

Procedural Justice and Self Governance on Twitter: Unpacking the Experience of Rule Breaking on Twitter

Matthew Katsaros, Tom Tyler, Jisu Kim, and Tracey Meares

“ *This study demonstrates empirically that there is an association between online platform user’s experiences with a content moderation action and their later rule violating behavior. If users experience the moderation action as involving more just procedures, they are more likely to adhere to platform rules in the future.* ”

✧ THE PROBLEM:

Deterrence—punishing and sanctioning “bad actors” who violate rules—is the dominant method for managing unwanted content and behaviors on many online platforms. Could an alternative approach be used to reduce recidivism?

✧ THE QUESTION:

Would a procedural justice approach—one which focuses on the process and the fairness through which content moderation decisions are made and communicated—be an effective approach to managing these problems?

✧ THE STUDY:

In collaboration with Twitter, a survey was sent to (n=10,487) users who had recently violated one of Twitter’s rules to try and understand:

- 1 - What led to the rule breaking incident?
- 2 - What was the relationship between people’s perceptions of Twitter’s enforcement process and their likelihood to violate Twitter’s rules in the future?

✧ THE FINDINGS:

1 - People break rules for many different reasons—from not knowing a platform has rules (~10% of those surveyed were not aware of Twitter rules) to impulsivity or self-defense. These findings present a more nuanced depiction of online harassment and challenge the idea that most of this behavior is driven by a few trolls and “bad actors”.

2 - There was strong support for a procedural justice approach as those who felt more fairly treated during the enforcement process were less likely to break Twitter’s rules in the future.

✧ THE CONCLUSION:

Taking these findings together, it is clear that there is an opportunity for platforms to put a stronger emphasis on the process by which rules are communicated and enforced as a potential avenue to build legitimacy and reduce rule breaking behavior on a platform. The data from this study shows support for a procedural justice approach which focuses on providing voice, transparency, and dignity to those that break rules. Furthermore, a deeper investment in better understanding the pathways and motivations for violating rules can be leveraged to build more upstream interventions or education for users.

[Read the full article in The Journal of Online Trust and Safety](#) ►