

## “At Least the Pizzas You Make Are Hot”: Norms, Values, and Abrasive Humor on the Subreddit r/RoastMe

Anna Kasunic, Geoff Kaufman

Carnegie Mellon University  
Human-Computer Interaction Institute  
Pittsburgh, PA, USA  
akasunic@andrew.cmu.edu, gfk@cs.cmu.edu

### Abstract

We present a mixed methods study of the online forum r/RoastMe, a comedy-focused subreddit of the parent site reddit.com, wherein members post photos of themselves to be ridiculed by other members; the site generally encourages harsh and offensive forms of humor in these interpersonal exchanges. We conducted semi-structured interviews with sixteen participants (both “roasters” and “roastees”) in the online forum to understand their motivations for participating, their experiences in the subreddit, and their perceptions of their and other members’ participation. To complement our qualitative analyses, we also analyzed a RoastMe data set of over 9,000 image posts and 230,000 comments from June-August of 2017. From our interviews, we found that, like other deviant online communities, RoastMe relies on a specific set of norms. In RoastMe, roasters rely heavily on perspective-taking rather than dissociation from their targets, roastees highly value the often scathing assessments offered by users on RoastMe, and, despite the salience of norms that enhance feelings of safety, there is lingering concern among participants about the potential for emotional or psychological harm. Our quantitative analyses confirm many of the statements made in our qualitative interviews and provide further insights into the specific nature of interactions on the subreddit. Our study directs us toward different vantage points from which to design online community spaces that account for or leverage users’ predilections for baiting behaviors, harsh judgments, and caustic humor.

### Introduction

Few people would likely characterize verbal attacks, decidedly unconstructive criticisms, and otherwise offensive comments based on one’s physical appearance as the foundation of a positive online experience. However, the 700,000+ subscribers to r/RoastMe, a humor-focused sub-community (subreddit) of the large online forum reddit.com, might beg to differ. On RoastMe, people willingly post photos of themselves to have other users ridicule, offering themselves up as the target of jokes and comments that are by standard norms offensive, impolite, or politically incorrect. For example, responses to photos often include markers of racism, sexism, violence, body-shaming, and homophobia; comments of the

ilk quoted in this paper’s title are relatively mild and innocuous examples of RoastMe’s typical content.

Roasting as a form of comedy is nothing new; dark humor and satire have a long history in comedy (take the works of ancient Greek playwrights and Shakespeare, for example), and roasting, or subjecting a singular “guest of honor” to insulting but generally good-natured jokes for an audience’s amusement, is a concept that predates both Reddit and the Internet. In the 20th century, it became common practice among (wealthy) clubs and organizations to hold testimonial dinners praising a guest of honor. The Friar’s Club, founded in 1904, began putting a spin on these testimonial dinners, introducing jabs and sarcastic humor. By the 1940s, insult humor became the defining features of the Friar’s Club roasts (Adams 1976). The practice has extended to modern times, with the television network Comedy Central hosting roasts of celebrities like singer Justin Bieber, actor James Franco, and even the current president (then television host) Donald Trump (Blauvelt 2011).

On RoastMe, within a particular roasting thread, one participant (the roastee) posts a photo of themselves, thereby offering themselves up to be the target of the roast. As in traditional roasts, on RoastMe the “roastees” ostensibly participate of their own volition, and the directive for all members of the community is “comedy, not hate.” However, in traditional roasts, roasters are typically friends and colleagues of the roastee, and roasters, the roastee, and the audience (also friends, acquaintances, and general supporters of the roastee) participate together in the event in person. In contrast, on RoastMe, the roasters are identified only by their anonymous Reddit handles (pseudonyms), and are strangers to the roastee. The community may not be everyone’s cup of tea, but its situation at the intersection of anonymity and exposition, the subversive behaviors and norms the community espouses, and the inter- and intra-personal dynamics it engenders make it a fascinating and instructive milieu for human-computer interaction research. RoastMe provides an opportunity to study the norms and values that drive and impact participation in a subversive humor community and in turn, to consider how the dynamics on such a site could inform the design of other online domains.

Seeking to understand the norms of r/RoastMe within the context of previously studied communities and interactions, including subversive online humor communities and photo

sharing communities, we embarked on a two-part mixed methods study of r/RoastMe. In the first part of the study, we conducted semi-structured interviews with 16 members of the RoastMe community. In line with previous work on unconventional or subversive online communities, the RoastMe community highly values abidance by community-specific norms. However, we found that RoastMe differentiates itself through roasters' heavy use of perspective-taking, roastees' values and presentation norms in relation to harsh judgment, and the high level of concern about harm generation in the community.

In the second part of the study, we extend our qualitative findings by quantitatively exploring a data set of RoastMe activity from June-August 2017 collected through the Reddit API. The data set consists of roughly 9,000 posts of photos to be roasted, and 230,000 comments on (i.e., roasts of) these posts. In this part of the study, we explore whether our interviewees' descriptions of the community cohere with the activity markers and participant attributes present in the data set, and we cluster topics present in the data set using Latent Dirichlet Allocation (LDA). In addition, we build a linear regression model to predict the popularity of posts based on roastee attributes, with our findings suggesting that those who appear to be female, and have Reddit activity that indicates possible mental health struggles are more likely to receive more comments (roasts). Together, our qualitative and quantitative explorations point to design changes that could improve user experiences in RoastMe, and suggest HCI opportunities to meet "deviant" desires that could extend to domains beyond the context of RoastMe.

## Background

RoastMe operates under a subversive set of norms, involves online self-presentation in a context that differs from more mainstream online photographic presentations, and encourages humor that would be considered offensive or inappropriate in other contexts. Thus, to understand RoastMe and parse the results of our exploration, we consider the extant literature in these three areas.

### Subversive Online Behaviors and Communities

Online anonymity can increase non-normative behaviors in online communities (Kraut, Resnick, and Kiesler 2016). However, as explained by the social identity model of deindividuation effects, group identity can supersede individual identity (Reicher, Spears, and Postmes 1995). This can result in behaviors that are consistent with community norms, even if the community itself identifies as subversive in nature and departs from more traditional norms. Multiple academic studies lend credence to this model and help to contextualize our investigation of norms in RoastMe. For example, a study of the collection of forums called Something Awful identified specific norms guiding community practices that deviate from the standard norms present in other online communities, such as the use of rough and abusive humor and a focus on high-quality content, regardless of how offensive that content might be (Pater et al. 2014). Reddit as a whole shares a loose set of norms (known as "reddiquette")

that differ from those that are prominent on other popular, more conventional popular social media platforms like Facebook, and there are many subreddits in addition to RoastMe that embrace aspects of *schadenfreude* (Massanari 2015). Research has found that people can be more disinhibited in online spaces (Suler 2004), thereby increasing the proclivity to engage in forms of more subversive or boundary-crossing humor and interaction when online. Scholars have also studied other spaces that would be considered far more deviant than RoastMe, such as online pedophilia networks (Holt, Blevins, and Burkert 2010), and persistently find a set of norms and standards—albeit aberrant—that guide community behaviors.

A highly relevant case in point: as part of a recent study of a 2015 ban of two self-identifying hate-speech oriented subreddits (r/CoonTown and r/fatpeoplehate) (Chandrasekharan et al. 2017), researchers tracked the subreddits to which participants in the banned subreddits migrated; RoastMe emerged as one the top ten migration destinations for both of the banned subreddits. Thus, despite RoastMe's adamant "comedy, not hate" stance, it appears that the site's brand of humor may be similar enough to hate speech that it appeals to users who aim to directly harm and provoke their targets. In our study, we investigate participants' relationships to deviant community norms, paying special attention to situations that suggest uncertainty about norms and norm violation to understand how we can better accommodate needs and concerns of participants in communities with subversive norms.

### Online Photographic Self-Presentation

Existing literature suggests that the particular context and the relationship between the person who posts a photograph and the audience who views it dictate self-presentational behaviors, goals, and outcomes. For social networks such as Facebook, where the assumed audience is comprised of friends and colleagues, studies have found that people are generally motivated to self-present authentically in online photographs. For example, a study that compared personality survey ratings and observer ratings of Facebook photos found that participants' photos reflected their actual personality rather than an idealized version of the self (Back et al. 2010), and a computer vision analysis of Facebook profile photos in another study revealed similar findings (Celli, Bruni, and Lepri 2014). However, dynamics change in other online contexts. For example, in online dating contexts, users may be more likely to present an idealized version of the self in the hopes of attracting interest (Ellison, Heino, and Gibbs 2006). The photographic presentation norms of RoastMe differ from the norms of previously studied contexts, for in RoastMe, people are specifically seeking ridicule based on their appearance, and as we will discuss in our results, may self-present "ridiculously" by emphasizing certain flaws or making absurd expressions rather than self-presenting in an authentic or idealized fashion. Thus, as part of our qualitative investigation, we studied the specific nature of photographic self-presentation norms in the RoastMe context and consider how this may inform the design of online photographic presentation contexts more broadly.

## Humor

The brand of humor celebrated in RoastMe could be said to fall into the camp of the “lulz,” internet-speak for *schadenfreude* that represents “amusement derived from others’ misfortunes” (Phillips 2015). Someone unfamiliar with the terrain may be unable to distinguish between, say, hateful or overtly harm-seeking comments, and the comments on RoastMe. However, as messages and guidelines on the site itself emphasize, and as our interview participants echoed, RoastMe is a forum dedicated to comedy, not hate. To study a form of humor that some might characterize as funny while others might characterize as hateful or offensive, we can draw from benign violation theory. This theory proposes that humor occurs only when the following two conditions are simultaneously met: (1) the situation violates a norm, and (2) the situation is considered to be safe or acceptable by those involved (Warren and McGraw 2014).

In a related study, the theory’s supporters argue that situations can be considered benign if a) alternative norms are in place that render the violation acceptable, if b) there is a weak commitment to the violated norm, and if c) there is psychological distance from the situation (McGraw and Warren 2010). The benign violation theory helps to explain why some people view certain content as very humorous, while others are appalled by the same material. As an example, “RIP trolls”—individuals who post purposefully offensive content on Facebook memorial pages for the deceased—may view their violations as benign and thereby humorous because they feel psychologically distant from their prey; indeed the literature on trolling has repeatedly found that trolls tend to emotionally dissociate from their targets (Phillips 2015), which we can view as a form of psychological distancing under the benign violation theory.

In RoastMe, we can expect that the degree to which participants view different violations as benign is likely to vary greatly by individuals, but community norms may enable certain violations to feel more safe and acceptable. Moreover, different forms of psychological distance from the situation may help enhance the perception of harmlessness. In our study, we sought to understand the contexts that designate roasts as benign for different individuals, and to identify where the lines become blurry. Our study of a comedy-focused online community that embraces deviant community norms thus builds upon previous work on norms in a) subversive online communities, b) online photographic presentation, and c) humor. By exploring how norms and values are defined, created, and grappled with in RoastMe, we contribute to existing literature and consider how to better design for the desires and values uncovered by our study.

## Methods

### Semi-Structured Interviews

For this study, we received IRB approval and recruited 16 participants from r/RoastMe. To do so, we contacted the Reddit account shared by all the group moderators, and a moderator assisted us by pinning a description of the project to the front page of the subreddit. Interested parties were

instructed to contact us via Reddit and/or email. Each interview lasted about 1 hour and took place over Skype or Google Hangouts. Specific questions we asked as part of our interview protocol included: “What, if anything, do you find challenging about participating in RoastMe? Rewarding? How would you describe the RoastMe community? How did you come to be a roaster/roastee (probing on specific circumstances)? What motivates you to participate as a roaster/roastee?” We compensated all participants with a \$15 Amazon gift card.

As part of our study, we included a brief, write-in demographic questionnaire; participants in our study skewed young, white, male, and North American. The average age was 22.5, with a median of 19, 13/16 were male, and 9/16 identified as white or Caucasian. Of the remaining seven, three identified as Latino/a or Hispanic, one identified as brown, one identified as African, and two identified as Asian and/or Chinese. Twelve participants hailed from the US, and one each came from the UK, Canada, Russia and South Africa, respectively. For 15/16, English was a primary language; three of these native English speakers also held a second native tongue (one participant each also spoke Chinese, Spanish, and Afrikaans). There was also one Russian speaker who did not identify English as a primary language. For occupation, five identified as students, three as employed in sales, two as self-employed, and two as artists/musicians, with the remaining six employed in consulting, teaching, IT, and penetration testing/ethical hacking. Two of the participants were RoastMe moderators who had also participated as members, while the rest were general members.

There are two primary ways for members in the community to participate in r/RoastMe: either 1) by posting as roastees, in which they upload a picture of themselves with an adjoining title asking for others to roast them, or by 2) commenting as roasters, in which they provide insulting/humorous comments in forum threads about the photos and titles that other members have posted as roastees. Two of our study participants had exclusively posted as roastees, six had exclusively commented as roasters, and the remaining eight participants had engaged both as roasters and roastees.

We audio-recorded and transcribed all interviews, and a single member of the team conducted open coding on the transcripts, starting with high-level emergent themes and then iteratively refining the codes. A preliminary codebook and a sample of 20% of the responses were shared with the other team member for independent coding. Based on comparisons and discussion of disagreements in the resulting codes, the original coder revised and refined the codebook and shared a new sample of quotes reflecting about 20% of the data set. Based on this final test set, the Cohen’s Kappa measure of inter-rater reliability from this subset was 0.98. Patterns and findings emerged around depictions of the community and interactions and behaviors therein, revealed in low-level codes such as: (1) “outlet.haven” to describe RoastMe as a safe place to let out non-politically correct rhetoric, frustration, and other types of potentially offensive commentary; (2) “perspective-taking,” to reflect ways in which roasters adopted the point-of-view of others while roasting (e.g., putting themselves in the poster’s shoes, iden-



tifying with other participants, or thinking about the poster's feelings or desires); and (3) reasons why roasts were seen as negative or undesirable, such as "direct\_support," reflecting disapproval of comments offering support or flattery to the poster instead of roasting him/her, which violates RoastMe rules.

### Analysis of Reddit API Data

Our qualitative findings resulted in several further questions that we subsequently explored through quantitative analyses. The Reddit API allows users to access historical comments and posts. Although we initially used the Reddit API in conjunction with the Python wrapper PRAW to collect data, latency and limitations on the number of calls led us to a publicly accessible version of Reddit's historical data on Google BigQuery. Using Google BigQuery's SQL interface, we collected all the image posts and comments (roasts) from RoastMe between June 1, 2017 through August 31, 2017, totalling over 290,000 comments (roasts) and over 12,000 posts (photos to be roasted). The data set included features such as (for roatee posts) time of posts, number of comments received, author of the post, and text title accompanying the post, and (for roaster comments) body of the comment, and parent ID of the post with which it is associated. Using the implementation of the Latent Dirichlet Allocation (LDA) in NLTK (Natural Language Toolkit) for Python, we then performed topic modeling on the roatee titles.

We conducted additional queries to mark each roaster and roatee with the number of posts (roatees) and roasts/comments (roasters) they had made since the subreddit's birth over two years ago to more comprehensively determine the nature of roasters' and roatees' participation in the subreddit. Lastly, because several of our interview participants had expressed uncertainty about the mental stability of some of the roatees, and the morality of roasting in such cases, we curated a list of 34 popular subreddits devoted to mental health struggles drawn from various subreddit guides and other online recommendations. This list included subreddits such as "suicidewatch," "suicidenotes," "depression," "eatingdisorder", "stopselfharm" and "survivorsofabuse." We marked any individuals in our data set who had submitted posts or comments to the list of mental-health related subreddits as possibly vulnerable to mental health issues.

Following these steps, we removed untidy, corrupt, and instances with key data missing; for example, we excluded from our analysis posts and associated comments that were missing the column for the number of comments received, and excluded comments that were not associated with any posts from the time period (e.g., comments on posts that had been submitted prior to June 1, 2017). This left us with 234,475 comment instances representing 53,885 unique roasters, and 9,032 posts representing 6,013 unique roatees. Next, we employed Microsoft's Cognitive Face and Emotion APIs to estimate the age, gender, and emotions of roatees' photos in our data set. Because all Reddit users are anonymous/pseudonymous, the Reddit API does not make available demographic information such as age or gender. However, by applying Microsoft's Cognitive Services to the data,

we can estimate these demographic features and incorporate them into our descriptions and analyses. We acknowledge that this is a rough and highly imperfect measure. For example, Microsoft's Face API will predict gender based on one's physiognomy, not one's self-presentation or internal identity. Thus, an individual who self-presents and/or self-identifies as female or non-binary but has physical markers of a masculine facial structure will be classified as male, not female, by the Face API. Nonetheless, human perception shares some of the same flaws as algorithmic approaches, and many viewers on RoastMe may assume the gender of roatees based on their physical features, and roast them accordingly. For the purposes of our analyses, we thus deem our gender estimate to be a flawed but still reasonable proxy for gender. Microsoft does not offer clear metrics about its services' reliability, so we conducted an informal check on the data by taking a random sample of 100 image post instances and manually coding them as "male," "female," or "unknown"; when we compared this to the Microsoft Face API results for these instances, we found a 99% match. We incorporated these newly derived features into our data set.

Lastly, we built a linear regression machine learning model to predict the estimated popularity of a post using WEKA software as well as the ADAMS workflow tools to assist with converting data sets to ARFF format. Here, we used a subset of our data that included only posts for which we were able to obtain age and gender information; posts with low quality, blurry, or deleted photos were thus excluded from the analysis, leaving us with 4,710 instances.

Reddit has features to upvote and downvote comments and posts on any subreddit, but it does not display the actual number of upvotes and downvotes on the site (numbers are "fuzzed" to avoid spambotting) and returns only null values for upvotes and downvotes in its API. The API does make scores available, which are defined as the number of downvotes subtracted from the number of upvotes, but this measure provides limited information; for example, a post with 1,000 upvotes and 1,000 downvotes will share the same score as a post with one upvote and one downvote. Thus, we turn to the number of comments a post receives as a flawed but reasonable choice for our model's dependent variable, given that it approximates the amount of traffic that a post receives. We first divided our post data into train and test data sets using an 80:20 split, and iteratively refined the model. After tuning our model, we evaluated the final model on our complete data set using 10-fold cross-validation.

### Qualitative Results

First (1), we present findings that, in keeping with previous literature, show that the RoastMe community supports a specific set of norms. Next, we define how RoastMe's norms are unique, highlighting a) roasters' heavy use of perspective-taking, b) roatees' values and presentation norms in relation to harsh judgment, and c) the high level of concern about harm generation in the community. In the Discussion section, we will explain how these findings about RoastMe's norms and values are instructive for the design of online communities.

## Norm Adherence in RoastMe

Here, we present an overview of the ways in which adherence and commitment to norms is manifested on RoastMe, including the high value placed on humor and originality, the expectations around consent, and community- and administration-driven moderation practices. Gaining a general understanding of RoastMe normative guidelines and behaviors will help the reader to see how our study extends previous findings about adherence to norms in subversive communities, as well as provide context for comprehending the unique aspects of RoastMe's norms that can inform novel design directions for online communities and online self-presentation that we will discuss in the following sub-sections.

As stated directly on the site, RoastMe is about “comedy, not hate.” Our participants echoed this value in their discussion of the site. For example a participant who was also a moderator (p12) called the phrase their “rule number one.” They also expected other participants in the site to adhere to the guiding principle. For example, p16 explained, “RoastMe is about comedy; it’s about making people laugh. It’s not about starting a hatred thread. . . It’s just that: I wish people would understand the rules about RoastMe.”

The community also values creative, unique, and genuinely funny content; to post unoriginal content, then, is for many one of the highest RoastMe offenses, as investigations of other subversive humor forums have similarly shown (Pater et al. 2014). Taboo and offensive comments are perfectly acceptable within RoastMe; being unoriginal, however, falls outside the RoastMe norms and violates RoastMe values. One participant explains that roasts that encourage suicide are not off limits, but do violate the rules if they do so in a generic, uncreative way. He explains that such roasts are unacceptable, saying, “If it’s in the generic rules, like ‘Kill yourself,’ because this is against the rules—roasts like that I don’t think are allowed anymore. Because it’s just too flat-out and generic; if it was more creative then it would be allowed” (p1).

Another participant (p11) directed us towards a previous post he had made outlining the kinds of roasts that were not acceptable; here, he advocated against certain stereotypes because they were generic, overused, and therefore unfunny, violating the value in RoastMe of creative, humorous, and original roasts. These included racist, homophobic, and violent roasts; the key here is that the offensive nature of these roasts was not the racism, homophobia, or violence per se, but rather, the lack of originality in their deployment, in violation of the stipulations of RoastMe. Six participants actively helped enforce RoastMe rules, stating that they had either downvoted or directly reported comments to the moderators because they considered them to be unoriginal and of low quality, thereby violating the norms of RoastMe. As one participant explained, this can be a problem with newcomers who haven’t yet learned the RoastMe rules: “Most people, like, when new people join the community, it’s basically just racist comments and stuff like that until they realize these just get downvoted a lot, so then they come up with original content” (p9).

Of course, under the “comedy, not hate” rule, any com-

ments that are specifically intended to produce harm and not humor are not permitted either. To this end, six participants also shared that they had reported or downvoted comments that they felt were unnecessarily cruel. As one participant explained, “You know, if you’re saying like, ‘You look like your grandma just died,’ that’s funny. But if you say ‘I hope your grandma dies. Go burn in hell,’ that’s a little weird. Those comments usually get downvoted into oblivion, just gone” (p6). Another participant (p9) said they were not okay with people roasting people in the background of the photo, as this goes against RoastMe rules—people in the background did not directly give their consent to be roasted.

Participants also spoke negatively of actions that might be viewed positively in other contexts, but that violated the RoastMe rules and values and were therefore deemed to be inappropriate. For example, one participant (p11) said he disapproved when people offered direct, encouraging support to a roastee such as “I hope you’re okay,” explaining, “What bothers me is when people mess with the process of RoastMe.” A moderator (p13) explained that flattery is not permitted: “One of the rules of RoastMe is that you can’t hit on any of the Roastees, so like, nobody can be like, ‘How can I Roast you? You’re actually kind of hot’ or whatever.”

In keeping with the literature on community norms and benign violation theory, seven of our participants specifically viewed RoastMe as an outlet, a safe space where they can release built up frustrations, “satisfy urges” as one participant (p11) put it, and give voice to taboo thoughts that they can’t express elsewhere. P11 compared RoastMe to trolling that happens on other internet sites and forums, saying that because so many social spaces encourage us to repress taboo and potentially harmful thoughts, we might have “explosions” that occur on trolling threads, RoastMe, or “any anonymous outlet that people can help unleash that.” Others echoed this urge to unleash, saying, for example, “sometimes it’s fun to say mean things that you can’t say in person” (p3).

Fourteen of our sixteen participants cited the expectations of the community to explain why comments they or others might deem as offensive in other contexts were copacetic on RoastMe. Acknowledging that people who post photos of themselves are willingly doing so, and know to expect harsh and insulting feedback, produced the sentiment that everything is “fair game” (p2). As p7 explained, “. . . it’s in a controlled environment where it’s expected that you’re going to be mean and you just be as funny about it as you can be.” Another participant (p15) contrasted it with social media cyberbullying, explaining that even if the comments look similar, the intents and expectations of RoastMe create a new, safer context in which such comments would not constitute bullying.

This feeling of safety within RoastMe was further supported by the forum’s moderation, as revealed by moderators themselves, as well as opinions other RoastMe participants expressed about moderators. Moderation helped participants feel that this was a safe space to engage in taboo or trolling-like behaviors without causing too much harm. Moderation of RoastMe includes efforts to ensure consent, such as stipulating that all roastees must hold up a hand-written sign

reading “RoastMe,” and further monitoring the posts to see if PhotoShop may have been used to forge consent. Participants who have posted a photo of themselves and then later regret it or are upset by the comments can contact moderators directly to request deletion, and moderators reported that they will immediately delete the thread; one moderator stated that such a situation has happened on several occasions. Moderators also try to protect youth, requiring that roastees be at least sixteen years of age. While much of this process is self-reported and verifications of age largely subjective, moderators also use bots, such as a bot that helps protect people with poor mental health by crawling through each roastie’s posting history to learn if they’ve posted in any pro-self-harm subreddits.

Understanding that RoastMe operates under a specific, defined set of norms is important context for delving into the unique nature of RoastMe’s norms that we will now discuss.

### Perspective-Taking

Where other target-perpetrator dynamics of subversive humor emphasize emotional dissociation (Phillips 2015), in our interviews, we learned that perspective-taking is a critical component of engagement in RoastMe. This finding indicates that there are norms of perspective-taking that are unique to the context presented in RoastMe. We saw perspective-taking emerge particularly often during the roasting process. Ten of the 14 participants who had commented as roasters discussed directly engaging in perspective-taking as part of the roast process. Participants often imagined themselves in the place of the person who had posted a photo, and thought about what they would want to hear in their place. For example, “I want them [‘roastee’] to feel as content as I would want to feel” (p3). Others (p12) asked themselves questions like, “What would I want to hear about that [physical] feature? What would I think that’s funny about myself based on that?” This might include taking on the perspective of the general RoastMe audience such as, “What would I think if I wasn’t in the picture and reading the comments?”

Roaster participants sometimes sought out roastees with whom they personally identified; for example, one of the three female participants we interviewed explained that she specifically tries to comment on other women because she shares their perspective. Identifying with roastees also can have benefits for the roasters. As p3 explains, “Being able to see someone that has the same flaws as me makes me realize that I’m not alone, and then being able to criticize someone else about it makes it a little bit better. . .” P15 spoke of how he hoped that by making roastees laugh, he could make them feel valuable and give them hope. Another participant (p11) explained, “Yeah, I really believe that laughter is the best thing you can do for someone. . . if it’s terrible and you can get someone to laugh, that’s a couple seconds at least where you can distract them from something bad and they can enjoy themselves.” Others spoke of tempering their roast so as to “zing” without being too harmful, indicating a mindfulness of other participants’ feelings.

The use of perspective-taking within RoastMe is further reflected in the high value placed on participating as both

a roaster and a roastie within the community. Participants characterized RoastMe as a “two way experience.” As p15 explained, “It’s not just people posting photos and getting roasted, it’s also giving the chance for someone else to anonymously critique or to make a comment about someone else. . .” This participant (p15) had not yet posted his own photo to be roasted, but still felt there was value in the two-way process. He explained, “I might do it [post] in the near future, because I think that it’s unfair for someone in the community to give out these comments that seem kind of hurtful—even if they’re funny on some level—I think it’s quid pro quo. If you made comments, have gotten upvotes, then you should probably subject yourself to the same thing.” Another participant (p3) explained that participating as a roastie helped him become a better roaster: “. . . being a roastie made it so I could see the other person’s point of view when I’m writing comments about them, so I know what they’re kind of going through.”

Although less prominently represented in our participants’ responses, perspective-taking may even take the form of encouraging others to perspective-take (what we’ll call “perspective-sharing”). This was true of one participant (p11) who often targets roasties who appear in some way privileged (e.g., those who are judged to be more attractive or generally happier or more well-off than average). As he explained, “And I’ll admit that I think that I’ve had more than my fair share of suffering, and for whatever reason, this comes into play in my roasts; that I want other people to feel that as well.” Thus, we can see that unlike other, more dissociative contexts in which individuals are provoking or ridiculing a target, perspective-taking is integral to roasting on RoastMe.

### Concern about Harm

We also found that many of our RoastMe participants were quite concerned about the well-being of other participants. From the perspective of benign violation theory, there were indications that the level of benignity of the humor violations was uncertain for many participants; they expressed uncertainty as to whether it was truly safe and okay to engage in certain forms of humor on the site. For example, four of our participants expressed guilt, secrecy, or regret about their participation in RoastMe. The lines between cruelty and humor were not always clear, and participants worried how their actions might affect others negatively. One participant (p15) discussed how he sometimes deletes comments after writing them, such as when he removed a self-authored comment that he felt reinforced an unfair, negative stereotype about gay women.

Participants saw people who posted photos as thick-skinned, that is, as willing and prepared to take on whatever the RoastMe community will throw at them. Indeed, the interviewees who had posted photos of themselves all described themselves as equipped to “handle the heat” in one way or another. Yet some questioned whether they really could trust this assumption. This was especially true if the participant self-identified as insecure or “thin-skinned,” as was the case for four of the participants who had chosen only to engage as a roaster, not a roastie. For example,



one participant conjectured that perhaps some people post photos of themselves because they have low, not high self-esteem (p7). Another participant stated that some comments on RoastMe can be exceedingly harsh, and that, “I kind of worry about destroying people’s self-esteem” (p6). Another worried that there could be site passersby who could be negatively affected by roasts “by proxy,” explaining, “It may actually just hurt the people there that are viewing the subreddit based on what they may think is unattractive and they may feel self-conscious about it” (p1). The participant, who stated his race as African in our demographic questionnaire, said he found himself feeling self-conscious about how others view him after reading the plethora of racist roasts of others on the site, wondering, “Is that really what people think the first time they see me?” Others worried that even though there are age cutoffs for the site, there might be young people participating that could get seriously injured. “Honestly, I don’t like it when kids come onto RoastMe because I don’t think they can handle it” (p5).

Participants also expressed the desire to remain within ethical lines. However, descriptions of where these lines actually fell varied greatly by participant. Many expressed that they took no issue with any offensive comments, but others felt certain kinds of comments, such as those that were directly racist, insulting of one’s religion or one’s sexual preferences, or treating rape or child abuse as a joke, crossed the line, and subsequently would either downvote or report the comment. For example, participants had mixed reactions to posts that referenced suicide or self-harm. As participant p7 explained, “There are a couple where the title is something depressing like, maybe they even reference being depressed, or self-harm or even suicide, and to be honest, those ones make me a little uncomfortable so I don’t post on those ones. Because you never know: someone might be joking, but a lot of times even when someone’s joking there might be some degree of truth to it.”

Although first impressions of RoastMe might lead one to assume that roasters are not particularly concerned about the feelings or well-being of their roastee targets, our interviews suggested otherwise. On RoastMe, desires to engage in humor that borders on the offensive and cruel coexist with desires to protect and support other community members.

## Benefits for Roastees

Heretofore, much of our focus has been on the values of roasters. Now, we zone in on how being a target of ridicule creates value for roastee participants. Participants cited “skin thickening” as a benefit to posting as a roastee; by subjecting themselves to criticism in the RoastMe space, they felt they would be able to handle “real life” insults and abuse. As p13 explains, “It just makes life so much better because even when somebody now insults you in the real world, you can laugh at it. Even though everyone’s trying to tear each other down, there’s still a feeling you came out as... almost a newer version of yourself. You have this newfound respect for yourself, and thicker skin...”

In fact, seven participants cited finding insights into themselves from anonymous strangers, and learning ways in which they might improve themselves, as key reasons for

posting. Some even changed their behavior after posting, such as one participant (p4) who modified how he dressed and wore his hair, and another participant (p6) who said she was glad someone had ridiculed her dark under-eye circles because she hadn’t previously noticed them, and had started wearing more under-eye makeup as a result. Another participant (p10) posted soon before going on a date. After receiving jokes about his hair, he decided to get a haircut, noting that, “I actually needed to get a haircut, so it was good for the date.” This same participant, who was very nervous prior to the date, found that the process helped give him perspective, stating that it helped him “not take things too seriously.”

Participants also enjoyed having people they didn’t know being able to accurately pinpoint aspects of their true selves; nine participants found this type of comment particularly humorous. Whereas, as discussed in related work, photographic self-presentation on certain stranger-dominated online contexts such as online dating platforms can encourage idealized self-presentation, in RoastMe, only one participant directly strove to hide what they deemed to be physical flaws (in this case, the participant stated he had purposely occluded his teeth). Seven participants chose to self-present authentically, many with the hopes of garnering honest feedback. In addition, we also saw a third type of photographic self-presentation that previous literature on online photographic self presentation has not yet discussed. Several participants actually *amplified* ways in which they might be considered flawed or aberrant in order to provide more roasting fodder, such as p12, who stated he tried to look dead inside in his photo in order to elicit better posts, or p3, who advertised in his posting title that he is Jewish, an artist, and bisexual in order to elicit funnier comments. In RoastMe, value is not just created for the roasters who want to engage in humor or unleash behaviors that might otherwise be considered taboo. Roastees, too, discussed how they benefited from RoastMe, gaining resilience and useful insights about themselves.

From our interview study, we learned that RoastMe participants abide by a set of norms that suggest a different lens through which to view and design for subversive humor in online communities. In RoastMe, perspective-taking rather than dissociation from targets guides roasting behaviors; concerns about harm coexist with desires to engage in offensive humor; and benefits accrue to those who are the butt of the joke. In the next section, we present the quantitative results that extend these qualitative findings.

## Quantitative Results

### Descriptive Findings

In the qualitative portion of this study we just discussed, saturation rather than representation was our aim, but we will note that our sample ended up being fairly representative of the RoastMe community at large. Of the 4,710 posts in our cleaned quantitative data set, the Microsoft Face API classified 79.9% as male by the Microsoft Face API, and estimated the average age as 26.2 (instances with unknown age and/or gender values omitted from the data set). As several of our interviewees spoke of taking a plain, authentic photo,

we might expect neutral expressions to be common; the data confirms this. Over half of the participants in the data set displayed an expression that was classified as neutral (using a threshold of 0.5/1 or above).

Our interviewees also discussed the value of reciprocity in the community (participating as both a roastee and a roaster). However, our data set suggests that although reciprocity may be a perceived norm and shared value in the community, it may not be quite as common in practice. Whereas a high proportion—87%—of the 3,964 unique roastees in our data set also contributed a roast at some point, a much lower 9.6% of roasters (5,183 of 53,885) had proffered photos of themselves for the community’s ridicule, indicating that many RoastMe participants are only experiencing one side of the RoastMe experience. Meanwhile, a relatively small but far from invisible proportion of roastees—5.7%, or 224 roastees—had submitted content to mental health oriented subreddits that might indicate cause for concern, validating some of the participants’ fears about whether the humor-intended violations enacted on RoastMe are truly benign.

### Topic Modeling of Post Titles

Iteratively tweaking our parameters to arrive at semantically meaningful results, we used Latent Dirichlet Allocation (LDA) to conduct topic modeling on the unfiltered data set of RoastMe posts (12,046). The most semantically meaningful model was built using 20 passes, 3 topics, and 5 words per topic. As we expected from informally browsing through the forum, RoastMe post titles are very homogeneous; the term “roast,” for example, emerged as the highest or second-highest weighted term in all three models. Nonetheless, our model allowed us to generate three broad categories of titles: 1) titles stressing the age of the roastee (e.g. “I just turned 19 years old, roast me!”), 2) titles asking roasters to do or give “their best” (e.g. “Go ahead, roast me. Give me your best roast”) and 3) titles indicating that the post is made on behalf of a friend, and that they want roasters to “do their worst” (e.g. “My friend thinks he can’t be roasted. Do your worst, RoastMe”). At least on the surface level, such titles give the appearance of confidence and “thick skin,” as our interview participants discussed.

### Linear Regression Model

Lastly, we iteratively trained a linear regression model in WEKA to predict the number of comments a given post can expect to receive using the filtered post data set of 4,710 posts (only including posts for which we were able to derive numeric and binary estimates of age and gender, respectively). Given the homogeneity we witnessed in the post title topic modeling, we did not incorporate features extracted from text mining to our data set. We applied the WEKA unsupervised attribute removal filter to the data; this machine learning algorithm weighs and eliminates less useful attributes from the model. After iteratively training and testing, we evaluated the model using 10-fold cross validation. Results are shown in Tables 1 and 2.

The model accounts for 86% of the variation in our data, and predicts that female participants will receive about 18

score	mental_health	age	female	sadness
0.24	9.41	0.49	17.95	-14.97

Table 1: Independent Variable Coefficients

corr coeff	MAE	RMSE	RAE	RRSE
0.86	22.49	41.94	66.26%	51.41%

Table 2: Other Model Metrics

more comments than their male counterparts, suggesting that the RoastMe community may have an especially high interest in roasting women. Activity in mental health related subreddits also positively predicts the number of comments (roasts), suggesting that those with potentially unstable mental conditions may attract more interest and roasts from RoastMe participants. While this might indicate some cause for concern about harm generation, we also note that sadness (derived from the Emotion API) has a high negative coefficient, implying that RoastMe participants may steer away from insulting those who appear more sad in their photos. Although “score” (downvotes subtracted from upvotes on the original post) is also present as a feature in this model, it is difficult to assign any real world value to the finding given the black-box nature of the measure, as described previously.

## Discussion & Conclusion

In accordance with previous literature, although the specific behaviors enacted on RoastMe might be considered non-normative, participants highly value and strictly abide by the community-specific norms that govern interactions in the subreddit. The shape these norms take on RoastMe highlight distinct, relatively unexplored needs and desires that people may seek from a niche online humor and self-presentation community. Where dissociation from targets (roastees) would be expected, RoastMe roasters instead relied heavily on perspective-taking. In photographic presentation, rather than self-present to elicit approval or affirmation, roastees reported learning and growing from the harsh judgment and ridicule they experienced on RoastMe. Unlike previously studied communities that encourage subversive humor or potentially offensive behaviors, participants in RoastMe are also quite sensitive to the emotional well-being of targets (roastees), desiring to mitigate harm.

Viewing RoastMe not as a bizarre, niche community, but as part of a larger confederation of communities displaying similar, non-normative desires, directs us to broader-reaching design implications. The specific desires expressed by participants in RoastMe, a fairly large and still-growing subreddit, implies that there may be other venues, settings, and audiences for which beneficial aspects of RoastMe’s design could apply.

First, we consider how our findings can inform Reddit’s community leaders and administrators as they develop rules and guidelines for Reddit as a whole, and for particular subreddits. As prior work has shown, bans on Reddit can be effective in curtailing hate speech on the platform at large



(Chandrasekharan et al. 2017). For example, when participants from banned subreddits migrated to relatively more innocuous subreddits like RoastMe, they did not sour the community by blasting it with hate speech or shifting the norms of the community towards pure hate or violence; rather, it appears that they conformed to existing norms. In light of the slew of Reddit bans and policy changes over the past few years (Hauser 2017), (Hudson 2015), it appears that Reddit leaders are shifting their stances and cracking down on subreddits with blatantly hateful or violent goals. They may lose support of some of their users as a result, calling some to question whether Reddit still holds true to its values of free and emboldened speech and off-color humor (Bokhari 2015). Amid these crackdowns, Reddit leaders would be wise to address not just what they will curtail, but what they will support. Although some people may still have qualms with the types of joke-insults bandied about on RoastMe—including current participants in the subreddit—“comedy, not hate” is an ostensible improvement on intentionally aggressive acts of hate speech. As our study’s findings reveal, there are ways in which we can further improve existing subversive humor communities like RoastMe to support freedom of speech and expression while simultaneously protecting and caring for its participants.

As we discussed, participants in RoastMe struggled with certain aspects of the community’s design. Our interview participants expressed uncertainty and anxiety about the level of harm generated in RoastMe. The commenting patterns revealed by our quantitative analyses also suggest that certain types of users—namely, women and those who may struggle with mental health issues—are disproportionately targeted. Professed values and realities in the community do not always align. For example, among our participants who placed high value on the reciprocal (“roast and be roasted”) nature of RoastMe, not all had actually posted as roastees themselves. Thus, the community space could be redesigned to allow for different levels of self-exposition and exposure to critique. Such design modifications could permit more fluid and comfortable levels of reciprocity. For example, the current design of Reddit does not encourage many one-to-one interactions; directly reciprocal exchanges (where a dyadic pair could serve as both roaster and roastee to one another) might have appeal to subsets of the community, and allow for currently exclusive roasters who are hesitant about submitting themselves to a roast to engage more fully in the community. We might also consider the role and value of anonymity in the forum. Here, we can envision a RoastMe-identical space that differs in one key aspect: all participants, including roasters, are photographically identifiable. To be clear, we make no claims that such a design would be “successful”. However, by exploring such a design as a digital probe, we could both deepen our understanding of the current role of anonymity and exposure in RoastMe, and inform alternative community designs. Another digital probe could explore the change in dynamics if the Reddit sorting algorithm for displaying new posts and popular posts were weighted and reconfigured to discourage excessive roasting of certain demographic subsets of users (in this case, women and those who had posted on mental health-related subred-

dits).

The brand of subversive, offensive humor celebrated on RoastMe is already integrated into the culture of Reddit. However, the types of interactions we see in RoastMe could function similarly in other domains that don’t currently embrace subversive sensibilities. We thus propose the design and implementation of digital probes to explore how aspects of RoastMe might translate to other domains. Below, we provide two examples to illustrate how such explorations could play out.

First, we’ll take an example that might be especially salient for academics. Under the mindset of reciprocal, purposefully unconstructive criticism, we can envision communities where researchers choose to upload abstracts, research ideas, or full papers, and then ask other researchers to tear apart their work, providing brutally honest or disruptive feedback, but all underneath the constraint of comedic intentions, or “lulz.” In exploring this digital probe in an academic domain, we might then employ experimental design to study the potential attitudinal and behavioral effects of engagement. For example, translated to this hypothetical academic “roasting” context, our RoastMe findings suggest that researchers whose work is roasted may value the feedback, which may allow them to take their own work less seriously and build resilience for future criticism. Similarly, those who roast may be able to unleash built up stress and pressure by saying the types of things they may often feel obligated to silence in their normal interactions with students and colleagues. How dynamics we observed on RoastMe may play out in different contexts is unclear, however. For example, would roasted researchers actually emerge from the experiences more resilient to academic critique? Would roasters find catharsis, or merely feel more comfortable in critiquing students, colleagues, and paper writers in more harsh, irrelevant, or offensive ways?

Another example of a topic-specific application could draw from our findings regarding photographic self-presentation in RoastMe, in which participants either self-presented authentically or even purposefully amplified their physical flaws, hoping to get harshly candid feedback about their appearance to build resilience and gain perspective on ways to improve. Online dating applications and other online communities where participants may be heavily judged by their photographs may cause attendant anxiety and pressures to self-present the idealized, rather than the true self. Under the roasting model, a community like RoastMe could essentially be re-branded to function as an “audition” process for the real site; dating sites or apps could even have opt-in roasting features to get unceremoniously blunt feedback on a photo or profile text before posting. Research questions with such manner of probes would revolve around whether participation actually protects participants, as our RoastMe interviews suggest, or only serves to increase anxiety.

As members of a subversive but norm-abiding community, RoastMe participants take the perspective of (rather than dissociate from) the targets of their humor, operate under non-traditional norms for photographic self-presentation and audience reception, and desire to unleash potentially in-

jurious behaviors while simultaneously aiming to mitigate harm. We find the interactions in RoastMe to be fascinating in their own right. However, the true value in this study lies in how the specific norms and interactions presented in RoastMe can inform and inspire new, experimental designs and digital probes in both similarly-oriented communities and disparate online domains that can meet unmet user needs and desires.

## Acknowledgements

A special thanks to the NSF Graduate Research Fellowship Program for supporting this work; to Sauvik Das, Saiph Savage, Joseph Seering, Joselyn McDonald, Michael Rivera, Nicholas Kasunic, and many other colleagues for their feedback and support; and of course, to our study participants, without whom this work would not have been possible.

## References

- Adams, J. 1976. *Here's to the Friars: The Heart of Show Business*. Crown Publishers.
- Back, M. D.; Stopfer, J. M.; Vazire, S.; Gaddis, S.; Schmukle, S. C.; Egloff, B.; and Gosling, S. D. 2010. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science* 21(3):372–374.
- Blauvelt, C. 2011. The Comedy Central roast of Donald Trump.
- Bokhari, A. 2015. #RedditRevolt rages as redditors condemn 'Chairman Pao's' crackdown.
- Celli, F.; Bruni, E.; and Lepri, B. 2014. Automatic personality and interaction style recognition from Facebook profile pictures. In *Proceedings of the 22nd ACM international conference on Multimedia*, 1101–1104. ACM.
- Chandrasekharan, E.; Pavalanathan, U.; Srinivasan, A.; Glynn, A.; Eisenstein, J.; and Gilbert, E. 2017. You can't stay here: The efficacy of Reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction* 1(CSCW):31.
- Ellison, N.; Heino, R.; and Gibbs, J. 2006. Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication* 11(2):415–441.
- Hauser, C. 2017. Reddit bans Nazi groups and others in crackdown on violent content.
- Holt, T. J.; Blevins, K. R.; and Burkert, N. 2010. Considering the pedophile subculture online. *Sexual Abuse: A Journal of Research and Treatment* 22(1):324.
- Hudson, S. 2015. Free speech vs. harassment: Why did Reddit ban five subreddits?
- Kraut, R. E.; Resnick, P.; and Kiesler, S. 2016. *Building successful online communities: evidence-based social design*. The MIT Press.
- Massanari, A. L. 2015. *Participatory Culture, Community, and Play*. Peter Lang.
- McGraw, A. P., and Warren, C. 2010. Benign violations: Making immoral behavior funny. *Psychological Science* 21(8):1141–1149.
- Pater, J. A.; Nadji, Y.; Mynatt, E. D.; and Bruckman, A. S. 2014. Just awful enough. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI 14*.
- Phillips, W. 2015. *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. MIT Press.
- Reicher, S. D.; Spears, R.; and Postmes, T. 1995. A social identity model of deindividuation phenomena. *European review of social psychology* 6(1):161–198.
- Suler, J. 2004. The online disinhibition effect. *CyberPsychology & Behavior* 7(3):321–326.
- Warren, C., and McGraw, A. P. 2014. Benign violation theory. In *Encyclopedia of Humor Studies*. SAGE Publications, Inc.