing community-driven social and behavior change intervention to improve maternal and neonatal health in rural India [36, 37, 58]. *Projecting Health* empowers community-based organizations to produce videos that feature local people discussing key health messages in a local dialect. Accredited social health activists (ASHAs) share the videos in group sessions with women via portable projectors. The project is currently operating in over 125 villages in Uttar Pradesh with 170 mother groups. Thus far, 80 videos have reached an estimated 100,000 people through 12,000 screenings.

A critical component of *Projecting Health* is to obtain feedback from stakeholders to ensure that videos are suitable for dissemination in rural areas. During the initial phase of the project, several participants attended video disseminations out of the curiosity to watch videos featuring people 'like them', and also because of the novelty of accessing health information via videos. Since these effects lead only to short-term engagement, the *Projecting Health* staff has aimed to design improved videos that low-income, low-literate women find engaging, interesting, informative, and entertaining. However, the staff has reported great difficulties in obtaining any critical feedback from rural women because of high levels of participant response bias. Often they receive positive feedback, or feedback that lack details. During an informal conversation in the field, the program manager of *Projecting Health* described:

> *"The biggest challenge [in Projecting Health] is to improve the quality of the videos. If a video is of good quality, useful, and entertaining, people will automatically watch it again and share it with others. However, it is almost impossible to get constructive feedback in rural areas. They [people in rural areas] always say the video is very nice and there is no need of improvement."*

The goal of our research is to contribute techniques for influencing response bias and encouraging participants to provide constructive, critical feedback. A key design consideration is to ensure that the intervention is easy to administer and generalizable to a variety of settings. To this end, we designed an intervention that uses social proof to persuade participants to provide substantive critical feedback. We conducted a between-subjects study where participants were randomly assigned to one of the three conditions: positive social proof, negative social proof, and no social proof (*i.e.*, baseline). Participants in the positive and negative conditions were subtly exposed to a set of positive and negative video reviews, respectively. In reality, we authored the reviews in collaboration with the *Projecting Health* team, and experimentally controlled their content and tone to provide participants with either *positive* or *negative* social proof. For example, a review that we created to provide participants with positive social proof is: *"It is very important for people to learn this information. The video content is great! The health messages are very easy to understand."* By contrast, an example of a review that we created to provide participants with negative social proof is: *"Nobody can understand the content of this video. The message is not clear. This will never help anyone."* We hoped that showing participants these 'reviews', which they perceived to have been given by other participants 'like them', would encourage them to provide their own feedback on the video. In particular, we hypothesized that if participants perceived that other people had contributed negative feedback, they may feel comfortable to critique the artifact being evaluated.

After participants received positive, negative, or no social proof, they watched a three-minute *Projecting Health* video about safe drinking water. The video featured a discussion between an ASHA, two representatives of a village-level committee, and a local doctor on how to keep ground water clean. The *Projecting Health* staff recommended this video since it had both strengths (*e.g.,* important topic and new knowledge for most people) and weaknesses (*e.g.,* unskilled actors and uninteresting storyline). After watching the video, participants completed a survey in which they provided quantitative ratings of the video along with unstructured qualitative feedback.

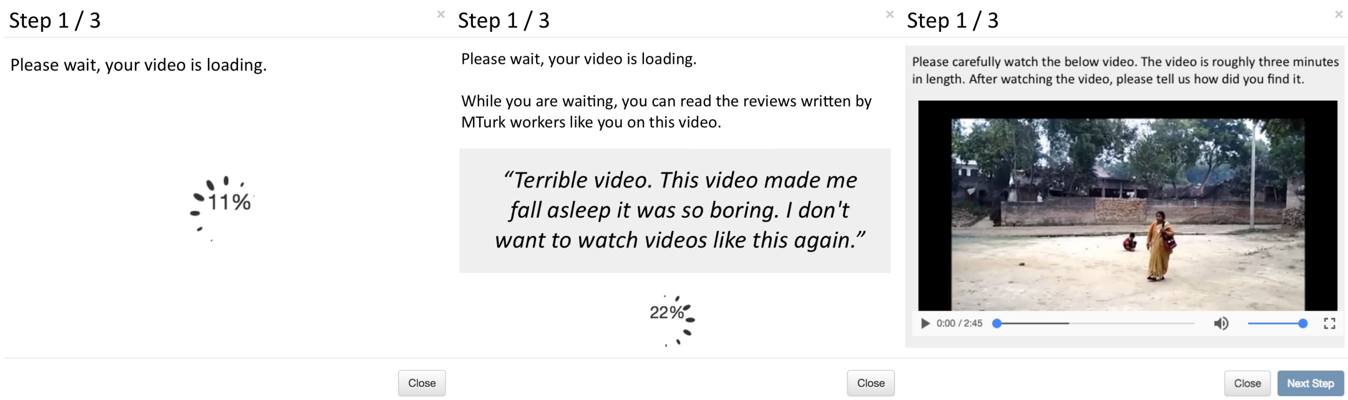
We conducted two experiments to evaluate the impact of our social proof intervention with participants in different contexts: (1) an online study with MTurk workers, and (2) a field study with low-income women in rural India. Each experiment focused on answering the following research questions:

RQ1: How does social proof impact participants' quantitative ratings of an intervention?
Many HCI studies evaluate new designs, products, or interventions by asking participants to rate their subjective experiences or opinions on the intervention using quantitative instruments such as a Likert scale [39]. We hypothesized that participants' quantitative ratings of a *Projecting Health* video would be influenced by the kinds of reviews that they saw before watching the video. For example, participants who were exposed to negative video reviews would submit more negative ratings than those who were exposed to positive reviews.

RQ2: How does social proof impact the qualitative feedback provided by participants?
We hypothesized that participants would be influenced to provide qualitative feedback of a tone similar to the reviews that they saw before watching the video. For example, participants who saw negative reviews would provide more negative qualitative feedback than those who saw positive reviews.

EXPERIMENT 1: STUDY ON MTURK

Our first experiment analyzed the impact of social proof in an experiment conducted with 245 participants recruited through MTurk—an online crowdsourcing marketplace where workers complete tasks such as categorization, translation, and surveys in exchange for small monetary payments [1]. An increasing number of HCI studies recruit MTurk workers as participants [31, 32, 42] since MTurk makes it easy to recruit large numbers of geographically distributed populations at a relatively low cost. Since the prevalence of HCI studies conducted on MTurk is rapidly increasing, we examined how social proof might impact the evaluation of an HCI artifact by MTurk workers.



(a) Video loading without any reviews (baseline). (b) Video loading with negative reviews (c) Video playing after it is loaded.

Figure 1: Screenshots from the MTurk experiment (shown in English for readability, although the experiment was in Hindi).

Authoring and Validating Reviews

In collaboration with the *Projecting Health* staff, we authored thirty positive and thirty negative reviews in Hindi that commented on the video's production quality, content, acting, storyline, duration, and entertainment value. The positive and negative reviews were similar in length and attributes being evaluated. The average length of reviews was 26 words (SD = 6 words). To ensure that the reviews were perceived as positive or negative, we recruited 125 MTurk workers from India. Each worker was randomly assigned ten reviews to read and rate on a five-point Likert scale from very negative to very positive. Since the reviews were in Hindi, we restricted participation to MTurk workers who could understand Hindi by providing the instructions and prompts in Hindi.

Workers who rated the reviews were 32 years old, on average. Eighty-eight workers were male, 34 were female, and three did not indicate their gender. One worker had completed secondary school, three had completed high school, 76 had finished a bachelor's degree, and 45 had finished a master's degree. The positive reviews received an average rating of 4.6 ($SD = 0.23$) while the negative reviews received an average rating of 1.7 ($SD = 0.31$). For the final experiment, we selected the ten highest rated and ten lowest rated reviews.

Procedure

Since the *Projecting Health* video as well as the reviews were in Hindi, we restricted participation to MTurk workers who were located in India, and were comfortable reading and understanding Hindi. To participate in our study, MTurk workers needed to answer a basic arithmetic question (*i.e.*, what is ten plus seven) displayed in Hindi. Workers who provided the correct response were directed to an external webpage that contained the study instructions and prompts in Hindi.

Each consenting MTurk worker was randomly assigned to one of the three experimental conditions: positive social proof, no social proof (*i.e.*, baseline), or negative social proof. We balanced these three groups on participants' income, age, and education. Before showing participants the *Projecting Health* video, we purposefully introduced a thirty-second delay that we told participants was due to the video loading. In the baseline condition, participants simply saw a progress bar that took thirty seconds to reach 100% (see Figure 1a). In the positive and negative conditions, we used the delay to show participants three randomly selected reviews, each for ten seconds (see Figure 1b). After the thirty-second period was over, participants in all three conditions watched the video and provided their feedback. We requested participants to rate the video using a five-point Likert scale on four parameters: how much they liked or disliked the video (*likeability*), how useful the video was (*usefulness*), how entertaining the video was (*entertainment value*), and how much the video could be improved (*scope of improvement*). We also asked participants to share their subjective feedback on the video. To filter participants that might not have paid attention to the video, we asked a simple validation question about the subject matter of the video. We also collected participants' demographic information. The experiment lasted for around ten minutes and participants received USD 1 for their participation.

Participant Demographics

We recruited 245 MTurk workers for our experiment, with 84, 73, and 88 participants in the positive, baseline, and negative conditions, respectively. Since seven participants in the positive condition, and ten participants each in the baseline and negative conditions answered the validation question incorrectly, we removed their responses from our analysis. Table 1 shows the demographic characteristics for the MTurk participants who answered the validation question correctly. Participants came from sixty cities in India. All participants had access to a mobile phone and 45% of them shared their phone with family members. Almost 90% of them watched videos regularly and 97% had access to mobile Internet.

Data Analysis

We conducted a single-factor, between-subjects experiment with three levels. The single factor was *type of social proof* with the levels *positive*, *baseline*, and *negative*. We used non-parametric Kruskal-Wallis tests [35] to analyze differences in participants' Likert-scale ratings on likeability, usefulness, entertainment value, and scope of improvement. Post-hoc pairwise comparisons were conducted using Dunn's tests [20] with Bonferroni correction [19] for multiple comparisons.

Condition	No of workers	Male (%)	Age (years)	Education (years)	Family Income (USD/year)
Baseline	63	68	31	15.6	1191
Positive	77	71	32	15.4	1116
Negative	78	75	33	15.6	1100

Table 1: Demographic characteristics of MTurk participants.

We analyzed participants' qualitative feedback along several dimensions, including the number of participants who submitted feedback, the length of the feedback, the tone of the feedback, and whether participants provided *substantive* feedback. We defined feedback as *substantive* if participants provided concrete details on what they liked or disliked about the video or suggested specific points for improving it. To analyze the qualitative feedback, we recruited three Hindi speakers (1 male and 2 female) who read each review independently in a random order, and classified the tone of the feedback as positive, negative, or mixed, and noted whether the feedback was substantive. The reviewers were blinded to the experimental conditions. We used majority voting to break ties, and analyzed differences between the experimental conditions using Pearson's Chi-squared tests [49] or Fisher's exact test.

Results of MTurk Experiment

RQ1: Impact on Participants' Quantitative Ratings

Our first research question focuses on understanding the impact of the social proof intervention on participants' quantitative ratings of the video. Table 2 shows that participants in the positive condition rated the video highest on *likeability*, *usefulness*, and *entertainment value*. In contrast, participants in the negative condition rated the video lowest on *likeability*, *usefulness*, and *entertainment value*. Participants in the negative condition found greater scope for improving the video than participants in the other conditions. Results of Kruskal-Wallis tests indicated that these differences were significant for all four parameters: *likeability* (p < .001), *usefulness* (p = .001), *entertainment value* (p < .001), and *scope of improvement* (p < .001). Post-hoc pairwise comparisons between experimental conditions indicated significant differences between the positive and negative conditions, and the negative and baseline conditions, for all parameters (see Table 3). These findings suggest that negative social proof effectively decreased participants' quantitative ratings of the video.

RQ2: Impact on Participants' Qualitative Feedback

Our second research question focuses on understanding the impact of social proof on the qualitative feedback provided by participants. We found that a greater percentage of participants provided feedback in the positive (69%) and negative (76%) conditions than in the baseline condition (63%). In addition, the average length of feedback submitted by participants in the positive condition (20 words) and negative condition (19 words) was greater than the baseline condition (17 words). This may indicate that participants who were exposed to other reviews wrote longer feedback since they wanted to conform to other workers who submitted the subjective feedback. However, these differences were not statistically significant for either the number of participants who gave feedback or the length of the feedback.

Condition	Likeability	Usefulness	Entertainment value	Scope of improvement
Baseline	3.7	3.8	3.4	3.1
Positive	4.1	3.9	3.6	2.9
Negative	3.2	3.2	2.8	3.7

Table 2: Average Likert-scale ratings of the video by participants in the MTurk experiment.

Condition	Baseline	Negative
Positive	$L^§$	$L^§$ $U^†$ $E^§$ $S^§$
Baseline		$L^§$ U* E* $S^†$

Table 3: Pairwise comparison of experiment conditions on (L)ikeability, (U)sefulness, (E)ntertainment value, and (S)cope of Improvement (* is p < .05, † is p < .01 and § is p < .001).

Table 4 shows the classification of participants' qualitative feedback as positive, negative, or mixed (*i.e.*, it contained both positive and negative elements). An example of a participant's negative feedback is, *"The conversation was very unnatural. The flow of ideas can be improved. Dialogue delivery can be improved."* By contrast, an example of mixed feedback is:

> *"This video contained good information most of which I was unaware of. It was useful for me, but the video could be improved using graphics and other video enhancing ways. The current video is plain and monotonous. "*

Participants in the positive condition submitted more positive and mixed comments, and fewer negative comments, than those in the baseline condition. In contrast, participants in the negative condition submitted more negative and mixed comments, and fewer positive comments, than those in the baseline condition. These differences were significant ($\chi^2(4, N = 152) = 23.2$, $p < .0001$), which indicates that negative social proof led participants to submit more negative qualitative feedback, and vice versa for the positive condition.

The qualitative feedback provided by participants was also classified as either being substantive (*i.e.*, containing concrete suggestions or discussion) or not. An example of feedback that was *not* substantive is, *"This is a good video,"* while an example of a substantive feedback is:

> *"Very nice video that gives us a very important message. Disease is spreading in village due to polluted water. Hand pumps should be very deep and we should try to keep the surrounding area very neat and clean."*

Table 4 shows that 74% of participants in the positive condition and 85% of participants in the negative condition provided feedback that was judged as substantive, compared to 68% of participants in the baseline condition. These differences were not statistically significant though. Analysis of the negative and mixed feedback indicated that participants provided several suggestions, such as improving the acting (N=48), creating interesting storyline (N=24), enhancing entertainment value (N=16), and adding graphics and examples (N=8), among others. Analysis of the comments that contained positive and mixed feedback indicated that 81 participants found the video useful and informative, seven liked the location

Condition	Total # comments	Positive feedback	Mixed feedback	Negative feedback	Substantive feedback
Baseline	40	65%	25%	10%	68%
Positive	53	68%	24%	8%	74%
Negative	59	29%	41%	30%	85%

Table 4: Classification of the feedback provided by participants in the MTurk experiment.

where it was shot, and five appreciated the acting skills of people with rural background featured in the video.

In summary, our experiment with MTurk workers demonstrated that social proof influenced participants' quantitative ratings and improved their qualitative feedback. Participants who were exposed to positive reviews perceived the video more positively, provided positive ratings, and supported their ratings with substantive positive comments. Similarly, participants who were exposed to negative video reviews were more critical of the video, submitted lower ratings, and wrote substantive negative and mixed feedback critiquing the video.

EXPERIMENT 2: FIELD STUDY IN RURAL INDIA

Our second experiment examined how social proof might impact a field study in which a researcher conducts a face-to-face evaluation of an HCI intervention with participants. In particular, prior research has shown that evaluations of HCI artifacts with participants in HCI4D contexts may suffer from high levels of participant response bias [15]. Our work directly engages with these contexts through an *in situ* experiment with low-income, low-literate women in rural India.

Authoring and Validating Reviews

Conducting our experiment within the context of an ongoing HCI project introduced a number of considerations. In particular, the *Projecting Health* staff requested that we create reviews for the experiment that do not critique key aspects of *Projecting Health*'s design such as the use of local dialect and actors. Thus, we authored a new set of 15 positive and 15 negative reviews that focused only on other video attributes like production quality, content, storyline, duration, and entertainment value. The positive and negative reviews were similar in length as well as the attributes being evaluated. The average length of the reviews was 30 words (*SD* = 6 words).

To ensure that the reviews were successfully perceived as positive and negative, we recruited three *Projecting Health* staff to read each review in a random order and rate it on a five-point Likert scale from very negative to very positive. The staff members (1 male and 2 females) were native Hindi speakers and had completed master's degrees. They had been associated with *Projecting Health* since its inception and had a deep understanding of the rural communities it serves. The positive reviews received an average score of 4.5 (*SD* = 0.33) and the negative reviews received an average score of 1.6 (*SD* = 0.4). For the field experiment, we selected the ten highest rated and ten lowest rated reviews.

Procedure

With the support of NYST, a grassroots organization implementing *Projecting Health* in rural Uttar Pradesh, we recruited 63 low-income, low-literate women to participate in the field experiment. Typically, the *Projecting Health* staff show videos to community members using a portable projector and request feedback from them to improve the videos. We designed our procedure to mimic this existing feedback collection routine.

To avoid contamination and confusion, we wanted to ensure that participants in one condition are unaware of the activities assigned to participants in other conditions. This was easy to execute in the MTurk experiment since the participants were geographically distributed and used their own personal computers. However, in rural India, which has a highly collectivist culture, assigning participants to different conditions without contamination and confusion was challenging, especially since the field staff reported that women often come together in a group to watch the videos. We also could not share the purpose of our research experiment with participants beforehand since doing so might have influenced the study outcome. Thus, to avoid any contamination and confusion among participants, we along with the field staff identified three villages that were comparable to each other in terms of socioeconomic status, education, population size and distribution, and availability of resources such as health centers and schools. We then randomly assigned each village to either positive social proof, no social proof, or negative social proof condition, with all participants in the village assigned to the same condition.

In each village, the local ASHA asked women to attend the screening of a new *Projecting Health* video. Once we had a quorum, a local staff member told women that a researcher will show a three-minute *Projecting Health* video one by one to each participant, and ask questions to understand the strengths and weaknesses of the video. Participation in the experiment was voluntary. We asked consenting participants to wait in a specific area for their turn (see Figure 2a). We set up a portable projector and speakers in a room for the researcher (male, 30 years, native Hindi speaker) to show the video and ask questions (see Figure 2c). To ensure that participants were *subtly* exposed to feedback that they perceived to be from people 'like them', we set up a staging area where we asked the participant who would be next to visit the researcher to wait for her turn (see Figure 2b). In the staging area, two staff members, acting as confederates, were tasked to social proof the participant by acting out the skit described below.

The first confederate (male, 26 years) invited the participant to the staging area and asked her to sit next to the second confederate (female, 32 years) while she waited for her turn to interact with the researcher. The second confederate pretended reading a stack of feedback questionnaires that, in reality, contained the video reviews we had authored. The first confederate then casually inquired what the second confederate was reading. The second confederate replied that she was reading the feedback received from women in a neighboring village where the same activity was conducted yesterday. She then randomly selected three reviews and shared them with the first confederate and the participant. After casually reading the three reviews, she asked the waiting participant to also share her honest feedback with the researcher. Table 5 shows the script used for the experiment. The skit lasted less than three

(a) Waiting area for participants. (b) Two confederates social proofing a participant. (c) A participant watching the video projected on wall.

Figure 2: The three stages in the field experiment.

The second confederate is reading the feedback forms. In front of the participant, the first confederate asks the second:
First confederate: *"Sister, what are you reading?"*
Second confederate: *"Brother, yesterday we went to* [neighboring village] *where the researcher showed the three-minute Projecting Health video. He asked women for their feedback on the video and noted it down. I was just reading the feedback women gave to him. See, this women told him* [the confederate randomly selects a form and reads the feedback]"
First confederate: *"Hmm. What did others say?"*
Second confederate: *"Several women gave feedback. See* [points at another page], *this woman said* [reads a second review]"
First confederate: *"Hmm . . . Okay . . . "* [appreciating nod]
Second confederate: *"Yes brother, another sister told* [confederate selects and reads a third review]"
Second confederate turns to the participant.
Second confederate: *"Such detailed feedback is very important to improve the project. You should give your feedback without any hesitation like these women in the neighboring village did. He will also ask you information to fill this form. You should tell him what you like and what you don't like freely."*

Table 5: Translation of script used by the confederates to social proof participants.

minutes. We conducted ten rehearsals with the confederates to ensure that the skit appeared natural and finished on time. We decided against sharing the reviews with all participants as a group to ensure that each participant experienced approximately the same amount of delay between exposure to the reviews and interacting with the researcher. Only participants in the positive and negative conditions were exposed to the reviews. Participants in the baseline condition just waited for their turn while sitting next to the confederate.

After the researcher finished the study with the previous participant, the confederates sent the waiting participant to the researcher's room. The researcher showed the video to the participant, and then requested her to rate, on a five-point Likert scale, how much she liked the video, and what is the scope of improvement in the video. We only requested ratings on *likeability* and *scope of improvement* because the *Projecting Health* staff considered these two questions to be critical for their feedback process, and because they suggested that we limit the number of questions to reduce the time required to participate in the study as well as the possibility of confusing participants. In addition to quantitative ratings, the researcher also recorded qualitative feedback and demographic details. The entire interaction lasted around ten minutes.

Participant Demographics

Overall, 63 low-income, low-literate rural women participated in the field experiment, with 20 in the positive condition, 22 in the baseline condition, and 21 in the negative condition.

Condition	No of people	Age (years)	Family size	Education (years)	Family Income (USD/year)
Baseline	22	36	5.6	5.9	96
Positive	20	31	6.2	5.4	104
Negative	21	29	7.4	5.7	119

Table 6: Demographics of participants in the field experiment.

The majority (78%) owned a mobile phone while the rest used phones of family members. About 25% of participants reported watching videos on their phone, and only four had Internet access. Although 80% of the participants had previously watched a *Projecting Health* video, none of them had seen the video we used in the experiment. About 75% of the participants were homemakers, and the rest were farmers (N=8), laborers (N=3), domestic helpers (N=2), a cook (N=1), tailor (N=1), and teacher (N=1). Table 6 shows that participants possessed low levels of education and family income.

Data Analysis

We used the same statistical tests and procedures as the online experiment, including non-parametric Kruskal-Wallis tests to analyze differences in Likert-scale ratings, and Dunn's tests with Boneferroni correction for post-hoc pairwise comparisons. Qualitative feedback provided by participants were classified as containing positive, negative, mixed, and substantive feedback, and differences between conditions were analyzed using Pearson's Chi-squared tests or Fisher's exact test based on the values obtained in different conditions.

Condition	Likeability	Scope of improvement
Baseline	4.3	1.8
Positive	4.6	1.3
Negative	3.1	2.3

Table 7: Average Likert scale ratings for *likeability* and *scope of improvement* by participants in the field experiment.

Condition	Positive feedback	Mixed feedback	Negative feedback	Substantive feedback
Baseline	21	1	0	13
Positive	17	3	0	19
Negative	7	11	3	20

Table 8: Classification of the content and tone of participant feedback in the field experiment.

Results of the Field Experiment

RQ1: Impact on Participants' Quantitative Ratings

Table 7 shows that participants in the positive condition rated the video higher on *likeability* and lower on *scope of improvement* than the other two conditions. Conversely, participants in the negative condition rated the video lower on *likeability* and higher on *scope of improvement* than the other two conditions. A Kruskal-Wallis test also indicated significant differences in three conditions on *likeability* (H(2) = 22.5, $p < .0001$) and *scope of improvement* (H(2) = 7.6, $p = .02$). Post hoc tests with a Bonferroni correction indicated a significant difference ($p < .001$) in *likeability* for pairwise comparisons of all three conditions, and a significant difference ($p = .02$) in *scope of improvement* between the positive and negative conditions. These findings show that social proof effectively impacted participants' quantitative ratings of the video.

Comparing Quantitative Ratings Online vs. in the Field

We compared the ratings obtained in the field with those obtained in the online experiment. We found that the average *likeability* rating in the baseline condition of the field experiment was 4.3, which was significantly higher than the equivalent rating of 3.7 in the online experiment (H(1) = 5.1, $p = .02$). Moreover, the average score for *scope of improvement* in the baseline condition of the field experiment was 1.8, significantly lower than the equivalent score of 3.1 in the MTurk experiment (H(1) = 18.3, $p < .0001$). This suggests that either participants in the field genuinely liked the video more than the participants on MTurk, or that the response bias was much higher in a face-to-face field study with low-income, low-literate participants.

RQ2: Impact on Participants' Qualitative Feedback

Since we asked participants to provide qualitative feedback face-to-face, every participant provided at least some feedback, albeit with varying length and quality. Although some participants just said one word (*e.g., "good"*), many others gave detailed responses (*e.g.,* the longest feedback had 91 words). The average length of feedback was greater in the negative condition (45 words) than the positive condition (32 words) and the baseline condition (16 words). This difference was significant (H(2) = 25.4, $p < .001$), and post-hoc pairwise comparisons with Bonferroni correction showed significant differences between all conditions (all $p < .001$). This suggests that social proof, particularly negative social proof, successfully encouraged participants to provide more qualitative feedback.

Table 8 summarizes the classification of the content and tone of participants' qualitative feedback. Participants in the negative condition provided more mixed and negative feedback than those in the baseline and positive conditions. These differences were significant ($p < .0001$, Fisher's exact test), with post-hoc pairwise comparisons yielding significant differences between the positive and negative conditions ($p = .002$), and the baseline and negative conditions ($p < .001$). These findings indicate that negative social proof successfully encouraged participants to provide critical feedback on the video.

With respect to sharing concrete ideas for improving the video, Table 8 shows that more participants in the positive and negative conditions provided substantive feedback than participants in the baseline condition. These differences were statistically significant ($p = .002$, Fisher's exact test), with post-hoc pairwise comparisons revealing significant differences between the positive and baseline conditions ($p = .01$), and the baseline and negative conditions ($p = .01$). These findings show that exposure to reviews prompted participants to provide substantive suggestions for how to improve the video.

Our analysis of positive and mixed comments revealed that a majority of participants (N=43) found the video informative. Eight participants appreciated the local actors and efforts they *"put in to provide information while working tirelessly."* Other participants appreciated the production quality of the video and its entertainment value. A 28-year-old low-literate homemaker who was assigned to the positive condition suggested:

> *"I liked that information about diseases was given. The video taught us that we should not drink unsafe water and consume only clean water. I learned that we should use borewells that are deeper. I liked this information. You should also add songs. You should also add information about what precautions to take with tap water."*

Our analysis of negative and mixed reviews revealed that participants' comments contained actionable suggestions for how to make the video better, with many comments suggesting that *"of course, the video can always be improved."* Nine participants found the key health messages to be overwhelming since they felt that the video was *"rushed"* because of its short duration. Five participants suggested adding demonstrations (*i.e.*, acting things out instead of talking) to make the video more appealing. A 31-year-old low-literate homemaker who was assigned to the negative condition stated:

> *"A lot of information was not given in the video. The information was shared quickly in three minutes, making it difficult to remember. If they demo actions, show clean places, then it would be easier for us to understand."*

Another five participants suggested adding information about other related health subjects. Four participants recommended adding songs, dances, comedy, and photos of children to make the video more entertaining. Another four participants complained about the production quality, two did not like the acting, and one suggested using a more refined Hindi dialect. In summary, our analyses show that the social proof intervention effectively encouraged participants to submit greater amounts of qualitative feedback that contained useful and actionable suggestions for how to improve the video.

Although a few participants echoed the reviews they saw during the social proof exercise, most provided valuable feedback, including detailed suggestions for improving particular video attributes, new topics for future videos, detailed information on high-level themes they heard via social proof reviews, and concrete suggestions to improve the video. For example, although none of the social proof reviews mentioned demonstrations, several participants noted how demonstrations could improve their learning of the subject matter. Similarly, several participants recommended creating videos on new topics like nutritious food and waterborne diseases. Such suggestions were absent from the reviews we used for the social proof exercise. Although some participants did give feedback based on themes they heard via social proofing, they often shared specific details that expanded on these themes. For example, a participant who was exposed to a positive review containing *"I liked the information shared in the video"*, explicitly mentioned in her feedback that she *"liked the information that the hand-pump should not be broken and the house area should not be littered."* Finally, the social proof intervention gave agency to participants to make concrete suggestions that were absent from the social proof reviews, such as including songs or dances to make the video more engaging.

Comparing Qualitative Feedback Online vs. in the Field

We compared the qualitative feedback received in the field with that obtained in the MTurk experiment. Our findings show that the average length of the feedback received in the baseline conditions for both experiments was comparable: 16 words (*SD*=10) in the field experiment vs. 17 words (*SD*=11) in the online experiment. However, the length of the feedback received was significantly higher for the positive condition in the field (32 words, *SD*=14) vs. online (20 words, *SD*=9), ($H(1) = 12.2$, $p < .001$)). We found a similar trend for the negative condition: 45 words (*SD*=25) in the field vs. 19 words (*SD*=10) online, ($H(1) = 23.5$, $p < .0001$)). These differences could be due to the obligation the field participants may have felt to provide more feedback since they were face-to-face with the researcher and because others 'like them' were also providing feedback. Similar to the field participants, the feedback from the MTurk participants highlighted new topics and issues not present in the social proof reviews they saw.

DISCUSSION AND CONCLUSION

The goal of our research is to contribute a technique for influencing response bias and encouraging participants to provide constructive, critical feedback to researchers. We created an intervention that introduces social proof by subtly exposing participants to different kinds of feedback that they perceived to have been provided by other participants 'like them'. We evaluated the impact of our intervention through two controlled experiments: an online experiment with 245 MTurk workers and a field experiment with 63 low-income, low-literate women in India. At a high level, our findings show that social proof had a significant effect on participants' evaluations of an HCI artifact, both online and in the field. We found statistically significant differences between positive social proof, negative social proof, and no social proof conditions in *both* the quantitative ratings *and* the qualitative feedback provided by participants. Participants who were negatively social proofed provided lower ratings along with substantive, critical comments, while participants who were positively social proofed provided higher ratings and substantive, positive comments. These findings confirm that social proof is an effective way to influence response bias and, in particular, that negative social proof is an effective way to elicit critical feedback from participants, both online and in the field.

The feedback enabled the *Projecting Health* staff to understand the strengths and weaknesses of different video attributes, such as the production style, choice of accent, and informational content. In addition, the staff learned about specific topics for follow-up videos that would be of interest to their target population. Most importantly, the intervention made it easier for participating women to feel comfortable providing constructive, specific, actionable, and critical feedback. Such feedback had previously been very challenging for the *Projecting Health* staff to obtain, probably because the women were thankful for their efforts and did not want to hurt their feelings. The *Projecting Health* staff found the feedback they received very valuable. For example, based on the numerous suggestions from participants to make the videos more engaging, ten staff members of the grassroots organization took part in a three-day video production training in November 2017.

There were two main differences in our findings between the online and the field experiments. First, the baseline ratings in the field experiment were significantly higher for *likeability* and lower for *scope of improvement* than the corresponding ratings in the MTurk experiment. Since the *Projecting Health* video was exclusively designed for low-income, low-literate women, these differences in ratings could either because participants in the field genuinely liked the video more than participants online, or could be attributed to higher levels of response bias in the field. Second, although the amount of qualitative feedback provided by participants in the baseline condition was comparable between the field experiment and the online experiment, participants in the positive and negative conditions in the field provided significantly more feedback than the corresponding conditions online. This suggests that the social proof intervention, combined with the face-to-face nature of the interaction (*e.g.,* asking questions in-person), encouraged participants to provide more qualitative feedback.

Our social proof intervention has a number of key benefits that make it practical for researchers and practitioners to implement. One of our aims was to create an intervention that is generalizable and reproducible. We demonstrated that our intervention can be applied in two distinct contexts—an online experiment and a field study with low-literate participants in resource-constrained settings. In both experiments, we used the same experimental procedure with minor variations and received similar results that prove the efficacy of our social proof intervention. Compared to other techniques that aim to reduce response bias (*e.g.,* randomized response [60] and unmatched count [50]), our intervention is low-cost, practical, easy to understand for organizations and participants, reproducible in different contexts (as we demonstrated), effective for both quantitative and qualitative feedback, and elicits critical feedback even when participants are evaluating a single artifact that is known to be associated with the researcher [15]. Taken together, these benefits suggest that, with a small amount of adaptation (described below) the intervention could be used by HCI researchers in a wide range of contexts and domains.

Challenges and Design Recommendations

We now discuss the challenges we faced in executing our experiments, and recommendations for researchers interested in using our intervention. Although the intervention is designed to be an add-on to any artifact being evaluated, its efficacy is dependent on how *subtly* participants are exposed to social proof. On MTurk, we introduced a fake delay in loading the video and used that time to show video reviews. In the field, we created a skit that exposed participants to feedback that they perceived to have been provided by women in neighboring villages. Researchers in other contexts will need to find new ways to subtly expose participants to social proof.

We encountered several practical challenges in the field. For example, it was difficult to find sufficient space to conduct the experiment. We needed three physical spaces (waiting area, staging area, and researcher's room) to avoid contamination and confusion among the participants. We coped with the space challenge by using verandahs, porches, and lawns as waiting or staging areas. We were also concerned about participants returning to the waiting area after interacting with the researcher and sharing their experience with other participants waiting for their turn. Although there is no foolproof plan for such scenarios, we simply asked participants to not return to the waiting area. Similarly, although we limited MTurk workers to participate in our study only once, we do not know whether they were aware of the other conditions since prior work has shown that MTurk workers in India frequently communicate with each other [24]. However, although such strict controls were necessary due to the controlled nature of our experiment, organizations who are simply trying to elicit critical feedback from participants could allow participants to share their experiences, with the expectation that this sharing would increase the amount of social proof that they experience.

Another practical challenge was to determine who to select as confederates. We chose staff members of the grassroots organization because they were available, understood our research, and were trusted by women in the villages. Moreover, since researchers in HCI4D contexts are often accompanied by staff of local organizations who introduce them to communities, we anticipate that other researchers could follow our lead by arranging for local staff to act as confederates. Future work could compare the efficacy of other people playing the role of confederates, such as a health worker or village head.

A key strength of our work is the field evaluation with marginalized women in resource-constrained settings. We designed an intervention to influence response bias and collected strong evidence to demonstrate how social proof could be used to elicit critical feedback on a real, large-scale HCI project deployed in rural India. Although situating our work within the *Projecting Health* project provided several benefits (*e.g.,* access to field locations) there were some disadvantages as well. For example, we were mindful that our intervention must not negatively affect either the *Projecting Health* project or the grassroots organization. For this reason, we rewrote the reviews for the field experiment to ensure that we do not critique key elements of *Projecting Health* such as inviting local people to act in the video and using the local dialect. Other researchers will need to create thoughtful ways to social proof participants without causing damage to existing interventions, local culture and practices, and grassroots organizations.

Ethics

The use of a confederate approach in our experiments introduces important ethical considerations. Specifically, we made participants believe that the reviews they saw had been provided by other participants 'like them' when, in reality, we wrote the reviews. This deception was necessary because, for experimental validity, we needed to control the content and length of reviews across conditions. Although we told participants the purpose of the study, we did not tell them about the use of deception. We made this decision after careful thought and prolonged discussion with the *Projecting Health* staff, who thought that disclosure may introduce significant confusion and ultimately cause more harm than good. Although the use of confederates in scientific experiments is well-established in psychology [3, 25], medical [61] and HCI research [28, 30, 33, 40], it should be used with extreme caution. It would be much better to not deceive participants at all. Moving forward, we hope that researchers using our intervention do not need to use deception for the sake of controlling an experiment. Instead, they could seed the intervention with real feedback from real participants and incorporate additional critical feedback into the intervention as it is received.

Limitations

Our work has several limitations. For example, although our intervention is clearly influencing response bias in different ways, it is not necessarily providing researchers with any objective truth. In addition, since we only exposed participants to social proof *before* watching the video, additional research is needed to study the effects of post-exposure. The field experiment was also conducted with a male Indian researcher, and the results may change (*i.e.,* greater bias) if a foreign researcher conducted the evaluation [15]. Moreover, all participants in the field experiment were women, and additional research is needed to examine the impact of ethnicity and gender of researchers and participants on the response bias. Finally, our experiments were done with people in India, in part because prior work [15] demonstrated high-levels of response bias in India. Future research is needed to understand how our results might differ with people in different countries, cultures, and contexts.

ACKNOWLEDGMENTS

We are grateful to PATH and NYST for their incredible support. We thank Radha Akangire, Rashmi Kanthi, Shrirang Mare, and the anonymous reviewers for their helpful suggestions.

REFERENCES

1. Amazon. 2017. Mechanical Turk. https://www.mturk.com. (2017). [Online; accessed September 10, 2017].
2. Yaw Anokwa, Thomas N Smyth, Divya Ramachandran, Jahanzeb Sherwani, Yael Schwartzman, Rowena Luk, Melissa Ho, Neema Moraveji, and Brian DeRenzi. 2009. Stories from the field: Reflections on HCI4D experiences. *Information Technologies & International Development* 5, 4 (2009), pp–101.
3. Diana Baumrind. 1985. Research using intentional deception: ethical issues revisited. *American psychologist* 40, 2 (1985), 165.
4. Nicola J Bidwell, Thomas Reitmaier, Gary Marsden, and Susan Hansen. 2010. Designing with mobile digital storytelling in rural Africa. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1593–1602.
5. Rod Bond and Peter B Smith. 1996. Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological bulletin* 119, 1 (1996), 111.
6. Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 7415 (2012), 295–298.
7. Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1657–1666.
8. 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.
9. Apala Lahiri Chavan. 2005. Another culture, another method. In *Proceedings of the 11th International Conference on Human-Computer Interaction*, Vol. 21. Erlbaum Mahwah, NJ.
10. Robert B Cialdini, Wilhelmina Wosinska, Daniel W Barrett, Jonathan Butner, and Malgorzata Gornik-Durose. 1999. Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin* 25, 10 (1999), 1242–1253.
11. Dan Cosley, Shyong K Lam, Istvan Albert, Joseph A Konstan, and John Riedl. 2003. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 585–592.
12. Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. 2014. Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. ACM, 739–749.
13. Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. 2015. The role of social influence in security feature adoption. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1416–1426.
14. Nicola Dell and Neha Kumar. 2016. The ins and outs of HCI for development. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2220–2232.
15. Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. Yours is better!: participant response bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1321–1330.
16. Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology* 51, 3 (1955), 629.
17. Francis J Di Vesta. 1959. Effects of confidence and motivation on susceptibility to informational social influence. *The Journal of Abnormal and Social Psychology* 59, 2 (1959), 204.
18. Tawanna R Dillahunt. 2014. Fostering social capital in economically distressed communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 531–540.
19. Olive Jean Dunn. 1961. Multiple comparisons among means. *J. Amer. Statist. Assoc.* 56, 293 (1961), 52–64.
20. Olive Jean Dunn. 1964. Multiple comparisons using rank sums. *Technometrics* 6, 3 (1964), 241–252.
21. Brittany Fiore-Silfvast, Carl Hartung, Kirti Iyengar, Sharad Iyengar, Kiersten Israel-Ballard, Noah Perin, and Richard Anderson. 2013. Mobile video for patient education: the midwives' perspective. In *Proceedings of the 3rd ACM Symposium on Computing for Development*. ACM, 2.
22. Adrian Furnham. 1986. Response bias, social desirability and dissimulation. *Personality and individual differences* 7, 3 (1986), 385–400.
23. Mrunal Gawade, Rajan Vaish, Mercy Nduta Waihumbu, and James Davis. 2012. Exploring employment opportunities through microtasks via cybercafes. In *Global Humanitarian Technology Conference (GHTC), 2012 IEEE*. IEEE, 77–82.
24. Neha Gupta, David Martin, Benjamin V. Hanrahan, and Jacki O'Neill. 2014. Turk-Life in India. In *Proceedings of the 18th International Conference on Supporting Group Work (GROUP '14)*. 1–11.
25. Ralph Hertwig and Andreas Ortmann. 2008. Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior* 18, 1 (2008), 59–92.

26. Melissa R Ho, Thomas N Smyth, Matthew Kam, and Andy Dearden. 2009. Human-computer interaction for development: The past, present, and future. *Information Technologies & International Development* 5, 4 (2009), pp–1.

27. Jessica Hullman, Eytan Adar, and Priti Shah. 2011. The impact of social information on visual judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1461–1470.

28. Jennifer Hyde, Sara Kiesler, Jessica K Hodgins, and Elizabeth J Carter. 2014. Conversing with children: cartoon and video people elicit similar conversational behaviors. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1787–1796.

29. Lilly Irani. 2010. HCI on the move: methods, culture, values. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2939–2942.

30. Hee-Tae Jung, Yu-kyong Choe, and Roderic Grupen. 2015. Extended virtual presence of therapists through home service robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, 209–210.

31. Aniket Kittur, Ed H Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–456.

32. Steven Komarov, Katharina Reinecke, and Krzysztof Z Gajos. 2013. Crowdsourcing performance evaluations of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 207–216.

33. Peter M. Krafft, Michael Macy, and Alex "Sandy" Pentland. 2017. Bots As Virtual Confederates: Design and Ethics. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. 183–190.

34. Adam DI Kramer. 2012. The spread of emotion via Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 767–770.

35. William H Kruskal and W Allen Wallis. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 47, 260 (1952), 583–621.

36. Neha Kumar and Richard J Anderson. 2015. Mobile phones for maternal health in rural India. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 427–436.

37. Neha Kumar, Trevor Perrier, Michelle Desmond, Kiersten Israel-Ballard, Vikrant Kumar, Sudip Mahapatra, Anil Mishra, Shreya Agarwal, Rikin Gandhi, Pallavi Lal, and others. 2015. Projecting health: Community-led video education for maternal health. In *The International Conference on Information and Communication Technologies and Development*. ACM, 17.

38. Jonathan Ledlie. 2010. Huzzah for my Thing: Evaluating a Pilot of a Mobile Service in Kenya. *Qual Meets Quant, London, UK* (2010).

39. Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of psychology* (1932).

40. Steve Love and Mark Perry. 2004. Dealing with mobile conversations in public places: some implications for the design of socially intrusive technologies. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1195–1198.

41. Meethu Malu, Nikunj Jethi, and Dan Cosley. 2012. Encouraging personal storytelling by example. In *Proceedings of the 2012 iConference*. ACM, 611–612.

42. David Martin, Benjamin V. Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a Turker. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 224–235.

43. Indrani Medhi, Aman Sagar, and Kentaro Toyama. 2006. Text-free user interfaces for illiterate and semi-literate users. In *Information and Communication Technologies and Development, 2006. ICTD'06. International Conference on*. IEEE, 72–82.

44. Tamir Mendel and Eran Toch. 2017. Susceptibility to Social Influence of Privacy Behaviors: Peer versus Authoritative Sources. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 581–593.

45. Maletsabisa Molapo, Melissa Densmore, and Limpho Morie. 2016. Apps and Skits: Enabling New Forms of Village-To-Clinic Feedback for Rural Health Education. In *Proceedings of the 7th Annual Symposium on Computing for Development*. ACM, 10.

46. Martin T Orne. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist* 17, 11 (1962), 776.

47. Joyojeet Pal, Anandhi Viswanathan, Priyank Chandra, Anisha Nazareth, Vaishnav Kameswaran, Hariharan Subramonyam, Aditya Johri, Mark S Ackerman, and Sile O'Modhrain. 2017. Agency in assistive technology adoption: Visual impairment and smartphone use in Bangalore. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5929–5940.

48. Jennifer Pearson, Simon Robinson, Matt Jones, Amit Nanavati, and Nitendra Rajput. 2013. Acqr: acoustic quick response codes for content sharing on low end phones with no internet connectivity. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*. ACM, 308–317.

49. Karl Pearson. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50, 302 (1900), 157–175.

50. D. Raghavarao and W. T. Federer. 1979. Block Total Response as an Alternative to the Randomized Response Method in Surveys. *Journal of the Royal Statistical Society. Series B (Methodological)* 41, 1 (1979), 40–45. DOI:**http://dx.doi.org/10.2307/2984720**

51. Nithya Sambasivan and Edward Cutrell. 2012. Understanding negotiation in airtime sharing in low-income microenterprises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 791–800.

52. Nithya Sambasivan, Ed Cutrell, Kentaro Toyama, and Bonnie Nardi. 2010. Intermediated technology use in developing communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2583–2592.

53. Muzafer Sherif. 1935. A study of some social factors in perception. *Archives of Psychology (Columbia University)* (1935).

54. Shari Trewin, Diogo Marques, and Tiago Guerreiro. 2015. Usage of Subjective Scales in Accessibility Research. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, 59–67.

55. Aditya Vashistha, Edward Cutrell, Gaetano Borriello, and William Thies. 2015b. Sangeet Swara: A Community-Moderated Voice Forum in Rural India. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 417–426. DOI: **http://dx.doi.org/10.1145/2702123.2702191**

56. Aditya Vashistha, Edward Cutrell, Nicola Dell, and Richard Anderson. 2015c. Social media platforms for low-income blind people in india. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, 259–272.

57. Aditya Vashistha, Edward Cutrell, and William Thies. 2015a. Increasing the reach of snowball sampling: The impact of fixed versus lottery incentives. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1359–1363.

58. Aditya Vashistha, Neha Kumar, Anil Mishra, and Richard Anderson. 2016. Mobile Video Dissemination for Community Health. In *Proceedings of ICTD*.

59. Aditya Vashistha, Neha Kumar, Anil Mishra, and Richard Anderson. 2017. Examining Localization Approaches for Community Health. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. ACM, 357–368.

60. S. L. Warner. 1965. Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* 60, 309 (March 1965), 63–66.

61. David Wendler and Franklin G Miller. 2004. Deception in the pursuit of science. *Archives of Internal Medicine* 164, 6 (2004), 597–600.

62. Niall Winters and Kentaro Toyama. 2009. Human-computer interaction for development: mapping the terrain. *Information Technologies & International Development* 5, 4 (2009).

63. Fang Wu and Bernardo Huberman. 2008. How public opinion forms. *Internet and Network Economics* (2008), 334–341.

64. Susan P Wyche, Sarita Yardi Schoenebeck, and Andrea Forte. 2013. Facebook is a luxury: An exploratory study of social media use in rural Kenya. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 33–44.