Moral Judgments and Punishment Decisions on Social Media

Sarah Vahed[1], Catalina Goanta[2], Pietro Ortolani[3] & Alan Sanfey[1]

1. Faculty of Social Sciences, Radboud University, Nijmegen, The Netherlands 2. Faculty of Law, Utrecht University, Utrecht, The Netherlands 3. Faculty of Law, Radboud University, Nijmegen, The Netherlands

DONDERS INSTITUTE

#INTRODUCTION

The spread of harmful and inappropriate social media content is a pertinent issue necessitating the need to understand how users respond to different types of online content and how they wish to mitigate the spread of harm on social media.

Empirical studies have consistently identified a bystander's perceived responsibility as a key determinant of action versus inaction against harm.[1]

It has also been found that situational context such as intention of an actor, can affect moral judgments.[2]

In line with developments in EU law on content governance, we designed an online study to investigate the impact of perceived responsibility, poster intention and moral image category on decisions to report content and preferences to punish other social media users.

RESEARCH QUESTIONS

- What **content** do social media users want to report and punish on social media?
- What is the role of a poster's intention on users' reporting and punishment decisions?
- How does a user's sense of perceived responsibility impact their decisions to moderate content?

#METHODS

N = 294 social media users residing across EU Member States (57.82% Male, M_{age}=28.33, SD_{age}=8.42), recruited on Prolific

Judgments and decisions captured in two online tasks
Task 1: Report DV: 'Like', 'Dislike', 'Report'
Task 2: Punish DV: Assign Length of Ban: 0 - 30+ days

Between-Subjects:

2 Groups
No Adjudication: Instructed to respond as they would usually on social media

Adjudication: Assigned the role of 'user content moderator' with responsibility of identifying inappropriate online content

Within-Subjects:

3 Image Categories
Morally negative, neutral and positive images from *Socio-Moral Image Database*[3]

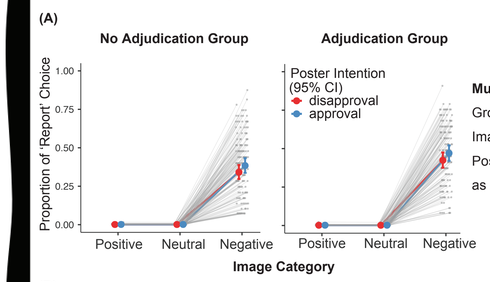
2 Poster Intention
Ostensible poster approval or disapproval

Comments and feedback are welcome! Please contact:
sarah.vahed@donders.ru.nl

RESULTS

TASK 1: REPORT

Perceived Responsibility, Image Category and Poster Intention each impact decisions to report on social media



Multilevel logistic regression found: Group (No Adjudication - Adjudication; ***p=0.01***), Image (Positive - Negative; ***p<.001***) and Poster Intention (approval - disapproval; ***p<.001***) as significant factors in decisions to Report.

Figure 1: (A) Proportion of decisions to 'Report' across Images and Poster Intention, between Groups. Error bars represent 95% Confidence Intervals; (B) Model estimates expressed as odds ratio for each of the model parameters. Significance ***p<.001, *p<.05.

TASK 2: PUNISH

Image Category and Poster Intention impact punishment decisions on social media (but Perceived Responsibility does not)

Multilevel linear regression (with Poission distribution) found Image (Positive - Negative; ***p<.001***) and Poster Intention (approval - disapproval; ***p<.001***) as significant factors in punishment preferences.

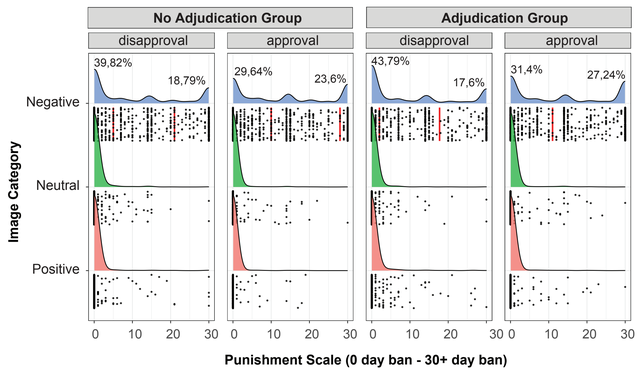


Figure 2: Distribution of Punishment Amount between Groups, and across Images and Poster Intention

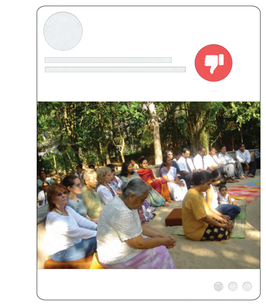
KEY TAKEAWAYS

Our research suggests that user judgments about what should and should not be on social media are complex. Various factors play a role in their online decisions. These factors should be borne in minde by policymakers developing content moderation policies and strategies.

#EXAMPLE STIMULI

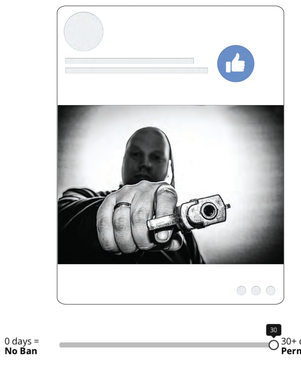
TASK 1: REPORT

Morally neutral image shared with poster disapproval



TASK 2: PUNISH

Morally negative image shared with poster approval



#CONCLUSIONS

- Together our findings suggest that the **perceived responsibility** of users, the **images** they view as well as the **context** behind which images are shared online are significant considerations in the moderation of content by users.
- Future behavioural studies which seek to impact law and policymaking can benefit from **interdisciplinary collaborations** between researchers in law and psychology.

#REFERENCES

1. Butler, L. C., Graham, A., Fisher, B. S., Henson, B., & Reyns, B. W. (2022). Examining the effect of perceived responsibility on online bystander intervention, target hardening, and inaction. Journal of interpersonal violence, 37(21-22), NP20847-NP20872.
2. Li, J., Hou, W., Zhu, L., & Tomasello, M. (2020). The development of intent-based moral judgment and moral behavior in the context of indirect reciprocity: A cross-cultural study. International Journal of Behavioral Development, 44(6), 525-533.
3. Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. PloS one, 13(1), e0190954.

RADBOUD CENTER FOR DECISION SCIENCE

Radboud University

Radboudumc