

‘That’s Not Funny!’ Standing Up Against Disparaging Humor

Emma F. Thomas, *Flinders University*

Craig McGarty, *Western Sydney University*

Russell Spears, *the University of Groningen*

Andrew G. Livingstone, *University of Exeter*

Michael Platow, *the Australian National University*

Girish Lala, *Western Sydney University*

Kenneth Mavor, *the University of St. Andrews*

ACCEPTED FOR PUBLICATION IN: *Journal of Experimental Social Psychology* special
issue on ‘moral courage’.

Please address correspondence to: Dr Emma F Thomas. College of Education,
Psychology & Social Work, Flinders University, Bedford Park, Adelaide, Australia. Email:
Emma.Thomas@flinders.edu.au. This research was supported by an Australian Research
Council Discovery Early Career Researcher Award (DE120101029) to the first author, and a
Murdoch University Distinguished Collaborator Award to the third author. The authors wish
to thank Geraldine O’Brien, Cassandra Barnes and Robert Grimsey for their assistance with
the execution of this research.

Abstract

The current article addresses bystander action to confront disparaging humor as a form of moral courage. We ask: When is disparaging humor seen as harmless fun or as a pernicious form of prejudice? What are the social and psychological processes through which bystanders confront, evade, or collaborate in disparaging humor? Three experiments ($Ns = 95, 213, 220$), involving a novel paradigm ('the shared media paradigm') test the role of bystander emotional responses (anger/amusement) in shaping action to confront disparagement humor, through emotion-based social influence. Study 1 demonstrates that bystander action to confront disparagement humor as prejudice is shaped by the angry (but not amused) responses of co-present others. Study 2 considers a moderator of the influence process: the role of one's own emotional reaction to disparagement humor (angry/amused). Bystander confrontation was more intense when one's own angry reaction was validated by that of other bystanders but there was otherwise mixed evidence that the two interacted to promote collaboration/confrontation. Study 3 tests the claim that disparagement humour is especially challenging to confront because humor disarms opposition. Intergroup commentary was seen as more amusing and confrontation was more strongly resisted when humor was used (vs. a non-humorous control remark). Overall, the results show that the reactions of bystanders play an important role in shaping what is (or is not) *perceived to be* prejudice. Courageous action to confront the disparagement of members of minority groups is enabled by the emotional signals of others who are co-present.

Keywords: emotion norms; social influence; disparagement humor; bystander action; confrontation; social appraisal; prejudice; collective action.

‘That’s Not Funny!’ Standing Up Against Disparaging Humor

“Everyday” prejudice is all too common. Although legal and social sanctions can limit overt or blatant prejudice, members of minority groups report that they continue to experience frequent disparaging remarks about their group, with detrimental outcomes for the individual (e.g., Mellor, 2003) and society (Elias & Paradies, 2016). One of the most common and insidious forms of prejudice is *disparaging humor*. Disparagement humor includes remarks which elicit amusement through the denigration or belittlement of a target (Ferguson & Ford, 2004, p. 283). Approximately one in four online comedy videos contain some form of anti-gay, sexist, or racist humor (Parrott, 2016). Such humor is typically seen as a more ‘acceptable’ form of intergroup commentary because the levity with which the remark is delivered suggests that it should not be taken seriously – after all, it is “just a joke” (Ford, Boxer, Armstrong, & Edel, 2008). However, disparaging humor, even where it trades on ostensibly benign stereotypes, is deeply problematic (see Ford, Breeden, O’Connor, & Banos, 2017; Ferguson & Ford, 2008 for reviews). Such humor reinforces intolerant attitudes and derogatory stereotypes (e.g., Ford, 1997) and shifts the normative context to one that supports discrimination (Ford et al., 2008; Ferguson & Ford, 2004; Saucier, O’Dea, & Strain, 2016), normalizing harm (Strain, Martens, & Saucier, 2016).

Part of the solution to the insidious effects of disparagement humor may be to encourage bystanders (i.e., people who are present when the remark is delivered but not targets of the remark) to confront it; that is, to take *bystander action* (Pedersen, Paradies, Hartley, & Dunn, 2011). However, this raises its own challenges. Research shows that the most common response to incidents of “everyday” prejudice is to ignore it (e.g., Hyers, 2010). Even those bystanders who are offended by disparaging humor may remain silent because the context itself remains highly ambiguous (Swim & Hyers, 1998). Simply put, the remarks are often not interpreted *to be* prejudice: even when they are highly counter-

normative, pejorative and unjustified, the social context of the humor makes confrontation appear inappropriate (“it was only a joke... where’s your sense of humor?”).

The current article addresses bystander action to confront disparaging humor as a form of moral courage. Moreover, we tackle one of the complexities of moral violations more generally: the idea that what “is” and “is not” a moral violation is often subjective, “in the eye of the beholder” and, as such, a key site of social influence (see Mikula & Wenzel, 2000; van den Bos, 2003). Accordingly, our research addresses two key questions: When is disparaging humor seen as harmless fun or as an unacceptable form of prejudice? What are the social and psychological processes through which bystanders confront, evade, or collaborate in disparaging humor? Our analysis focuses on the role of bystander anger and amusement in shaping responses to disparaging humor.

Bystander action to confront disparagement humor as a group process

Bystander action to confront everyday prejudice, including disparaging humor, serves multiple important social functions: it provides social and emotional support to the targets of prejudice, challenges the acceptability of such humor, and reduces the likelihood that it will be repeated (Nelson, Paradies, & Dunn, 2011). Bystander action may reduce distress in the targets of prejudice but also challenge the false consensus that prejudice is acceptable, thereby reducing the likelihood that prejudicial behavior will be repeated (see Pedersen et al., 2011). Indeed, Czopp, Monteith, and Mark (2006; Czopp & Monteith, 2003) demonstrated that although confrontation elicits negative affect and cognitions in perpetrators of bias (the joke teller), it also promotes feelings of guilt and self-reflection, indirectly reducing their use of stereotypes and attenuating prejudice. For Baumert, Halmburger, and Schmitt (2013), bystander action to confront prejudice, in spite of the potential for negative consequences for

oneself, is a quintessential form of moral courage (see also Osswald, Greitemeyer, Fischer, & Frey, 2010).

What are the factors that shape perceptions of disparagement humor? Existing research has primarily focused on the role of individual differences in attitudes towards the target group and/or prejudice (see Woodzicka & Ford, 2010, for a review in the context of sexist humor). For instance, Hodson, Rush, and MacInnis (2010) demonstrated that cavalier humor beliefs – beliefs that endorse the characterization of disparagement humor as light-hearted and not serious – are associated with favourable reactions to jokes that disparage outgroups, as well as generalized prejudice and prejudice-correlates such as social dominance orientation. Ford et al. (2008) showed that for sexist men (i.e., those high in hostile sexism), exposure to sexist humor promoted support for discriminatory resource allocations. Other research highlights the challenges to interpersonal relationships posed by “everyday prejudice”: confronting prejudice can be costly. Swim and Hyers’ (1998; Hyers, 2010) research on interpersonal confrontation, for example, suggests that members of victim groups will weigh up the personal costs and benefits of confronting a prejudiced perpetrator. Thus, responses to everyday prejudice, such as disparaging humor, have primarily been studied in terms of the characteristics of the perceiver that shape recognition (i.e., pre-existing prejudiced or socially dominant motives), and/or the interpersonal costs of confrontation.

We take a different tack to focus on confrontation and collaboration as processes of social influence (see Turner, 1991).¹ Our starting point is the observation that disparagement humor is highly ambiguous in nature and that it is the reactions of others that will therefore tell us what is normal, acceptable and “right” (Platow et al., 2005). Just as the impact of a live comedy act or stage play is influenced by the laughter, applause, groans and hisses from the audience, we propose that responses to everyday ‘performances’ of prejudice are influenced in similar ways. Indeed, Condor and colleagues (2006; Condor, Figgou, Abell, Gibson, &

Stevenson, 2006) argued that public expressions of prejudice (including, we suggest, disparagement humor) constitute ‘collaborative accomplishments’, a product of joint action amongst a number of individuals (see also Durrheim, Quayle, & Dixon, 2016). It follows that it is the reactions of other bystanders – those co-present when the disparagement is delivered – that will (partly) inform responses to the statement as prejudice promoting angry confrontation, or as harmless fun to be affirmed through amusement and enjoyment.

This confrontation need not occur only in face-to-face situations. Disparagement humor online is ubiquitous and, in this digital age, one of the primary ways in which stereotypes are encountered, disseminated, reinforced or contested (see Parrott, 2016). In an online setting, overt prejudice or hate speech can be reported to administrators. However, disparagement humor is covert and subtle, and so we would not expect formal reporting to be a common response. Rather, there are other opportunities to express dissent to content through such things as comments, (dis)likes and shares. Bystanders might be able to confront disparagement online without risk of physical danger (cf., Fischer et al., 2011) but there are nevertheless high potential costs in a setting where the effects of confrontation are no longer limited to one’s immediate interpersonal network (Crockett, 2017). Malicious trolling, online shaming or ‘pile ons’ – mass, anonymous harassment or denigration directed at those who are perceived to have transgressed – are also realities of the online environment (Cheung, 2014). People who object to disparaging humor can themselves become subjected to online victimization and, as such, *online confrontation* represents an important, everyday form of morally courageous action.

It is important to note that, in an online environment, the written word can connote approval or disapproval in a narrow sense, but can itself also be behavioral (reflecting confrontation/collaboration). Speech act theory (Austin, 1962; Searle, 1969) suggests that utterances should not be seen as somehow separate from behavior but are, in and of

themselves, ways of telling others “how things are”, influencing others, committing oneself, expressing values, and bringing about changes in the world (see also Fiedler, 2008). In this case, then, ‘disliking’ is indeed disapproval but is also public and carries risks in an online setting, where other friends or users in the social network can see that you have indicated your disapproval of the content. Similarly, ‘liking’ is merely approval in a narrow sense, but in an online environment can also function as de facto collaboration because of how ‘likes’ empower an online users’ position. Comments represent written and traceable evidence of commitment or antagonism and are, in many ways, more committal than more fleeting and/or deniable offline interactions. It is notable also that the targets of such confrontation include the creator of the content (the ostensible ‘perpetrator’) but also a potentially large past and future audience (some of whom have been, or could become, accomplices to the disparagement; Condor, 2006).

In short, online environments move confrontation from an interpersonal setting to a (potentially) global one (see Sawaoka & Monin, 2018). In this way, action to confront disparagement humor is both the domain of mundane “everyday” social interaction and at the same time requires significant moral courage (Gearhart & Zhang, 2010). On the other hand, inaction contributes to the *spiral of silence*, whereby opinions are not expressed due to a fear of social isolation, despite inaction indirectly maintaining opposition as a minority perspective (Noelle-Neumann, 1974).

There is also a methodological benefit to studying disparagement humor and confrontation online. Specifically, the online setting allows us to present disparagement remarks in a way that imitates exposure in everyday social interaction, overcoming a lack of experimental realism that characterizes many empirical tests of the effects of disparagement humor (see also Saucier, Strain, Miller, O’Dea & Till, 2018) and bystander action (which has tended to rely on self-report or scenarios; e.g., Pedersen et al., 2011).

Emotion and emotion communication in bystander responses to disparagement humor

To understand bystander action against disparaging humor, and the moral courage to confront it, our analysis draws on the joint insights of the social identity approach (Tajfel & Turner, 1979; Turner et al., 1987) and social appraisal theories of emotion (Manstead & Fischer, 2001). These theories allow us to test the role of bystander emotional reactions and group-based processes of social influence in shaping perceptions of, and responses to, disparagement humor. Disparagement humor that goes unchallenged may well promote a normative climate that supports prejudice (e.g., Ford et al., 2008; Hodson et al., 2010), but the reverse may also be true: the perceived emotional reactions of other bystanders, those co-present when the disparaging remarks are delivered, may also shape perceptions of disparagement humor itself, either as prejudice or playful levity. This seems especially likely given that disparagement humor is inherently ambiguous and, in situations of uncertainty, we are likely to look to the reactions to other people to inform our perceptions (Asch, 1955; McGarty, Turner, Oakes, & Haslam, 1993; Sherif, 1935). It is through perceiving and interpreting the reactions of others that aspects of social reality comes to be experienced as objectively “real and true” (Hardin & Higgins, 1996; Wittenbrink & Henly, 1996). These processes of social influence, social validation and co-construction are inherently transmitted through social communication and interaction (Echterhoff, Higgins, & Levine, 2009).

Theoretically, we conceive of this as a process of *emotion-based social influence* (see also Livingstone, Shepherd, Spears, & Manstead, 2016; Livingstone, Spears, Manstead, Bruder, & Shepherd, 2011; Thomas & McGarty, 2009). Social appraisal theory recognises that the expression of emotion is a powerful form of social communication (Manstead & Fischer, 2001; Parkinson, 1996; Parkinson, Manstead, & Fischer, 2005; van Kleef, de Dreu, & Manstead, 2010). Visible emotional reactions do not just inform us about another person’s internal state; they also provide information about their intended actions in relation to a given

event, and make claims to others about possible and proper actions in a given situation (see Niendenthal & Brauer, 2012). For example, Parkinson (1996) argues that expressing anger not only reflects one's own disapproval of an event, but can also communicate to others that this is an event of which they should also disapprove. In this way anger can be instrumental in enlisting collective support in resisting the situation of which one disapproves. Conversely, the expression of amusement communicates to others that this is a situation of which one should approve, reflecting intentions to affiliate with the source of amusement (Hess, Blairy & Kleck, 2000). We follow the earlier work of Livingstone et al. (2011; Livingstone et al., 2016) and Thomas and colleagues (Thomas & McGarty, 2009; Thomas et al., 2009a) to suggest that emotions (specifically: anger and amusement), as with cognitions and behavior, can act as sources of social influence to shape intergroup processes.

The social identity account of social influence has also highlighted the prominent role of the social frame of reference in shaping these influence processes (who is in one's ingroup vs. the outgroup; Abrams et al., 1990) and bystander intervention specifically (Levine, Cassidy, Brazier, & Reicher, 2002; Levine & Crowther, 2008). Platow et al. (2005), for example, asked participants to listen to an audio recording of jokes, informing participants that the audible ("canned") laughter was produced by ingroup members (fellow students) or outgroup members (members of a political party with whom participants did not identify). Data revealed more smiling and laughter, and more favourable humor ratings when participants heard ingroup laughter than outgroup laughter or no laughter at all. This suggests that people may well be more likely to find disparaging humor funny if ingroup members laugh and find it offensive where ingroup members do not. In the present research, we control for the effects of the referent group by telling participants that the other bystanders were all test users of the site (i.e., people fulfilling the same role as them), creating a nominal ingroup

affiliation as a pre-condition of social influence (see also Bourhis, Gadweld, Giles, & Tajfel, 1977).

Overview of Studies

The current research considers reactions to disparaging humor as an outcome of (emotion-based) social influence. In three experiments, we test the role of bystander emotion (anger/amusement) in shaping bystander actions to confront disparagement humor. Our experiments introduce the ‘shared media paradigm’ as a novel method through which to examine bystander action as a form of moral courage in online settings. Participants were told that they were participating in a study on online humor and that their task was to evaluate video clips. They were presented with a series of ‘humorous’ videos (Study 1) or images (memes; Studies 2-3) housed on a bogus site and were asked to interact with the webpages as they would typically interact with content on social media. One of the clips (the target clip) presented a piece of anti-gay disparagement humor: a ‘joke’ that could be dismissed as harmless fun, or perceived as perpetuating offensive stereotypes about gay people. Specifically, the content drew on stereotypes about gay people as superficial and/or fashion obsessed (although we note that any stereotypical portrayal of minority group members has been linked with increased negative perceptions of the group as a whole; e.g., Ford, 1997). We systematically varied the information conveying the emotional reactions of other bystanders (ostensible users of the site); and took behavioral measures of outcome variables (‘likes’, ‘star ratings’). The comment on the clip was content analysed for evidence of angry confrontation and/or amused collaboration. The paradigm therefore has a strong analogue with video sharing website YouTube (Study 1) and Facebook (Studies 2-3).

Study 1 tests the effects of emotion-based social influence on confrontation of disparagement humor, as one socially consequential form of moral courage. We also move the literature on disparagement humor beyond a singular focus on amusement, to consider the

role of other bystander emotions, in combination with one's personal anger (Study 2) in promoting confrontation (see Woodzicka & Ford, 2010). Finally, although it is recognized that disparagement humor may be uniquely able to denigrate its targets whilst stifling criticism, it is not yet clear whether this also flows on to undermine bystander confrontation (relative to non-humorous forms of intergroup commentary; Study 3).

Study 1

Participants were presented with a disparagement video clip and were asked to view all the content on the page before leaving their own ratings and comment. The bystander response was manipulated either to reflect anger, amusement or, in a control condition, there was no bystander response (and participants were led to believe that they were the first to review the content). We also took self-reported measures of enjoyment and the perceived appropriateness of user comments on the site. We expected that confrontation – evidenced in lower ratings, greater 'dislikes', and angry commentary – would be greatest where the bystander response was socially appraised as angry, relative to both the 'amused' and control conditions. Conversely, collaboration in the disparagement – evidenced in higher ratings, greater 'likes', and supportive commentary – would be greatest where the bystander response was socially appraised as amused, relative to both the 'anger' and control conditions. The control condition provided a baseline against which to compare the effects of angry or amused bystander responses, respectively.

Method

Design and Participants

Across all three studies, we report all measures and exclusions. Study 1 adopted a 3-cell between-groups design (no bystander response / angry / amused bystander response) and compared the effects on behavioral ratings of the clip (thumbs up/down, star ratings, report

button, comment) as well as ratings of the appropriateness of others' comments and self-reported enjoyment of the clip. Sample size was determined based on prevailing standards at the time that we collected the data of approximately $n = 30$ per cell. Participants ($N = 94$) were students or other members of the community recruited on the campus of an Australian university. They were primarily female (59.6%, 2 participants did not report their gender), with an average age of 25.19 ($SD = 9.01$). On average, participants reported accessing clips on YouTube and other social media several times per week. They were reimbursed with \$10 for their time or received partial course credit. A sensitivity analysis using G*power 3.1 (Faul, Erdfelder, Land, & Buchner, 2007) indicated that this sample size provides 80% power to detect an effect as small as $\eta_p^2 = .096$ ($f = .325$) in the present design ($df_{\text{num}} = 2$; groups = 3). We return to the matter of power in the discussion of study findings.

Procedure

Participants attended the laboratory and, having provided consent, were seated in front of an iPad. They were told that researchers were conducting market research on a new website which has content relating to online humor ('Random Vids'). Participants were told that their task was to view a series of randomly-selected humorous clips, read all the comments of other users (described as people who, like them, were also evaluating the site) and complete their own ratings and comment. They were not given any information about specific future interactions with other users but understood that their comments would be there for future participants to read. They were each assigned a gender neutral, personal user ID (e.g., Alex19) and logged into the Random Vids site (published on a local implementation of WordPress) using the iPads. They then viewed three clips, the second of which represented the target stimuli: it was a clip of a North American comedian talking about gay men in ways that perpetuated offensive stereotypes (it can be viewed here: <https://www.youtube.com/watch?v=5TQ04KsHNi8>).

We manipulated bystander responses to the clip in several ways (Figure 1 displays the user-interface for the ‘angry’ condition). In the angry bystander condition 16 participants had given the clip a thumbs down (6 thumbs up); rated the clip 3/10 stars; and a snapshot of the ‘angry’ comments appeared next to the clip. The user comments on the clip represented an angry consensus with 12 comments reflecting anger (e.g., ‘honestly, I found this pretty disgraceful’, ‘this makes me ANGRY’), and six indicating a neutral (e.g., ‘hmmmm’) or humorous response (e.g., ‘this made me lmao’). Conversely, in the amused condition 16 participants had rated the clip thumbs up (six thumbs down); rated the clip 7/10 stars; and a snapshot of ‘amused’ comments appeared next to the clip. The user comments reflected amusement with 12 comments reflecting amusement (e.g., ‘so funny’, ‘I love her facial expressions... It makes it so much funnier!’), and six indicating a neutral (e.g., ‘Ok I guess’) or angry response (e.g., ‘ANGRY :\$’). Many of these comments were adapted from user comments that appear on the clip on YouTube. Having viewed the clip and read the other user comments (there was no information from other users in the ‘no bystander response’ control condition), participants were invited to leave their own ratings (stars, thumbs up/down, report the content as offensive) and their own comment. This comment was content analysed. Participants then completed the same process with a third clip (Jedi Kittens; which also acted as a positive mood induction) and were directed to a secure online server (hosted by SurveySelect) to complete an online questionnaire.

Measures

In addition to key behavioral measures (star rating, thumbs up/down, report the content as offensive, comment), participants completed measures checking their appraisal of the other users’ reactions (manipulation checks), as well as measures of their own enjoyment of the clip/s and the perceived appropriateness of bystander comments. We also took measures of social dominance orientation, anti-gay prejudice, use of humor, and importance

of confronting; the results for these latter measures are not described here but can be found in a supplementary file. We also took a single-item measure of the degree to which the participant identified as gay, lesbian, bisexual, or transsexual; this variable is included as a covariate in the analyses reported below (its inclusion did not change the pattern of findings). Responses were recorded on a 1 (*Strongly disagree*) to 7 (*Strongly agree*) scale, unless otherwise indicated.

Manipulation check. The social appraisal of the other users' reactions of amusement and anger was checked with one item for each emotion: "It seemed like most people found the clip funny [outrageous]".

Behavioral measure. The comments left by the participants were content analysed by two independent coders who coded for the occurrence (1 = occurred, 0 = did not occur) and intensity (1 = mild, 2 = moderate, 3 = strong) of the expression of confrontation (overt expressions of discontent, anger) and collaboration (overt expressions of enjoyment, amusement), respectively. In these data, an example of the strong (coded 3) occurrence (coded 1) of confrontation is: 'OFFENSIVE !!!'; an example of moderate (coded 2) occurrence (coded 1) of confrontation is: 'She spent the whole time stereotyping gay men. She was annoying and it was poorly written'. An example of the strong (coded 3) occurrence (coded 1) of collaboration is: 'Soooooo funny...are you a uniform?'; an example of the moderate (coded 2) occurrence (coded 1) of collaboration is: 'Quite funny, more towards sexual jokes. But is pretty good'.

Enjoyment. Self-reported enjoyment was assessed with two items: "The video was easy to watch" and "The video was enjoyable". The two items were highly correlated, $r = .81$, $p < .001$, and were averaged to form a scale.

Perceived appropriateness of bystander commentary. Two items assessed the degree to which the bystander (user) responses were seen as appropriate: “I thought the other participants’ comments were appropriate” and “I thought the other participants’ comments were tasteless” (reverse-scored). The two items were correlated, $r = .53, p < .001$, and were averaged to form a scale.

Results and Discussion

Preliminary Analyses

A small amount of missing data (less than 5.3% for any variable) was Missing Completely at Random, $\chi^2(6) = 1.81, p = .94$ and was addressed using listwise deletion within each analysis. To first assess the effectiveness of the bystander manipulation, we compared the responses of the two conditions who viewed bystander responses (we did not include the control as they did not view other user comments). Results demonstrated that the manipulation was successful: participants agreed that the other users had found the clip funny to a greater extent in the amusement condition ($M = 4.55, SD = 1.55$) than the anger condition ($M = 2.03, SD = 1.00$), $F(1, 56) = 57.52, p < .001, \eta^2_p = .51$; whilst participants agreed that other users had found the clip outrageous to a greater extent in the anger condition ($M = 6.37, SD = .77$) than the amusement condition ($M = 5.00, SD = 1.16$), $F(1, 56) = 30.69, p < .001, \eta^2_p = .35$. The relatively high mean for endorsement of angry commentary in the amusement condition ($M = 5.00$) is a point that we return to below.

Main analyses

Table 1 displays the frequencies (for categorical variables) and/or means (for continuous variables) across the three conditions. Although the thumbs down ratings trended in the expected direction, these differences were not significant, $\chi^2(4) = 7.43, p = .12$. However, binomial tests showed that the proportions of people selecting thumbs down were

not different from 50% (chance) in the amused ($p = .35$) or control conditions ($p = .25$) but a significantly greater proportion of people selected thumbs down in the anger condition ($p = .004$). Moreover, there were differences in the star ratings of the clip, $F(2, 85) = 4.88, p = .010, \eta^2_p = .10$. Follow-up contrasts revealed that differences in ratings were driven by reductions in the anger condition relative to the amused, $t(86) = 2.69, p = .009, d = 0.75$, and control conditions, $t(86) = 2.71, p = .008, d = 0.73$. There were (marginal) differences in the occurrence of confrontational commentary, $\chi^2(2) = 5.91, p = .052$, reflecting greater frequency of confrontational comments in both the anger and amusement conditions (compared to the control); there was also greater intensity of confrontation in the angry and amused conditions, relative to the control, $F(2, 86) = 3.70, p = .04, \eta^2_p = .07$. Conversely, there were no differences in the occurrence, $\chi^2(2) = 1.31, p = .52$, or intensity, $F(2, 86) = 0.58, p = .56, \eta^2_p = .01$, of collaborative commentary. Only one person reported the content as offensive (that person was in the ‘angry’ condition) and, as such, this measure did not yield sufficient cases for analysis. There were no differences in self-reported enjoyment, $F(2, 85) = 0.98, p = .38, \eta^2_p = .02$; however, the commentary of the other users was perceived as more appropriate in the anger condition relative to the other conditions, $F(2, 85) = 6.84, p = .002, \eta^2_p = .14$.

This study provides initial support for our key prediction: bystander action to confront disparagement humor as prejudice is shaped by the perceived responses of others. When the same content was presented with no bystander response or with a normative (humorous) response, approximately half of participants rejected the clip; however, when the content was presented alongside angry bystander feedback nearly three quarters of participants rejected the clip (thumbs down) and overall ratings of the clip were lower.

The coding of the comments provides a congruent yet slightly different pattern whereby there was more frequent occurrence, and greater intensity, of confrontational

rejection in both the anger and amusement conditions. The manipulation checks showed that participants perceived bystander responses as relatively angry even in the amusement condition ($M = 5.00$, above the scale mid-point). It may be that the small number of angry comments (4 of 22) that were presented in this condition produced this effect (a form of minority influence, discussed further below; Moscovici, Lage, & Naffrechoux, 1969). That is, because the expression of anger runs counter to the broader normative project of a site about online humor (where amused or neutral commentary is the norm), these minority expressions of anger were still impactful in shaping the views of participants in so far as they suggested reduced consensus or even a bimodal distribution of reactions (see Asch, 1956, for a similar example of how a dissenting minority can reduce the influence of a majority on individual perceivers). By contrast, in the no feedback (control) condition there is ostensibly less reason to deviate from the assumption that amusement for such a site is the dominant norm. Viewed this way, the amusement condition could also be seen as a ‘weak anger’ condition suggesting that even a small minority of bystanders expressing anger at disparagement humor may help to promote its rejection.

It is also the case that our manipulation of bystander response may have been compromised by the audible audience reactions in the video clip; the laughter of the audience in the clip may have obscured the effects of the manipulation. Finally, the power analysis reported above suggests that this initial study may have been under powered to detect effects of medium magnitude or smaller, such as those observed on the comment variables. It also meant that significant effects in this study were also large, and could possibly represent an overestimation of the population effect. For both of these reasons, Study 2 tested our hypotheses with a larger sample size.

Study 2

Study 1 demonstrated the important role of (emotion-based) social influence in shaping perceptions of disparagement humor as either necessitating confrontation, or harmless levity. But what about the role of one's own emotional reactions to the remarks? Prior to perceiving feedback from the social environment about the emotional responses of other bystanders (social appraisal), perceivers are also likely to experience their own emotional reactions that might also inform their understanding of the situation (e.g., Livingstone et al., 2011, 2016; van den Bos, 2003; van Kleef, 2009). Appraisal theories of emotion highlight that appraisals, such as the meaning of an event to the self (primary appraisal), and coping ability (secondary appraisal), produce specific emotional reactions to events and situations (Lazarus, 1991). In turn, emotions such as anger or amusement elicit situationally-relevant action tendencies (Frijda, 1986; Parkinson, 1996). For example, anger arises from appraisals of an event as being inconsistent with one's own moral framework, and with one's ability to act effectively in response (Lazarus, 1991). This anger, in turn, generates action tendencies to confronting the source of the anger (e.g., the person who told an inappropriate joke).

Conversely, amusement is characterised, at least in part, by appraisals that the object is consistent with a moral framework, ultimately eliciting more positive emotional responses (e.g., smiling and laughter) and has a relationship-building function (e.g., Fraley & Aron, 2004). Prior research suggests that those high in pre-existing prejudice (hostile and ambivalent sexism, Ford et al., 2008; social dominance motive, Hodson et al., 2011) are more likely to find disparagement humor amusing. We are unaware of experimental work examining feelings of anger about disparagement humor (see Woodzicka & Ford, 2010, who note that this is an important but hitherto un-explored direction for future research). Nevertheless, because these two emotions (anger and amusement) indicate very different

initial reactions to disparagement humor, they should *qualify* the effects of others' emotional reactions to disparagement humor on confrontation. Put differently, the moral courage to act on one's own reaction may well be enabled by the emotional signals of others in bystander situations.

Study 2 therefore tests the role that one's own emotion (anger /amusement) plays in shaping responses to disparaging humor both directly, and also in combination with bystander emotional responses (anger /amusement). We test the idea that when confronted with instances of disparaging humor, people use the responses of other bystanders to inform their responses of the remarks, but this is qualified by one's own emotional reaction (Livingstone et al., 2011, 2016; van den Bos, 2003). To the extent that the reactions of other bystanders (other users of the site and therefore nominal ingroup members) affirm one's own reaction, then that social appraisal should *validate* and enhance one's emotional response (van Zomeren, Spears, Fischer, & Leach, 2004; see also Thomas, McGarty, & Mavor, 2009a). However, when one's own and others' emotional reactions to disparaging humor differ (i.e., own anger, other amusement; own amusement, other anger), then this is likely to lead to a *re-calibration* of one's own initial emotional response, so that we come to feel (respectively) less angry or amused due to the influence of others (Abrams et al., 1990; Manstead & Fischer, 2001; Turner, 1991). We therefore expected to observe statistical interactions between the individual's own emotional response (continuous measures of anger or amusement) and the responses of bystanders (manipulated anger or amusement) such that confrontation is greatest where personal anger and bystander anger match; and collaboration in the disparagement 'project' is greatest where personal amusement and bystander amusement match. Disparities between one's own and the socially-appraised emotion should diminish both confrontation and collaboration, resulting in evasion.

Study 2 also addresses two limitations of Study 1. The first is that we removed the audience “canned” laughter that was present in the video clip by using memes, or images (see Weng, Flammini, Vespignani, & Menczer, 2012). There is little work on the social psychology of memes but these represent an important part of the social media landscape and are increasingly used to contest unequal intergroup relations (Leach & Allen, 2017). The paradigm was therefore adapted to reflect a change to ‘motivational memes/statements’ in social media, analogous to interacting with content on Facebook or Twitter. To measure individual emotional reactions, participants were presented with a pop-up box asking them to rate their emotional reactions before they were exposed to bystander information. The second priority was to recruit a larger sample to ensure sufficient power to detect small-to-medium-sized effects.

Method

Design and Participants

Study 2 extended the three-cell between-groups design (no response / angry / amused bystander response) of Study 1 with two continuous moderators (individual amusement and anger) on behavioral ratings of the clip (thumbs up/down, report button, comment) as well as ratings of the appropriateness of others’ comments and self-reported enjoyment of the clip. Participants ($N = 220$) were North American residents recruited through online crowdsourcing site Prolific Academic (<https://prolific.ac/>). They were primarily male (54.1%; 3 participants did not report their gender), with an average age of 31.09 ($SD = 10.58$). 90% of participants had a Facebook account, which they reported accessing an average of 10-30 minutes per day. Sample size was primarily determined by resources available to conduct the study, and we sought to maximize the sample size given available resources. A sensitivity

analysis using G*power 3.1 indicated that this sample size provides 80% power to detect an effect as small as $\eta_p^2 = .04$ ($f = .21$) in the present design (largest $df_{\text{num}} = 2$).

Procedure

The procedure and set-up were similar to that described in Study 1, but participants were exposed to images (memes) instead of videos. The website and cover story were adapted to reflect an ostensible interest in understanding motivational content on social media and participants were told that they would be randomly presented with five motivational statements (memes) from a database of over 600 statements. Participants read that the statements would be similar to those that they would encounter on social media, where some would be “philosophical statements about life, others are observations about aspects of social life”. The site was called ‘motivational memes’ and was implemented in a custom built Wordpress site. Participants were presented with five memes, the fourth of which represented the target disparaging humor stimuli. Specifically, the target meme featured an image of a man speaking to an audience with the captions: “Everyone’s always saying how well gay men dress. I guess they didn’t spend all that time in the closet doing nothing”. The target stimuli therefore reflected the disparaging stereotypes that gay men hide their sexuality and are vain, excessively interested in outward appearances.

After participants had viewed each meme, a pop-up box would appear and participants were asked to quickly rate the content on how: amused, inspired, outraged, entertained, and angry it made them feel. ‘Inspired’ was a filler item (to maintain the cover story regarding motivational content); amused and entertained, $r = .77$, $p < .001$, and outraged and angry, $r = .79$, $p < .001$, were strongly correlated and these items were averaged to form measured responses of own amusement and anger, respectively.

Having completed the ratings, participants were then presented with the information about other users of the site (ostensible ‘bystanders’). As in Study 1, the content around the meme was systematically varied to reflect bystander reactions. In the angry bystander condition, participants saw that only eight people (of 42 votes) had liked (thumbs up) the meme, and the comments reflected a majority anger response (12 of 18 comments; three amused, three neutral). In the amused bystander condition, participants saw that 36 people had liked the meme (of 42 votes), and the comments reflected a majority amused response (12 of 18; three angry, three neutral). We retained the small proportion of people who presented an opposing view (i.e., three angry in the amused condition; three amused in the angry condition) because doing so provides a more conservative test of hypotheses, with greater realism (most online environments do not present a completely unanimous view). In the no response control condition, participants saw that no people had liked the meme because no one had viewed it yet and there were no comments. Participants were instructed to complete their rating (thumbs up/down) and leave a comment. Once they had completed these tasks for all five memes, participants were directed to a secure online server (hosted by Qualtrics) to complete an online questionnaire.

As in Study 1, in addition to key behavioral measures (thumbs up/down and comment), we took measures of the social appraisal (which serve as manipulation checks, $r = .73, p < .001$ for angry, $r = .93, p < .001$ for amused), self-reported enjoyment, $r = .87, p < .001$, and the appropriateness of bystander comments, $r = .50, p < .001$. The measurement approach was identical to Study 1 but was adapted to refer to the memes instead of videos. The comments left by the participants were again content analysed by two independent coders who coded for the occurrence (1 = occurred, 0 = did not occur) and intensity (1 = mild, 2 = moderate, 3 = strong) of the expression of confrontation (overt expressions of discontent, anger) and collaboration (overt expressions of enjoyment, amusement),

respectively. Inter-rater reliability was acceptable, $\kappa = .43-.92$, $p < .001$. In these data, an example of the strong (coded 3) occurrence (coded 1) of confrontation is: ‘This is actually offensive. Gay men can do as they please’; an example of moderate (coded 2) occurrence (coded 1) of confrontation is: ‘Not funny, demeaning’. An example of the strong (coded 3) occurrence (coded 1) of collaboration is: ‘Hilarious!’; an example of the moderate (coded 2) occurrence (coded 1) of collaboration is: ‘Hah, got a chuckle, so thumbs up’. An item assessing the degree to which the participant identified as lesbian, gay, bisexual, or transsexual (LGBT) was again included as a covariate in the analyses reported below. As in Study 1, we included several supplementary variables – the results for these are available in the supplementary file.

Results and Discussion

Preliminary Analyses

There were some missing data on the behavioral variables (between 3-9%) and self-report variables (~1%); this data was Missing Completely at Random, $\chi^2(64) = 67.90$, $p = .35$ and was addressed using listwise deletion within each analysis. We again assessed the effects of the bystander manipulation by comparing the responses of the two conditions who viewed bystander responses (we did not include the control as they did not view other user comments). The manipulation was successful: participants agreed that the other users had found the clip funny to a greater extent in the amusement condition ($M = 4.79$, $SD = 1.39$) than the anger condition ($M = 2.14$, $SD = 1.18$), $F(1, 140) = 140.56$, $p < .001$, $\eta^2_p = .50$. Participants also agreed that other users had found the clip outrageous to a greater extent in the anger condition ($M = 5.67$, $SD = 1.08$) than the amusement condition ($M = 3.39$, $SD = 1.40$), $F(1, 141) = 116.61$, $p < .001$, $\eta^2_p = .45$. We conclude that the manipulation was successful.

Main Analyses

Table 2 displays the descriptive statistics (proportions, mean and standard deviation) for the key variables by manipulated experimental condition. To provide an overview of effects that may be of interest to readers beyond the scope of our hypotheses, the sub-scripts denote where there were differences between conditions ($p < .05$ obtained via Analysis of Variance with pairwise comparisons). It can be seen that there were significant overall differences between conditions such that the angry bystander response reduced the occurrence and intensity of collaboration, as well as ‘thumbs down’ ratings. Bystander commentary was seen as less appropriate in the angry condition. Unlike Study 1 (see Table 1), confrontational comments occurred relatively infrequently overall (in ~13% of cases). Although the differences between conditions were not significant, it is notable that there were twice as many confrontational comments in the angry bystander condition than there were in either of the other two conditions.

We used Hayes (2019) PROCESS tool v3.3 (Model 2) with 5000 bias-corrected bootstrap samples to test the hypotheses about the effect of bystander responses (manipulated independent variable) and one’s own emotional reactions (two continuous, centred, moderator variables: own anger and amusement) on responses to disparaging humor (ratings, confrontational or collaborative commentary, self-reported enjoyment, and perceived appropriateness of bystander reactions). Given that the manipulated independent variable had three levels we used the multi-categorical tool to conduct two comparisons for each outcome variable. Effect code 1 tested the effect of an angry bystander response (coded 1) relative to no bystander response (control condition; coded -1). Effect code 2 tested the effect of an angry bystander response (coded 1) relative to an amused bystander response (coded -1).² The supplementary file contains the results comparing the effect of an amused bystander response (coded 1) relative to no response (coded -1). Moderators were centred and examined

at +/- one standard deviation from the mean; simple slopes for one moderator were examined at the average level of the other moderator. LGBT identification was included as a covariate; its inclusion did not alter any of the findings. We conducted regression for continuous variables (intensity of comment; enjoyment and appropriateness) and logistic regression for the categorical variables (thumbs down; occurrence of comment); differences in the reporting below reflect the two forms of analysis. Table 3 displays the unstandardized regression coefficients (standard error in brackets, 95% confidence interval in square brackets) and significance level for the direct effects and interaction terms. No one 'reported' the content as offensive.

Thumbs up/down. Table 3 shows that participants were more likely to rate the disparagement clip with a thumbs down when they were low in amusement, and also when assigned to the angry bystander condition (relative to the amused condition). There was no direct effect of angry bystanders relative to the control (effect code 1) but there was a significant interaction between effect code 1 and own anger ($p = .04$) such that the bystander manipulation did not affect ratings when anger was low, $b = -0.53$, $p = .32$, 95% CI [-0.51, 1.57], or average, $b = 0.19$, $p = .61$, 95% CI [-0.91, 0.54]. Unexpectedly, however, exposure to bystander anger was associated with greater 'thumbs up' ratings when personal anger was high, $b = 1.07$, $p = .046$, 95% CI [0.02, 2.13]. Figure 2 displays the interaction.

Confrontational commentary. Table 3 shows that participants were more likely to make a confrontational comment, and with greater intensity, when their initial response was angry (effect of own anger) and less likely to do so when they found it amusing (effect of own amusement). The combination of one's own reaction and those of bystanders influenced the intensity of confrontation (i.e., an interaction between effect code 2 and own anger; see Table 3, Figure 3a). When participants were angry (1 standard deviation above the mean) and were exposed to the anger of others (relative to the amusement of others; effect code 2), the

intensity of the confrontation was greater, $b = 0.28$, $p = .001$, 95%CI [0.11, 0.44]. When participants' own anger was low, $b = -0.09$, $p = .22$, 95%CI [-0.23, 0.05], or average, $b = -0.07$, $p = .18$, 95%CI [-0.03, 0.18], bystander responses did not affect the intensity of the confrontational comment (Figure 3a).³ A second interaction between effect code 1 and own amusement suggested that, when participants were unamused (low) and there was no bystander response (relative to an angry response; effect code 1), the intensity of the confrontation was greater, $b = -0.22$, $p = .007$, 95%CI [-0.38, -0.06]. Conversely, when participants were highly amused and exposed to an angry bystander response the intensity of confrontation increased, $b = 0.18$, $p = .03$, 95%CI [0.02, 0.33]. See Figure 3b. The manipulation did not affect intensity of confrontation when participants' own amusement was at the average, $b = -0.02$, $p = .69$, 95%CI [-0.13, 0.08].

The occurrence of confrontation was also explained by an interaction between effect code 1 (anger v control conditions) and own amusement ($p = .03$). None of the simple slopes were significant, however: the manipulation did not appear to affect confrontation at low, $b = -0.49$, $p = .29$, 95%CI [-1.39, 0.42], average levels of amusement, $b = 0.36$, $p = .46$, 95%CI [-0.60, 1.32], or at high levels of amusement, $b = 1.21$, $p = .11$, 95% CI [-0.28, 2.70].

Collaborative commentary. Participants collaborated in the disparagement when they reported the meme as amusing (effect of own amusement) and were less likely to do so when they reported that the meme was enraging (effect of own anger; Table 3). There was no direct effect of the bystander manipulation; however the effects of the manipulation on collaboration were qualified by one's own anger (i.e., an interaction between effect code 2 and anger; $p = .048$). For those who reported high levels of anger, an angry bystander response (relative to an amused response) was associated with significant reductions in collaborative commentary, $b = -1.13$, $p = .008$, 95%CI [-1.97, -0.29]. When anger was at average, $b = -0.49$, $p = .08$, 95%CI [-1.03, 0.06], or low levels, $b = 0.02$, $p = .95$, 95% CI [-

0.72, 0.77] bystander responses did not affect the likelihood of collaboration. Figure 4 displays the interaction.

Self-reported enjoyment. Explicit enjoyment followed a related pattern to collaborative commentary: enjoyment was positively predicted by own amusement and negatively predicted by own anger (Table 3). There was also evidence that self-reported enjoyment was shaped by the combination of one's initial level of amusement and the bystander response (i.e., an interaction between effect 1 and own amusement; $p = .02$). Angry bystander commentary (relative to the control condition, effect 1) was (marginally) associated with increases in self-reported enjoyment only when levels of amusement were high, $b = 0.36$, $p = .08$, 95% CI [-0.05, 0.77], but not when levels of amusement were low, $b = -0.30$, $p = .16$, 95% CI [-0.71, 0.12] or average, $b = 0.03$, $p = .82$, 95% CI [-0.24, 0.31]. A figure depicting the interaction is available in the supplementary file.

Appropriateness of bystander reactions. Finally, an amused bystander response was seen as more appropriate (relative to an angry response, effect code 2; Table 3) but this effect was qualified by an interaction between the manipulation (effect code 2) and measured amusement (i.e., an interaction between effect 2 and own amusement; $p = .03$). There were no differences in the perceived appropriateness of bystander responses when amusement was low, $b = -0.01$, $p = .93$, 95% CI [-0.33, 0.30], but angry responses were seen as less appropriate (relative to amused responses, effect code 2) when the participants found the clip amusing on average, $b = -0.27$, $p = .02$, 95% CI [-0.49, -0.05], and were highly amused, $b = -0.52$, $p = .001$, 95% CI [-0.84, -0.21]. A figure depicting the interaction is available in the supplementary file.

Summary. Overall, the results suggest that bystander reactions shaped how people responded to disparaging humor (see Table 2), as did self-reported own anger and amusement

(Table 3, main effects of anger and amusement). However, the findings provide mixed support for our hypotheses about the interactive effects of own and bystander reactions. We expected that confrontation would be greatest where there was a match between one's own anger and the anger of bystanders (i.e., two-way interactions between own anger and effect codes 1 and 2). Table 3 shows that this was supported in the context of the intensity of confrontation, such that when participants were angry and were exposed to the anger of others (but only relative to the amusement of others, effect code 2), the intensity of confrontation was greater (Figure 3a). It was also the case that for those who reported high levels of anger, an angry bystander response (relative to an amused response) was associated with significant reductions in collaboration (Figure 4). However, this pattern was not evident in the other analyses, suggesting that the interactive effects of one's own and bystander anger affect the intensity of the response (but not its occurrence per se); and are only expressed in comparison to the active collaboration of other bystanders.

In the context of collaboration, the matching hypothesis was not supported and none of the interactions between one's own amusement and the amusement of bystanders was significant. Although unexpected, this latter finding is consistent with other analyses of emotion-based social influence: Livingstone et al., (2011, Study 1) observed increased willingness to challenge injustice only when there was shared anger (which has clear, action-associated tendencies in response to illegitimate behaviour; Frijda, 1986), but not shared happiness (which has less clear implications for mobilizing group action). Because the broader project (and normative climate) of shared media sites is to entertain and amuse, it may be that bystander amusement is not required to license collaboration, especially for those who are already highly amused. More broadly, the findings suggest that in addition to whether our own and others' emotional reactions to disparaging humor 'match' per se, the

specific content of those emotional reactions matters too when it comes to understanding resistance to such humor.

There were several other unexpected, counterintuitive results that suggest that the social psychological processes at play are highly complex. One of the observed effects appears to run counter to the effects described above: specifically, participants were more likely to rate the clip with a ‘thumbs up’ if they were angry and so were other bystanders (Figure 2). One speculative explanation for this pattern is that people in this condition were not endorsing the clip, so much as the angry response of other bystanders (in the same way that Facebook and YouTube allow the user to like specific comments on contentious content). Moreover, the interaction in Figure 3b suggests that an angry bystander response increased confrontation amongst those who were highly amused (consistent with theorizing above); at the same time as it undermined confrontation for those low in amusement. This latter effect may also indicate a form of a diffusion of responsibility whereby those low in amusement (but not high in anger per se) were content to let angry others do the work of confrontation (Ling et al., 2005).

Finally, and perhaps attesting to a persistent online conflict between those who enjoy or condone such humor, and those who ostensibly oppose it, there was also evidence that people who were highly amused and exposed to bystander anger actively increased their self-reported enjoyment (Table 3; Figure S1). Thus, for those who are amused at the outset, a disparaging clip is even funnier if it makes other people angry – akin to the aggravating pleasures of trolling (see Coles & West, 2016).

Although several of the interactions were non-significant, it is also worth noting that the high mundane realism of the procedure and the high-threshold nature of the behavioral dependent variables are relatively rare in a controlled, experimental study such as this. It is

therefore perhaps not surprising that the effect sizes are not large or consistent across all dependent variables. Future research will need to disaggregate multiple, competing social psychological processes around (emotion-based) social influence (Livingstone et al., 2011), leadership (especially in the absence of a response), bystander diffusion of responsibility (especially in the presence of a strong response; Ling et al., 2005), and trolling (i.e., pleasure experienced by virtue of aggravating other people).

One further limitation here is that we tested the effects of own anger and amusement as measured variables rather than manipulated variables. It is notoriously difficult to manipulate emotions (van den Bos, 2003) and especially fraught in cases like this where we sought to maintain realism in the paradigm. The findings here should therefore be considered in light of the methodological limitations that correlational analyses entail. In the interests of transparency, we report in the supplementary materials an additional study that reports an unsuccessful attempt to manipulate individual emotion using the bogus pipeline procedure (Jones & Sigall, 1971).

Study 3

The final study tested the effects of emotion-based social influence on the confrontation of prejudice in the context of disparaging humor versus an equivalent, non-humorous disparaging remark. In addition to the importance of bystander emotion-based social influence (Study 1 and 2), and one's own emotional reactions (Study 2) in shaping confrontation of disparagement humor, we have claimed that there is something particularly insidious about disparaging humor, beyond other instances of more overt "everyday prejudice" (see also Ferguson & Ford, 2004; Saucier, O'Dea & Strain, 2016). Humor implies levity – that the remark is not to be taken seriously – but it is also an affiliative emotion, one that functions to build relationships (see e.g., Thomae & Pina, 2015, in the context of sexist

humor). It follows that confrontation of disparagement humor may require greater moral courage than confrontation of more explicit forms of prejudice. If this is the case, the bystander response may be relatively more important in allowing disparagement humor to be interpreted *as prejudice* necessitating confrontation, relative to a non-humorous statement.

On the other hand, given the relational challenges of confronting prejudice generally (e.g., Czopp et al., 2006; Hyers, 2010; c.f., Mallett & Wagner, 2011), it is possible that appraising social support for confrontation is equally important in inspiring action for both forms of “everyday prejudice” (van Zomeren et al., 2004). To test the relative importance of emotion-based social influence in the context of disparagement humor and non-humor, we decomposed the stereotypes of a “joke” into statement form and assessed how bystander reactions (of anger/amusement) interact to promote confrontation, collaboration, or evasion. We test whether the emotion-based social influence process is more important in eliciting the courage to confront disparagement humor (relative to an equivalent disparaging remark); or whether the bystander response is equally important in shaping action in both cases.

Method

Design and participants

Study 3 had a 3 (bystander response: no response / angry / amused) X 2 (prejudice: humor / non-humor) between-groups design, and dependent variables included behavioral ratings of the clip (thumbs up/down, comment) as well as ratings of the appropriateness of others’ comments and self-reported enjoyment of the clip. Participants ($N = 213$) were North American residents recruited through online crowdsourcing site Prolific Academic (<https://prolific.ac/>). They were primarily female (60.1%, 9 participants did not report their gender), with an average age of 35.93 ($SD = 12.63$). Most (81.7%) participants had a Facebook account, which they reported accessing an average of 10-30 minutes per day.

Sample size was primarily determined by resources available to conduct the study, and sample size was maximized given available resources. A sensitivity analysis using G*power 3.1 indicated that this sample size provides 80% power to detect an effect as small as $\eta_p^2 = .044$ ($f = .214$) in the present design (largest $df_{\text{num}} = 2$); a smaller effect size than most of the relevant effects observed in Studies 1 and 2.

Procedure

The procedure and set-up were similar to that described in Study 2. Participants were presented with five memes, the fourth of which represented the target stimuli. As in Study 2, in the humor condition, the target meme featured an image of a man speaking to an audience with the captions: “Everyone’s always saying how well gay men dress. I guess they didn’t spend all that time in the closet doing nothing”. In the statement condition, the target meme featured the same image but the caption read: “Gay men hide their sexuality for many years. Why are they so obsessed with fashion?” The statement thus decomposed the two key stereotypical “claims” of the humor clip into a non-humorous form. Unlike Study 2, where participants were presented with a pop-up rating box, participants were able to view the meme as well as the information about other users of the site simultaneously (as in Study 1). The content around the meme was systematically varied to reflect bystander reactions, identical to Study 2 and participants were instructed to complete their rating (thumbs up/down) and leave a comment. Once they had completed these tasks for all five memes, participants were directed to a secure online server (hosted by Qualtrics) to complete an online questionnaire.

As in Studies 1 and 2, in addition to key behavioral measures (thumbs up/down and comment), we took measures of the social appraisal (which serve as manipulation checks, $r = .73, p < .001$ for angry, $r = .93, p < .001$ for amused), self-reported enjoyment, $r = .87, p <$

.001, and the appropriateness of bystander comments, $r = .50, p < .001$. As in Studies 1-2, the comments left by the participants were content analysed by two independent coders who coded for the occurrence (1 = occurred, 0 = did not occur) and intensity (1 = mild, 2 = moderate, 3 = strong) of the expression of confrontation (overt expressions of discontent, anger) and collaboration (overt expressions of enjoyment, amusement). Inter-rater reliability was acceptable, $\kappa = .57-.83, p < .001$. In these data, an example of the strong (coded 3) occurrence (coded 1) of confrontation is: 'I don't get offended easily but this was pretty bad'; an example of moderate (coded 2) occurrence (coded 1) of confrontation is: 'Casual homophobia doesn't really fly anymore. Not funny'. An example of the strong (coded 3) occurrence (coded 1) of collaboration is: 'Don't be oversensitive. It's a good joke'; an example of the moderate (coded 2) occurrence (coded 1) of collaboration is: 'This is a solid pun'. An item assessing the degree to which the participant identified as LGBT was again included as a covariate in the analyses. As in Studies 1-2, we included several supplementary variables; the results for these are available in the supplementary file.

Results and Discussion

Preliminary Analyses

Perhaps unsurprisingly given the online method of data collection, there were some missing data (no more than 5.6% for any variable); this data was Missing Completely at Random, $\chi^2(77) = 61.94, p = .89$ and was addressed using listwise deletion within each analysis. We again assessed the effects of the bystander manipulation by comparing the responses of the conditions who viewed bystander responses (we did not include the control as they did not view other user comments) using a 2 x 2 factorial ANOVA. The manipulation was successful: participants agreed that the other users had found the clip funny to a greater extent in the amusement condition ($M = 4.78, SD = 1.50$) than the anger condition ($M = 2.44,$

$SD = 1.61$), $F(1, 127) = 72.36$, $p < .001$, $\eta^2_p = .36$; the effect of the clip (humor/non-humor) was not significant, $p = .28$, and nor was the interaction between the two variables, $p = .79$. Participants also agreed that other users had found the clip outrageous to a greater extent in the anger condition ($M = 5.32$, $SD = 1.52$) than the amusement condition ($M = 3.56$, $SD = 1.43$), $F(1, 125) = 43.18$, $p < .001$, $\eta^2_p = .26$; the effect of the clip (humor/non-humor) was not significant, $p = .93$, and nor was the interaction term, $p = .78$. As in Studies 1-2, the manipulation was successful.

Main Analyses

Table 4 displays the means (standard deviations) or proportions for the six experimental conditions. It can be seen that the clip was ‘liked’ more when it was presented in humor format and there was greater occurrence and intensity of collaborative commentary in the humor format. Consistent with the pattern of effects identified in Study 2, there was a low incidence of confrontational commentary (~9% of overall comments) but confrontational comments occurred more often, and with greater intensity, in the angry bystander conditions.

Given that our outcome variables are composed of a combination of continuous and categorical variables, we used Hayes (2012) PROCESS tool (Model 1) with 5000 bias corrected bootstrap samples to test the hypotheses about the effect of bystander responses (manipulated independent variable) and the format of the clip (manipulated moderator variable: effect coded 1 = humor, -1 = non-humor) on behavioral indicators of collaboration and confrontation, self-reported enjoyment, and the perceived appropriateness of other bystander comments. We conducted regression for continuous variables (intensity of comment; enjoyment and appropriateness) and logistic regression for the categorical variables (thumbs down; occurrence of comment). Moreover, in order to disaggregate the effects within the bystander condition (i.e., beyond the main effects displayed in Table 4) and

test the unique effects of each set of comparisons we again used the multi-categorical tool to conduct two comparisons for each outcome variable. Effect code 1 tested the effect of an angry bystander response (coded 1) relative to no bystander response (control condition; coded -1). Effect code 2 tested the effect of an angry bystander response (coded 1) relative to an amused bystander response (coded -1). The supplementary file contains the results of a comparison of the effect of an amused bystander response (coded 1) relative to no response (coded -1). LGBT identification was included as a covariate; its inclusion did not alter any of the findings. Table 5 displays the unstandardized regression coefficients (standard error in brackets, 95% confidence interval in square brackets) for the direct effects and interaction terms.

Thumbs up/down. Table 5 shows that the two manipulations affected whether the clip was ‘liked’ or not (thumbs down). Specifically, participants were more likely to reject the clip when an angry response was presented (relative to an amused response; effect code 2) but there were no differences between the anger and control conditions. The presentation of the clip in non-humorous form was associated with significantly lower endorsement (thumbs down), relative to presentation of the clip in a humorous format (an effect of presentation of the clip).

Confrontational commentary. An angry bystander response (relative to an amused one; effect code 2) had a marginal unique effect ($p = .06$) on the intensity of confrontational commentary. There was also some evidence that the presentation of the remark in humor form was associated with greater occurrence of confrontation, though this effect was marginal ($p = .08$).

Collaborative commentary. Consistent with Studies 1 and 2, participants were more likely to collaborate in the amusement, and with greater intensity, when an amused bystander

response was present (relative to an angry response; effect code 2), but also when the remark was presented in humor form (effect of presentation; Table 5). The interaction between effect 2 and the presentation of the clip (marginally) explained the intensity of the collaborative comment ($p = .08$). Figure 5 displays the interaction. Simple slopes confirmed that when the clip was in non-humor form the bystander response did not affect intensity of amused collaboration, $b = -0.09$, $p = .48$, 95%CI [-0.35, 0.16] but when the clip was presented in humorous form, the angry bystander response (relative to an amused bystander response, effect code 2) was associated with weaker collaboration, $b = -0.40$, $p = .002$, 95% CI [-0.65, -0.16].

Self-reported enjoyment. Participants reported that they enjoyed the clip more when it was in humor form than when it was a non-humorous statement. Participants in the angry bystander condition reported greater enjoyment than the control condition (effect 1), but less enjoyment than the amused condition (effect 2; Table 5).

Appropriateness of bystander reactions. Table 5 shows that there was also a significant interaction between effect code 2 and presentation of the clip on the perceived appropriateness of the bystander comments ($p = .04$). However, the simple slopes were not significant when a non-humorous clip was presented, $b = 0.22$, $p = .16$, 95% CI [-0.09, 0.54] or when a humorous clip was presented, $b = -0.25$, $p = .11$, 95% CI [-0.55, 0.06].

Summary. Study 3 provides support for the claims about the uniquely insidious role of disparagement humor, with some caveats. Participants found the clip more enjoyable, and rated it ‘thumbs up’ more frequently, when it was presented in joke form: since prejudice is no laughing matter, this suggests that disparagement humor is less likely to be recognized as a form of marginalizing intergroup commentary, relative to a non-humorous equivalent. There was also evidence that the perceived appropriateness of bystander commentary is

conditioned upon whether they are responding to a joke or a statement, but the direction of this effect was difficult to interpret because the simple slopes were non-significant. Contrary to our expectation that social support may be especially important in the ambiguous context of disparagement humor, an angry bystander response was equally impactful in promoting confrontational commentary for both forms of disparaging remark. However, angry bystander responding did mitigate the intensity of the collaborative commentary when disparaging humor was presented (but not in the context of non-humorous remarks; Figure 5). This finding was again identified only in the context of the comparison between the anger and amusement conditions (and not in the context of anger and a no response control), suggesting that, in these data, it is the active presence of amused commentary (relative to anger) that licenses collaboration (Platow et al., 2005).

General Discussion

The current research addresses the social and psychological antecedents of bystander action to confront disparagement humor as a quintessential form of “everyday” moral courage (Baumert et al., 2013). Consistent with observations about bystander intervention and helping more generally (Latane & Darley, 1968; see also Fischer et al., 2011), people must first perceive disparagement remarks *as* prejudice in order to intervene, but this is not straightforward. Much of the existing literature pre-supposes that prejudice exists independent of social context: it either “is” or “is not” prejudice (see Mikula & Wenzel, 2000; Platow et al., 2019; van den Bos, 2003, for similar critiques). However, disparaging humor is about managing, preserving, and challenging relationships between groups (Hodson et al., 2010; also Billig, 2005). It follows that, in addition to individual differences (Woodzicka & Ford, 2010) and interpersonal factors (Hyers, 2010), we need to understand the social processes through which people *come to respond to* a given interaction or remark as disparagement necessitating redress, rather than playful levity (Condor, 2006; Condor et

al., 2006; see also Durrheim et al., 2016). Building on the insights of the social identity approach and social appraisal theories of emotion, we tested the hypothesis that people will develop the moral courage to confront disparagement humor when the angry responses of other bystanders indicate a majority perception of the remarks as prejudicial in nature.

Three studies broadly support this hypothesis, tested in the context of online confrontation and involving a novel ‘shared media’ experimental paradigm which recorded behavioral measures of confrontation versus collaboration. Study 1 showed that the same clip can be viewed in very different ways depending on the reactions of other bystanders. When those co-present were perceived to be enraged by the clip, participants rated the clip lower and gave it more ‘thumbs down’, relative to a condition in which there was no bystander information provided, or the bystanders were amused. There was greater prevalence and intensity of confrontational commentary when the bystanders were angry and amused (relative to no-bystander information), the latter effect likely due to the presence of a small but distinct minority who reported finding the content enraging (discussed further below).

Studies 2 and 3 provide evidence of the boundary conditions of the effects of social appraisal and social influence on bystander confrontation, but with some twists. Study 2 showed that one’s own emotional responses (anger, amusement; van den Bos, 2003) as well as those of bystanders (anger, amusement, no response), independently shape responses to disparaging humor. Study 2 also provided mixed support for the idea that confrontation and collaboration are driven by an interaction (fit or match) between the one’s own response and those of others. On the one hand, participants who were angry and had their anger validated by bystanders recorded a more intense confrontation, as well as reduced occurrence of collaboration. On the other hand, the validating effect of a ‘fit’ between one’s own anger and others’ anger appeared to be limited to intensity of confrontation and occurrence of collaboration, in the context of a comparison between angry and amused bystanders.

Moreover, we did not observe a validation effect on collaboration associated with a fit between one's own amusement and those of bystanders. It may be that in this context, shared amusement does not necessarily invoke the need for active collaboration in the same way that shared anger may implicate the need for confrontation (consistent also with Livingstone et al., 2011). Study 2 also suggested a role for other social psychological processes in understanding complex reactions to disparaging humor (detailed in the summary of that study).

Finally, Study 3 tested whether the effects of emotion-based social influence were greater in the context of disparagement humor, relative to an equivalent non-humorous remark. The findings showed that disparagement humor was more enjoyable, 'liked' and collaborated in via comment, relative to a non-humor equivalent, confirming suggestions that disparagement humor can uniquely denigrate a target whilst stifling criticism (Ferguson & Ford, 2004). Although an angry bystander response appeared to mitigate some of the intensity of the collusion with disparagement humor, angry bystander responses promoted confrontation across both forms of disparagement remark, attesting to the important role of social support in promoting action to challenge injustice (van Zomeren et al., 2004; Thomas, et al., 2009a).

In this digital age, online interaction is one of the primary means through which stereotypes about groups are disseminated, perpetuated, and/or contested. Shifman and Thelwall (2009), for instance, demonstrated how one meme (a joke about men, women, and computers) was extensively circulated, evolving to incorporate small, local adaptations. Inaction may contribute to a spiral of silence (Noelle-Neumann, 1974), allowing those who denigrate or subjugate members of minority groups to go unchallenged (see Geaerhart & Zhang, 2014, for a discussion in the context of anti-gay prejudice). Worse still, active collaboration (by 'sharing', 'liking', and commentary that endorses the humor) may serve to

validate and reinforce stereotypes – fostering group cohesion whilst indirectly fostering hostility towards minority group members (e.g., Ford, 1997; Thomae & Pina, 2015).

Confronting disparaging remarks (even, if not especially, when humorous) can be personally costly and requires significant moral courage (Hyers, 2010). However, hitherto, studies of disparagement humor have not considered the role of other bystanders, co-present, in shaping responses to the humor project. Nor has that literature considered online environments, or the role of individual emotions beyond amusement in shaping responses to disparagement humor (Study 2; Woodzicka & Ford, 2010). We consider in detail below the implications of having done so.

Bystander action to confront disparaging humor at the intersection of moral courage, prejudice reduction, and collective action

Theoretically, bystander action to confront prejudice represents the intersection between literatures on moral courage, prejudice reduction, and collective action. Prejudice reduction and collective action represent two frameworks for approaching social justice and equality that have, until relatively recently, tended to be considered separately (Wright & Baray, 2012). The literature on moral courage has evolved separately altogether, and moral courage has been differentiated from helping and other forms of pro-social behavior but not compared to collective action or activism per se (Greitemeyer, Fischer, Kastenmüller, & Frey, 2006; Osswald et al., 2010). It is therefore worth considering how these findings articulate across the three broad traditions.

Moral courage is traditionally defined as “brave behavior accompanied by anger and indignation, which intends to enforce societal and ethical norms without considering one’s own social costs” (Osswald et al., 2010, p.150). The current paradigm fits with this definition in providing a situation in which there is an imbalance of power: a direct perpetrator and a

group of victims, but also a number of bystanders (in this case, other potential confronters or enablers) who comprise a third party or audience. However, our approach also differs from other treatments of moral courage by considering intervention in contexts where the harm is not overt and the societal norms about the expression of such humor are contested (Condor, 2006; cf. Fischer, Greitemeyer, Pollozek, & Frey, 2006). Our focus on online confrontation means that the inferred ‘costs’ are not in terms of immediate physical safety but can instead be counted as potentially enduring effects on reputation, status, and emotional health: Ronson’s (2016) monograph on online shaming attests to the serious emotional and life consequences that contentious online encounters can have.

Prior research suggests that morally courageous action is not affected by the presence of bystanders per se (Fischer et al., 2006) but that anger (Greitemeyer et al., 2006) and situational pro-social norms (see Osswald et al., 2010) foster moral courage. Our approach complements these insights to suggest that both one’s own anger (Study 2), as well as the perceived anger of those co-present (i.e., an anger or outrage norm; Thomas & McGarty, 2009; Studies 1-3) also shapes confrontation because, in the context of disparagement humor, it allows the situation to be perceived as necessitating (potentially costly) confrontation. Thus, it is not the presence or absence of bystanders per se that matters (Fischer et al., 2006): rather, the substantive reactions of those bystanders have the potential to shape responses to the situation through processes of (emotion-based) social influence (see also Levine & Crowther, 2008). The finding that confrontation was more intense when one’s personal anger was matched by the anger of other bystanders (Study 2) suggests a role of anger at both the individual and group level, indicating that moral confrontation is driven by both intra-personal moral judgements (feelings of anger), as well as social context. Thus, although anger may be linked with aggression and/or seen as an undesirable expressive emotion

(Deffenbacher, Oetting, Lynch, & Morris, 1996) in some contexts it is a critical pro-social, moral, emotion (Thomas, McGarty, & Mavor, 2009b).

Indeed, our findings here concur with those in the literature on collective action more generally (that is, the literature on when people will act as members of groups to challenge inequalities): people need to perceive collective support for their worldviews in order to develop the commitment to act (van Zomeren et al., 2004; Smith, Thomas, & McGarty, 2015; Thomas et al., 2009). In situations where the inequality or injustice is clear to see, this support comprises an important part of emotion-focused coping around injustice (van Zomeren et al., 2012). However, in situations where *the inequality is itself not recognized or is contested*, bystander responses – those co-present at the moment when the inequality is raised and discussed – may be more important still. A denial of inequality is one of the core markers of modern sexism (e.g., Swim, Aikin, Hall, & Hunter, 1995) and racism (e.g., McConahay, 1986): people often reject that discrimination exists at all. It is likely that bystander responses – online and offline – play an important role in shaping perceptions of inequality as a precondition to resistance more broadly (see also Leach & Livingstone, 2015). Thus, although confrontation is often conceived of as interpersonal in nature (e.g., Hyers, 2010), it is nevertheless shaped by group processes.

It is notable that in Study 1 in particular even a small number of angry comments (4 of 22) that were presented in the ‘humor’ condition produced weaker forms of confrontation (see also Asch, 1956). That is, even a small dissenting minority created fertile ground for an alternative construction of the remark (Moscovici et al., 1969). It is also possible that the different emotions (humor/anger) delineated an emotion-based intergroup context, in which people had to decide who they stand “with” (social categorization) and “against” (see Kessler & Hollbach, 2005; Livingstone et al., 2011, 2016). Emotions provide powerful markers of intergroup relations; exploring the ways that people can form common cause to resist

injustice through processes of emotion-based social influence is a critical direction for future research (see also Thomas et al., 2009b). Finally, there also seems to be an inherent tension between, on the one hand, encouraging online confrontation of disparagement as a normatively ‘good’ or worthwhile method for contesting everyday prejudice; and, on the other, the promotion of online shaming – whereby the online confrontational response becomes disproportionate to the initial injury.

Limitations and Future Directions

Experimental tests of actual bystander confrontation are rare (Nelson et al., 2011). Our approach developed an experimental method that was rich in mundane realism and germane to understanding the “everyday” ways in which people come to act as moral agents to perceive disparagement as prejudice necessitating confrontation. Nevertheless, our approach entails a number of complexities and limitations that should be acknowledged. First, given the behavioral nature of the paradigm, and that several of the focal variables were relatively infrequent and categorical, the effects were not observed consistently across all measures (occurrence, intensity) in each of the studies. Second, in our paradigm, participants may or may not have felt that they were addressing the perpetrator directly. That is, as in most online interactions and other instances of moral confrontation more generally, the intended audience of the confrontation in this case was not only the creator of the disparaging content (the perpetrator) but also an audience of past and future consumers of that content – some of whom corroborated the content and are therefore, in Condor’s (2006) terms, accomplices to the act. It will be important for future research to disentangle the effects of confrontation directed at perpetrators (as considered by Czopp et al., 2003; Czopp et al., 2006) or as a tactic to influence the positions of other bystanders (see Postmes & Brunsting, 2002, for a discussion of the distinction between confrontational versus conversionary forms of action).

Moreover, thirdly, participants in our paradigm had no reason to expect further or on-going interactions with other users of the site. Although people do appear to invest emotionally in their online commentary, this experimental setting may have been experienced as rather less risky than would be the case in everyday online interactions. Since social costs are central to definitions of moral courage, future research should directly measure and/or manipulate the perceived risks of online confrontation to assess whether these qualify effects on emotion-based influence (or not). Finally, as described above, findings from Study 1 suggest that the ‘amusement’ condition was, in effect, an anger minority condition. As such, these studies constitute a conservative test of hypotheses and effects may be stronger still where there was unanimous support amongst bystanders for humorous or angry responses (e.g., as may be the case with online ‘echo chambers’, e.g., Barbera, Jost, Nagler, Tucker & Bonneau, 2015; or viral outrage, Sawaoka & Monin, 2018).

The current findings also raise a number of more general suggestions for future research. First, the theorizing presented above suggests that bystander responses may be particularly important in the context of disparagement humor (as opposed to blatant disparagement) because of its inherent ambiguity (Ford et al., 2008; Ford et al., 2017). Any disambiguating effect of bystander action may be operating in parallel, or interaction, with the emotion-based social influence process articulated here. Research shows that norms may be particularly powerful sources of social influence when the stimuli generate uncertainty (Sherif, 1935; Latane & Darley, 1968) but also when there is subjective disagreement in judgements about that stimuli (e.g, McGarty et al., 1993). Future research might directly test the disambiguating effects of bystander commentary via use of experimental manipulations which inform bystanders about the pernicious effects of disparagement humour and therefore reduce uncertainty about the appropriate response; and/or use of covert psychophysiological measures of arousal as a proxy for uncertainty in-situ (e.g., galvanic skin response; Epstein &

Roupenian, 1970). Finally, future research may also investigate the processes identified here using archival data sourced from social networking services to reveal online confrontation in-situ.

Concluding Comments

There is increasing recognition that the very essence of prejudice can, itself, be contested (e.g., Condor, 2006). As Durrheim et al. (2016, p.18) argue:

it [prejudice] may be contested precisely because group members want to shape social actions, events, and definitions of reality, to encourage alignments with and against particular causes, influence our treatment of others, and impel social movements of various kinds.

Few would agree that prejudice is a laughing matter; applying the reverse logic, it follows that a disparaging remark cannot reasonably be construed as prejudice (necessitating morally courageous confrontation), if we are laughing. The results reported here demonstrate that the same clip may be evaluated very differently depending on the normative climate in which it is presented (see also Hsueh, Yogeewaran, & Malinen, 2015). Far from being minor definitional or conceptual points, these issues are core to social psychological attempts to grapple with the contested nature of intergroup relations, moral courage, and collective efforts to challenge inequality.

Footnotes

¹ We adopt the term ‘social influence’ rather than ‘normative influence’ throughout to refer to the impact of knowing what co-present others feel about the stimuli (following Turner, 1991).

² Note that, by default, the PROCESS effect coding was reversed such that anger condition was coded (-1) relative to the amused (1) and control (1) conditions. However, given that it is more intuitive to examine effects of anger (coded 1) relative to the other two conditions, we reversed the signs in our reporting of the results for ease of interpretation (Study 2 and Study 3).

³ A small negative predicted score of -.06 was truncated to zero in Figure 3a, to reflect the practical range of the intensity scale, where zero means no action.

References

- Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. *European Review of Social Psychology, 1*, 195-228.
- Abrams, D., Wetherell, M., Cochrane, S., Hogg, M.A. & Turner, J.C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, and group polarization. *British Journal of Social Psychology, 29*, 97-119.
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*, 31-35.
- Austin, J. L. (1962). *How to do things with words*. Cambridge, Mass.: Harvard University Press.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science, 26*, 1531-1542.
- Baumert, A., Halmburger, A., & Schmitt, M. (2013). Interventions against norm violations: Dispositional determinants of self-reported and real moral courage. *Personality and Social Psychology Bulletin, 39*, 1053-1068.
- Billig, M. (2005). *Laughter and ridicule: Towards a social critique of humour*. Sage: London.
- Bourhis, R. Y., Gadweld, N. J., Giles, H., & Tajfel, H. (1977). Context and ethnic humour in intergroup relations. In A. J. Chapman & H. C. Foot (Eds.), *It's a funny thing, humour* (pp. 261–265). New York: Pergamon Press
- Cheung, A. S. (2014). Revisiting privacy and dignity: online shaming in the global e-village. *Laws, 3*(2), 301-326.

Coles, B. A., & West, M. (2016). Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior, 60*, 233-244.

Condor, S. (2006). Public prejudice as collaborative accomplishment: Towards a dialogic social psychology of racism. *Journal of Community & Applied Social Psychology, 16*, 1-18.

Condor, S., Figgou, L., Abell, J., Gibson, S., & Stevenson, C. (2006). 'They're not racist...' Prejudice denial, mitigation and suppression in dialogue. *British Journal of Social Psychology, 45*, 441-462.

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour, 1*(11), 769-771.

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin, 29*, 532-544.

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90*, 784-803.

Deffenbacher, J. L., Oetting, E. R., Lynch, R. S., & Morris, C. D. (1996). The expression of anger and its consequences. *Behaviour Research and Therapy, 34*, 575-590.

Durrheim, K., Quayle, M., & Dixon, J. (2016). The struggle for the nature of "prejudice": "Prejudice" expression as identity performance. *Political Psychology, 37*, 17-35.

Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science, 4*, 496-521

- Elias, A., & Paradies, Y. (2016). Estimating the mental health costs of racial discrimination. *BMC Public Health, 16*, 1205, doi.org/10.1186/s12889-016-3868-1.
- Epstein, S., & Roupelian, A. (1970). Heart rate and skin conductance during experimentally induced anxiety: The effect of uncertainty about receiving a noxious stimulus. *Journal of Personality and Social Psychology, 16*, 20-28. doi.org/10.1037/h0029786
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.
- Ferguson, M.A. & Ford, T.E. (2008). Disparagement humour: A theoretical and empirical review of psychoanalytic and social identity theories. *HUMOR: International Journal of Humor Research, 21*, 283-312.
- Fiedler, K. (2008). Language: A toolbox for sharing and influencing social reality. *Perspectives on Psychological Science, 3*(1), 38-47.
- Ford, T. E., & Ferguson, M. A. (2004). Social consequences of disparagement humor: A prejudiced norm theory. *Personality and Social Psychology Review, 8*, 79-94.
- Ford, T. E. (1997). Effects of stereotypical television portrayals of African-Americans on person perception. *Social Psychology Quarterly, 60*, 266-275.
- Ford, T.E., Boxer, C.F., Armstrong, J., & Edell, J.R. (2008). More than “Just a joke”: The prejudice-releasing function of sexist humor. *Personality and Social Psychology Bulletin, 34*, 159-170.
- Ford, T., Breedon, C., O'Connor, E., & Banos, N. (2017, September 26). Jokes and humor in intergroup relations. *Oxford Research Encyclopedia of Communication*. Ed. Retrieved 20 Sep. 2018, from

<http://communication.oxfordre.com/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-431>.

Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology*, *36*(2), 267-278.

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., ... & Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*(4), 517-537.

Fraley, B. & Aron, A. (2004) The effect of shared humorous experience on closeness in initial encounters. *Personal Relationships*, *11*, 61-68.

Frijda, N. H. (1986). *The Emotions*. Cambridge, UK: Cambridge University Press.

Gearhart, S., & Zhang, W. (2014). Gay bullying and online opinion expression: Testing spiral of silence in the social media environment. *Social Science Computer Review*, *32*, 18-36.

Greitemeyer, T., Fischer, P., Kastenmüller, A., & Frey, D. (2006). Civil courage and helping behavior: Differences and similarities. *European Psychologist*, *11*(2), 90-98.

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition. Handbook of motivation and cognition* (pp. 28-84). New York, NY, US: Guilford Press.

- Hess, U., Blairy, S. & Kleck, R.E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior, 24*, 265-283.
- Hodson, G., Rush, & MacInnes, C.C. (2010). A joke is just a joke (except when it isn't): Cavalier humor beliefs facilitate the expression of group dominance motives. *Journal of Personality and Social Psychology, 99*, 660–682.
- Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). “Leave your comment below”: Can biased online comments influence our own prejudicial attitudes and behaviors?. *Human Communication Research, 41*, 557-576.
- Hyers, L. (2010). Alternatives to silence in encounters with everyday heterosexism: Activism on the interpersonal front. *Journal of Homosexuality, 57*, 539-565.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349-364. <http://dx.doi.org/10.1037/h0031617>
- Kelley, H. H., & Lamb, T. W. (1957). Certainty of judgment and resistance to social influence. *Journal of Abnormal and Social Psychology, 55*, 137.
- Kessler, T., & Hollbach, S. (2005). Group-based emotions as determinants of ingroup identification. *Journal of Experimental Social Psychology, 41*, 677-685.
- Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology, 10*(3), 215-221.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York, NY: Oxford University Press.
- Leach, C. W., & Allen, A. M. (2017). The social psychology of the Black Lives Matter meme and movement. *Current Directions in Psychological Science, 26*(6), 543-547.

- Leach, C.W. & Livingstone, A.G. (2015). Contesting the meaning of intergroup disadvantage: Towards a psychology of resistance. *Journal of Social Issues, 71*, 614-632.
- Lee, G.C., Platow, M.J., Augustinos, M., Van Rooy, D., Spears, R., & Bar Tal, D. (in press). When are anti-fat attitudes understood as prejudice vs. truth? An experimental study of social influence effects. *Obesity*.
- Levine, M., Cassidy, C., Brazier, G., & Reicher, S. (2002). Self-Categorization and Bystander Non-intervention: Two Experimental Studies 1. *Journal of Applied Social Psychology, 32*, 1452-1463.
- Levine, M., & Crowther, S. (2008). The responsive bystander: How social group membership and group size can encourage as well as inhibit bystander intervention. *Journal of Personality and Social Psychology, 95*, 1429-1439.
- Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., ... & Resnick, P. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication, 10*, 00-00.
- Livingstone, A. G., Shepherd, L., Spears, R., & Manstead, A. S. (2016). "Fury, us": Anger as a basis for new group self-categories. *Cognition and Emotion, 30*, 183-192.
- Livingstone, A. G., Spears, R., Manstead, A. S., Bruder, M., & Shepherd, L. (2011). We feel, therefore we are: emotion as a basis for self-categorization and social action. *Emotion, 11*, 754-767.
- Mallett, R. K., & Wagner, D. E. (2011). The unexpectedly positive consequences of confronting sexism. *Journal of Experimental Social Psychology, 47*, 215-220.

- Manstead, A. S. R., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (pp. 221 – 232). New York: Oxford University Press.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). San Diego, CA, US: Academic Press.
- McGarty, C., Turner, J. C., Oakes, P. J., & Haslam, S. A. (1993). The creation of uncertainty in the influence process: The roles of stimulus information and disagreement with similar others. *European Journal of Social Psychology, 23*, 17-38.
- Mellor, D. (2003). Contemporary racism in Australia: The experience of Aborigines. *Personality and Social Psychology Bulletin, 29*, 479-486.
- Mikula, G., & Wenzel, M. (2000). Justice and social conflict. *International Journal of Psychology, 35*, 126–135.
- Moscovici, S., Lage, E., & Naffrechoux, M. (1969). Influence of a consistent minority on the responses of a majority in a color perception task. *Sociometry, 32*, 365-380.
- Nelson, J., Paradies, Y. & Dunn, K. (2011). Bystander anti-racism: A review. *Analyses of Social Issues and Public Policy, 11*, 263-284.
- Niedenthal, P. & Brauer, M. (2012). Social functionality of human emotion. *Annual Review of Psychology, 63*, 259-285.
- Noelle-Neumann, E. (1974). The spiral of silence a theory of public opinion. *Journal of Communication, 24*, 43-51.

- Osswald, S., Greitemeyer, T., Fischer, P., & Frey, D. (2010). What is moral courage? Definition, explication, and classification of a complex construct. In C. L. S. Pury & S. J. Lopez (Eds.), *The psychology of courage: Modern research on an ancient virtue* (pp. 149-164). Washington, DC, US: American Psychological Association.
doi.org/10.1037/12168-008
- Parkinson, B. (1996). Emotions are social. *British Journal of Psychology*, *87*, 663–683.
- Parkinson, B., Fischer, A. H., & Manstead, A. S. R. (2005). *Emotion in social relations: Cultural, group, and interpersonal processes*. New York, NY: Psychology Press.
- Parrott, S. (2016). When everyone is laughing: The presence, characteristics, and enjoyment of disparagement humor in online TV. *Mass Communication and Society*, *19*, 49-73.
- Pedersen, A., Paradies, Y., Hartley, L. K., & Dunn, K. M. (2011). Bystander antiprejudice: Cross-cultural education, links with positivity towards cultural ‘outgroups’ and preparedness to speak out. *Journal of Pacific Rim Psychology*, *5*, 19-30.
- Platow, M.J., Haslam, S.A., Both, A., Chew, I., Cuddon, M. et al., (2005). “Its not funny if they’re laughing: Self-categorization, social influence, and responses to canned laughter. *Journal of Experimental Social Psychology*, *41*, 542-550.
- Platow, M.J., Van Rooy, D., Augustinos, M., Spears, R., Bar-Tal, D., Grace, D.M. (2019). Prejudice is about collective values, not a biased psychological system. *New Zealand Journal of Psychology*, *48*, 16-22.
- Postmes, T., & Brunsting, S. (2002). Collective action in the age of the Internet: Mass communication and online mobilization. *Social Science Computer Review*, *20*(3), 290-301.
- Ronson, J. (2016). *So You've Been Publicly Shamed*. New York, US: Riverhead Books.

- Saucier, D. A., O'Dea, C. J., & Strain, M. L. (2016). The bad, the good, the misunderstood: The social effects of racial humor. *Translational Issues in Psychological Science, 2*, 75-85.
- Saucier, D. A., Strain, M. L., Miller, S. S., O'Dea, C. J., & Till, D. F. (2018). "What do you call a Black guy who flies a plane?": The effects and understanding of disparagement and confrontational racial humor. *Humor, 31*, 105-128.
- Sawaoka, T., & Monin, B. (2018). The paradox of viral outrage. *Psychological Science, 29*(10), 1665-1678.
- Searle, J.R. (1969). *Speech Acts*. Cambridge, United Kingdom: Cambridge University Press.
- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology, 187*, 60.
- Strain, M. L., Martens, A. L., & Saucier, D. A. (2016). "Rape is the new black": Humor's potential for reinforcing and subverting rape culture. *Translational Issues in Psychological Science, 2*, 86-95.
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68*, 199-214.
- Swim, J. K., & Hyers, L. L. (1998). Excuse me-What did you say?! Women's public and private response to sexist remarks. *Journal of Experimental Social Psychology, 35*, 68-88.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In S. Worchel & W. G. Austin (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

- Thomae, M., & Pina, A. (2015). Sexist humor and social identity: The role of sexist humor in men's in-group cohesion, sexual harassment, rape proclivity, and victim blame. *Humor, 28*, 187-204.
- Thomas, E. F. & McGarty, C. (2009). The role of efficacy and moral outrage norms in creating the potential for international development activism through group-based interaction. *British Journal of Social Psychology, 48*, 115-134.
- Thomas, E. F., McGarty, C., & Mavor, K. I. (2009a). Aligning identities, emotions and beliefs to create commitment to sustainable social and political action. *Personality and Social Psychology Review, 13*, 194-218.
- Thomas, E. F., McGarty, C., & Mavor, K. I. (2009b). Transforming "apathy into movement": The role of prosocial emotions in motivating action for social change. *Personality and Social Psychology Review, 13*, 310-333.
- Turner, J. C. (1991). *Social influence*. Milton Keynes: Open University Press.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the Social Group: A self-categorization theory*. Oxford: Blackwell.
- Van den Bos, K. (2003). On the subjective quality of social justice: the role of affect as information in the psychology of justice judgments. *Journal of Personality and Social Psychology, 85*, 482-498.
- Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18*, 184-188.
- Van Kleef G.A., De Dreu C.K.W. & Manstead A.S.R. (2010). An interpersonal approach to emotion in social decision making: the emotions as social information model. In MP

Zanna (Ed.), *Advances in Experimental Social Psychology*, pp. 45–96. Burlington, MA: Academic

Van Zomeren, M., Spears, R., Fischer, A. H., & Leach, C. W. (2004). Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy. *Journal of Personality and Social Psychology*, *87*, 649-664.

Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific Reports*, *2*, 335.

Wittenbrink, B., & Henly, J. R. (1996). Creating social reality: Informational social influence and the content of stereotypic beliefs. *Personality and Social Psychology Bulletin*, *22*, 598-610.

Woodzicka, J. A., & Ford, T. E. (2010). A framework for thinking about the (not-so-funny) effects of sexist humor. *Europe's Journal of Psychology*, *6*, 174-195.

Wright, S. C., & Baray, G. (2012). Models of social change in social psychology: Collective action or prejudice reduction? Conflict or harmony. In J. Dixon & R.M. Levine (Eds.) *Beyond prejudice: Extending the social psychology of conflict, inequality and social change* (pp. 225-247). Cambridge, UK: Cambridge University Press.

Table 1. Means and proportions for key variables, by condition (Study 1).

Outcome variables	No bystander response control (<i>n</i> = 30)	Angry bystander response (<i>n</i> = 30)	Amused bystander response (<i>n</i> = 29)
Behavioral measures			
Thumbs down	56% (17) ^a	73% (22) ^a	58% (17) ^a
Star ratings	3.90 (2.77) ^a	2.17 (1.90) ^b	3.90 (2.65) ^a
Occurrence of confrontational comment	16% (5) ^a	43% (13) ^b	41% (12) ^b
Intensity of confrontational comment	0.27 (.69) ^a	0.93 (1.23) ^b	0.79 (1.15) ^b
Occurrence of collaborative comment	43% (13) ^a	30% (9) ^a	41% (12) ^a
Intensity of collaborative comment	0.77 (1.04) ^a	0.50 (.90) ^a	0.76 (1.09) ^a
Self-reported measures			
Enjoyment	3.80 (1.87) ^a	3.12 (1.80) ^a	3.64 (1.99) ^a
Appropriateness of user comments	4.15 (.49) ^a	5.03 (1.39) ^b	4.14 (1.22) ^a

Note. The values for thumbs down and comment (confrontational/collaborative) variables represents the proportion of people (percentage) selecting 'thumbs down' or for whom a confrontational/collaborative response was recorded, with the number of people who did so (*n*) in brackets. Super-scripts represent values where means/proportions differ at $p < .05$.

Table 2. Means (standard deviations) for key variables, by experimental condition (Study 2).

Outcome variables	No bystander response control (<i>n</i> = 68)	Angry bystander response (<i>n</i> = 71)	Amused bystander response (<i>n</i> = 63)
Behavioral measures			
Thumbs down	33% (22) ^a	44% (31) ^b	23% (15) ^a
Occurrence of confrontational comment	13% (5) ^a	17% (12) ^a	8% (6) ^a
Intensity of confrontational comment	0.25 (0.70) ^a	0.31 (0.79) ^a	0.12 (0.41) ^a
Occurrence of collaborative comment	62% (42) ^a	49% (34) ^b	70% (48) ^a
Intensity of collaborative comment	1.38 (1.23) ^a	0.93 (1.12) ^b	1.46 (1.10) ^a
Self-reported measures			
Enjoyment	4.45 (2.04) ^a	4.07 (1.93) ^a	4.81 (1.82) ^a
Appropriateness of user comments	4.38 (0.99) ^a	4.13 (1.33) ^b	4.62 (1.16) ^a

Note. Super-scripts represent values where means or proportions differ at $p < .05$.

Table 3. Unstandardized regression coefficients (standard errors) [95% confidence intervals] for test of effect of angry bystander reaction relative to control (effect 1) or bystander amusement (effect 2) on key outcome variables, qualified by own levels of anger and amusement (Study 2).

Outcome measures	Effect 1 (angry =1 control =-1)	Effect 2 (angry = 1, amusement =-1)	Effect of own anger	Effect of own amusement	Interaction between effect 1 and own anger	Interaction between effect 2 and own anger	Interaction between effect 1 and own amusement	Interaction between effect 2 and own amusement
Behavioral measures								
Thumbs up (1)/ Thumbs down (-1)	0.19 (.38) [-0.54, 0.93]	-0.95 (.46)* [-1.84, -0.06]	-0.13 (.34) [-0.79, 0.53]	2.21 (.33) *** [1.57, 2.85]	1.05 (.51)* [0.06, 2.05]	-0.61 (.51) [-1.61, 0.40]	0.27 (.44) [-0.60, 1.14]	-0.07 (.46) [-0.98, 0.84]
Occurrence of confrontational comment	0.37 (.49) [-0.59, 1.33]	-0.13 (.42) [-0.95, 0.69]	0.88 (.30)** [0.30, 1.45]	-0.43 (.24) † [-0.89, 0.03]	0.16 (.40) [-0.63, 0.95]	0.49 (.43) [-0.35, 1.33]	0.74 (.35)* [0.06, 1.42]	-0.06 (.32) [-0.68, 0.57]
Intensity of confrontational comment	-0.02 (.05) [-0.12, 0.09]	0.07 (.05) [-0.03, 0.18]	0.23 (.05)*** [0.13, 0.32]	-0.09 (.04)* [-0.16, -0.02]	0.01 (.07) [-0.13, 0.14]	0.24 (.07)*** [0.09, 0.38]	0.17 (.05)*** [0.07, 0.27]	-0.03 (.05) [-0.13, 0.06]
Occurrence of collaborative comment	-0.09 (.29) [-0.67, 0.49]	-0.49 (.28) [-1.03, 0.06]	-0.78 (.30)** [-1.37, -0.20]	1.32 (.23)*** [.089, 1.77]	0.41 (.45) [-0.48, 1.29]	-0.76 (.38)* [-1.51, -0.01]	-0.58 (.36) [-1.28, 0.12]	0.20 (.28) [-0.36, 0.75]
Intensity of collaborative comment	-0.15 (.10) [-0.34, 0.04]	-0.13 (.10) [-0.32, 0.06]	-0.13 (.09) [-0.30, 0.04]	0.55 (.06)*** [0.42, 0.67]	0.08 (.12) [-0.16, 0.32]	-0.17 (.13) [-0.42, 0.08]	-0.04 (.09) [-0.22, 0.14]	-0.06 (.09) [-0.24, 0.11]
Self-report measures								
Enjoyment	0.03 (.14) [-0.25, 0.31]	-0.24 (.14) [-0.52, 0.04]	-0.33 (.13)** [-0.59, -0.08]	1.04 (.09)*** [0.86, 1.23]	-0.15 (.18) [-0.51, 0.21]	0.19 (.19) [-0.18, 0.57]	0.29 (.13)* [0.02, 0.56]	-0.08 (.13) [-0.34, 0.18]

Appropriateness of	-0.008 (.11)	-0.26 (.11)*	-0.15 (.10)	-0.10 (.07)	0.10 (.14)	0.21 (.15)	-0.10 (.10)	-0.22 (.10)*
user comments	[-0.23, 0.21]	[-0.49, -0.04]	[-0.35, 0.05]	[-0.24, 0.05]	[-0.19, 0.38]	[-0.08, 0.51]	[-0.31, 0.11]	[-0.43, -0.02]

Note: *** denotes significant at $p < .001$, ** denotes significant at $p < .01$, * denotes significant at $p < .05$, † denotes $p \leq .07$. Analyses for thumbs up/down and occurrence of confrontational/collaborative comment are based on logistic regression.

Table 4. Means and proportions for key variables, by condition (Study 3).

Outcome variables	No bystander response control / Non-humorous clip (n = 34)	Angry bystander response / Non-humorous clip (n = 36)	Amused bystander response / Non-humorous clip (n = 35)	No bystander response control / Humorous clip (n = 39)	Angry bystander response / Humorous clip (n = 28)	Amused bystander response / Humorous clip (n = 39)
Behavioral measures						
Thumbs down	90% (28) ^a	91% (31) ^a	73% (24) ^a	36% (14) ^b	26% (7) ^b	21% (8) ^b
Occurrence of confrontational comment	3% (1) ^a	17% (6) ^b	3% (1) ^a	13% (5) ^a	21% (6) ^b	10% (4) ^a
Intensity of confrontational comment	0.03 (0.17) ^a	0.42 (0.97) ^b	0.09 (0.51) ^a	0.26 (0.75) ^a	0.46 (0.96) ^b	0.15 (0.49) ^a
Occurrence of collaborative comment	6% (2) ^a	8% (3) ^a	14% (5) ^a	59% (23) ^b	54% (15) ^b	79% (31) ^c
Intensity of collaborative comment	0.12 (0.48) ^a	0.19 (0.71) ^b	0.29 (0.75) ^a	1.18 (1.14) ^c	1.11 (1.20) ^d	1.74 (1.04) ^c

Self-reported measures						
Enjoyment	1.88 (1.23) ^a	2.29 (1.49) ^a	2.78 (1.93) ^b	4.19 (2.13) ^c	4.39(1.92) ^c	4.96 (1.74) ^d
Appropriateness of user comments	4.16 (0.50) ^a	4.84 (1.12) ^c	4.18 (1.40) ^a	4.69 (0.94) ^b	4.54 (1.49) ^b	4.99 (1.06) ^c

Note. The values for thumbs down and comment (confrontational/collaborative) variables represents the proportion of people (percentage) selecting 'thumbs down' or for whom a confrontational/collaborative response was recorded, with the number of people who did so (*n*) in brackets.

Table 5. Unstandardized regression coefficients (standard errors) [95% confidence intervals] for test of effect of angry bystander reaction relative to control (effect 1) or bystander amusement (effect 2) on key outcome variables, qualified by presentation of the remark (humorous/statement) (Study 3).

Outcome measures	Effect 1 (angry =1, control =-1)	Effect 2 (angry = 1, amusement =-1)	Effect of presentation (1 = humor, -1 = non- humor)	Interaction between effect 1 and presentation	Interaction between effect 2 and presentation
Behavioural measures					
Thumbs up (1)/	0.42 (.28)	-0.62 (.26)*	1.37 (.19)***	0.03 (.28)	0.27 (.25)
Thumbs down (-1)	[-0.13, 0.97]	[-1.11, -0.11]	[0.99, 1.75]	[-0.52, 0.58]	[-0.23, 0.76]
Occurrence of confrontational comment	0.30 (.44)	0.62 (.45)	0.52 (.29) [†]	-0.32 (.44)	-0.03 (.45)
	[-0.56, 1.16]	[-0.27, 1.51]	[-0.05, 1.09]	[-1.17, 0.54]	[-0.91, 0.86]
Intensity of confrontational comment	0.08 (.07)	0.13 (.07) [†]	0.06 (.05)	-0.06 (.07)	0.03 (.07)
	[-0.05, 0.22]	[-0.007, 0.27]	[-0.04, 0.16]	[-0.20, 0.07]	[-0.10, 0.18]
Occurrence of collaborative comment	0.27 (.31)	-0.56 (.28)*	1.48 (.21)***	-0.13 (.31)	-0.11 (.28)
	[-0.34, 0.89]	[-1.11, -0.01]	[1.06, 1.89]	[-0.74, 0.49]	[-0.66, 0.44]
Intensity of collaborative comment	0.11 (.09)	-0.25 (.09)**	0.57 (.07)***	0.03 (.09)	-0.15 (.09) [†]
	[-0.07, 0.29]	[-0.43, -0.06]	[0.44, 0.69]	[-0.15, 0.21]	[-0.33, 0.03]
Self-reported measures					
Enjoyment	0.38 (.18)*	-0.46 (.18)**	1.09 (.12)***	-0.06 (.17)	0.02 (.18)
	[0.04, 0.72]	[-0.81, -0.11]	[0.85, 1.34]	[-0.41, 0.28]	[-0.33, 0.36]

Appropriateness of user comments	0.14 (.11) [-0.08, 0.36]	-0.01 (.11) [-0.23, 0.21]	0.17 (.08)* [0.01, 0.33]	-0.09 (.11) [-0.31, 0.13]	-0.24 (.11)* [-0.46, -0.02]
-------------------------------------	-----------------------------	------------------------------	-----------------------------	------------------------------	--------------------------------

Note: *** denotes significant at $p < .001$, ** denotes significant at $p < .01$, * denotes significant at $p \leq .05$, † denotes $p \leq .08$. Analyses for thumbs up/down and the occurrence of a confrontational/collaborative comments are based on logistic regression.

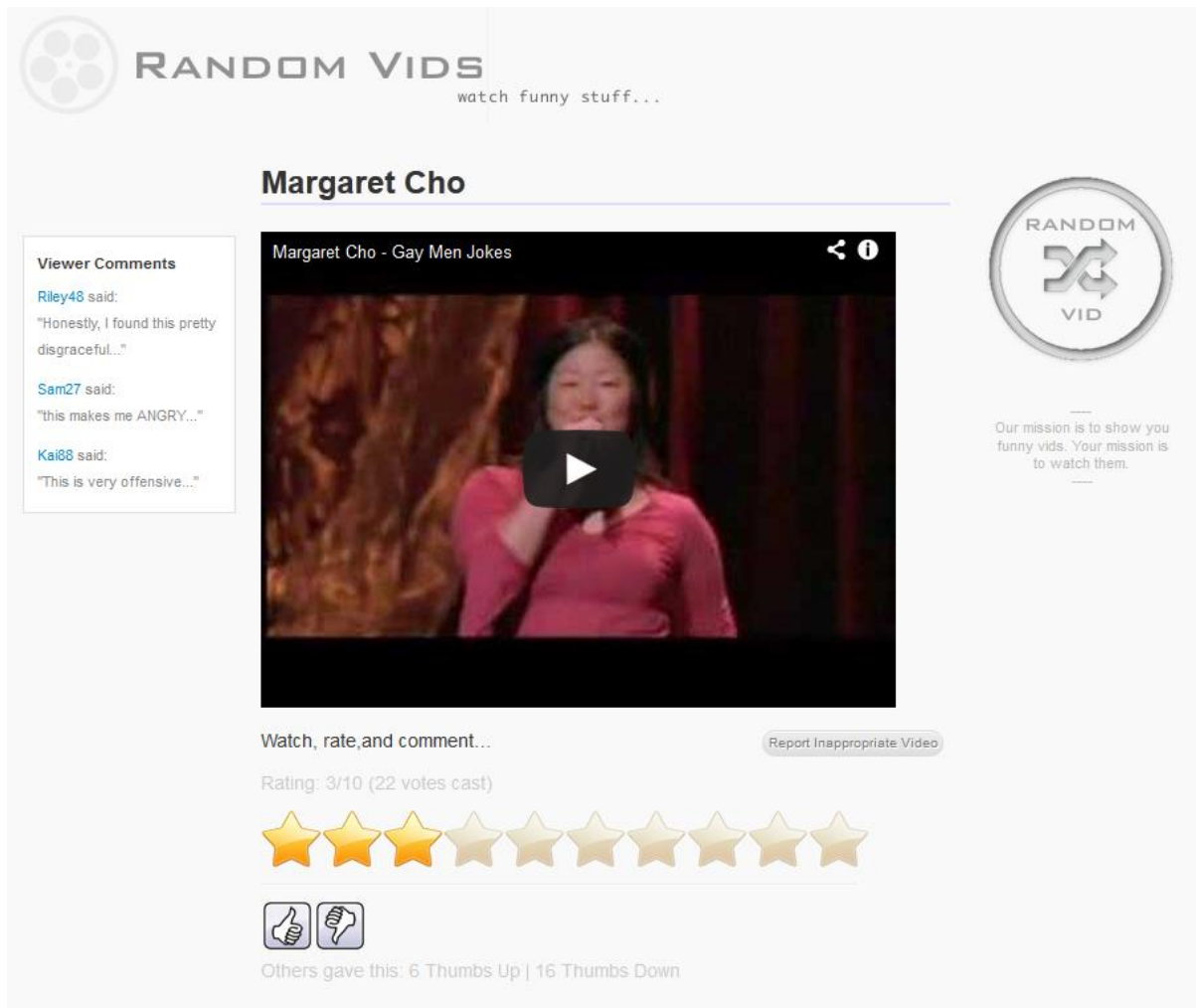


Figure 1. *User interface for the shared media paradigm (angry bystander condition). Ratings (3/10) and Thumbs Down (16) denote dislike; viewer comments capture bystander appraisals of anger (top left provided a snapshot, full 'bystander' comments were viewed when the participant scrolled down). Participants indicated their own ratings (stars, thumbs up/down), could report the content as inappropriate, and left a comment.*

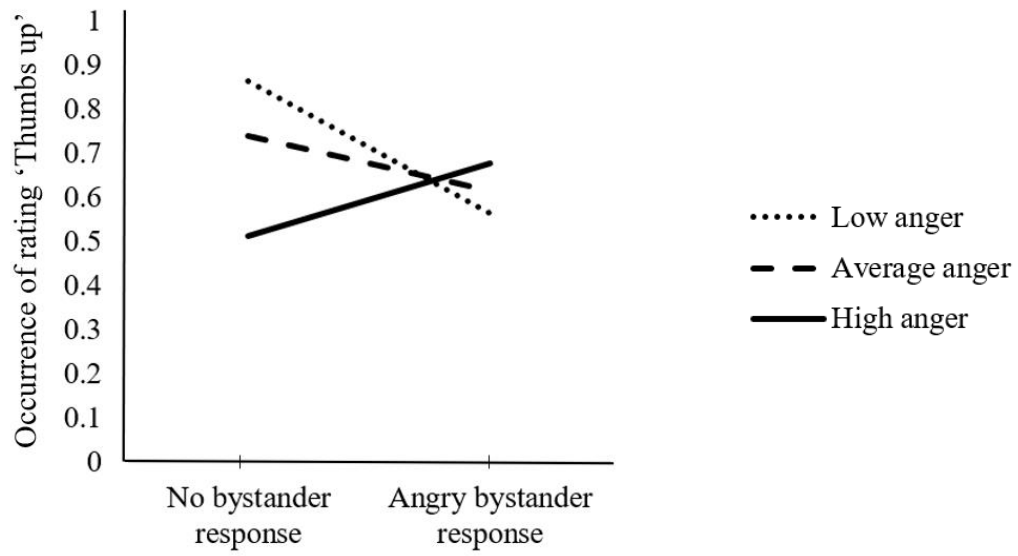


Figure 2. *The effect of bystander reactions on ratings is qualified by levels of anger*

(Study 2).

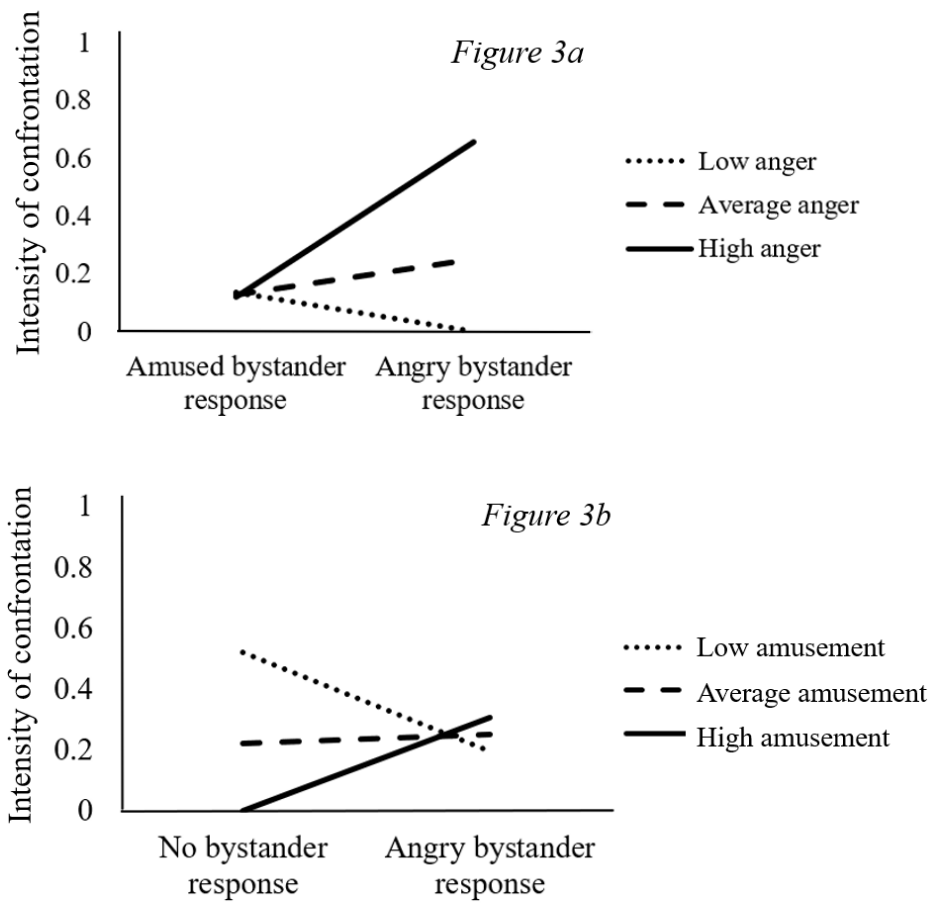


Figure 3. *The effect of bystander responses on intensity of confrontation is qualified by own emotional responses (Study 2).*

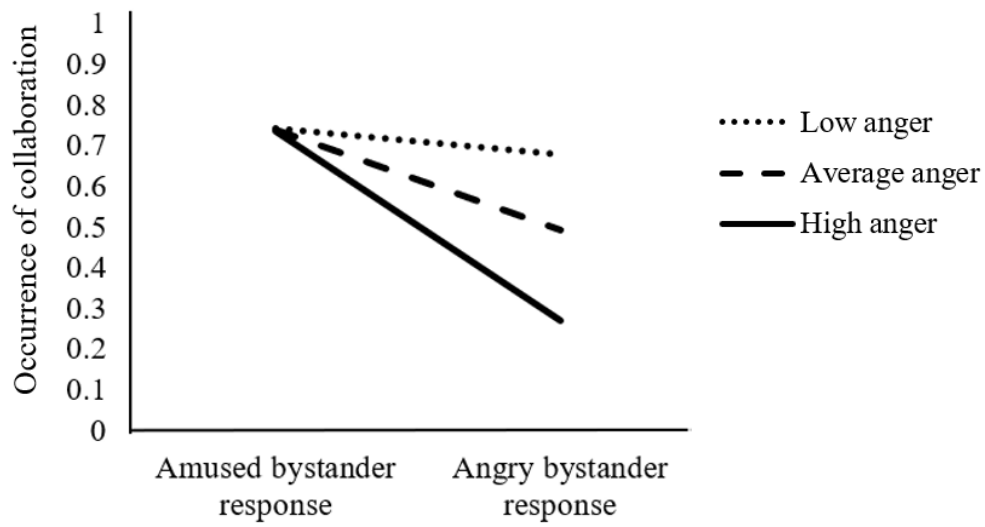


Figure 4. *The effect of bystander responses on the occurrence of collaboration is qualified by levels of anger (Study 2).*

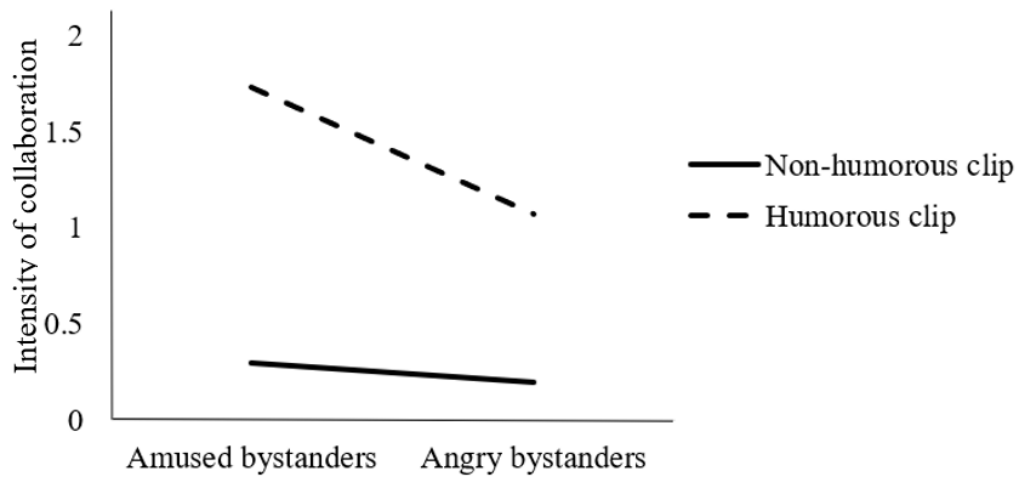


Figure 5. *The effect of bystander responses on the intensity of collaboration is qualified by whether the clip was presented in a humorous/non-humorous form (Study 3).*

Supplementary Materials

In addition to the focal variables (reported in the main analysis), in Studies 1-3 we took additional measures of: the degree to which participants identified with other users of the site, prejudice towards gay people (Morrison & Morrison, 2003), social dominance orientation (Pratto, Sidanius, Stallworth & Malle, 1994), the importance of confronting (Rasinski, Geers & Czopp, 2013) and humor style (Thorson & Powell, 1993). The tables below describe the key findings on these variables across all studies, using the identical analytical approach to that described in the main analysis. We also provide additional figures which were not contained in the primary manuscript due to space considerations.

Study 1

Table S1. Means (standard deviations) for supplementary variables, by condition (Study 1).

Outcome variables	No bystander response control (<i>n</i> = 30)	Angry bystander response (<i>n</i> = 30)	Amused bystander response (<i>n</i> = 29)
Identification with other users	3.93 (1.16) ^a	3.83 (1.32) ^a	3.91 (1.00) ^a
Prejudice towards gay people	2.53 (1.26) ^a	2.00 (1.00) ^a	2.51 (1.14) ^a
Social dominance orientation	2.53 (1.25) ^a	2.60 (1.03) ^a	2.53 (1.16) ^a
Importance of confronting	4.62 (.97) ^a	4.91 (1.15) ^a	5.09 (1.23) ^a
Humor style	5.31 (.78) ^a	5.70 (.55) ^a	5.39 (.98) ^a

Note. Super-scripts represent values where means differ at $p < .05$.

Study 2

Figures S1-S2 display the interactions between the bystander response and own amusement on self-reported enjoyment (Figure S1) and appropriateness of bystander commentary (Figure S2).

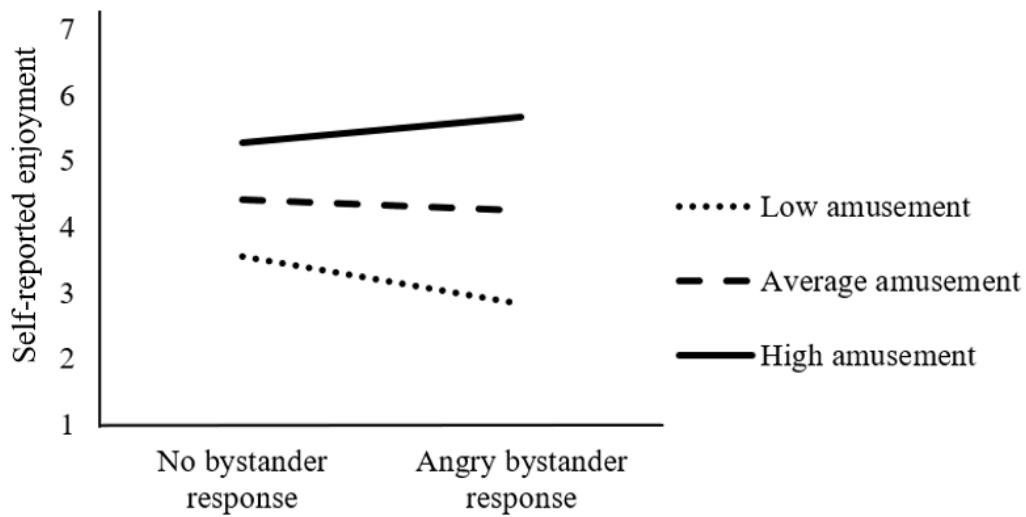


Figure S1. *The effect of bystander reactions on self-reported enjoyment is qualified by levels of amusement (Study 2).*

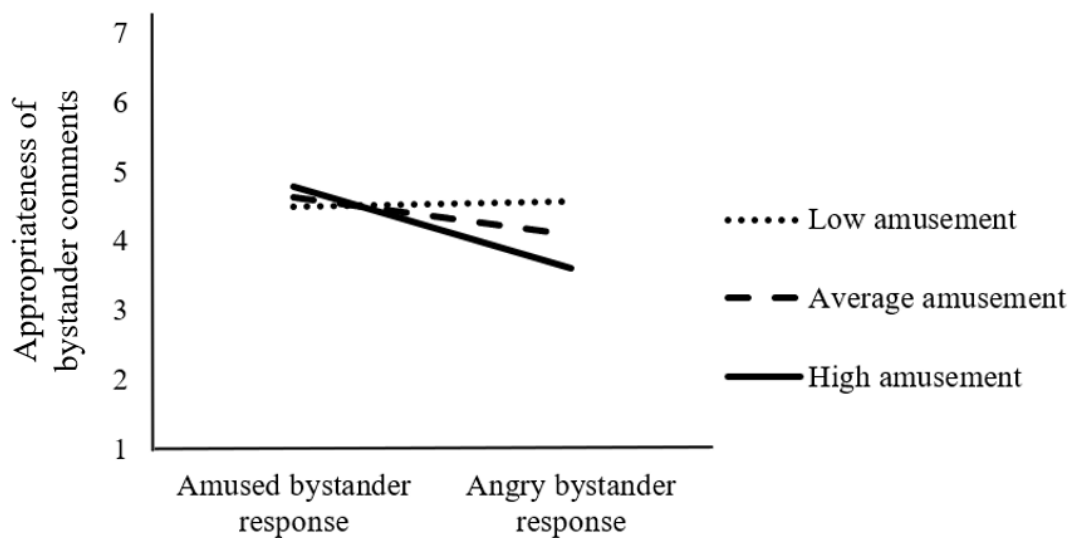


Figure S2. *The effect of bystander reactions on appropriateness of bystander commentary is qualified by levels of amusement (Study 2).*

During the review process it was suggested that we conduct a comparison between the amused and control conditions and include this in the supplementary results for completeness. Table S2 displays the results for the test of the difference between the amused bystander and control conditions (effect 2) on the key outcome variables, controlling also for the difference between the anger and control conditions (effect 1). We focus our discussion here (in the supplementary section) on the findings in relation to the comparison between amusement and control, and interaction terms.

Table S2 shows that an amused bystander response (relative to control; effect 2) was associated with greater proportion of thumbs up ratings, and amused commentary was also seen as more appropriate (relative to no commentary). There were three significant interactions between effect 2 and one's own anger on the intensity of confrontation, occurrence of collaboration and perceived appropriateness of bystander commentary; and one significant interaction between effect 2 and own amusement on the appropriateness of bystander commentary. These can be summarised as follows:

- When anger was low, $b = 0.10$, $p = .14$, or average levels, $b = -0.08$, $p = .40$, the manipulation did not affect the intensity of confrontation; when one was high in anger, exposure to amused commentary (relative to no commentary) reduced the intensity of confrontation, $b = -0.26$, $p = .004$.
- When one is high in anger, participants were more likely to collaborate in the amused condition (than the control condition), $b = 1.04$, $p = .03$, but there were no differences between conditions at average ($p = .80$) or low levels of anger ($p = .39$).
- When one is low, $b = 0.50$, $p = .001$, or average in anger, $b = 0.39$, $p = .001$ the bystander respondents in the amused condition seem more appropriate. However, when one is high in anger, there was no difference in the perceived appropriateness of bystander responses in the amused and control conditions. The interaction between effect 2 and own amusement expresses a similar pattern: when one is low in amusement, there were no differences between conditions ($p = .78$), however, when one is at average, $b = 0.39$, $p = .001$, or high levels of amusement, $b = 0.60$, $p = .001$, then the amused bystander response is seen as more appropriate.

Table S3 displays the means (standard deviations) for the supplementary variables, by condition, whilst Table S4 reports the results of the test of the effect of the bystander manipulation (effect 1 and effect 2) by own emotion (anger and amusement) for the supplementary variables.

Table S2. Unstandardized regression coefficients (standard errors) for test of effect of amused bystanders relative to control (effect 2) on key variables, qualified by own levels of anger and amusement (Study 2).

Outcome measures	Effect 1 (anger =1 control =-1)	Effect 2 (amused = 1, control = -1)	Effect of own anger	Effect of own amusement	Interaction between effect 1 and own anger	Interaction between effect 2 and own anger	Interaction between effect 1 and own amusement	Interaction between effect 2 and own amusement
Behavioral measures								
Thumbs up (1)/	-0.76 (.36)*	0.95 (.46)*	-0.13 (.34)	2.21 (.33)***	0.44 (.41)	0.61 (.51)	0.20 (.48)	0.07 (.46)
Thumbs down (-1)								
Occurrence of confrontational comment	0.24 (.43)	0.13 (.42)	0.88 (.30)**	-0.43 (.24)	0.65 (.42)	-0.49 (.43)	0.68 (.34)*	0.06 (.32)
Intensity of confrontational comment	0.05 (.05)	-0.07 (.05)	0.23 (.05)***	-0.09 (.04)*	0.25 (.07)***	-0.24 (.07)***	0.14 (.05)	0.03 (.05)
Occurrence of collaborative comment	-0.58 (.27)*	0.49 (.28)	-0.78 (.30)**	1.33 (.23)***	-0.35 (.42)	0.76 (.38)*	-0.38 (.29)	-0.20 (.28)
Intensity of collaborative comment	-0.27 (.10)**	0.13 (.10)	-0.13 (.09)	0.55 (.06)***	-0.08 (.12)	0.17 (.13)	-0.10 (.09)	0.06 (.09)
Self-report measures								
Enjoyment	-0.21 (.14)	0.24 (.14)	-0.33 (.13)*	1.04 (.09)***	0.05 (.18)	-0.20 (.19)	0.21 (.13)	0.08 (.13)

Appropriateness of	-0.27 (.11) *	0.26 (.11) *	-0.15 (.10)	-0.10 (.07)	0.31 (.14) *	-0.21 (.15) *	-0.33 (.10) **	0.22 (.10) *
user comments								

Note: *** denotes significant at $p < .001$, ** denotes significant at $p < .01$, * denotes significant at $p < .05$. Analyses for thumbs up/down and occurrence of confrontational/collaborative comment are based on logistic regression

Table S3. Means (standard deviations) for supplementary variables, by experimental condition (Study 2).

Outcome variables	No bystander response control (<i>n</i> = 73)	Angry bystander response (<i>n</i> = 76)	Amused bystander response (<i>n</i> = 69)
Identification with other users	3.56 (1.56) ^a	3.30 (1.56) ^a	3.62 (1.47) ^a
Identification as anti- prejudice	5.01 (1.31) ^a	4.95 (1.45) ^a	4.89 (1.50) ^a
Prejudice towards gay people	2.93 (1.52) ^a	2.74 (1.47) ^a	2.97 (1.72) ^a
Social dominance orientation	2.74 (1.23) ^a	2.59 (1.07) ^a	2.55 (1.16) ^a
Importance of confronting	4.84 (1.06) ^a	4.84 (1.10) ^a	4.67 (1.23) ^a
Humor style	5.13 (1.13) ^a	5.16 (0.79) ^a	5.30 (0.99) ^a
Political correctness	3.74 (1.57) ^a	3.62 (1.75) ^a	3.61 (1.82) ^a

Note. Super-scripts represent values where means differ at $p < .05$.

Table S4. Unstandardized regression coefficients (standard errors) for test of effect of angry bystander reaction relative to control (effect 1) or bystander amusement (effect 2) on supplementary variables, qualified by own levels of anger and amusement (Study 2).

Outcome measures	Effect 1 (angry =1 v control =-1)	Effect 2 (angry = 1, amusement =-1)	Effect of own anger	Effect of own amusement	Interaction between effect 1 and own anger	Interaction between effect 2 and own anger	Interaction between effect 1 and own amusement	Interaction between effect 2 and own amusement
Identification with other users	-0.05 (.14)	-0.15 (.14)	0.41 (.13)***	0.39 (.09)***	0.16 (.18)	-0.04 (.19)	-0.15 (.13)	-0.10 (.13)
Identification as anti-prejudice	-0.06 (.13)	0.20 (.14)	0.08 (.12)	0.18 (.09)*	-0.12 (.18)	0.03 (.18)	0.17 (.13)	-0.08 (.13)
Prejudice towards gay people	-0.03 (.15)	-0.27 (.15)	0.10 (.13)	-0.11 (.10)	0.02 (.19)	0.19 (.20)	-0.14 (.14)	0.02 (.14)
Social dominance orientation	-0.12 (.12)	0.02 (.12)	0.12 (.11)	0.0002 (.07)	0.15 (.15)	-0.07 (.15)	0.01 (.11)	-0.02 (.11)
Importance of confronting	-0.13 (.11)	0.15 (.11)	0.08 (.10)	0.12 (.07)	0.13 (.14)	0.10 (.15)	0.05 (.11)	-0.20 (.10)*
Humor style	0.07 (.10)	-0.10 (.10)	-0.07 (.09)	0.19 (.06)**	0.06 (.12)	0.02 (.13)	0.07 (.09)	-0.11 (.09)
Political correctness	-0.14 (.17)	0.13 (.17)	0.36 (.15)*	0.13 (.11)	-0.02 (.22)	0.11 (.22)	0.08 (.16)	-0.18 (.15)

Note: *** denotes significant at $p < .001$, ** denotes significant at $p < .01$, * denotes significant at $p < .05$.

Study 3

During the review process it was suggested that we conduct a third comparison between the amused and control conditions and include this in the supplementary results for completeness. Table S5 displays the results for the test of the difference between the amused bystander and control conditions (effect 2) on the key outcome variables, controlling also for the difference between the anger and control conditions (effect 1). We focus our discussion here (in the supplementary section) on the findings in relation to the comparison between amusement and control, and interaction terms.

Table S5 shows that an amused bystander response (relative to control; effect 2) was associated with greater proportion of thumbs up ratings, self-reported enjoyment, and amused commentary also lead to greater occurrence and intensity of collaborative commentary. There was one significant interaction between effect 2 and own amusement on the perceived appropriateness of bystander commentary. However, neither of the simple slopes were significant ($p = .16$ for the statement clip and $p = .11$ for the humorous clip).

Table S6 displays the means (standard deviations) for the supplementary variables and Table S7 displays the results of the test of the effect of the bystander manipulation (effect 1 and effect 2) by presentation of the clip (humor/non-humor) for the supplementary variables.

Table S5. Unstandardized regression coefficients (standard errors) for test of effect of amused bystanders relative to control on key outcome variables, qualified by presentation of the clip (Study 3).

Outcome measures	Effect 1 (angry =1, control =-1)	Effect 2 (amusement =1, control = -1)	Effect of presentation (1 = humor, -1 = non- humor)	Interaction between effect 1 and presentation	Interaction between effect 2 and presentation
Behavioural measures					
Thumbs up (1)/ Thumbs down (-1)	-0.19 (.29)	0.61 (.26)*	1.37 (.19)***	0.30 (.29)	-0.27 (.25)
Occurrence of confrontational comment	0.92 (.35)**	-0.62 (.45)	0.52 (.29)	-0.34 (.35)	0.03 (.45)
Intensity of confrontational comment	0.21 (.07)**	-0.13 (.07)	0.06 (.05)	-0.03 (.07)	-0.03 (.07)
Occurrence of collaborative comment	-0.29 (.29)	0.56 (.28)*	1.48 (.21)***	-0.24 (.29)	0.11 (.28)
Intensity of collaborative comment	-0.13 (.09)	0.25 (.09)**	0.57 (.06)***	-0.13 (.09)	0.15 (.09)
Self-reported measures					
Enjoyment	-0.08 (.18)	0.46 (.18)*	1.09 (.12)***	-0.05 (.18)	-0.02 (.18)

Appropriateness of	0.13 (.11)	0.01 (.11)	0.17 (.08) *	-0.32 (.11) **	0.24 (.11) *
user comments					

Note: *** denotes significant at $p < .001$, ** denotes significant at $p < .01$, * denotes significant at $p < .05$. Analyses for thumbs up/down and occurrence of confrontational/collaborative comment are based on logistic regression.

Table S6. Means (standard deviations) for supplementary variables, by condition (Study 3).

Outcome variables	No bystander response control / Non-humorous clip (<i>n</i> = 34)	Angry bystander response / Non- humorous clip (<i>n</i> = 36)	Amused bystander response / Non- humorous clip (<i>n</i> = 35)	No bystander response control / Humorous clip (<i>n</i> = 39)	Angry bystander response / Humorous clip (<i>n</i> = 28)	Amused bystander response / Humorous clip (<i>n</i> = 39)
Identification with other users	4.01 (1.35) ^a	3.52 (1.57) ^a	3.43 (1.61) ^a	3.72 (1.43) ^a	3.83 (1.31) ^a	3.62 (1.49) ^a
Identification as anti- prejudice	4.70 (1.69) ^a	4.67 (1.64) ^a	5.41 (1.44) ^b	5.31 (1.22) ^b	4.73 (1.63) ^a	4.63 (1.50) ^a
Prejudice towards gay people	3.06 (1.65) ^a	3.16 (1.66) ^a	2.32 (1.23) ^a	2.53 (1.50) ^a	2.96 (1.66) ^a	3.41 (1.77) ^a
Social dominance orientation	2.39 (1.09) ^a	2.72 (1.43) ^b	2.10 (.96) ^a	1.91 (1.22) ^a	2.68 (1.12) ^b	2.40 (1.25) ^a
Importance of confronting	4.79 (1.23) ^a	4.67 (1.18) ^a	5.06 (1.07) ^a	4.96 (1.07) ^a	4.82 (1.12) ^a	4.15 (1.12) ^a

Humor style	5.30 (.91) ^a	5.24 (.99) ^a	5.45 (.71) ^a	5.20 (.64) ^a	5.29 (1.12) ^a	5.07 (1.03) ^a
Political correctness	4.79 (1.35) ^a	4.51 (1.20) ^a	4.99 (1.36) ^a	5.18 (1.35) ^a	4.83 (1.37) ^a	4.45 (1.27) ^a

Table S7. Unstandardized regression coefficients (standard errors) for test of effect of angry bystander reaction relative to control (effect 1) or bystander amusement (effect 2) on supplementary variables, qualified by presentation of the remark (humorous/statement) (Study 3).

Outcome measures	Effect 1 (angry =1, control =-1)	Effect 2 (angry = 1, amusement =-1)	Effect of presentation (1 = humor, -1 = non- humor)	Interaction between effect 1 and presentation	Interaction between effect 2 and presentation
Identification with other users	-0.16 (.15)	0.13 (.15)	0.03 (.10)	0.16 (.15)	-0.03 (.15)
Identification as anti-prejudice	-0.17 (.15)	0.01 (.15)	-0.01	-0.29*	0.32*
Prejudice towards gay people	0.20 (.15)	-0.10 (.15)	0.05 (.11)	0.28 (.15)	-0.42 (.15)**
Social dominance orientation	0.25 (.12)*	0.05 (.12)	-0.04 (.08)	0.19 (.12)	-0.16 (.12)
Importance of confronting	0.16 (.11)	0.18 (.11)	-0.10 (.08)	-0.17 (.11)	0.34 (.11)**
Humor style	0.007 (.09)	-0.007 (.09)	-0.06 (.07)	-0.01 (.09)	0.10 (.09)
Political correctness	-0.25 (.13)	0.16 (.13)	0.03 (.09)	-0.14 (.13)	0.26 (.13)*

Note: *** denotes significant at $p < .001$, ** denotes significant at $p < .01$, * denotes significant at $p \leq .05$. Analyses for thumbs up/down and the occurrence of a confrontational/collaborative comments are based on logistic regression.

Study 4: Bogus pipeline study

In addition to the studies reported in the manuscript we conducted an additional study in which we attempt an experimental manipulation of individual anger and amusement prior to the bystander (social appraisal) manipulation of anger and amusement. Specifically, we used an adaptation of the bogus pipeline technique (see Jones & Sigall, 1971) to experimentally manipulate individual, personal emotional reactions of anger and amusement.

Method

Design and participants. The study was conducted in the laboratory. Upon arrival participants were randomly allocated to one of four experimental conditions comprising a 2 (individual emotion: anger/amusement) x 2 (bystander emotion: anger/amusement) between persons design. Participants ($N = 121$) were university students or persons recruited on the campus of an Australian university. They were remunerated with course credit or AUD\$10. Participants were primarily female (76%) with an average age of 27 ($SD = 10.82$).

Procedure. The participants were greeted by an experimental assistant and were given the same cover story as in Studies 1-3. In addition, they were told that we were going to use a galvanic skin response to record their implicit/automatic emotional responses to the clip. Electrodes were placed on the hands of participants and an experimental assistant sat on the other side of a partition to ostensibly administer the psychophysiological measure, but whilst maintaining the anonymous responding of participants.

As in Study 1, participants first watched a filler clip before watching the target disparagement clip. Having finished the clip, participants were presented with a graphic that suggested that the computer was calculating their response. They were then given the bogus feedback that they had either registered an "angry" response or an "amused" response. Participants then went on to view the content of the other users. As in Study 1, those

comments either reflected amusement (bystander amusement condition) or anger (bystander anger condition). Participants left their own ratings and comments on the Wordpress site and were then sent to an online survey (Survey Select) to complete other measures of self-reported enjoyment, anger/amusement, appropriateness of comments and supplementary measures.

Unfortunately, significant technical problems occurred throughout the study. Specifically: the webpages that we were using had problems loading, necessitating frequent intervention from the experimental assistant. These problems may have obscured the effects of the independent variables.

Results

Manipulation checks. Measures of individual emotions revealed that the bogus pipeline procedure did not appear to affect self-reported emotional states ("I found the clip outrageous [amusing]"). The differences in anger were not reliable, $F(1, 119) = 2.89, p = .09, \eta^2_p = .025$, and neither were the differences in amusement, $F(1, 119) = .08, p = .78, \eta^2_p = .001$. However, as in Studies 1-3, the bystander manipulation was effective. Those in the 'angry' bystander condition reported that bystanders found the clip more outrageous ($M = 5.50, SE = .19$), than those in the 'amused' bystander condition ($M = 4.36, SE = .19$), $F(1, 119) = 18.05, p < .001, \eta^2_p = .14$. Conversely, those in the 'amused' bystander condition reported that bystanders found the clip more amusing ($M = 4.42, SE = .18$) than those in the 'angry' condition ($M = 4.42, SE = .18$), $F(1, 119) = 61.06, p < .001, \eta^2_p = .35$.

Main analysis. Table S8 displays the means and standard deviations for the behavioral measures, self-reported measures and supplementary measures. We used PROCESS (Model 1) to test the direct and interactive effects of the manipulations on the outcomes. The super-scripts denote where means were different ($p < .05$) It can be seen that there was a main effect of individual anger on the rating, $b = -.46, SE = .22, p = .04$. There

was also a main effect of bystander emotion on the perceived appropriateness of those responses such that angry responses were seen as more appropriate than amused responses, $b = .24$, $SE = .12$, $p = .05$, but this was qualified by a significant interaction, $b = .25$, $SE = .12$, $p = .05$. Simple slopes revealed that bystander comments were judged to be more appropriate when the individual was told that they were angry, and they were also exposed to the anger of other bystanders, $b = .49$, $SE = .18$, $p = .006$, but not when the bystanders were told that they were personally amused, $b = -.007$, $SE = .18$, $p = .97$. This provides converging support for the findings of Study 2, although those effects did not (here) affect confrontation or collusion.

Table S8. Means and standard deviations for the behavioral measures, self-reported measures and supplementary measures

Outcome variables	Own anger/ bystander anger (<i>n</i> = 30)	Own amusement / bystander anger (<i>n</i> = 29)	Own amusement / bystander amusement (<i>n</i> = 30)	Own anger/ bystander amusement (<i>n</i> = 29)
Behavioral measures				
Star ratings	2.63 (1.82) ^a	3.32 (2.50) ^b	3.92 (2.50) ^a	2.96 (2.37) ^b
Occurrence of confrontational comment	36% (11) ^a	38% (11) ^a	28% (9) ^a	59% (17) ^a
Intensity of confrontational comment	.66 (1.04) ^a	.93 (1.31) ^a	.66 (1.11) ^a	1.04 (1.07) ^a
Occurrence of collaborative comment	39% (212) ^a	35% (10) ^a	53% (17) ^a	41% (12) ^a
Intensity of collaborative comment	.69 (1.00) ^a	.59 (.95) ^a	1.00 (1.10) ^a	.75 (1.00) ^a
Self-reported measures				

Enjoyment	2.89 (1.65) ^a	2.83 (1.88) ^a	3.53 (1.95) ^a	3.28 (2.04) ^a
Appropriateness of user comments	5.16 (1.34) ^a	4.86 (1.33) ^b	4.88 (1.39) ^c	4.18 (1.36) ^c
Supplementary measures				
Identification with other users	4.24 (.91) ^a	3.98 (1.25) ^a	4.22 (1.14) ^a	3.85 (1.16) ^a
Identification as anti- prejudice	4.98 (1.22) ^a	5.18 (1.46) ^a	5.24 (1.36) ^a	5.01 (1.50) ^a
Prejudice towards gay people	2.95 (1.22) ^a	2.29 (1.11) ^b	2.54 (1.05) ^b	2.84 (1.41) ^a
Social dominance orientation	2.19 (.77) ^a	2.04 (.79) ^a	2.19 (.90) ^a	2.61 (1.13) ^a
Importance of confronting	5.11 (1.03) ^a	5.25 (.90) ^a	5.40 (.77) ^a	5.25 (.75) ^a
Humor style	5.22 (.94) ^a	5.26 (.84) ^a	5.19 (.96) ^a	5.31 (.60) ^a
Political correctness	4.93 (.98) ^a	4.66 (.97) ^a	5.09 (.92) ^a	4.72 (.87) ^a

References

- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*, 349-364.
- Morrison, M. A., & Morrison, T. G. (2003). Development and validation of a scale measuring modern prejudice toward gay men and lesbian women. *Journal of Homosexuality*, *43*, 15-37.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, *67*, 741-763.
- Rasinski, H. M., Geers, A. L., & Czopp, A. M. (2013). "I guess what he said wasn't that bad" dissonance in nonconfronting targets of prejudice. *Personality and Social Psychology Bulletin*, *39*, 856-869.
- Thorson, J. A., & Powell, F. C. (1993). Development and validation of a multidimensional sense of humor scale. *Journal of Clinical Psychology*, *49*, 13-23.