Peer Community In Registered Reports

Benevolent correction may provide a promising antidote to online toxicity

*A recommendation by **Chris Chambers** based on peer reviews by **Marcel Martončik** and **Corina Logan** of the STAGE 2 REPORT:*

Alison I. Young Reusser, Kristian M. Veit, Elizabeth A. Gassin, and Jonathan P. Case (2023) Responding to Online Toxicity: Which Strategies Make Others Feel Freer to Contribute, Believe That Toxicity Will Decrease, and Believe that Justice Has Been Restored? OSF, ver. 2, peer-reviewed and recommended by Peer Community in Registered Reports.
https://osf.io/k46e8

Submitted: 03 August 2023, Recommended: 08 November 2023

Cite this recommendation as:
Chambers, C. (2023) Benevolent correction may provide a promising antidote to online toxicity. *Peer Community in Registered Reports*, 100537. 10.24072/pci.rr.100537

Published: 08 November 2023

Social media is a popular tool for online discussion and debate, bringing with it various forms of hostile interactions – from offensive remarks and insults, to harassment and threats of physical violence. The nature of such online toxicity has been well studied, but much remains to be understood regarding strategies to reduce it. Existing theory and evidence suggests that a range of responses – including those that emphasise prosociality and empathy – might be effective at mitigating online toxicity. But do such strategies work in practice? In the current study, Young Reusser et al (2023) tested the effectiveness of three types of responses to online toxicity – *benevolent correction* (including disagreement), *benevolent going along* (including joking/agreement) and *retaliation* (additional toxicity) – on how able participants feel to contribute to conversations, their belief that the toxicity would be reduced by the intervention, and their belief that justice had been restored. The results showed the benevolent correction – while an uncommon strategy in online communities – was most effective in helping participants feel freer to contribute to online discussions. Benevolent correction was also the preferred approach for discouraging toxicity and restoring justice. Overall, the findings suggest that responding to toxic commenters with empathy and understanding while (crucially) also correcting their toxicity may be an effective intervention for bystanders seeking to improve the health of online interaction. The authors note that future research should focus on whether benevolent correction actually discourages toxicity, which wasn't tested in the current experiment, and if so how the use of benevolent corrections might be encouraged. Following one round of review and revisions, the recommender judged that the manuscript met the Stage 2 criteria and awarded a positive recommendation. **URL to the preregistered Stage 1 protocol:** https://osf.io/hfjnb **Level of bias control achieved:** Level 6. *No part of the data or evidence that was used to answer the research*

question was generated until after IPA.
List of eligible PCI RR-friendly journals:

- Collabra Psychology
- F1000Research
- International Review of Social Psychology
- Peer Community Journal
- PeerJ
- Royal Society Open Science
- Studia Psychologica
- Swiss Psychology Open

References:

1. Young Reusser, A. I., Veit, K. M., Gassin, E. A., & Case, J. P. (2023). Responding to Online Toxicity: Which Strategies Make Others Feel Freer to Contribute, Believe That Toxicity Will Decrease, and Believe that Justice Has Been Restored? [Stage 2 Registered Report] Acceptance of Version 2 by Peer Community in Registered Reports. https://osf.io/k46e8

Reviews

Evaluation round #1

DOI or URL of the preprint: https://osf.io/hnds5?view_only=2b45b35cf37e46e5818a40bf79fc981d
Version of the preprint: 1

Authors' reply, 12 October 2023

Download author's reply

Decision by Chris Chambers, posted 08 September 2023, validated 11 September 2023

Revision invited

Your Stage 2 manuscript has now been evaluated by the two reviewers from Stage 1. Overall the assessments are positive and I foresee no major hurdles to achieving final Stage 2 acceptance. However, the comments do note a range of issues that need to be addressed to fully meet the Stage 2 criteria, including addressing points of clarity (including structure of certain parts of the manuscript) and interpretration of the findings. I hope you find the reviews constructive and look forward to receiving your revised manuscript in due course.

Reviewed by Corina Logan, 04 August 2023

Dear authors,

Congratulations on completing your study! The results are really interesting and will be useful for people when navigating online interactions. You did such a good job with your Stage 1 that it was very straightforward to review your Stage 2 using the track changes document. Thank you for such clarity!

My comments are as follows (using page numbers from the track changes .docx file)...

Abstract - "pre-registered experiment" and "Preregistered Stage 1 protocol" - a preregistration is different from a registered report, so please change the language to registered report.

Abstract - "benevolently correcting the toxicity was seen as the most helpful option for all three dependent measures" - the word "for" makes me think that the dependent measures were something other than benevolently correcting, etc. If so, please clarify what the dependent measures were. If not, you can replace "for" with "of".

Table 1 - you could add a column that states the results. If you do, perhaps it would be more useful to move the table to the results section. I think this would be useful because there are so many results and they are hard to keep track of, so having them in one place would be really handy.

p28 - your planned sample size was 1122 participants and you mention "several hundred additional people completed the study, resulting in a sample of 1360". Several hundred people implies that 500+ additional people participated. Did hundreds of people fall out of the sample or were only 238 additional people tested?

p32 - I'm finding it difficult to follow the reasoning starting with "Using Rosseel's (2012) lavaan R package in jamovi". I'm not clear why "measuring freedom to contribute and toxicity dissuaded" needed to be combined "into a single confirmatory factor" - doesn't that combine these initially separate dependent variables into one? If so, how can they be thought of as separate variables after this? What is Item 2 and what does it mean to have removed it from this analysis? Adding a summary sentence in this paragraph or the next could help clarify whether you found "evidence of unidimensionality" (is unidimensionality the same as strongly loading onto one factor?) and for which dependent variables (it looks like both of them?).

p35 First impression of toxic commenter - "While the initial comments did not differ between the two benevolent conditions (p = .70), those in the Retaliatory condition were rated as less toxic". Does this mean that it was the same set of initial comments (across both benevolent conditions and the retaliatory condition) that participants were reading, and that the participants in the retaliatory condition rated these same comments as less toxic? I'm having a hard time figuring out why this would be the case. Perhaps add an example here or provide a bit more clarification to help the reader follow?

p36 - "However, inconsistent with Hypothesis 1, participants did not feel less free to contribute in the benevolent going-along condition than the retaliatory condition; the opposite was found". The double negatives and reference to the opposite was confusing me and I had to go back to the study design table to figure out what the original prediction was. It would be easier if you rephrased to "However, inconsistent with Hypothesis 1, participants felt more free to contribute in the benevolent going-along than the retaliatory condition". Along the same lines, I would add ", rather than there being no difference between these conditions as we initially predicted" to the end of "Also inconsistent with Hypothesis 1, participants felt more free to contribute in the benevolent correction condition than in the retaliatory condition according to a post hoc comparison".

p38 - "According to a post hoc comparison, it was also significantly, though only slightly, lower than the retaliatory condition mean". Please add clarifications to this sentence similar to how I did above by explicitly reminding the reader of what the hypothesis states and how the results are consistent or not.

Figures - jittering the blue dots would allow the whole data set to be seen. As it is now, the blue dots cover the whole range of y-axis values and it is impossible to judge from this presentation where one would expect the mean to be.

p42 - "Note that these effect sizes were all small or less than small." - please explain what this means for interpreting the results...should they have less confidence in them until future studies can replicate the findings?

Also, did you have the power to detect differences in these small effect sizes? If so, then it shouldn't matter that they are so small and you can add a note about this here. I tried to figure this out from the manuscript, but it is difficult. In the Results > Free to Contribute section where the results for Hypothesis 1 are, there is only 1 F-statistic and it is 17.84. The other two analyses for Hypothesis 1 don't have F-statistics so how are you determining whether you were able to detect differences? It would be good to explicitly state throughout your Results sections whether you were able to detect each result given the sampling plan detection levels in Table 1.

p42 - "Replies which benevolently corrected the toxicity resulted in a mean near the scale midpoint" - please clarify what variable the "mean" was of (i.e., perception that toxicity had been dissuaded).

p44 - "Benevolent Corrections as Injunctive Norms" - Something to consider adding to the discussion... The fact that the benevolent correcting strategy has been shown to be more effective, but also used less indicates that there must be some external pressure causing the reduction in the use of this strategy. One reason people might not use this strategy is that they aren't aware that it is more effective. So a lack of awareness of what works, along with some external pressure could also explain why this strategy isn't used as often.

In the PCI RR questionnaire, you state that you provided the analysis code and pointed readers to this URL: https://osf.io/6vkqz/?view_only=2b45b35cf37e46e5818a40bf79fc981d. I went to the URL, but couldn't find any code. Please add the code to the OSF repo and provide a direct link to this document so people don't have to click through every file to try to find it.

All my best,
Corina Logan

Reviewed by Marcel Martončik, 15 August 2023

Download the review