



The University of Texas at Austin
Center for Media Engagement
Moody College of Communication

MODERATING UNCIVIL COMMENTS HURTS TRUST IN NEWS

Martin J. Riedl, Gina Masullo Chen, and Kelsey N. Whipple

SUMMARY

The Center for Media Engagement's previous research showed that uncivil comments can taint the perception of a news site. Now, we look at how those comments affect the moderators of a news site. This study found that uncivil comments:

- Hurt moderator trust in the news outlet.
- Made moderators more emotionally exhausted.

SUGGESTED CITATION:

Riedl, Martin J., Chen, Gina Masullo, and Whipple, Kelsey N. (2019, July). Moderating uncivil comments hurts trust in news. *Center for Media Engagement*. <https://mediaengagement.org/research/moderating-uncivil-comments>

THE PROBLEM

News organizations are increasingly turning to comment moderators – or their own journalists or other employees – to remove incivility that often mars online discussions. Our [previous research](#) looked at the effect of these negative comments on a news site. Now we turn our attention to the moderators. This project considered two main questions in the often-troublesome task¹ of comment moderation:

- Does moderating uncivil comments decrease how much moderators trust the news site?
- What emotional toll does moderating uncivil comments have on the moderators?

KEY FINDINGS

Moderators who focused exclusively on uncivil content:

- Perceived the news outlet they were working for as less trustworthy,²
- Were more emotionally exhausted,³
- Were less satisfied with the task of moderation,⁴
- Had a less immersive work experience.⁵ An immersive work experience is called a sense of “flow” in a task that makes it feel less effortful.⁶

SOLUTIONS FOR NEWSROOMS

Our results suggest that moderating only uncivil comments diminishes perceived trust in a news outlet. It also affects moderators on a very personal level, leading to emotional exhaustion and a less positive work experience. Considering most comment moderators work for much longer than the 24-minute shift used in this experiment, it’s particularly notable that moderating even for a short time had effects. The toll of moderating uncivil comments may be much stronger for moderators putting in several hours or a full day.



The toll of moderating uncivil comments may be much stronger for moderators putting in several hours or a full day.

News organizations should focus on keeping the tone in comments sections civil, so less moderation of uncivil comments is needed. News sites can do this by having journalists

engage in comments sections and by encouraging higher-quality conversations.⁷

Breaking up the task can make moderation less harmful for workers. They may have a better experience on the job if they moderate a mix of civil and uncivil comments, rather than focusing exclusively on uncivil content. We also suggest that comment moderators be given other tasks, along with comment moderation, to improve the experience.

THE EXPERIMENT

Participants were asked to moderate 78 comments each – a task that took them on average 24 minutes. They were told they were moderating the comments for an unidentified news site and were randomly assigned to moderate either all uncivil comments, all civil comments, or a mix of civil and uncivil comments. For each comment, participants decided if it should be accepted for publication on a news site or rejected based on how uncivil the comment was. Then they answered a series of questions about how much they trusted the news outlet, as well as how they felt about having to do this task. We found that moderating only uncivil comments led to negative effects.

Specifically, moderating only uncivil comments:

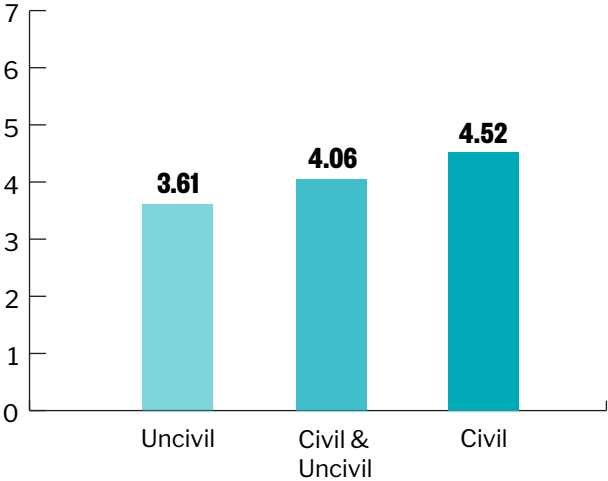
Diminished trust in the news outlet.

- People who moderated only uncivil comments viewed the news organization as less trustworthy⁸ than those who moderated only civil comments or a mix of civil and uncivil.

Took an emotional toll.

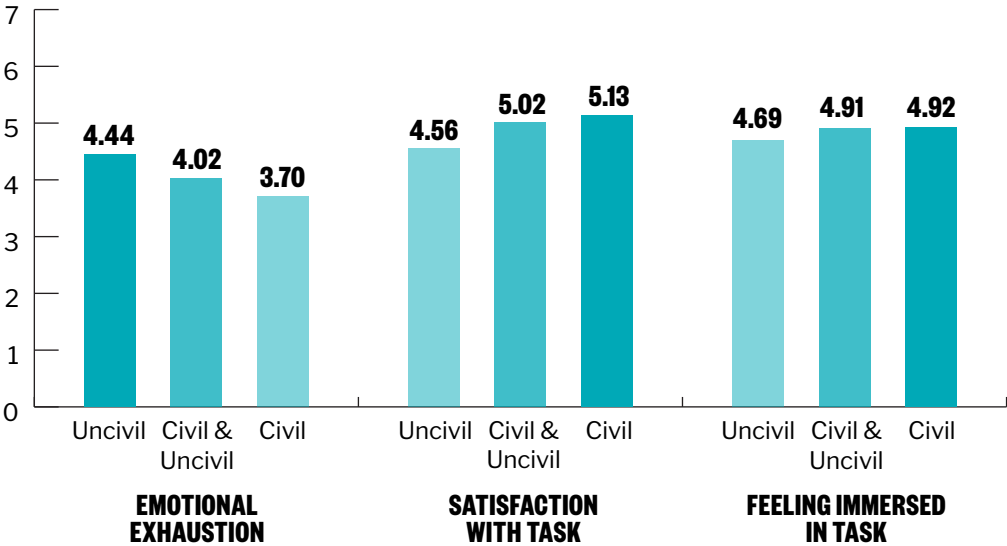
- People who moderated exclusively uncivil comments were more emotionally exhausted⁹ and less satisfied¹⁰ with the task than those who moderated only civil comments or a mix of civil and uncivil.
- People who moderated exclusively uncivil comments felt less immersed¹¹ in their work than those who moderated only civil comments or a mix of civil and uncivil.

TRUST IN NEWS



Notes: Average scores. Higher scores indicate more trust. Responses significantly different at the $p < .001$ level. 1 – 7 scale. Data from the Center for Media Engagement.

FEELINGS ABOUT TASK



Notes: Average scores. Higher scores indicate more exhaustion, satisfaction, or feeling of immersion. Responses between all uncivil and all civil and between mixed and all uncivil differ significantly at the $p < .05$ level. 1 – 7 scale. Data from the Center for Media Engagement.

METHODOLOGY

This experiment was embedded in a survey using Qualtrics' survey platform. The 747¹² participants were recruited via Amazon's Mechanical Turk. After answering demographic questions, participants were randomly assigned to moderate either all uncivil comments, all civil comments, or a mix of civil and uncivil comments for an unspecified news organization. Then they answered a series of questions about the news outlet and their experience moderating the comments.

Comments used in the experiment were real comments posted on *Breitbart*, *MSNBC*, *The New York Times*, and *Yahoo*. All comments were presented in the same style, so participants could not discern which platform comments came from. Uncivil comments contained profanity, insults, name-calling, and words in all capital letters,¹³ and civil comments lacked all such features. Before launching the experiment, 51 people who were not involved in the experiment rated whether they perceived the comments to be civil or uncivil.

PARTICIPANT DEMOGRAPHICS

	Experiment Participants <i>N</i> = 747	U.S. Adult Internet Population
Age		
18-29 years	33.5%	24.0%
30-49	54.2	37.0
50-64	10.3	26.0
>64	2.0	13.0
Race		
White	64.0	74.3
Black	12.0	9.4
Other	24.0	16.3
Gender		
Male	37.5	51.0
Female	62.5	49.0
Education		
High school or less	11.5	37.0
Some college	28.6	32.0
College +	59.8	31.0
Income		
< \$30,000	19.4	30.0
\$30,000-\$50,000	27.0	19.0
\$51,000-\$75,000	29.6	16.0
>\$75,000	24.0	35.0

Note: U.S. adult Internet population is based on data from Pew Research Center when data were collected in December 2018 and January 2019.

¹ Caplan, R. (2018). Content or context moderation? Artisanal, community-reliant, and industrial approaches. New York. Data & Society. Retrieved from https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf; Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press; Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. New Haven, CT: Yale University Press.

² The effect on trustworthiness of the news outlet was tested using an ANOVA that examined the main effect of what type of comments were being moderated (all uncivil, all civil, or a mix) and political beliefs, $F(2, 738) = 38.31$, $p < .001$, $\eta^2 = .09$, with Sidak post-hoc corrections. An interaction between type of comments being moderated and political beliefs was also tested, but it was not significant.

³ The effect on emotional exhaustion was tested using an ANOVA that examined the main effect of what type of comments were being moderated (all uncivil, all civil, or a mix) and political beliefs, $F(2, 738) = 11.40$, $p < .001$, $\eta^2 = .03$, with Sidak post-hoc corrections. An interaction between type of comments being moderated and political beliefs was also tested, but it was not significant.

⁴ The effect on task satisfaction was tested using an ANOVA that examined the main effect of what type of comments were being moderated (all uncivil, all civil, or a mix) and political beliefs, $F(2, 732) = 15.88$, $p < .001$, $\eta^2 = .04$, with Sidak post-hoc corrections. An interaction between type of comments being moderated and political beliefs was also tested, but it was not significant.

⁵ The effect on feeling immersed in the task was tested using an ANOVA that examined the main effect of what type of comments were being moderated (all uncivil, all civil, or a mix) and political beliefs, $F(2, 738) = 4.61$, $p = .01$, $\eta^2 = .01$, with Sidak post-hoc corrections. An interaction between type of comments being moderated and political beliefs was also tested, but it was not significant.

⁶ Bucher, E., & Fieseler, C. (2017). The flow of digital labor. *New Media & Society*, 19(11), 1868–1886; Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row; Csikszentmihalyi, M. (2014). *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*. New York: Springer.

⁷ Stroud, N. J., Scacco, J. M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' Facebook sites. *Journal of Computer-Mediated Communication*, 20(2), 188–203.

⁸ Participants rated trustworthiness using 10 items adapted from Gaziano, C., & McGrath, K. (1986). Measuring the concept of credibility. *Journalism Quarterly*, 63(3), 451–462; Johnson, T. J., & Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the internet influence credibility perceptions of weblogs among blog users. *Journalism and Mass Communication Quarterly*, 81(3), 622–642; Kioussis, S. (2003). Job approval and favorability: The impact of media attention to the Monica Lewinsky Scandal on public opinion of President Bill Clinton. *Mass Communication & Society*, 6(4), 435–451. Using a 1 (strongly disagree) to 7 (strongly agree) scale, participants rated how well 10 items described the news outlet they were moderating for with a higher number indicating more trust. The items were *transparent, informative, credible, accurate, unbiased, fair, can be trusted, reliable, does not have an agenda, and reputable*. Responses to the 10 items were averaged together.

⁹ To measure emotional exhaustion, people were asked to respond as if comment moderation were their day job. Using a 1 (*strongly disagree*) to 7 (*strongly agree*) scale, participants rated their level of agreement or disagreement on five statements, such as “I would feel emotionally drained from this work” and “I would feel frustrated by this job.” Responses were averaged together. Statements were adapted from Coates, D. D., & Howe, D. (2015). The design and development of staff wellbeing initiatives: Staff stressors, burnout and

emotional exhaustion at children and young people's mental health in Australia. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(6), 655–663; Wharton, A. S. (1993). The affective consequences of service work: Managing emotions on the job. *Work and Occupations*, 20(2), 205–232.

¹⁰ Task satisfaction reflected how satisfied people were with the task of moderating comments. Using a 1 (*strongly disagree*) to 7 (*strongly agree*) scale, they rated their agreement or disagreement on five statements, such as “In general, I like doing this task” and “I am satisfied with the important aspects of this task.” Responses were averaged together. Statements were adapted from Bowling, N. A., & Hammond, G. D. (2008). A meta-analytic examination of the construct validity of the Michigan Organizational Assessment Questionnaire Job Satisfaction Subscale. *Journal of Vocational Behavior*, 73(1), 63–77; Cammann, C., Fichman, M., Jenkins, G. D., & Klesh, J. R. (1983). Assessing the attitudes and perceptions of organizational members. In S. E. Seashore, E. E. Lawler, P. H. Mirvis, & C. Cammann (Eds.), *Assessing organizational change: A guide to methods, measures, practices* (pp. 71–138). New York: Wiley; Janssen, O. (2001). Fairness perceptions as a moderator in the curvilinear relationships between job demands, and job performance and job satisfaction. *The Academy of Management Journal*, 44(5), 1039–1050; Morris, M. G., & Venkatesh, V. (2010). Job characteristics and job satisfaction: Understanding the role of enterprise resource planning system implementation. *Management Information Systems Quarterly*, 34(1), 143–161.

¹¹ For the measurement of flow, which describes an immersive work experience, we used a scale adapted from Engeser, S. (2012). Appendix: Flow short scale. In *Advances in flow research* (pp. 201–202). New York: Springer; Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3), 158–172; Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept (Tests und Trends N.F. 2)* (pp. 261–279). Göttingen: Hogrefe. Using a 1 (*strongly disagree*) to 7 (*strongly agree*) scale, people rated their agreement or disagreement with 10 statements, such as “I am totally absorbed in what I am doing” and “I feel just the right amount of challenge.” Responses to the 10 statements were averaged together.

¹² A total of 818 people completed this experiment, but some responses were not used in our analysis. Data from people ($n = 53$) who took the survey more than once and who answered questions in a way that indicated they were not reading carefully ($n = 2$) were removed. The survey was constructed so that participants ($n = 16$) who tried to complete the survey using a mobile phone would be blocked from participating. This resulted in 747 participants.

¹³ Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679; Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. New York: Palgrave Macmillan.