Better Feedback from Nicer People: Narrative Empathy and Ingroup Framing Improve Feedback Exchange

Y. WAYNE WU, University of Illinois at Urbana-Champaign, USA
BRIAN P. BAILEY, University of Illinois at Urbana-Champaign, USA

Online feedback exchange platforms enable content creators to collect a diverse set of design feedback quickly. However, creators can experience low quality and harsh feedback when using such platforms. In this paper, we leverage the empathy of the feedback provider to address both these issues. Specifically, we tested two narrative-based empathy arousal interventions: a negative experience and a design process narrative. We also examined whether ingroup framing further enhances the effects of empathy arousal. In a 3x2 online experiment, participants (n=205) wrote feedback on a poster design after experiencing one of the intervention conditions or a control condition. Our results show both the design process narrative and ingroup framing conditions significantly increased the feedback quality and effort invested in the task. The negative experience narrative condition had similar effects and participants reported significantly increased disapproval towards harsh feedback. We discuss the implications of our results for the design of feedback exchange platforms.

CCS Concepts: • **Human-centered computing → Human-computer interaction (HCI)**; *HCI design and evaluation methods; Laboratory experiments*

KEYWORDS: Empathy; ingroup framing; design feedback

ACM Reference format:
Y. Wayne Wu, Brian P. Bailey. 2020. Better Feedback from Nicer People: Narrative Empathy and Ingroup Framing Improve Feedback Exchange. *In Proceedings of the ACM on Human-Computer Interaction, Vol. 4, CSCW,* Article 236 (December 2020). 20 pages. https://doi.org/10.1145/3432935

1 INTRODUCTION

Novice content creators often leverage online feedback exchange platforms to quickly gather feedback on their creative projects from diverse audiences. However, feedback received online often has variable quality. Users may be unwilling to provide high-quality feedback because of the time and effort required [75]. Peers on a platform also provide lower quality feedback than experts do [80]. Moreover, content creators may receive feedback delivered in an overly negative tone, which we refer to as harsh feedback in this paper. For example, on a popular feedback exchange platform, a novice content creator seeking constructive feedback was told the work was "trash" and "no one wants to critique it." Such harsh feedback significantly reduces content creators' affective states and task performance [72]. This issue is not uncommon, as users with antisocial tendencies are disproportionally eager to generate negative valence content [4,12]. Furthermore, harsh feedback snowballs [14]. One piece of harsh feedback can incite more people to provide similar feedback. Due to the above-mentioned issues, feedback exchange platforms frequently fail to deliver constructive feedback and may even discourage content creators from practicing design further.

In this paper, we study interventions that operationalize empathy to improve feedback quality and discourage harsh feedback. Providing constructive feedback is a type of prosocial behavior. One empirically validated way to promote prosocial behavior is by eliciting empathy [3,27,31]. In our experiment, we examined two empathy-based approaches: narrative empathy and ingroup framing, and whether they could improve feedback quality and mitigate harsh feedback. In the context of feedback exchange, empathy may encourage effort and thus higher quality feedback. It also reduces the likelihood of aggressive behaviors and promotes intervening action when harsh feedback occurs. While there are many ways to arouse empathy, in our experiment, we chose narratives as they are applicable for most online platforms. Reading narratives from content creators encourages perspective-taking and improves the effectiveness of feedback [68]. Prior work indicates that empathy has two core aspects, affective empathy and cognitive empathy [49]. In our work, we examined the effectiveness of both aspects, testing narratives about experience receiving harsh feedback (affective empathy) and design process of the target work (cognitive empathy).

We also examined how ingroup framing interacts with the empathy narrative interventions. Ingroup framing is one of the factors that affect empathy and there is extensive literature studying the interaction between these two [10,15,29,32,51,64–66]. Prior work shows perceiving others as ingroup members promotes prosocial behaviors [71]. In addition, people are more likely to have empathetic feelings towards ingroup members. Following prior work, we implemented ingroup framing by establishing interdependency between the feedback provider and the designer and assigning a label to the pair [44]. Then we examined how ingroup framing interacted with narrative empathy interventions and influenced feedback composition.

We conducted a 3x2 (n=205) full factorial experiment with two factors: narrative empathy and ingroup framing. Narrative empathy has three levels: design process, negative experience, and control article. Ingroup framing has two levels: ingroup and control framing. In the experiment, participants reviewed a poster design, then read a passage with different content based on the narrative empathy condition, and later wrote feedback for the poster. Participants performed the task under a group framing or independently based on the ingroup framing condition. Afterward, we recruited domain experts to evaluate the quality of the collected feedback and used feedback length as a heuristic for effort. We also measured changes in participants' attitudes towards harsh feedback in pre and post-study surveys.

Our results showed that both the design process and ingroup framing interventions significantly increased the effort invested in feedback composition by ~40% and the final feedback quality by ~30%. The negative experience condition had similar effects, increasing both measures by ~20%. Also, the pre and post-study surveys showed participants experiencing the negative experience narrative condition reported a significantly more disapproving view towards harsh feedback and significantly more likely to accept responsibility to intervene if harsh feedback occurred, which could reduce the snowballing of harsh feedback.

Our work makes three contributions to the CHI community: i) empirical evidence that empathy arousal and team dependency improve feedback exchange; ii) a deeper understanding of the underlying theories of narrative empathy, ingroup framing, and their interactions; iii) practical guidelines regarding how platform designers could use these interventions to help users to receive higher quality feedback while mitigating the prevalence of harsh criticism.

2 RELATED WORK

We situate the contribution of our work in the context of prior methods for controlling harsh feedback and improving feedback quality. We then discuss theories of empathy arousal to motivate the interventions used in our experiment.

2.1 Existing Approaches to Control Harsh Feedback

A prior empirical study showed that receiving even one piece of harsh feedback negatively impacts content creators and their willingness to act on the feedback [72]. Prior work has therefore explored techniques to limit the production of harsh feedback and its effect on content creators. We categorize these techniques into recipient- and provider-oriented techniques. For recipient-oriented techniques, content creators could perform an explicit coping activity such as self-affirmation to mitigate the effects of harsh feedback [72]. A second technique in this category is presenting the feedback based on its valence such that a content creator consumes the most positive feedback first. Prior work shows that this technique mitigates the effects of harsh feedback on content creators' affective states relative to consume the feedback in a different order [74]. A third technique is visually aggregating the feedback to create a buffer between the harsh feedback and the content creator [78].

For provider-oriented techniques, prior work has explored scaffolding to reduce the potential negativity of the content by constraining the type of feedback generated [42,76]. Online platforms can rely on central/distributed moderation [39,53] and learning-based filtering [13,56,57] to identify and remove harsh feedback or block malicious providers from future participation in feedback exchange [36]. In our work, we explore a new provider-oriented intervention based on empathy arousal to encourage the composition of helpful feedback.

2.2 Existing Approaches to Improve Feedback Quality

High-quality feedback should identify the perceived goal of the creative project, the status of the current prototype in relation to that goal, and what actions the content creator should take to close the gap [59]. Prior work has tested the effects of presenting feedback that models the desired characteristics of good feedback and relied on the feedback providers to mimic and learn these characteristics [28]. While this method can be effective, it may also lead to homogeneous feedback based on the models. Alternatively, platform designers could offer rubrics to guide feedback providers [80]. Prior work has explored how specific feedback content, such as justification [26], positive framing [46], and rubric style [17] influences feedback quality. Furthermore, prior work has reported success using machine learning models to provide applicable design guidelines based on design content [37]. Researchers have also built interactive systems to allow content creators to effectively follow provided guidelines and track their own processes [8]. In our work, we examine three interventions to increase feedback providers' motivation to provide quality feedback. These interventions are not exclusive and could be used in conjunction with methods described in prior work.

2.3 Empathy Arousal

Empathy is a vicarious response to others' emotional states [22]. Prior work has explored different empathy arousal methods [5,7,25,33,62,69]. A common and online-appropriate method for eliciting empathy is having a user read a narrative. Prior work shows habitual fiction readers report higher than average levels of empathy [20]. Reading narratives also has immediate effects on people's empathy. Prior work shows reading a short narrative essay can lead to higher

empathy and higher prosocial behavior tendency immediately after performing the task [11,20]. Empathy towards people in distress causes an emotional appraisal and leads to prosocial behaviors [38]. While researchers haven't reached a consensus regarding the definition of empathy, most agree that it includes both a cognitive part and an affective part [19,52]. The former functions as people analyzing the target's experience and current situation to deduce his or her emotions in the moment [19]; the latter functions as people intuitively recognizing the target's emotions [19]. In our experiment, we test how two types of narratives corresponding to these two categories of empathy: negative experience and design process affect the composition of feedback.

2.4 Narrative Interventions

For the negative experience narrative condition in our experiment, participants review the design accompanied by a narrative about the designer's recollection of one episode of receiving harsh feedback. Prior work shows sharing an unpleasant experience induces empathy and encourages helping behaviors [6]. Also, reading narratives about unpleasant experiences discourages people from inflicting similar experiences to others and make them less tolerant about the offensive behaviors [23]. In our study, the negative experience narrative should encourage participants to perform more helping behaviors, which is to provide more useful feedback in the context of feedback collection and improve their attitude to intervene when they observe offensive behaviors from other users, such as providing harsh feedback. In the design process condition, the narrative includes a description of the goal of the design and a series of explained design decisions made in the process. Prior work shows information about the protagonist of the narrative, including their background, goal, and their journey so far, cultivates empathetic feelings in the readers [16,35,48]. In our study, we try to frame the design process in a similar way to arouse empathy toward the content creator. The narrative describes the goal of the design project, along with how the creator planned to achieve it and current progress. In addition, the design process narrative has the potential to increase the usefulness of the feedback. Prior feedback theory argues a piece of feedback needs to recognize the project goal, the current state of the work, and actionable advice to reach the goal to be most helpful to the content creators [58]. The design process narrative helps the feedback providers to judge these aspects of the design more accurately.

2.5 Ingroup Framing

Another intervention we plan to examine is ingroup framing. Perceiving others as ingroup members makes it easier for people to develop empathetic feelings and show prosocial behaviors. Prior work shows people feel more empathetic and have stronger prosocial tendencies towards ingroup members [65]. Empathy towards ingroup members may even induce costly helping behaviors, such as choosing to endure physical pains for other people [32]. Prior empirical studies also suggest people may show empathy only towards ingroup members [10,29]. In extreme cases, people may not only fail to feel empathy towards outgroup members but instead gain pleasure from the suffering of outgroup members who they dislike [15]. Fortunately, such differences in attitudes are not immutable. Prior work shows changes in social categorization influence group membership perception and the likelihood of empathy arousal [66]. Researchers had also shown empathy arousal interventions could be used to improve intergroup relationships [64]. In our study, we focus on using ingroup framing to arouse empathy and increase the likelihood of prosocial behaviors. Prior work argues group labeling and interdependent relationships foster a sense of group membership [44]. The results of prior experiments also demonstrate that group

framing interventions may stimulate participants to perceive computer agents as their teammates [51]. Our experiment contributes to this literature by testing the effects of group framing in conjunction with narratives for arousing empathy and improving feedback quality.

2.6 Research questions

We focused on answering two research questions:

· R1: How do narratives such as reading about the designer's experience receiving harsh feedback or the design process of the project influence feedback composition and attitudes towards harsh feedback?

· R2: How does ingroup framing influence participants and interact with the effects of empathy-arousal narratives for the same measures?

We answered these two research questions through an online experiment.

3. METHODOLOGY

We conducted a 3x2 full factorial experiment with two factors: narrative empathy and ingroup framing. Narrative empathy had three conditions: negative experience, design process, and control article. Ingroup framing had two conditions: ingroup framing and control framing.

3.1 Experimental Task

In the online experiment, we asked participants to review a poster design and then provide feedback to help the designer to improve it (see Figure 1). We collected the design from a designer (Asian female with 2.5 years of experience at the time of the experiment) recruited from UpWork, a popular freelancing platform. Participants reviewed the design together with its background information, including its purpose, target audience, explanations of design elements, and where it would be displayed. Later, we instructed participants to write about the strengths and weaknesses of the design and provide actionable and specific advice to help the designer improve the design. We also promised a bonus for high-quality feedback to incentivize participants.

3.2 Narrative Empathy Factor

Narrative empathy had three conditions: negative experience, design process, and control article. For all three conditions, participants read a 300-word passage with content based on experimental conditions (see Table 1). For negative experience, the passage described a prior episode of the designer receiving harsh feedback on a creative project. From a first-person perspective, the designer recalled who commissioned the design, how s/he received the harsh feedback, and how he s/felt at that moment. We recruited a designer (not the one who provided the design for the study) from UpWork to compose the passage based on personal experiences.

For the design process condition, the passage described the purpose of the design, design decisions made in the process, the reason to seek feedback, and how the designer felt about feedback collection. We asked the designer of the poster to provide the narrative. For both narratives, we edited the text for brevity and revised passive sentences into active forms, as prior work showed this style is more effective for arousing empathy [9,11].

For the control article, the passage described a technological concept, a topic that was orthogonal to the task, from a third-person perspective. We selected the passage from a news site and revised it to match the length of the other two narratives. For all three passages, we rephrased some sentences so they had the same level of readability (negative experience: 6.18; design process: 6.30; control article: 7.00 evaluated by Automated Readability Index [63]).

Your teammate created the poster for the Boys & Girls Clubs of Philadelphia to promote their annual spring fundraiser, the Philly Showcase of Wine, Cheese & Beer. Attendees have access to hundreds of fine wines, cheeses, beers and local food vendors. The four people in the center of the photo were the honoree and chairs of the event. The poster was intended to be displayed on walls throughout the city including public transportation.

> **Bonus**: To reward high performance, we offer a bonus opportunity in this HIT. We have grouped the designer of the poster and you as Team Orange and you two will collaborate on this task. After you submit the HIT, the the designer of the poster will review your feedback and try to revise the design accordingly. A domain expert will rate the success of your collaboration by evaluating the improvement in the design based on the feedback that you provide. The teams who score in the top 10 will share a cash bonus of $8, $4 per team member.

Please write your feedback in the text-box below. The feedback should incorporate the strengths and weaknesses of the poster, and provide specific and actionable suggestions for improvement.

NEXT

Figure 1. A screengrab of the experimental task. Participants reviewed the poster and background information for the poster provided by the designer (e.g., the people shown were the honoree and chairs of the event). They then read a bonus opportunity statement, which was phrased differently based on the group framing conditions. Participants then wrote feedback targeting the perceived strengths and weaknesses of the poster and suggestions for improvement.

Negative Experience Narrative
One day, I was told to create a design for an upcoming event hosted by a local client. The client preferred a minimalistic style, meaning it'd be a simple yet visually-appealing design. I quickly cranked out a couple mock-ups as this was a style that I had created similar designs before. Later I visited the client to present my designs. I knew something was off when I saw their smile turn into a frown. Before I could ask for suggestions, I was immediately interrupted. "This is not at all what I'm looking for," they said. I've dealt with criticism in the past, but I was not prepared for their rude words. "It's pathetic and weak-looking." I turned my gaze from my designs to their face to see if they would laugh and say "just kidding," but their tone became more offensive as they began to harshly analyze my designs further. "Have you not learned anything from your time working with us? A rookie with no experience could have made a better mock-up than what I'm looking at." I was taken aback. Not quite yelling, but still louder than necessary, they continued in great detail while the others watched. "It's juvenile and low-quality. If you can't handle these simple projects, then maybe I need to find someone else who can." These condescending words were coming from someone that I respected. The fact that it came from someone close to me who generally supported me and my work made the situation worse. I was visibly hurt, and embarrassed by the way that I was being addressed. It was unexpected and uncalled for. It's one thing to criticize my work, but to go after my skills and abilities and imply that I'm not good enough is extreme.
Design Process Narrative
As a member of the company, I had a significant influence on its design. The poster needs to include the logos of the company and the sponsors. The event information and logo also need to be there. Another need was to have photos of the honoree and chairs of the event. They had ties to a local sports team and could help promote the event. My biggest concern with the poster is that it's difficult to determine what type of event it is just by looking at it. The only way to fully understand the event is to go the ticket website link and learn more. I believe that's what my manager hoped would happen. The people at the center would drive traffic to our ticket website. Because the design is Philly themed we went with a picture of the skyline for the background. The company used the original image for previous years' promotional material. So I altered the colors and blurred it to make it somewhat indistinguishable. The edits also allowed the text to stand out and the individuals' headshots to be a focal point. I choose to render all sponsor logos in white. This more consistent color profile also helps to play down that section. The font is Futura, which I like to use because of it is minimal and includes many weights. Weight options are key to creating a hierarchy of importance with text. I'm happy with how it turned out given the circumstances I was working under. The poster was intended to be displayed throughout the city including public commute. It's simple and eye-catching in those settings. I did a good job at updating the original material to the current advertising style.
Control Article
When an object is cool and warm air touches the cool object, the air cools and droplets of water forms on the outside of the object. This is the result of the hot and cold air coming into contact with each other. This water in the air is called water vapor. Water vapor is in the form of a gas. Characteristics of water vapor include it being colorless, odorless, invisible, and has no taste. Humidity is the amount of water vapor in the air. When the in the air turns into a gas it is called evaporation. Water vapor gets into the air day through the process of evaporation. Ocean water, and other bodies of water, is turned into water vapor using the energy from the sun. The molecules of the water is absorbed by the Sun's energy near the surface of the water which then evaporates into the air. The changing of a gas into a liquid is called condensation. An example of condensation is the water which covers a mirror following a hot shower. Another large source of water vapor in the air is when the plants absorb water through their roots and stems into their leaves. The leaves then give off water. The process of plants releasing water into the air is called transpiration. All of the water in the air, whether it is from the world's ocean and other bodies of water, the water on a mirror following a hot shower, or the water a plant releases into the air; it is all called humidity because it is the amount of water vapor in the air.

Table 1: Participants read one of the above passages corresponding to the narrative empathy condition after reviewing the poster. In the negative experience and design process conditions, participants were instructed to imagine how the designer would feel in these situations [6].

3.3 Group Framing Factor

Group framing had two conditions: ingroup framing and control framing. Participants read different instructions throughout the task based on the conditions (see Table 2). For the ingroup framing condition, we informed the participants that we had grouped them and the poster designer as a team. Following prior work, the team was assigned an arbitrary and neutral label, namely "Team Orange". We also created interdependency between the participant and the designer by promising a team-based bonus. We informed the participants that a domain expert would rate the usefulness of their feedback by evaluating how much the designer improved the design using their feedback. The top 10 teams with the highest rating would share a bonus of $8, $4 for each person. For the control framing condition, we instructed participants to complete the task independently. A domain expert will rate how useful their feedback was by evaluating its potential to help the designer improve the design. The participants writing the feedback rated in the top 10 would receive a bonus of $4. Throughout the experiment, we referred to the provider of the design as "your teammate" or "the designer" respectively based on the group framing condition.

3.4 Procedure

In total, 205 participants (see Table 3 for condition breakdown) finished the experiment. Figure 2 shows the experimental flow of the task. All participants went through an informed consent process. After that, participants reviewed an overview page describing the workflow of the task. Part of the instruction was composed differently based on the group framing conditions. Before reviewing the poster and the narratives, participants filled out a pre-study survey measuring their empathy quotients and attitudes towards harsh feedback. Later, participants reviewed the design and read a passage based on narrative empathy conditions. Last, participants composed feedback for the design and filled out a post-study survey regarding their perception of the designer and their attitudes towards harsh feedback.

Ingroup Framing Condition	**Control Framing Condition**
To reward high performance, we offer a bonus opportunity in this HIT. We have grouped the designer of the poster and you as Team Orange and you two will collaborate on this task. After you submit the HIT, the designer of the poster will review your feedback and try to revise the design accordingly. A domain expert will rate the success of your collaboration by evaluating the improvement in the design based on the feedback that you provide. The teams who score in the top 10 will share a cash bonus of $8, $4 per team member.	To reward high performance, we offer a bonus opportunity in this HIT. We ask you to complete this task independently. Your feedback should help the designer improve the poster. A domain expert will rate how useful your feedback is by evaluating its potential for helping the designer improve the poster. The participants whose feedback is ranked in the top 10 will earn a cash bonus of $4.

Table 2: Participants read statements about a bonus opportunity corresponding to the *group framing* conditions. Other task instructions also referred to the content creator as "the designer" or "your teammate" based on *group framing* conditions throughout the task.

3.5 Participants

Due to the scale of the experiment, we conducted the experiment on Amazon Mechanical Turk (AMT). To ensure participants were representative of feedback exchange platform users, we adopted a screening process where they answered a question regarding their experience in

providing design feedback. To warrant truthful answers, we asked the same screening question again in the post-study survey. Participants with inconsistent answers to these two questions were excluded from the final analysis.

A common issue on AMT was workers' satisfying behaviors [30]. To minimize this behavior, we implemented a series of confirmation checks to ensure participants were performing the tasks as requested throughout the experiment. Participants answered questions about the assigned team label and the content of the narratives. We also added a confirmation check in the pre and post-study surveys about their opinions on an issue unrelated to the experimental manipulation. Participants estimated the popularity of feedback exchange platforms and the participants with notably different answers between the pre and post-study surveys (more than 2 point difference on a 7-point scale) were excluded from the data set. We also excluded participants who repeatedly attempted to skip experimental tasks and participants who spent an unusually long time on the task (two standard deviations higher than the average) from the final analysis.

Table 3. Participant counts by experimental condition. There were 205 participants in total.

	Control Article	Negative Experience	Design Process
Control Framing	34	34	38
Ingroup Framing	34	32	33

In the final participant pool, 57% were female, 43% male; 13% were 18-24 years old, 44% 25-34 years old, 25% 35-44 years old, 18% 45-65 years old; 10% had higher school or lower degree, 41% some college or associate degree, 38% bachelor's degree, 12% graduate degree. Regarding the feedback collection experience, 22% received feedback daily, 34% monthly, 32% weekly, and 11% yearly. For the frequency of receiving harsh feedback, 11% had never received harsh feedback, 19% daily, 28% weekly, 31% monthly, and 12% yearly. We paid each participant $4 ($13.5/hr) upon task completion. Workers who failed attention checks received a payment proportional to HIT duration up to $4. All participants had finished more than 500 HITs on AMT and had a pass rate higher than 98%.

3.6 Measurement

The main measures included feedback quality, feedback length, and attitudes towards harsh feedback. Prior work has used feedback quality and invested effort (measured by feedback length) to evaluate feedback because they directly impact how much content creators would benefit from the feedback [21,43,70,72–74,77,79]. For feedback quality, we hired two domain experts from UpWork to rate the quality of the collected feedback separately. We share the instructions feedback providers received with the experts, and asked them to use their own judgment to decide the quality of the feedback. Each expert started with a calibration phase where they rated 30 pieces of randomly sampled feedback. We instructed the experts to use the entire 7-point scale in calibration and rate the rest of the feedback set using the same standard. Both experts gave similar ratings to the feedback (Pearson's r=0.53), and we averaged the ratings as the final measure. We also measured feedback length as a heuristic of the level of invested effort. To examine whether interventions would lead to changes in feedback content and sentiment, one researcher coded the collected feedback at a sentence level using an established feedback schema [45]. Following prior work, we used LIWC to analyze the feedback and examined ratings in relevant categories [60,67].

We also created an 8-question survey measuring participants' attitudes toward harsh feedback to gauge how likely they would take proactive interventions against harsh feedback. The survey was crafted based on prior survey work about harassment and bystander intervention [47]. As

Table 4 shows, the survey had three sections, focusing on participants' attitudes toward the recipients of harsh feedback, the occurrence of harsh feedback, and their tendency to intervene. At the beginning of the survey, participants reviewed the definition of harsh feedback and an example to avoid confusion about later survey questions. To measure changes in their attitudes, participants answered the same set of questions twice before and after the experimental task.

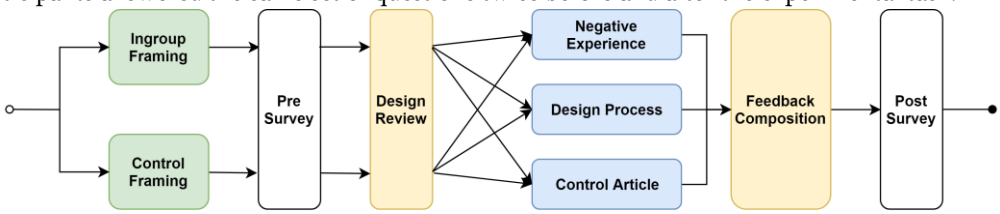

Figure 2. Experimental flow chart of the study. At the beginning of the study, all participants were evenly divided into two groups, one reviewing the ingroup framing instructions and the other control framing. Then all participants filled the same pre-study survey and reviewed the poster. Afterward, participants read different passages based on *narrative empathy* conditions. Participants then wrote feedback for the poster and completed a post-study survey.

General Attitudes Towards Harsh Feedback Recipients
○ I feel very sorry for people when they receive harsh feedback.
○ I have tender, concerned feelings for people who receive harsh feedback.
General Attitudes Towards Harsh Feedback
○ It is evident to me that people who receive harsh feedback need support from other members on the same online platform.
○ If someone writes harsh feedback, people should realize it is a necessary experience for them to grow.
○ I think such harsh feedback is hurtful and damaging to people.
General Tendency of Intervention
○ I feel personally responsible to intervene and offer support to people when they receive harsh feedback.
○ Even if I am not the one providing the harsh feedback, it is still my responsibility to try to discourage others from doing so.
○ I believe that my actions can help to reduce the occurrence of harsh feedback.
Perspective Taking
○ I tried to make my feedback more useful by imagining how the designer of the poster would react to it.
○ I was more concerned about whether my feedback would be useful than how the designer of the poster would react to it.

Table 4: Questions in the post-study survey. Participants rated their level of agreement for each statement on a scale from 1 to 7 (strongly agree). Except for the two perspective-taking questions, the other eight questions were also asked in the pre-study survey.

For confirmation checks of the *narrative empathy* manipulation, participants answered two questions about to what degree they had tried to provide feedback from the designer's perspective (see Table 4). Since only participants in the negative experience condition read about the harsh feedback the designer had received before, we used perspective taking as a heuristic for their empathy towards the designer. The question was adapted from prior work on interpersonal empathy [18]. For confirmation checks of the *ingroup framing* manipulation, we used the classic Inclusion of Other in the Self (IOS) scale [2]. We also included an 8-question survey to measure

the empathy quotient of the participants and use it as a covariate in the analyses to address individual differences [41].

4 RESULTS

Below we report the significant patterns in our results.

4.1 All Interventions Increased Feedback Quality

A two-way ANOVA showed *narrative empathy* improved feedback quality ($F[2, 198]=2.47$, $p=.088$). Participants wrote higher quality feedback in the design process condition ($\mu=3.81$, $sd=1.03$) than in the control article condition ($\mu=3.38$, $sd=1.42$; $p=.072$ after Tukey's HSD adjustment; Cohen's $d=0.35$). Participants in the negative experience condition wrote feedback of similar quality ($\mu=3.56$, $sd=1.07$; adj. $p=.638$; $d=0.14$) in comparison to the control article condition. We also observed a significant interaction between the two factors ($F[2, 198]=4.05$, $p=.019$). Ingroup framing tended to increase the feedback quality in the control article condition ($F[1, 198]=1.85$, $p=.176$; $d=0.65$). There was no significant difference between the ingroup framing ($\mu=3.70$, $sd=1.22$) and the control framing condition ($\mu= 3.48$, $sd=1.16$; adj. $p=.175$; $d=0.19$).

As shown in Figure 3, all five intervention conditions resulted in higher feedback ratings compared to the control framing/control article condition. For the five pair-wise planned comparisons, we adjusted the p-value threshold using Holm's Bonferroni method to minimize familywise error [81]. In comparison with the control framing/control article condition ($\mu=2.94$, $sd=1.19$), participants provided significantly higher quality feedback in the control framing/design process ($\mu=3.88$, $sd=0.97$; adj. $p=.002$; $d=0.87$), ingroup framing/design process ($\mu=3.73$, $sd=1.10$; adj. $p=.026$; $d=0.69$), and ingroup framing/control article ($\mu=3.82$, $sd=1.51$; adj. $p=.028$; $d=0.65$) conditions. The ingroup framing/negative experience ($\mu=3.55$, $sd=1.00$; adj. $p=.057$; $d=0.55$) and control framing/negative experience ($\mu=3.57$, $sd=1.14$; adj. $p=.057$; $d=0.54$) conditions also increased feedback quality.

4.2 Design Process and Ingroup Framing Increased Levels of Effort Invested

A two-way ANOVA showed *narrative empathy* had a main effect ($F[2, 198]=3.04$, $p=.050$). Participants in the design process condition ($\mu=938.0$, $sd=433.4$) wrote longer feedback than the ones in the control article condition ($\mu=800.7$, $sd=433.4$; adj. $p=.17$; $d=0.29$). Negative experience narrative ($\mu=757.1$, $sd=403.6$; adj. $p=.84$; $d=0.09$) had no effect. We also observed a weak interaction effect between the two factors ($F[2, 198]=2.85$, $p=.060$). Ingroup framing led to significant differences when participants read the control article. We report the pairwise differences in detail below. *Ingroup framing* had no effects ($F[1, 198]=0.53$, $p=.468$; $d=0.10$). There was no significant difference between the ingroup framing ($\mu=858.1$, $sd=445.8$) and the control framing condition ($\mu=812.0$, $sd=476.6$; adj. $p=.468$).

As Figure 4 shows, all five intervention conditions reported longer feedback lengths than the control framing/control article condition. In comparison with the control framing/control article condition ($\mu=668.6$, $sd=450.1$), participants provided significantly longer feedback in the control framing/design process condition ($\mu=961.0$, $sd=478.5$; adj. $p=.047$; $d=0.63$), and notably longer feedback in the ingroup framing/design process ($\mu=911.5$, $sd=380.5$; adj. $p=.080$; $d=0.58$) and ingroup framing/control article condition ($\mu=932.8$, $sd=566.6$; adj. $p=.111$; $d=0.52$) conditions. Participants tended to write more in the control framing/negative experience ($\mu=788.7$, $sd=464.9$; adj. $p=.573$; $d=0.26$) and ingroup framing/negative experience ($\mu=723.5$, $sd=330.4$; adj. $p=.566$; $d=0.14$) conditions.

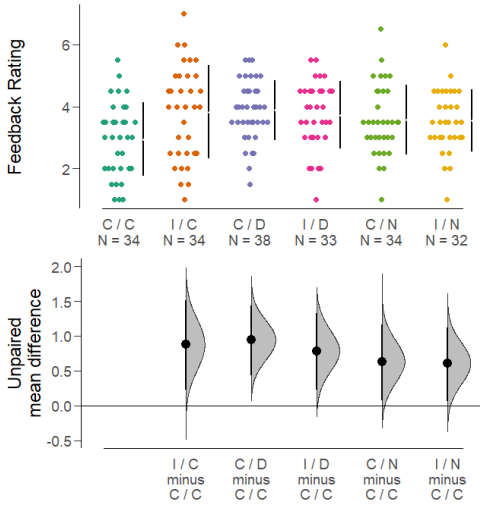

Figure 3. Feedback quality across conditions. In comparison with the leftmost control framing/control article condition, both ingroup framing and design process increased ratings of feedback quality. Negative experience had a similar effect. Here we label all group conditions in the format of "A / B". A indicates the group framing condition: Ingroup framing or Control framing, and B indicates the narrative empathy condition: Design process, Negative experience, or Control article.

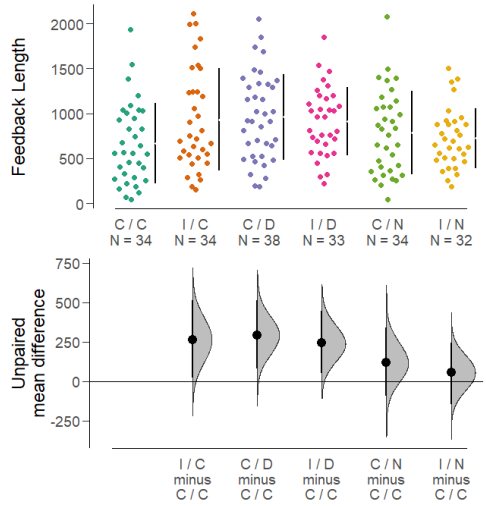

Figure 4. Feedback length was used as a heuristic for the effort invested in feedback composition. The leftmost bar represents the control framing/control article condition. Here we observed a similar trend to the one in feedback quality. Ingroup framing and design process increased feedback length; negative experience had a similar pattern but less pronounced.

	control article		design process		negative experience	
	control	ingroup	control	ingroup	control	ingroup
analytical	81.3 (19.1)	79.4 (22.1)	77.3 (18.7)	80.3 (12.6)	77.2 (17.5)	82.5 (16.5)
social	6.3 (3.7)	6.0 (2.6)	6.5 (2.9)	5.7 (2.6)	7.1 (4.4)	6.0 (2.9)
tone	*71.0 (26.7)*	**76.2 (24.4)**	**78.2 (21.0)**	**78.6 (19.9)**	**80.4 (22.8)**	**75.4 (24.3)**

Table 5. Mean and (standard deviations) for LIWC analysis across experimental conditions. All three output variables range from 0 to 100. Feedback in the intervention conditions had higher positive tone (in bold) than in the control framing/control article condition (in italics). For reference, intervention conditions reported a tone level similar to the one of natural speech (79.29) while the control framing/control article condition's level is slightly lower than the one of Twitter posts (72.24) [50].

4.3 Feedback Tends to Have More Positive Tones

We performed an LIWC analysis on the collected feedback (see Table 5). Following prior work, we used analytical, social, and tone as three main categories because of their relevance to the task [61]. Our results showed feedback across conditions had similar levels of analytical and social ratings. Feedback in all the intervention conditions had higher positive tone than in the control framing/control article condition. We did not observe statistical differences regarding feedback

categories and LIWC ratings. One reason might be we conducted the study in a realistic setting without offering strong stimulus to harsh feedback as in prior work [14,61]. Instead, the interventions had significant influences on how participants perceived harsh feedback.

4.4 No Differences Between Conditions for Feedback Categories

After coding the feedback using a scheme developed in prior work [45], we summed the length of feedback written in each feedback category in each experimental condition. Then we normalized the counts to calculate the ratios. There were no significant differences between the conditions in terms of the ratios of feedback categories. On average, 7.6% of the feedback was positive only, 18.9% positive and specific, 31.6% problem only, 29.4% solution only, 9.4% problem and solution, and 3.1% others. On average, participants expanded each feedback category proportionally when they wrote longer feedback in the intervention conditions.

4.5 Negative Experience Led to More Disapproving Stance Against Harsh Feedback

For participants' pre and post-study differences in attitudes towards harsh feedback, we summed the differences into three measures: attitudes towards recipients of harsh feedback, attitudes towards harsh feedback itself, and tendency to accept responsibility to help.

Narrative empathy had a main effect on participants' tendency to accept responsibility to help ($F[2, 198]=3.12$, $p=.046$; see Figure 5). In comparison with the control article condition ($\mu=0.10$, $sd=2.22$), negative experience ($\mu=0.83$, $sd=1.92$; adj. $p=.076$; $d=0.35$) tended to make participants more likely to accept responsibility to help. In particular, the ingroup framing/negative experience condition ($\mu=1.31$, $sd=1.71$) reported significantly stronger tendency in comparison with the control framing/control article condition ($\mu=0.21$, $sd=1.67$; adj. $p=.049$; $d=0.66$).

Narrative empathy also had a main effect over participants' attitudes towards harsh feedback ($F[2, 198]=6.32$, $p=.002$; see Figure 5). In comparison with the control article condition ($\mu=-0.68$, $sd=2.00$), negative experience ($\mu=0.59$, $sd=2.05$; adj. $p=.001$; $d=0.63$) made participants take a significantly more disapproving stance towards harsh feedback. Interestingly, when participants performed the task without any intervention they reported more tolerance of harsh feedback (paired t-test $p=.013$; $d=0.15$). The experience of providing feedback might make participants more aligned with the perspectives of feedback providers and more lenient with the behaviors.

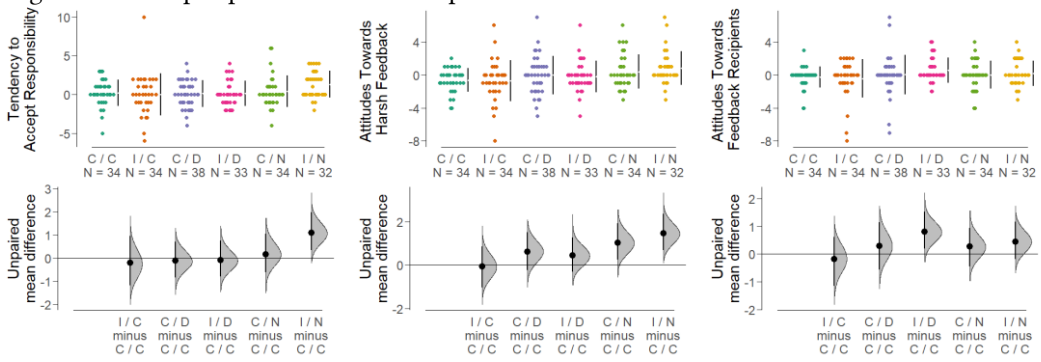

Figure 5. The above charts show how interventions influenced participants' attitudes towards harsh feedback and its recipients. Among the three interventions, negative experience was most effective. It led to a significantly more disapproving stance towards harsh feedback and made

participants significantly more likely to intervene.

For attitudes towards recipients of harsh feedback, *narrative empathy* did not have a main effect (F[2, 198]= 2.11, p=.124; see Figure 5). But both the design process (μ=0.27, sd=2.02; adj. p=.114; d=0.32) and negative experience (μ=0.09, sd=1.61; adj. p=.343; d=0.26) conditions led to more empathetic feelings towards the recipients than the control article condition (μ=-0.35, sd=1.85). In comparison with the control framing/control article condition (μ=-0.26, sd=1.29), participants in the ingroup framing/design process condition were notably more supportive to the feedback recipients (μ =0.55, sd=1.50; adj. p=.105; d=0.58).

4.6 Manipulation Checks

Ingroup framing had a main effect on group perception (F=5.879; p=.016). Participants in the ingroup framing condition reported significantly higher IOS (μ= 3.45, sd=1.51) than participants in the control framing (μ=2.90, sd=1.76; adj. p=.016; d=0.34). No significant effect was detected regarding participants' perspective-taking. The ingroup framing/negative experience condition (μ=8.34, sd=2.72) tended to encourage more perspective-taking in comparison with the control framing/control article condition (μ=7.03, sd=2.54; adj. p=.235; d=0.50).

4.7 Empathy Quotient Caused No Difference

We used participants' empathy quotient as a covariate to control individual differences in personalities. This measure did not influence the significance levels of any analysis. The effects we uncovered in our study were applicable to all participants with various levels of empathy capabilities.

5 DISCUSSION

All the proposed interventions were effective at increasing feedback quality and effort invested in the task (measured by feedback length). While all participants received the same payment and task instructions, ingroup framing increased feedback quality by 30% and feedback length by 40%; the design process narrative by 32% and 44%; and the negative experience condition by 21% and 18%, respectively. Regarding attitudes towards harsh feedback and its recipients, participants in the negative experience condition reported a more negative attitude toward harsh feedback and were more likely to accept responsibility to intervene when harsh feedback is observed relative to participants in the other conditions. Participants in the design narrative condition became more supportive to the recipients of harsh feedback.

Both the ingroup framing and design process narrative conditions achieved similar effects, i.e., increasing feedback quality and efforted invested in feedback composition. Platform designers should prefer to use ingroup framing when this intervention is applicable, as it imposes minimal, if any, additional task load on either the recipient or provider of feedback. In our study, we tested two common ways of ingroup framing: group interdependency and group labeling. For interdependency, platform designers could promise various rewards for collaboration. While monetary rewards may not always be suitable, platform designers could use reward points or badges as alternatives [1,55]. Future work could also test platforms where successful collaboration

helps users to earn privileges related to feedback exchange, such as longer exposure in the content feed so they could collect more feedback, or ability to invite experienced members to provide expert reviews.

For the implementation of team labeling, platform designers could ask users to generate their own names or select among provided options after forming teams on-demand [54]. Alternatively, platform designers could consider using labels based on an existing relationship, such as shared interests in a design genre/style, similar years of experience, or adjacent time of joining the platform. Future work could also test encouraging feedback providers to proactively form groups with designers. In our study, we tested short-term ingroup framing for one design-feedback cycle. Future work could examine long-term framings spanning multiple projects and how the effects of the framing might change over time. Future work could also explore what proportions of the observed effects could be attributed to interdependency and team labeling respectively.

Ingroup framing may be inapplicable in some scenarios. For example, when it is critical for feedback providers to provide an objective analysis of the work, an ingroup framing may bias their evaluation of the work. In these scenarios, the design process intervention may serve as an alternative. Platform designers could provide guidelines and templates to help content creators to write an effective design process narrative. Since some content creators may be unwilling to invest the effort, platform designers could offer this as a suggestion and highlight the benefits of receiving higher quality feedback. Prior work has explored scaffolding processes that help users to craft effective help-seeking emails [34]. Future work could explore similar scaffolding that makes it easier for content creators to write an effective narrative. Researchers have also explored recording design processes via design editor add-ons and re-creating the process using action logs [24,40]. Afterward, a content creator could annotate key frames to quickly compose a design process narrative.

Negative experience narrative is most effective at encouraging feedback providers to take a more disapproving stance against harsh feedback on the platform. This is particularly important as our results showed participants were more tolerant of harsh feedback after performing the task in the control framing/control article condition. The experience of providing feedback might have made them more inclined to justify harsh feedback. Negative experience helps to reverse this trend. Platform designers could selectively present this feature if harsh feedback reaches an undesirable level. Since asking the designer to write about past harsh feedback incurs additional work, platform designers could consider alternative methods to mitigate the costs. One way is to use harsh feedback the content creator received previously to showcase the negative experience. Meanwhile, content creators could choose to paraphrase the exchange and add their emotional responses to make the intervention more effective. Previous work shows negative valence posts may make users feel this type of content is acceptable and further incite more posts of similar valence [14]. Content creator's comments may negate this influence by conveying how such feedback is undesirable. Platform designers could also consider creating a pool of negative experience narratives and present them during an onboarding process of the community. We also observed that the negative experience narrative condition stimulated participants to accept responsibility to intervene when harsh feedback occurred. Future work should explore if this reported attitude translates into intervening action beneficial for the feedback exchange community.

We observed mixed results regarding the interaction of the interventions. The ingroup framing/negative experience condition led to the highest level of perspective-taking, a core aspect of empathy. Platform designers could use these two conditions together if empathy towards the designer is most important. Such a scenario may occur if the affective states of a content creator

had recently been affected by unfavorable interactions, such as receiving harsh feedback. On the other hand, using ingroup framing together with other narrative empathy conditions did not further enhance the effects of the narratives in terms of feedback quality and invested effort level. Platform designers could use a single intervention to maximize these measures while minimizing the overhead in implementation. One possible explanation of this pattern is diminishing marginal utility. With the same payment, a second intervention might not be enough to elicit meaningfully more effort in the task. Future work could test these interventions with participants under different incentive schemes and examine whether the combinations of the interventions could indeed lead to even higher feedback quality and effort levels.

6 LIMITATIONS

We used a single poster design in our experiment. Future work should evaluate the generalizability of our results to other genres of creative work. We may observe different patterns for artistic expressions with more abstract goals, where the standard for high-quality feedback is more subjective and less clear. Also, for designs that require a substantially longer time to review, such as a feature movie or a book, our interventions may have different effect sizes. Most participants in our experiment had a moderate amount of experience in design. Domain experts with extensive experience may also react differently to the interventions.

7 CONCLUSION

The results of the experiment reported in this paper showed that the narrative empathy interventions, including reading narratives about a prior negative experience and reading about the design process related to the current project, and ingroup framing both improved feedback quality and increased effort invested in the task. When reading a negative experience narrative, participants reported a more disapproving stance towards harsh feedback and more likely to intervene if observed in practice. Our work contributes new interventions that platform designers can implement to further improve the feedback exchange process.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under awards NSF CMMI 14-62693 and IIS 15-30818.

REFERENCES

[1] Judd Antin and Elizabeth F. Churchill. 2011. Badges in social media: A social psychological perspective. *Chi 2011* (2011), 1–4.

[2] Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *J. Pers. Soc. Psychol.* 63, 4 (1992), 596–612. DOI:https://doi.org/10.1037/0022-3514.63.4.596

[3] Richard P. Bagozzi and David J. Moore. 1994. Public Service Advertisements: Emotions and Empathy Guide Prosocial Behavior. *J. Mark.* 58, 1 (January 1994), 56. DOI:https://doi.org/10.2307/1252251

[4] Paul Baker. 2006. Moral panic and alternative identity construction in Usenet. *J. Comput. Commun.* 7, 1 (June 2006), 0–0. DOI:https://doi.org/10.1111/j.1083-6101.2001.tb00136.x

[5] C. Daniel Batson, Johee Chang, Ryan Orr, and Jennifer Rowland. 2002. Empathy, Attitudes, and Action: Can Feeling for a Member of a Stigmatized Group Motivate One to Help the Group? *Personal. Soc. Psychol. Bull.* 28, 12 (December 2002), 1656–1666. DOI:https://doi.org/10.1177/014616702237647

[6] C. Daniel Batson, Shannon Early, and Giovanni Salvarani. 1997. Perspective Taking: Imagining How Another Feels Versus Imaging How You Would Feel. *Personal. Soc. Psychol. Bull.* 23, 7 (July 1997), 751–758. DOI:https://doi.org/10.1177/0146167297237008

[7] C Daniel Batson and Nadia Ahmad. 2001. Empathy induced altruism in a prisoner's dilemma: What if the target of empathy has defected? *Eur. J. Soc. Psychol.* 36, January 2000 (2001), 25–36. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/ejsp.26/full

[8] Aditya Bharadwaj, Pao Siangliulue, Adam Marcus, and Kurt Luther. 2019. Critter. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. DOI:https://doi.org/10.1145/3290605.3300769

[9] Marisa Bortolussi and Peter Dixon. 2002. *Psychonarratology*. Cambridge University Press, Cambridge. DOI:https://doi.org/10.1017/CBO9780511500107

[10] Lisa M. Brown, Margaret M. Bradley, and Peter J. Lang. 2006. Affective reactions to pictures of ingroup and outgroup members. *Biol. Psychol.* 71, 3 (March 2006), 303–311. DOI:https://doi.org/10.1016/j.biopsycho.2005.06.003

[11] Tad T. Brunyé, Tali Ditman, Caroline R Mahoney, Jason S Augustyn, and Holly a Taylor. 2009. When You and I Share Perspectives. *Psychol. Sci.* 20, 1 (January 2009), 27–32. DOI:https://doi.org/10.1111/j.1467-9280.2008.02249.x

[12] Erin E. Buckels, Paul D. Trapnell, and Delroy L. Paulhus. 2014. Trolls just want to have fun. *Pers. Individ. Dif.* 67, (September 2014), 97–102. DOI:https://doi.org/10.1016/j.paid.2014.01.016

[13] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. 2017. The Bag of Communities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 3175–3187. DOI:https://doi.org/10.1145/3025453.3026018

[14] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, 1217–1230. DOI:https://doi.org/10.1145/2998181.2998213

[15] Mina Cikara, Emile G. Bruneau, and Rebecca R. Saxe. 2011. Us and Them: Intergroup Failures of Empathy. *Curr. Dir. Psychol. Sci.* 20, 3 (June 2011), 149–153. DOI:https://doi.org/10.1177/0963721411408713

[16] Jonathan Cohen. 2001. Defining Identification: A Theoretical Look at the Identification of Audiences With Media Characters. *Mass Commun. Soc.* 4, 3 (2001), 245–264. DOI:https://doi.org/10.1207/S15327825MCS0403_01

[17] Amy Cook, Jessica Hammer, Salma Elsayed-Ali, and Steven Dow. 2019. How guiding questions facilitate feedback exchange in project-based learning. *Conf. Hum. Factors Comput. Syst. - Proc.* (2019), 1–12. DOI:https://doi.org/10.1145/3290605.3300368

[18] Mark H. Davis. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 1 (1983), 113–126. DOI:https://doi.org/10.1037/0022-3514.44.1.113

[19] Mark H. Davis. 1995. *Empathy: A Social Psychological Approach*. Westview Press.

[20] Maja Djikic, Keith Oatley, and Mihnea C. Moldoveanu. 2013. Reading other minds: Effects of literature on empathy. *Sci. Study Lit.* 3, 1 (2013), 28–47. DOI:https://doi.org/10.1075/ssol.3.1.06dji

[21] Steven P Dow, Elizabeth M Gerber, and Audris Wong. 2013. A pilot study of using crowds in the classroom. *CHI* (2013), 227. DOI:https://doi.org/10.1145/2470654.2470686

[22] Nancy Eisenberg and Richard A Fabes. 1990. Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motiv. Emot.* 14, 2 (June 1990), 131–149. DOI:https://doi.org/10.1007/BF00991640

[23] John D. Foubert and Bradford C. Perry. 2007. Creating Lasting Attitude and Behavior Change in Fraternity Members and Male Student Athletes. *Violence Against Women* 13, 1 (January 2007), 70–86. DOI:https://doi.org/10.1177/1077801206295125

[24] C. Ailie Fraser, Joy O. Kim, Hijung Valentina Shin, Joel Brandt, and Mira Dontcheva. 2020. Temporal Segmentation of Creative Live Streams. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. DOI:https://doi.org/10.1145/3313831.3376437

[25] Susan Gair. 2013. Inducing Empathy: Pondering Students' (In)Ability to Empathize With an Aboriginal Man's Lament and What Might Be Done About It. *J. Soc. Work Educ.* 49, 1 (January 2013), 136–149. DOI:https://doi.org/10.1080/10437797.2013.755399

[26] Sarah Gielen, Elien Peeters, Filip Dochy, Patrick Onghena, and Katrien Struyven. 2010. Improving the effectiveness of peer feedback for learning. *Learn. Instr.* 20, 4 (August 2010), 304–315. DOI:https://doi.org/10.1016/j.learninstruc.2009.08.007

[27] William G. Graziano, Meara M. Habashi, Brad E. Sheese, and Renée M. Tobin. 2007. Agreeableness, empathy, and helping: A person × situation perspective. *J. Pers. Soc. Psychol.* 93, 4 (2007), 583–599. DOI:https://doi.org/10.1037/0022-3514.93.4.583

[28] Michael D Greenberg, Matthew W Easterday, and Elizabeth M Gerber. 2015. Critiki. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition - C&C '15*, 235–244. DOI:https://doi.org/10.1145/2757226.2757249

[29] Jennifer N. Gutsell and Michael Inzlicht. 2012. Intergroup differences in the sharing of emotive states: Neural evidence of an empathy gap. *Soc. Cogn. Affect. Neurosci.* 7, 5 (2012), 596–603. DOI:https://doi.org/10.1093/scan/nsr035

[30] Tyler Hamby and Wyn Taylor. 2016. Survey Satisficing Inflates Reliability and Validity Measures. *Educ. Psychol. Meas.* 76, 6 (December 2016), 912–932. DOI:https://doi.org/10.1177/0013164415627349

[31] Christian Happ, André Melzer, and Georges Steffgen. 2011. Bringing Empathy into Play: On the Effects of Empathy in Violent and Nonviolent Video Games. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 371–374. DOI:https://doi.org/10.1007/978-3-642-24500-8_44

[32] Grit Hein, Giorgia Silani, Kerstin Preuschoff, C. Daniel Batson, and Tania Singer. 2010. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 1 (2010), 149–160.

DOI:https://doi.org/10.1016/j.neuron.2010.09.003

[33] Fernanda Herrera, Jeremy Bailenson, Erika Weisz, Elise Ogle, and Jamil Zaki. 2018. Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLoS One* 13, 10 (October 2018), e0204494. DOI:https://doi.org/10.1371/journal.pone.0204494

[34] Julie S Hui, Darren Gergle, and Elizabeth M Gerber. 2018. IntroAssist. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–13. DOI:https://doi.org/10.1145/3173574.3173596

[35] Suzanne Keen. 2006. A Theory of Narrative Empathy. *Narrative* 14, 3 (2006), 207–236. DOI:https://doi.org/10.1353/nar.2006.0015

[36] Peter Kollock and Marc Smith. 1996. Managing the virtual commons. . 109. DOI:https://doi.org/10.1075/pbns.39.10kol

[37] Markus Krause, Tom Garncarz, JiaoJiao Song, Elizabeth M Gerber, Brian P Bailey, and Steven P Dow. 2017. Critique Style Guide. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4627–4639. DOI:https://doi.org/10.1145/3025453.3025883

[38] Claus Lamm, C. Daniel Batson, and Jean Decety. 2007. The Neural Substrate of Human Empathy: Effects of Perspective-taking and Cognitive Appraisal. *J. Cogn. Neurosci.* 19, 1 (January 2007), 42–58. DOI:https://doi.org/10.1162/jocn.2007.19.1.42

[39] Cliff Lampe and Paul Resnick. 2004. Slash(dot) and burn. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 543–550. DOI:https://doi.org/10.1145/985692.985761

[40] Zipeng Liu, Zhicheng Liu, and Tamara Munzner. 2020. Data-driven Multi-level Segmentation of Image Editing Logs. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. DOI:https://doi.org/10.1145/3313831.3376152

[41] Peter John Loewen, Greg Lyle, and Jennifer S. Nachshen. 2010. An eight-item form of the Empathy Quotient (EQ) and an application to charitable giving. (2010), 1–14.

[42] Kurt Luther, Jari-lee Tolentino, Wei Wu, Amy Pavel, Brian P Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 473–485. DOI:https://doi.org/10.1145/2675133.2675283

[43] Xiaojuan Ma, Yu Li, Jodi Forlizzi, and Steven Dow. 2015. Exiting the design studio: Leveraging online participants for early-stage design feedback. *Proc. ACM Conf. Comput. Support. Coop. Work. CSCW* 2015-Janua, (2015), 676–685. DOI:https://doi.org/10.1145/2675133.2675174

[44] Clifford Nass, B.J. Fogg, and Youngme Moon. 1996. Can computers be teammates? *Int. J. Hum. Comput. Stud.* 45, 6 (December 1996), 669–678. DOI:https://doi.org/10.1006/ijhc.1996.0073

[45] Tricia J Ngoon, C Ailie Fraser, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive Guidance Techniques for Improving Creative Feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–11. DOI:https://doi.org/10.1145/3173574.3173629

[46] Thi Thao Duyen T. Nguyen, Thomas Garncarz, Felicia Ng, Laura A. Dabbish, and Steven P. Dow. 2017. Fruitful Feedback. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, 1024–1034. DOI:https://doi.org/10.1145/2998181.2998319

[47] Amanda B. Nickerson, Ariel M. Aloe, Jennifer A. Livingston, and Thomas Hugh Feeley. 2014. Measurement of the bystander intervention model for bullying and sexual harassment. *J. Adolesc.* 37, 4 (June 2014), 391–400. DOI:https://doi.org/10.1016/j.adolescence.2014.03.003

[48] Keith Oatley. 1995. A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics* 23, 1–2 (1995), 53–74. DOI:https://doi.org/10.1016/0304-422X(94)P4296-S

[49] Patricia A. Oswald. 1996. The Effects of Cognitive and Affective Perspective Taking on Empathic Concern and Altruistic Helping. *J. Soc. Psychol.* 136, 5 (October 1996), 613–623. DOI:https://doi.org/10.1080/00224545.1996.9714045

[50] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. *Austin, TX Univ. Texas Austin* (2015), 1–22. DOI:https://doi.org/10.15781/T29G6Z

[51] Irene Rae, Leila Takayama, and Bilge Mutlu. 2012. One of the gang. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 3091. DOI:https://doi.org/10.1145/2207676.2208723

[52] Renate L.E.P. Reniers, Rhiannon Corcoran, Richard Drake, Nick M. Shryane, and Birgit A. Völlm. 2011. The QCAE: A questionnaire of cognitive and affective empathy. *J. Pers. Assess.* 93, 1 (2011), 84–95. DOI:https://doi.org/10.1080/00223891.2010.528484

[53] Paul Resnick and Richard Zeckhauser. 2002. Trust among strangers in internet transactions: Empirical analysis of eBay' s reputation system. *Adv. Appl. Microeconomics* 11, (2002), 127–157. DOI:https://doi.org/10.1016/S0278-0984(02)11030-3

[54] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*, 75–85. DOI:https://doi.org/10.1145/2642918.2647409

[55] Annika Richterich. 2013. ' Karma , Precious Karma!' Karmawhoring on Reddit and the Front Page ' s

Econometrisation. *J. Peer Prod.* May (2013), 1–12. Retrieved from http://peerproduction.net

[56] Jeffrey Rzeszotarski and Aniket Kittur. 2012. CrowdScape. In *Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST '12*, 55. DOI:https://doi.org/10.1145/2380116.2380125

[57] Jeffrey M Rzeszotarski and Aniket Kittur. 2011. Instrumenting the crowd. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, 13. DOI:https://doi.org/10.1145/2047196.2047199

[58] D. Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instr. Sci.* 18, 2 (June 1989), 119–144. DOI:https://doi.org/10.1007/BF00117714

[59] D. Royce Sadler. 1998. Formative Assessment: revisiting the territory. *Assess. Educ. Princ. Policy Pract.* 5, 1 (March 1998), 77–84. DOI:https://doi.org/10.1080/0969595980050104

[60] Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong Cherie Chen, Likang Sun, and Geoff Kaufman. 2019. Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–14. DOI:https://doi.org/10.1145/3290605.3300836

[61] Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong Cherie Chen, Likang Sun, and Geoff Kaufman. 2019. Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–14. DOI:https://doi.org/10.1145/3290605.3300836

[62] Jellie Sierksma, Jochem Thijs, and Maykel Verkuyten. 2015. Ingroup bias in children's intention to help can be overpowered by inducing empathy. *Br. J. Dev. Psychol.* 33, 1 (March 2015), 45–56. DOI:https://doi.org/10.1111/bjdp.12065

[63] Edgar A. Smith and J. Peter Kincaid. 1970. Derivation and Validation of the Automated Readability Index for Use with Technical Materials. *Hum. Factors J. Hum. Factors Ergon. Soc.* 12, 5 (October 1970), 457–564. DOI:https://doi.org/10.1177/001872087001200505

[64] Walter G. Stephan and Krystina Finlay. 1999. The Role of Empathy in Improving Intergroup Relations. *J. Soc. Issues* 55, 4 (January 1999), 729–743. DOI:https://doi.org/10.1111/0022-4537.00144

[65] Stefan Stürmer, Mark Snyder, Alexandra Kropp, and Birte Siem. 2006. Empathy-Motivated Helping: The Moderating Role of Group Membership. *Personal. Soc. Psychol. Bull.* 32, 7 (July 2006), 943–956. DOI:https://doi.org/10.1177/0146167206287363

[66] Mark Tarrant, Sarah Dazeley, and Tom Cottom. 2009. Social categorization and empathy for outgroup members. *Br. J. Soc. Psychol.* 48, 3 (September 2009), 427–446. DOI:https://doi.org/10.1348/014466608X373589

[67] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J. Lang. Soc. Psychol.* 29, 1 (March 2010), 24–54. DOI:https://doi.org/10.1177/0261927X09351676

[68] Matthew J. Traxler and Morton Ann Gernsbacher. 1993. Improving written communication through perspective-taking. *Lang. Cogn. Process.* 8, 3 (August 1993), 311–334. DOI:https://doi.org/10.1080/01690969308406958

[69] Theresa K. Vescio, Gretchen B. Sechrist, and Matthew P. Paolucci. 2003. Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions. *Eur. J. Soc. Psychol.* 33, 4 (July 2003), 455–472. DOI:https://doi.org/10.1002/ejsp.163

[70] Helen Wauck, Yu-Chun (Grace) Yen, Wai-Tat Fu, Elizabeth Gerber, Steven P. Dow, and Brian P. Bailey. 2017. From in the Class or in the Wild? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 5580–5591. DOI:https://doi.org/10.1145/3025453.3025477

[71] Drika Weller and Kristin Hansen Lagattuta. 2013. Helping the In-Group Feels Better: Children's Judgments and Emotion Attributions in Response to Prosocial Dilemmas. *Child Dev.* 84, 1 (January 2013), 253–268. DOI:https://doi.org/10.1111/j.1467-8624.2012.01837.x

[72] Y. Wayne Wu and Brian P. Bailey. 2018. Soften the Pain, Increase the Gain. *Proc. ACM Human-Computer Interact.* 2, CSCW (November 2018), 1–20. DOI:https://doi.org/10.1145/3274455

[73] Y Wayne Wu and Brian P Bailey. 2016. Novices Who Focused or Experts Who Didn't? In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 4086–4097. DOI:https://doi.org/10.1145/2858036.2858330

[74] Y Wayne Wu and Brian P Bailey. 2017. Bitter Sweet or Sweet Bitter? In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition - C&C '17*, 137–147. DOI:https://doi.org/10.1145/3059454.3059458

[75] W Xiong, D. Litman, and C. Schunn. 2012. Natural Language Processing techniques for researching and improving peer feedback. *J. Writ. Res.* 4, 2 (November 2012), 155–176. DOI:https://doi.org/10.17239/jowr-2012.04.02.3

[76] Anbang Xu, Shih-wen Huang, and Brian Bailey. 2014. Voyant. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*, 1433–1444. DOI:https://doi.org/10.1145/2531602.2531604

[77] Anbang Xu, Huaming Rao, Steven P. Dow, and Brian P. Bailey. 2015. A Classroom Study of Using Crowd Feedback in the Iterative Design Process. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, 1637–1648. DOI:https://doi.org/10.1145/2675133.2675140

[78] Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong. 2011. Review spotlight. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, 1541. DOI:https://doi.org/10.1145/1978942.1979167

[79] Yu-Chun Grace Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 2016. Social Network, Web Forum, or Task Market? In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, 773–784. DOI:https://doi.org/10.1145/2901790.2901820

[80] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Björn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 1003–1015. DOI:https://doi.org/10.1145/2818048.2819953

[81] Holm's Sequential Bonferroni Procedure. In *Encyclopedia of Research Design*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States. DOI:https://doi.org/10.4135/9781412961288.n178

Received June 2020; accepted July 2020.