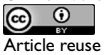
Article

Social Psychological and Personality Science
1–10



DOI: 10.1177/19485506251335373
journals.sagepub.com/home/spp


Moral Outrage Predicts the Virality of Petitions for Change on Social Media, But Not the Number of Signatures They Receive

Stefan Leach[1,2], Magdalena Formanowicz[3], Jan Nikadon[3,4], and Aleksandra Cichocka[1]

Abstract

Social media is a powerful tool for activists to share their perspectives, but concerns persist that the viral spread of online moral outrage may undermine collective causes in some ways. Analyzing posts on X (n = 1,286,442) with URLs to petitions on www.change.org (n = 24,785), we found that expressions of outrage were uniquely associated with the number of times posts were liked and reposted (virality). Mediation analyses showed that outrage was indirectly related to the number of signatures petitions received (via virality). However, outrage was associated with fewer signatures when controlling for virality. In contrast, expressions of agency, group identity, and prosociality were associated with more signatures but no more virality. The findings outline the factors linked to engagement with online petitions and describe how social media can amplify content which has no direct link to the sorts of effortful behaviors typically thought to be conducive to social change.

Keywords

online activism, clicktivism, social media, moral outrage

Introduction

Online spaces have changed what it means to engage with collective causes. By connecting billions of individuals across the globe, social media platforms amplify and propagate content in ways that were previously unimaginable. The accessibility and immediacy of online content allows activists, grassroots organizations, and individuals with limited resources or traditional means of influence to mobilize others. The democratization of information and the ability to rally support through hashtags, viral content, and online campaigns has fueled many recent movements, such as #MeToo and #BlackLivesMatter, allowing them to transcend geographical boundaries and muster unprecedented support (Lotan et al., 2011; Mendes et al., 2018).

Despite the significance of online spaces for recent social movements, there is a sense that they might limit engagement in some ways by encouraging immediate acts, such as sharing content to raise awareness, over more effortful ones, such as donating or signing petitions (Chayinska et al., 2021; Schumann & Klein, 2015). This has been referred to as the problem of "clicktivism" (Bennett & Segerberg, 2013; Morozov, 2009).

Moral outrage has become central to debates about the negative effects of social media. Many view it as performative and are concerned that the tendency for social media platforms to amplify moral outrage is doing more harm than good (Bloom & Jordan, 2018; Brady & Crockett, 2019; Crockett, 2017; Rathje et al., 2024; Van Bavel et al., 2024). An abundance of online outrage is thought to exacerbate problems surrounding "clicktivism" by shifting priorities to immediate forms of engagement in favor of more effortful and potentially impactful ones (Crockett, 2017). In this way, moral outrage may generate a lot of noise surrounding collective causes, but ultimately fail to translate into further action. This can be contrasted with other factors (e.g., appeals to common identities) which may draw less attention on social media, but ultimately be more conducive to further engagement. The present work seeks to contribute to these debates by exploring engagement with online petitions for social

[1]University of Kent, UK
[2]Lancaster University, UK
[3]SWPS University, Warsaw, Poland
[4]Nicolaus Copernicus University, Toruń, Poland

Corresponding Author:
Stefan Leach, Department of Psychology, Lancaster University, Lancaster LA14YF, UK.
Email: s.leach1@lancaster.ac.uk

change, considering both immediate actions (sharing) and more effortful ones (signing).

How Is Moral Outrage Related to Online Collective Causes?

Moral outrage is an emotional reaction to perceived norm violations involving strong feelings of anger and disgust (Brady et al., 2023). It is a remarkable indicator of online virality (Brady et al., 2021). This makes sense given that both moral and negative content grabs attention and receives more engagement online (Baumeister et al., 2001; Berger & Milkman, 2012; Brady et al., 2017; Gantman & Van Bavel, 2014). Brady et al. (2017) found that each additional negative moral-emotional word in social media posts about contentious social issues (e.g., gun control) was associated with 24% more reposts. This is partly why some believe that expressions of outrage can benefit online collective causes by helping content to go viral and increase awareness of relevant issues (see Spring et al., 2018).

A more persuasive defense of online moral outrage could be made if it could be linked to further effortful behaviors in service of collective causes (over and above sharing). A wealth of psychological theorizing suggests that negative emotional reactions to injustice mobilize people to engage in collective action (Drury & Reicher, 2009; Klandermans, 1997; van Zomeren et al., 2008; Wright, 2010). The social identity model of collective action (SIMCA) posits that anger about group injustices is built on a sense of identification with the group (as are beliefs about the group's ability to achieve positive change; van Zomeren et al., 2008). Numerous observations have revealed the links between offline collective action, such as participating in protests, and feelings of righteous anger, injustice, and outrage (Miller et al., 2009; Thomas et al., 2009). For example, one study found that anger at injustice positively predicted support for policies and behaviors aimed at wealth redistribution (Wakslak et al., 2007). The amount of moral language used on social media can be predictive of protests (Mooijman et al., 2018). There is also evidence that online engagement (often associated with moral outrage) is linked to more effortful action (Foster et al., 2021; Smith et al., 2023; cf. Chayinska et al., 2021; Schumann & Klein, 2015). These works suggest that moral outrage should also track more effortful collective action.

However, the expression of moral outrage can also be performative (Brady et al., 2023; Fincher & Tetlock, 2016; Jordan & Rand, 2020). Studies using offline dictator games find that expressing outrage paradoxically reduces the need to engage in more effortful action (Xiao & Houser, 2005). Whether expressions of moral outrage function in a similar way in the online contexts is therefore unclear. It remains an open question as to whether expressions of moral outrage only predict the virality of online causes, or whether this extends to further forms of more effortful engagement.

Additional Indicators of Online Engagement and Action

There are other potential factors associated with online collective action that are not subject to the same concerns as moral outrage. Classic psychological theorizing suggests that people engage in collective action when they feel that goals are achievable—that is, they feel a sense of efficacy (Drury & Reicher, 2009; Klandermans, 1997; Thomas et al., 2009; van Zomeren et al., 2008; Wright, 2010). In one study, Israelis were more likely to state that they would sign a petition against unreasonably high estate prices in Tel Aviv if they also believed that their actions would have a significant effect (van Zomeren et al., 2013). Calls-to-action often attempt to instill a sense of efficacy via the use of agentic language. We define agentic language as linguistic expressions that convey goal-oriented or action-oriented content (Formanowicz et al., 2024). Messages like "we act" clearly communicate a need for action, while messages such as "Vote leave" provide a concrete way for an action to manifest. Such messages are perceived to be effective campaign slogans (Formanowicz et al., 2021, 2024) and can prompt more charitable donations (Whillans & Dunn, 2018; see also Suitner et al., 2023). Agentic language is therefore likely to track engagement with online collective action.

People are likely to engage with collective causes if they identify with them. Individual helping behavior is shaped by identification with those in need (Levine et al., 2005) and collective behavior is much the same (Jost et al., 2018; Simon & Klandermans, 2001; Thomas et al., 2009; van Zomeren et al., 2008). One study found that Dutch farmers were more likely to participate in protests if they identified with other farmers (Klandermans, 2002). The use of words like "we" and "us" are relevant here as they denote a common group identity. They are often used in online calls-to-action (Suitner et al., 2023) and actual protests (Adam-Troian et al., 2021). There is also some evidence that using such language to cue a common sense of identity may be an effective method through which one can muster support for one's cause. For example, politicians' who rely more on words like "we" and "us" in their speeches are more likely to be elected compared with those who rely less on such language (Steffens & Haslam, 2013). Thus, references to common group identities may be linked to engagement with online collective action.

Warmth, kindness, and care are core aspects of prosociality which can foster group-oriented behavior, as they denote commitments to helping others and prioritizing collective well-being (Fiske et al., 2007; Penner et al., 2005). Prosociality is central to many conceptions of collective action. Definitions typically emphasize transcending self-interest to achieve outcomes that benefit the group or society at large. For example, van Zomeren and Iyer (2009) present a definition of collective action as: any action which aims to improve the conditions of a group rather than any specific individual. Indeed, many collective movements are

characterized by prosocial intentions. This is evident in solidarity-based collective actions, such as advantaged group members advocating for the rights of the disadvantaged, driven by shared moral convictions and a sense of responsibility (Becker, 2012; Iyer & Ryan, 2009; van Zomeren et al., 2011). Valuing the welfare of others is predictive of individual participation in movements—such as gender equality initiatives and pandemic-related safety behaviors (Boggio et al., 2024; Milesi & Alberici, 2018). Consistent with this, words denoting prosociality, such as "empathy" and "altruism," track the amount of engagement with collective online endeavors to crowdfund novel projects (Pietraszkiewicz et al., 2017). Thus, expressions of prosociality may be a key indicator of engagement with online collective action.

Overall, past work suggests that expressions of agency, group identification, and prosociality are likely to be important indicators of engagement with online collective action. This work documents the links between these factors and effortful actions, such as monetary contributions (Pietraszkiewicz et al., 2017; Whillans & Dunn, 2018). However, it is less clear whether they are associated with immediate forms of online engagement (such as sharing content). There is some evidence from offline experiments that people prefer to share social and positive-moral content with others (Mesoudi et al., 2006; Stubbersfield et al., 2019). In online discussions of contentious political issues, positive moral-emotional language is a minor indicator of engagement (Brady et al., 2017). However, it is dwarfed by moral outrage and negative moral-emotional language (Brady et al., 2017, 2021). Likewise, a broad take-away from the wider literature is that positive and low-arousal content is at a disadvantage in online spaces (Berger & Milkman, 2012; Brady et al., 2017, 2021; Rathje et al., 2024). This work suggests that agency, group identities, and prosociality may not be particularly strong indicators of immediate forms of online engagement with collective causes.

Present Work

We investigated collective action in the context of social media use by analyzing posts to X which contain URLs to petitions for social change. We focused on liking and reposting as modes of engagement indicative of endorsement and information diffusion in social networks (Berger & Milkman, 2012; Brady et al., 2017). We also examined online petition signing as a further form of more effortful engagement which is typically thought to be conducive to systemic change (van Zomeren et al., 2008). Examining liking and sharing in conjunction with petition signing provides insight into different forms of online behavior. They differ along two dimensions which characterize actions of clicktivism: immediacy and effort (Halupka, 2014; Skoric, 2012). We consider liking and reposting more immediate forms of action which require minimal effort (achieved on the present platform). We consider signing as a more distal form of action which requires more effort (achieved on a secondary platform).

We sought to address two primary questions: (1) which factors track engagement with online collective causes, and (2) whether some factors are uniquely associated with immediate forms of engagement (liking and reposting) compared with more distal and effortful ones (petition signing). That negative moral sentiments track virality (Berger & Milkman, 2012; Brady et al., 2017) led us to hypothesize that moral outrage would be associated with liking and reposting of posts containing petitions for social change. Making predictions about the relationship between moral outrage and petition signatures was less straightforward. On one hand, collective action is driven by feelings of anger at injustices (van Zomeren et al., 2008; Wakslak et al., 2007) and immediate forms of online engagement are often predictive of more effortful actions (Foster et al., 2021; Smith et al., 2023), which led us to expect that expressions of moral outrage would be associated with more petition signatures. On the other hand, work which views expressing moral outrage as substitutive of other forms of action (Brady & Crockett, 2019; Crockett, 2017) suggests that it may paradoxically be associated with fewer petition signatures.

Our thinking about the effects of agency, group identification, and prosociality was more straightforward. There is good reason to think that these factors should track engagement with typical forms of collective actions, such as petition signing, as they tend to do so in offline contexts (Boggio et al., 2024; Milesi & Alberici, 2018; van Zomeren et al., 2008). As such, we predicted that they would be associated with more petition signatures. However, the evidence for their potential associations with immediate forms of online engagement is more limited. High-arousal and negative sentiments are the strongest indicators of online engagement (Berger & Milkman, 2012; Brady et al., 2017). Expressions of agency, group identification, and prosociality are not of this sort. For this reason, we were agnostic about their associations with liking and reposting.

Method

The work received approval from the School of Psychology Ethics Board at the University of Kent (ID: 202316941706278627). The data and analysis script are available via the Open Science Framework (https://osf.io/2kmtp/). The work was not pre-registered.

To investigate online collective action, we compiled a large set of active petitions from www.change.org which had recently been shared on X (formerly Twitter). Change.org allows users to create and sign petitions to advance various social causes and influence decision-makers. Such endeavors are considered by many to be legitimate and democratic avenues for groups to voice grievances in hopes of social change (Bochel, 2013). For example, the U.S.

government previously committed to responding to petitions with sufficient support (Goldman, 2015). It actively encourages, and relies on, the use of social media to raise awareness of its petitions and increase support for them. All petitions include a direct link to "Tweet to your followers." As of 2023, the site reports to have over 500 million users from across the globe and petitions targeting causes such as social justice, human rights, education, environmental protection, animal rights, health, and sustainable food. X is a popular online social media and social networking service with over 300 million users, posting from primarily public accounts (Takhteyev et al., 2012). Users are able to post messages which can be engaged with and propagated to others by liking and reposting them.

We began by searching the X API (at the time, the Twitter academic API) for all English-language posts containing any URL to change.org or its subsites (e.g., change.org/protect-our-nhs/) in the last 4 months (January 1, 2023–June 1, 2023). This search was exhaustive and not guided by a sample target. It resulted in 24,785 unique petition URLs (derived from 254,213 posts). We scraped the number of signatures each petition received from these URLs.[1] Many of these petitions were created prior to our initial earliest search date: 23.8% in 2023, 51.1% in 2022, 6.5% in 2021, 5.2% in 2020, and 13.4% in or before 2019. As such, we then conducted a full historical search of the Twitter API, reaching back to the earliest available timepoint (March 2006), to compile all posts linking these 24,785 petitions. It resulted in 1,286,442 posts linking to the aforementioned petitions. The final dataset included each petition's description and the number of signatures it received, as well as each post's text content and the number of times they were liked and reposted.

We measured moral outrage using the Digital Outrage Classifier, a machine learning algorithm which can detect moral outrage in English-language posts in a manner consistent with trained human annotators (Brady et al., 2021). We measured agentic language via BERTAgent, a validated English-language transformer-based model for quantifying semantic agency in text (Nikadon et al., 2024). Following prior work, we used validated lists of words (Boyd & Schwartz, 2021; Pennebaker et al., 2003) to quantify the presence of group identification and prosociality. Group identity strength was indexed via the presence of words such as "we" and "us" minus words such as "sure" and "know" (Ashokkumar & Pennebaker, 2022) and prosociality was indexed via the presence of words such as "selfless" and "generous" (Pietraszkiewicz et al., 2017). Group identification and prosociality scores for each text were computed by dividing the number of relevant words by the total number of words in the post (Dehghani & Boyd, 2022). Each post received a score along all of these four dimensions (Further details on the methods are provided in the Supplementary Information). Through this approach, we arrived at a dataset which could reveal the factors associated with online collective action.

Results

Analytic Approach

We aggregated all indices within petitions to produce a set of scores reflecting the language used in posts which shared each petition, how often they were liked and shared, and how many signatures they received. This is the recommended approach when predicting level two variables (signatures) from Level 1 variables (post language, likes, reposts), as it controls for an otherwise inflated Type I error rate (Foster-Johnson & Kromrey, 2018). Analyses of the disaggregated data are almost identical (see Supplementary Information).

Following prior work (Brady et al., 2019; Rathje et al., 2021), we log-transformed the like, repost, and signature counts to better approximate a Gaussian distribution. Our main analyses were conducted by fitting a series of univariate ordinary least squares regression models to estimate the basic relationships between each type of language (moral outrage, agency, group identity, and prosociality) used in the posts and each metric of engagement (likes, reposts, and signatures). Analyses were conducted in R (version 4.0.1) via the *glm* function while specifying the Gaussian family of models. All variables were mean-centered and standardized using the *jtools* package (Long, 2024). Following prior work (Brady et al., 2019; Rathje et al., 2021), coefficients derived from these models were exponentiated so as to provide a more interpretable measure of the magnitude of the effects.

Main Analyses

We began by considering moral outrage. As can be seen in Figure 1, expressions of moral outrage were linked to immediate forms of engagement with collective causes. Petitions shared with more moral outrage received more likes and reposts compared with those shared with less, $\exp(\beta_{Likes}) = 1.07$, 95% confidence interval (CI) = [1.05, 1.10], $p < .001$; $\exp(\beta_{Reposts}) = 1.10$, 95% CI = [1.07, 1.12], $p < .001$. However, similar effects were not present for more effortful actions. Petitions shared with more moral outrage received no more signatures than those shared with less, $\exp(\beta_{Signatures}) = 1.00$, 95% CI = [0.97, 1.04], $p = .810$. Moral outrage is therefore linked to immediate forms of engagement, but less so with the sorts of effortful behaviors thought to be necessary for normative systemic change.

The pattern of results for agency, group identity, and prosociality were the opposite of moral outrage (see Figure 1). They were largely unrelated to immediate forms of engagement. Petitions shared with more agentic language received no more likes and slightly fewer shares compared with those shared with less agentic language, $\exp(\beta_{Likes}) = 0.99$, 95% CI = [0.97, 1.01], $p = .362$; $\exp(\beta_{Reposts}) = 0.97$, 95% CI = [0.95, 0.99], $p < .001$. Likewise, those shared with more expressions of group identity received no more likes and slightly more shares

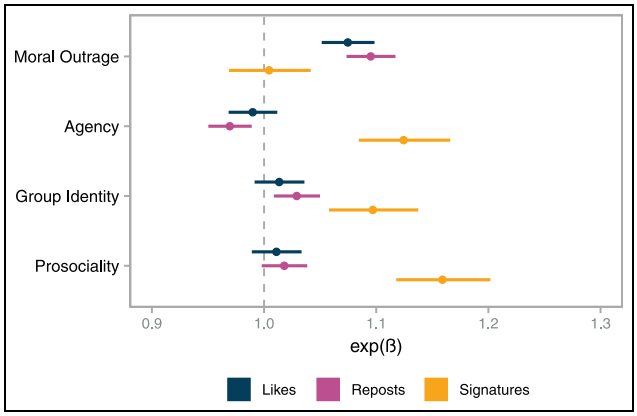


Figure 1. *Predictors of Online Petition Success*
Note. Coefficients are exponentiated. Scores above 1 indicate a relative increase, while scores below 1 indicate a relative decrease. Figure depicts coefficents (circles) and 95% CIs (whiskers).

compared with those shared with fewer expressions of group identity, $\exp(\beta_{Likes}) = 1.01$, 95% CI = [0.99, 1.04], $p = .232$; $\exp(\beta_{Reposts}) = 1.03$, 95% CI = [1.01, 1.05], $p < .001$. Expressions of prosociality were also largely unrelated to immediate forms of engagement with collective causes. Petitions shared with more prosocial language received no more likes and reposts compared with those shared with less prosocial language, $\exp(\beta_{Likes}) = 1.01$, 95% CI = [0.99, 1.03], $p = .332$; $\exp(\beta_{Reposts}) = 1.02$, 95% CI = [1.00, 1.04], $p = .084$.

The associations with agency, group identity, and prosociality were more noticeable in relation to petition signatures. Petitions advertised with more agentic language received more signatures than those with less, $\exp(\beta_{Signatures}) = 1.12$, 95% CI = [1.08, 1.17], $p < .001$. Petitions posted with more identity language garnered more signatures than those posted with less identity language, $\exp(\beta_{Signatures}) = 1.10$, 95% CI = [1.06, 1.14], $p < .001$, while those posted with more prosocial language received more signatures than those posted with less, $\exp(\beta_{Signatures}) = 1.16$, 95% CI = [1.12, 1.20], $p < .001$. These findings suggest that agentic, group identity, and prosocial language are associated with effortful online action in support of collective causes. At the same time though, posts with such content seem no more likely to propagate through social networks.

Exploratory Analyses

We sought to further explore how posts were shared on social media and the number of signatures they received. Given the emphasis change.org places on social media as a tool for raising awareness of its petitions, one would expect that those petitions that are shared more on social media would garner more signatures. We find this to be true, which suggests that virality is an indirect way to garner signatures, $\exp(\beta_{Reposts}) = 1.54$, $p < .001$. This is interesting in the case of moral outrage, as it could represent an important pathway through which it can foster more effortful behaviors in service of collective causes. This understanding is consistent with a statistical model in which the content of posts sharing petitions is indirectly associated with more signatures via the propensity to go viral. We fit such a mediation model via the *lavaan* package (Rosseel, 2012).

As can be seen in Figure 2, this analysis revealed a positive indirect effect of moral outrage through virality on the likelihood that petitions received signatures, $\exp(\beta_{Indirect}) = 1.03$, $p < .001$. After accounting for virality, the direct effect of moral outrage predicted fewer signatures, $\exp(\beta_{Direct}) = 0.98$, $p < .001$. These results show that, despite moral outrage having no direct association with effortful forms of engagement with collective causes (i.e., signatures), it may nevertheless be indirectly associated with more signatures, by virtue of its positive relationship to immediate forms of engagement (i.e., reposts).

We also analyzed the indirect effects of agency, group identity, and prosociality. As indicated by Figure 2, there was a small negative indirect effect of agency, $\exp(\beta_{Indirect}) = 0.99$, $p = .002$. Its tendency to reduce reposts had a detrimental effect on the number of signatures petitions received. Group identities had the opposite effect. Its minor capacity to increase reposts followed through to more signatures, $\exp(\beta_{Indirect}) = 1.01$, $p = .001$. Finally, there was no strong evidence of an indirect effect of prosociality, $\exp(\beta_{Indirect}) = 1.01$, $p = .057$, which is consistent with it having little capacity to increase reposts.

Robustness Checks

We sought to test the robustness of the observed effects. The effects of moral outrage, agency, group identity, and prosociality on virality and signatures were largely unchanged when controlling for the content of the petitions themselves. They were also robust to controlling for any potential covariation between the different types of language (see Supplementary Information). These results made us more confident that the central findings reflect how petitions are shared on social media, as opposed to other more general variations in the content of the petitions or covariation between different language types.

Discussion

Moral Outrage Is Associated With Virality But Not Petition-Signing

Motivated by concerns about the tendency for social media to propagate content which may undermine collective action (Brady & Crockett, 2019; Crockett, 2017; Van Bavel et al., 2024), we investigated the factors linked to engagement with petitions for social change. We found that moral

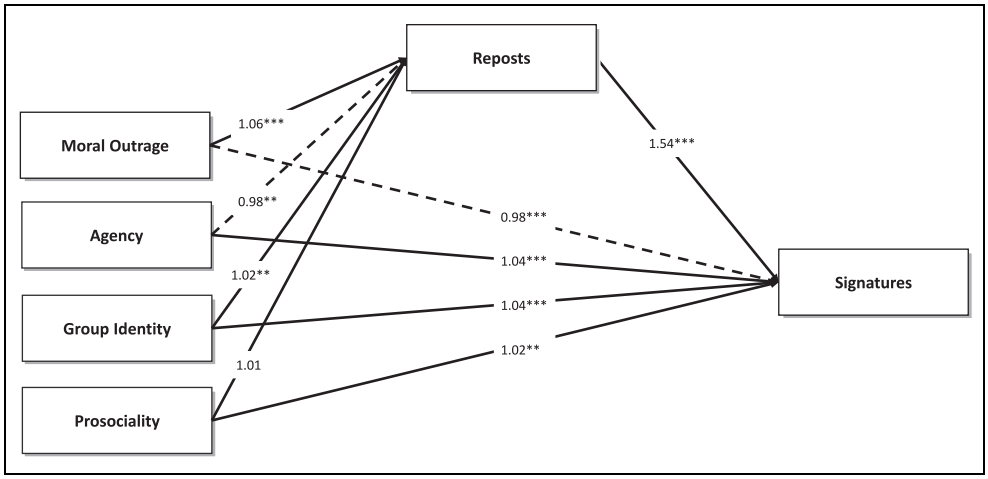


Figure 2. *Direct Effects of Petition Content on Signatures and Via Reposts (Virality)*
*Note.*** $p < .010$, *** $p < .001$. Coefficients are exponentiated. Scores above 1 indicate a relative increase in the number of reposts and signatures (solid lines), while scores below 1 indicate a relative decrease (dashed lines).

outrage was uniquely associated with the amount petitions were shared on social media—similar to how it tracks virality in other contexts (Brady et al., 2021). This helps explain why petitions eliciting strong moral sentiments garner so much attention. For instance, the death of Cecil the Lion in 2015 sparked a number of viral petitions calling for hunting reforms and punishment of the man who killed him, while the petition demanding justice for George Floyd is, as of writing, the most widely shared on the entire change.org platform (Kellen, 2020).

However, moral outrage was largely unrelated to the number of signatures petitions received, and when taking into account its capacity to go viral, it was associated with *fewer* signatures. This is consistent with a "clicktivism" effect, where immediate forms of online engagement, such as liking or sharing posts, replace more effortful actions (Greijdanus et al., 2020; Morozov, 2009). Although the present data are correlational and cannot corroborate causal claims, this is nevertheless the pattern of results one would expect if online expressions of outrage have a performative element that is supplanting the need for further action (Brady & Crockett, 2019; Crockett, 2017).

The findings further outline a number of key associations with online engagement with collective causes. Expressions of agency, group identity, and prosociality were all associated with more petition signatures. These findings align with prior observation (Formanowicz et al., 2021; Pietraszkiewicz et al., 2017) and theorizing (Drury & Reicher, 2009; Klandermans, 1997; van Zomeren et al., 2008; Wright, 2010). In addition, they support the broader applicability of the SIMCA model (van Zomeren et al., 2008) by demonstrating its effectiveness in predicting online dynamics of language use and engagement with collective causes. At the same time, expressions of agency, group identity, and prosociality were associated with no more online engagement. This highlights nuances in how collective behavior plays out in online spaces and the need to integrate psychological theorizing with understandings of engagement-maximizing algorithms (Brady et al., 2020).

Caveats and Constraints on Generalizability

The sample of X posts containing URLs to petitions on change.org ($n_{\text{Petitions}}$ = 24,785, n_{Posts} = 1,286,442) provides a comprehensive outline of the factors associated with engagement with collective endeavors for social change. However, they are constrained to a single social media platform and so may not generalize to others. X users tend to be younger and more liberal compared with, for example, Facebook users (Shearer & Mitchel, 2021). Moreover, there is evidence that engagement with misinformation (Vosoughi et al., 2018; cf. Bond & Garrett, 2023) and expressions of outgroup animosity (Rathje et al., 2021) vary across platforms. Whether the factors we focus on here explain engagement with collective causes on other platforms remains to be seen.

We studied two forms of engagement with collective causes, finding that petitions posted with moral outrage were more likely to be shared but no more likely to be signed. This suggests that moral outrage is neither uniformly beneficial nor detrimental to online collective causes (see Brady & Crockett, 2019; Spring et al., 2019). On one hand, it may be effective at increasing awareness. The way #MeToo spread through social networks was pivotal to cultivating an awareness of the scope of the issue (Mendes et al., 2018). However, it may be limited in its ability to translate the magnitude of an outcry into the sorts of further action which can influence decision-makers (Goldman, 2015). Whether moral outrage is ultimately an effective tool may therefore depend on the specific goals social movements seek to achieve (Brady & Crockett, 2019; cf. Spring et al., 2018).

Concluding Remarks

Taken as a whole, the findings indicate that at least one form of content (moral outrage) which tends to go viral is precisely that which has no association with the sorts of effortful behaviors thought to be conducive to legitimate and democratic social change (Bochel, 2013). Conversely, the content that struggles to go viral is instead that which shows the strongest connection to such behaviors. This result lends credibility to concerns about the role that social media platforms are playing in propagating inflammatory sentiments which have limited ability to further collective causes (Brady & Crockett, 2019; Crockett, 2017; Rathje et al., 2024; Van Bavel et al., 2024). However, the question remains open as to *why* moral outrage is not predictive of further forms of engagement. It could be, as experimental work suggests (Xiao & Houser, 2005), that expressing moral outrage is psychologically satiating, thereby reducing the impetus for further action. It could also be that outrage serves a specific function throughout the course of collective responses to injustices (Brady & Crockett, 2019; Lerner & Tiedens, 2006). The breaking news of George Floyd's murder was initially met with swift and decisive condemnation, with more organized and strategic collective responses to follow. Investigating these mechanisms will likely require careful experimental work and sophisticated modeling of samples which include both online engagement and offline behavior (e.g., Smith et al., 2023).

With the use of online platforms becoming more widespread and integral to social life, it is increasingly important to take serious concerns about their capacity to undermine constructive ways of relating to one another. The pattern of results we observed indicates that the content which tends to be amplified is surprisingly unrelated to further forms of engagement with collective causes. This points to a case of misalignment (Rathje et al., 2024), where the goal of social media platforms to maximize engagement may not be conducive to what promotes the sorts of effortful collective actions that are typically thought to be conducive to social change.

Author Contributions

Leach led all aspects of the work. Cichocka contributed to the conceptualization and reviewing/editing. Formanowicz contributed to the conceptualization and reviewing/editing. Nikadon contributed to the data collection and curation, validation of analyses, and reviewing/editing.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was supported by an OPUS 19 grant awarded to Magdalena Formanowicz and a Leverhulme Trust Research Project Grant (RPG-2021-241) awarded to Aleksandra Cichocka.

Ethical Approval

The work received approval from the School of Psychology Ethics Board at the University of Kent (ID: 202316941706278627).

Artificial Intelligence

ChatGPT was used to improve grammar and sentence structure.

Transparency

No aspects of the study were pre-registered. The data and analysis scripts are publicly available (https://osf.io/2kmtp/). Post texts are not shared to comply with X's Developer agreement.

ORCID iDs

Stefan Leach https://orcid.org/0000-0003-4065-3519
Magdalena Formanowicz https://orcid.org/0000-0001-5859-7100
Jan Nikadon https://orcid.org/0000-0002-2038-254X
Aleksandra Cichocka https://orcid.org/0000-0003-1703-1586

Supplemental Material

The supplemental material is available in the online version of the article.

Note

1. We also scraped each petition's description. These are not the focus of our analyses, but are included as covariates to test the robustness of the findings (see Additional Analyses).

References

Adam-Troian, J., Bonetto, E., & Arciszewski, T. (2021). "We shall overcome": First-person plural pronouns from search volume data predict protest mobilization across the United States. *Social Psychological and Personality Science*, *12*(8), 1476–1485. https://doi.org/10.1177/1948550620987672

Ashokkumar, A., & Pennebaker, J. W. (2022). Tracking group identity through natural language within groups. *PNAS Nexus*, *1*(2), pgac022. https://doi.org/10.1093/pnasnexus/pgac022

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Becker, J. C. (2012). The system-stabilizing role of identity management strategies: Social creativity can undermine collective action for social change. *Journal of Personality and Social Psychology*, *103*(4), 647–662. https://doi.org/10.1037/a0029240

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge University Press.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Bloom, P., & Jordan, M. (2018, August 9). Are we all "harmless torturers" now? *The New York Times*. https://www.nytimes.com/2018/08/09/opinion/are-we-all-harmless-torturers-now.html

Bochel, C. (2013). Petitions systems: Contributing to representative democracy? *Parliamentary Affairs*, *66*(4), 798–815. https://doi.org/10.1093/pa/gss005

Boggio, P. S., Nezlek, J. B., Alfano, M., Azevedo, F., Capraro, V., Cichocka, A., Pärnamets, P., Rego, G. G., Sampaio, W. M., Sjåstad, H., & Van Bavel, J. J. (2024). A time for moral actions: Moral identity, morality-as-cooperation and moral circles predict support of collective action to fight the COVID-19 pandemic in an international sample. *Group Processes & Intergroup Relations*, *27*(1), 178–195. https://doi.org/10.1177/13684302231153800

Bond, R. M., & Garrett, R. K. (2023). Engagement with fact-checked posts on Reddit. *PNAS Nexus*, *2*(3), pgad018. https://doi.org/10.1093/pnasnexus/pgad018

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, *40*(1), 21–41. https://doi.org/10.1177/0261927X20967028

Brady, W. J., & Crockett, M. J. (2019). How effective is online outrage? *Trends in Cognitive Sciences*, *23*(2), 79–80. https://doi.org/10.1016/j.tics.2018.11.004

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, *15*(4), 978–1010. https://doi.org/10.1177/1745691620917336

Brady, W. J., McLoughlin, K. L., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, *7*(33), eabe5641. https://doi.org/10.1126/sciadv.abe5641

Brady, W. J., McLoughlin, K. L., Torres, M. P., Luo, K. F., Gendron, M., & Crockett, M. J. (2023). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nature Human Behaviour*, *7*(6), Article 6. https://doi.org/10.1038/s41562-023-01582-0

Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, *148*(10), 1802–1813. https://doi.org/10.1037/xge0000532

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Chayinska, M., Miranda, D., & González, R. (2021). A longitudinal study of the bidirectional causal relationships between online political participation and offline collective action. *Computers in Human Behavior*, *121*, 106810. https://doi.org/10.1016/j.chb.2021.106810

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), Article 11. https://doi.org/10.1038/s41562-017-0213-3

Dehghani, M., & Boyd, R. L. (2022). *Handbook of language analysis in psychology*. Guilford Publications.

Drury, J., & Reicher, S. (2009). Collective psychological empowerment as a model of social change: Researching crowds and power. *Journal of Social Issues*, *65*(4), 707–725. https://doi.org/10.1111/j.1540-4560.2009.01622.x

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General*, *145*(2), 131–146. https://doi.org/10.1037/xge0000132

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Formanowicz, M., Beneda, M., Witkowska, M., Nikadon, J., & Suitner, C. (2024). Mobilize is a verb: The use of verbs and concrete language is associated with authors' and readers' perceptions of a text's action orientation and persuasiveness. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/01461672241238418

Formanowicz, M., Pietraszkiewicz, A., Roessel, J., Suitner, C., Witkowska, M., & Maass, A. (2021). Make it happen! Verbs as markers of agency increase message effectiveness. *Social Psychology*, *52*(2), 75–89. https://doi.org/10.1027/1864-9335/a000435

Foster, M. D., Tassone, A., & Matheson, K. (2021). Tweeting about sexism motivates further activism: A social identity perspective. *British Journal of Social Psychology*, *60*(3), 741–764. https://doi.org/10.1111/bjso.12431

Foster-Johnson, L., & Kromrey, J. D. (2018). Predicting group-level outcome variables: An empirical comparison of analysis strategies. *Behavior Research Methods*, *50*(6), 2461–2479. https://doi.org/10.3758/s13428-018-1025-8

Gantman, A. P., & Van Bavel, J. J. (2014). The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. *Cognition*, *132*(1), 22–29. https://doi.org/10.1016/j.cognition.2014.02.007

Goldman, J. (2015, July 28). How we're changing the way we respond to petitions. *The White House Blog*. https://obamawhitehouse.archives.gov/blog/2015/07/28/how-we-are-changing-way-we-respond-petitions

Greijdanus, H., de Matos Fernandes, C. A., Turner-Zwinkels, F., Honari, A., Roos, C. A., Rosenbusch, H., & Postmes, T. (2020). The psychology of online activism and social movements: Relations between online and offline collective action. *Current Opinion in Psychology*, *35*, 49–54. https://doi.org/10.1016/j.copsyc.2020.03.003

Halupka, M. (2014). Clicktivism: A systematic heuristic. *Policy & Internet*, *6*(2), 115–132. https://doi.org/10.1002/1944-2866.POI355

Iyer, A., & Ryan, M. K. (2009). Why do men and women challenge gender discrimination in the workplace? The role of group status and in-group identification in predicting pathways to collective action. *Journal of Social Issues*, *65*(4), 791–814. https://doi.org/10.1111/j.1540-4560.2009.01625.x

Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology*, *118*(1), 57–88. https://doi.org/10.1037/pspi0000186

Jost, J. T., Barberá, P., Bonneau, R., Langer, M., Metzger, M., Nagler, J., Sterling, J., & Tucker, J. A. (2018). How social

media facilitates political protest: Information, motivation, and social networks. *Political Psychology*, *39*(S1), 85–118. https://doi.org/10.1111/pops.12478

Kellen, S. (2020). Justice for George Floyd. *Change.org*. https://www.change.org/p/mayor-jacob-frey-justice-for-george-floyd

Klandermans, P. G. (1997). *The social psychology of protest*. Blackwell.

Klandermans, B. (2002). How group identification helps to overcome the dilemma of collective action. *American Behavioral Scientist, 45*(5), 887–900. https://doi.org/10.1177/0002764202045005009

Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making, 19*(2), 115–137. https://doi.org/10.1002/bdm.515

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*(4), 443–453. https://doi.org/10.1177/0146167204271651

Long, J. A. (2024). jtools: Analysis and presentation of social scientific data. *Journal of Open Source Software, 9*(101), 6610. https://doi.org/10.21105/joss.06610

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, *5*, 1375–1405.

Mendes, K., Ringrose, J., & Keller, J. (2018). #MeToo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women's Studies*, *25*(2), 236–246. https://doi.org/10.1177/1350506818765318

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, *97*(3), 405–423. https://doi.org/10.1348/000712605X85871

Milesi, P., & Alberici, A. I. (2018). Pluralistic morality and collective action: The role of moral foundations. *Group Processes & Intergroup Relations*, *21*(2), 235–256. https://doi.org/10.1177/1368430216675707

Miller, D. A., Cronin, T., Garcia, A. L., & Branscombe, N. R. (2009). The relative impact of anger and efficacy on collective action is affected by feelings of fear. *Group Processes & Intergroup Relations*, *12*(4), 445–462. https://doi.org/10.1177/1368430209105046

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour, 2*(6), 389–396. https://doi.org/10.1038/s41562-018-0353-0

Morozov, E. (2009, May 19). The brave new world of slacktivism. *Foreign Policy*. https://foreignpolicy.com/2009/05/19/the-brave-new-world-of-slacktivism/

Nikadon, J., Suitner, C., Erseghe, T., Dzanko, L., & Formanowicz, M. (2024). *BERTAgent: The development of a novel tool to quantify agency in textual data*. https://doi.org/10.31234/osf.io/qw6u3

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, *56*(1), 365–392. https://doi.org/10.1146/annurev.psych.56.091103.070141

Pietraszkiewicz, A., Soppe, B., & Formanowicz, M. (2017). Go Pro Bono: Prosocial language as a success factor in crowdfunding. *Social Psychology*, *48*(5), 265–278. https://doi.org/10.1027/1864-9335/a000319

Rathje, S., Robertson, C., Brady, W. J., & Van Bavel, J. J. (2024). People think that social media platforms do (but should not) amplify divisive content. *Perspectives on Psychological Science*, *19*, 781–795. https://doi.org/10.1177/17456916231190392

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Outgroup animosity drives engagement on social media. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 118*(26), Article e2024292118. https://doi.org/10.1073/pnas.2024292118

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schumann, S., & Klein, O. (2015). Substitute or stepping stone? Assessing the impact of low-threshold online collective actions on offline participation. *European Journal of Social Psychology*, *45*(3), 308–322. https://doi.org/10.1002/ejsp.2084

Shearer, E., & Mitchel, A. (2021). *News use across social media platforms in 2020*. Pew Research Center, Journalism and Media.

Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American Psychologist*, *56*(4), 319–331. https://doi.org/10.1037/0003-066X.56.4.319

Skoric, M. M. (2012). What is slack about slacktivism? *Methodological and Conceptual Issues in Cyber Activism Research*, *77*(7), 7–92.

Smith, L. G. E., Piwek, L., Hinds, J., Brown, O., & Joinson, A. (2023). Digital traces of offline mobilization. *Journal of Personality and Social Psychology*, *125*(3), 496–518. https://doi.org/10.1037/pspa0000338

Spring, V. L., Cameron, C. D., & Cikara, M. (2018). The upside of outrage. *Trends in Cognitive Sciences*, *22*(12), 1067–1069. https://doi.org/10.1016/j.tics.2018.09.006

Steffens, N. K., & Haslam, S. A. (2013). Power through 'us': Leaders' use of we-referencing language predicts election victory. *PLOS ONE*, *8*(10), e77952. https://doi.org/10.1371/journal.pone.0077952

Stubbersfield, J. M., Dean, L. G., Sheikh, S., Laland, K. N., & Cross, C. P. (2019). Social transmission favours the "morally good" over the "merely arousing." *Palgrave Communications*, *5*(1), 70. https://doi.org/10.1057/s41599-019-0269-y

Suitner, C., Badia, L., Clementel, D., Iacovissi, L., Migliorini, M., Salvador Casara, B. G., Solimini, D., Formanowicz, M., & Erseghe, T. (2023). The rise of #climateaction in the time of the FridaysForFuture movement: A semantic network analysis. *Social Networks*, *75*, 170–185. https://doi.org/10.1016/j.socnet.2022.06.003

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter networks. *Social Networks*, *34*(1), 73–81. https://doi.org/10.1016/j.socnet.2011.05.006

Thomas, E. F., McGarty, C., & Mavor, K. I. (2009). Aligning identities, emotions, and beliefs to create commitment to sustainable social and political action. *Personality and Social Psychology Review*, *13*(3), 194–218. https://doi.org/10.1177/1088868309341563

Van Bavel, J. J., Robertson, C. E., del Rosario, K., Rasmussen, J., & Rathje, S. (2024). Social media and morality. *Annual Review of Psychology*, *75*(1), 311–340. https://doi.org/10.1146/annurev-psych-022123-110258

van Zomeren, M., & Iyer, A. (2009). Introduction to the social and psychological dynamics of collective action. *Journal of Social Issues*, *65*(4), 645–660. https://doi.org/10.1111/j.1540-4560.2009.01618.x

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

van Zomeren, M., Postmes, T., Spears, R., & Bettache, K. (2011). Can moral convictions motivate the advantaged to challenge social inequality? Extending the social identity model of collective action. *Group Processes & Intergroup Relations*, *14*(5), 735–753. https://doi.org/10.1177/1368430210395637

van Zomeren, M., Saguy, T., & Schellhaas, F. M. H. (2013). Believing in "making a difference" to collective efforts: Participative efficacy beliefs as a unique predictor of collective action. *Group Processes & Intergroup Relations, 16*(5), 618–634. https://doi.org/10.1177/1368430212467476

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wakslak, C. J., Jost, J. T., Tyler, T. R., & Chen, E. S. (2007). Moral outrage mediates the dampening effect of system justification on support for redistributive social policies. *Psychological Science*, *18*(3), 267–274. https://doi.org/10.1111/j.1467-9280.2007.01887.x

Whillans, A. V., & Dunn, E. W. (2018). Agentic appeals increase charitable giving in an affluent sample of donors. *PLOS ONE*, *13*(12), Article e0208392. https://doi.org/10.1371/journal.pone.0208392

Wright, S. (2010). Collective action and social change. In J. F. Dovidio, M. Hewstone, P. Glick, & V. Esses (Eds.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 577–595). Sage. https://www.torrossa.com/en/resources/an/4913726

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences*, *102*(20), 7398–7401. https://doi.org/10.1073/pnas.0502399102

Author Biographies

Stefan Leach is a lecturer in social psychology at Lancaster University whose research focuses on the application of computational methods and natural language processing to political ideology and engagement.

Magdalena Formanowicz is an associate professor at the Center for Research on Social Relations, SWPS University in Warsaw, Poland. Her research focuses on social cognition and language. Her research relies on a multimethod approach that incorporates experimental studies and analyses of large textual data to investigate language pertinent to agentic language, mobilising communication, and discrimination and intergroup relations. She is also interested in dehumanization and its linguistic manifestations.

Jan Nikadon is a PhD student in the Department of Psychology at SWPS University, where he investigates the communicative and linguistic aspects of social cognition and the dynamics of social interactions. His research employs experimental studies and large-scale textual data analysis. Additionally, he develops computational tools to quantify psychological trait saturation in textual data. He also works at the fMRI lab of the Interdisciplinary Centre for Modern Technologies and the Department of Cognitive Science at Nicolaus Copernicus University in Toruń.

Aleksandra Cichocka is a Professor of Political Psychology at the University of Kent. Her research leverages the psychology of the self and the psychology of social groups to understand contemporary politics.

Handling Editor: Danny Osborne