Journal of Personality and Social Psychology

Moral Panics on Social Media Are Fueled by Signals of Virality

Curtis Puryear, Joseph A. Vandello, and Kurt Gray

Online First Publication, February 29, 2024. https://dx.doi.org/10.1037/pspa0000379

CITATION

Puryear, C., Vandello, J. A., & Gray, K. (2024, February 29). Moral Panics on Social Media Are Fueled by Signals of Virality. *Journal of Personality and Social Psychology*. Advance online publication. https://dx.doi.org/10.1037/pspa0000379

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

© 2024 American Psychological Association
ISSN: 0022-3514

https://doi.org/10.1037/pspa0000379

INNOVATIONS IN SOCIAL PSYCHOLOGY

Moral Panics on Social Media Are Fueled by Signals of Virality

Curtis Puryear[1], Joseph A. Vandello[2], and Kurt Gray[3]
[1] Department of Management and Organizations, Kellogg School of Management, Northwestern University
[2] Department of Psychology, University of South Florida
[3] Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill

Moral panics have regularly erupted in society, but they appear almost daily on social media. We propose that social media helps fuel moral panics by combining perceived societal threats with a powerful signal of social amplification—virality. Eight studies with multiple methods test a social amplification model of moral panics in which virality amplifies perceptions of threats posed by deviant behavior and ideas, prompting moral outrage expression. Three naturalistic studies of Twitter ($N = 237{,}230$) reveal that virality predicts moral outrage in response to tweets about controversial issues, even when controlling for specific tweet content. Five experiments ($N = 1{,}499$) reveal the causal impact of virality on outrage expression and suggest that feelings of danger mediate this effect. This work connects classic ideas about moral panics with ongoing research on social media and provides a perspective on the nature of moral outrage.

Keywords: social media, moral panic, outrage, politics, morality

Supplemental materials: https://doi.org/10.1037/pspa0000379.supp

Moral Panics on Social Media Are Driven by Virality

Perceived societal threats emerge regularly throughout history, causing eruptions of outrage and hostility to punish those responsible. In the 1970s, sociologist Stanley Cohen called these bursts of outrage "moral panics," pointing to frenzied reactions to marijuana use, Rock N' Roll music, and street crime. Widespread concern and outrage are often fueled by media and opinion leaders who focus the public's attention on some emerging threat, whether real or exaggerated. On social media, moral panics feel like a daily occurrence. In 2014, *Slate* magazine published a piece titled "The Year of Outrage," which documented an outraged firestorm each day of the year (Turner et al., 2014). Today, eruptions of outrage remain a daily occurrence, with moral panics over economic collapse, the end of democracy, the spread of injustice, or even the end of the world. Existing work has identified some features of social media that contribute to this new normal (e.g., the salience of group identities and increased exposure to threatening events). Here, we highlight another important but overlooked feature that might drive moral panics—virality.

It is well established that social media exposes users to threatening content (Crockett, 2017), but explicit metrics of "shares" or "likes" tell us which content is being shared and capturing widespread attention or going "viral" (Chung, 2017; J. W. Kim, 2018; J. Kim, 2021; Lee-Won et al., 2016). We suggest that these external virality metrics are powerful psychological signals of virality, telling people which threats warrant concern and helping to drive online moral panics. Synthesizing previous studies of moral panic with studies of how humans respond to threats, we outline a model of moral panics on social media—the social amplification model of moral panics.

The Social Amplification Model of Moral Panics

The social amplification model of moral panics is grounded in how humans evaluate and respond to societal threats. Humans are vigilant in detecting potential threats (Blanchard et al., 2011; Neuberg et al., 2011; Richards et al., 2014). We use social information (i.e., information about what others are saying and doing) to discern which potential threats warrant concern (Aguirre, 2005; Kasperson et al., 1988); we experience psychological feelings of danger, which prepare us to respond to threats (Gross, 1998; Steimer, 2002); and we deploy moral outrage to punish those who threaten society and to restore order (Henderson & Schnall, 2021; Rohloff, 2011; Rucker et al., 2004). These four elements of moral panics—potential threats, social amplification, feelings of danger, and moral punishment to mitigate the threat—are captured in the model in Figure 1 as they manifest on social media.

Curtis Puryear https://orcid.org/0000-0001-6067-4654

This work was supported by the Charles Koch Foundation (Grant 29201-29201-52459-319100-10000; Center for the Science of Moral Understanding) awarded to Kurt Gray.

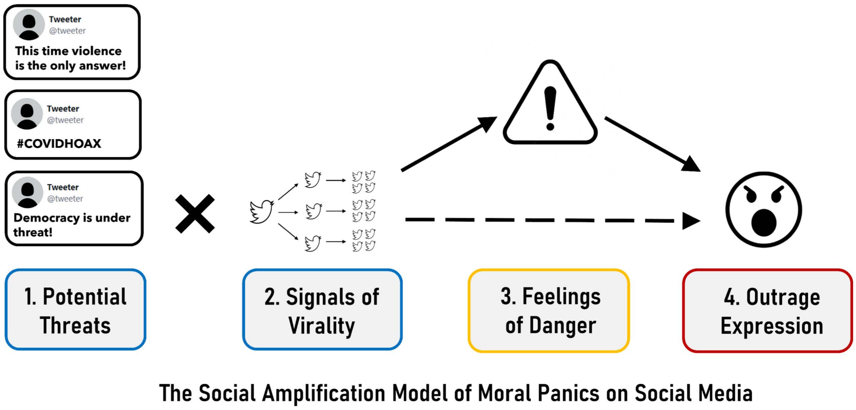
All data and analysis scripts in R are available on our open science framework page (https://osf.io/wr69q/). Studies 5–8 were all preregistered. Preregistration documents have been compiled on our open science framework page.

Curtis Puryear played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, resources, software, visualization, writing–original draft, and writing–review and editing. Joseph A. Vandello played a supporting role in conceptualization, methodology, and writing–review and editing and an equal role in supervision. Kurt Gray played a lead role in funding acquisition, a supporting role in conceptualization, investigation, methodology, and writing–review and editing, and an equal role in supervision.

Correspondence concerning this article should be addressed to Curtis Puryear, Department of Management and Organizations, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, United States. Email: cwpuryear12@gmail.com

Figure 1

The Social Amplification Model of Moral Panics Applied to Social Media



Note. Social media creates a limitless supply of potential threats (Step 1) and amplifies threats via signals of virality (Step 2), which then produces feelings of danger (Step 3) and causes outrage expression to punish those responsible (Step 4). See the online article for the color version of this figure.

On social media, we frequently encounter potential threats (Step 1), which are amplified by signals of virality (Step 2), which then increase feelings of danger (Step 3). To combat these viral threats—and help mitigate this sense of danger—people express outrage (Step 4). This model therefore includes design features of social media (Steps 1 and 2), psychological processes (Step 3), and behavior (Step 4). Although these steps are presented in the sequential order we investigate here, it is also likely that some paths in the model are bidirectional (e.g., outrage expression causing content to be more viral). We describe each element of our model, how they contribute to moral panics, and how they manifest on social media.

Potential Threats

Evolution has prepared us to be vigilant in detecting threats, but we encounter many behaviors, people, and even ideas that are potentially threatening. We must discern which of these potential threats warrants a response (Blanchard et al., 2011; Richards et al., 2014). Some threats stand out more than others. Events that involve obvious direct harm (e.g., mass shootings) are relatively rare, but humans can learn to fear many different threats based on their environments and experiences (LoBue & Rakison, 2013). Which societal threats are most harmful and relevant is often subjective (Darley & Pittman, 2003), and moral panics piggyback off of these subjective fears (Cohen, 1972; Goode & Ben-Yehuda, 1994). The fantasy game *Dungeons & Dragons* caused moral panic in the 1980s because many parents harbored fears of Satanism corrupting their children (BBC News, 2014). Although *Dungeons & Dragons* was an innocent game, moral panics can also stem from real threats. As fears of climate change have grown among the public, so too have eruptions of outrage in response to behaviors exacerbating this danger (Rohloff, 2011).

Because our minds pay close attention to potential threats, traditional media has always favored threatening content to drive engagement (Soroka et al., 2019). Social media content is no different (Brady et al., 2020), spreading content about COVID-19 conspiracies, climate change deniers, rigged elections, and "woke" organizations. People encounter this threatening content in at least two ways: by seeing trusted allies talk about threats (e.g., "Can you believe that these people are destroying America?") and by directly witnessing the threatening behavior from apparent moral deviants and ideological opponents ("Burn it all down!"; Bail et al., 2018; Bakshy et al., 2015; Barnidge, 2017). Because of the many weak ties on social media, these direct encounters with "moral deviants" are not uncommon (Barberá, 2020). The majority of frequent social media users also report encountering hate speech (Costello et al., 2016; Oksanen et al., 2014). Merely encountering a threat online can be frightening, but this alone is usually not enough to cause a panic (Kasperson et al., 1988). Our response to threats depends on additional information about their scope and how others are responding.

Social Amplification via Signals of Virality

People do not evaluate potential threats in isolation (Renn, 2011); they also look to other people (e.g., what our neighbors are saying) and institutions (e.g., what news media make salient) for information. The more that others seem to discuss, focus upon, or highlight a threat, the larger it looms in our minds. Traditionally, information about how many people in society were focusing on a specific threat was relatively implicit, with people gleaning other people's concerns by repeated exposure to conversations or news stories about threats. On social media, these implicit cues remain, as users may repeatedly see discussions about threats. However, sites like Twitter (recently rebranded as "X"), Reddit, and Facebook also include explicit signals of virality—defined as the social transmission or spread of content online (Berger & Milkman, 2012)—including tallies of shares and trending pages that display the most viral content. Virality metrics paired with threatening content may indicate that others consider the threat worth their attention, causing us to also infer that the threat must be dangerous (J. W. Kim, 2018; J. Kim, 2021; Lee-Won et al., 2017).

Any signal of virality may make threats loom larger, but there may also be nuances to this effect across different presentations of threats (e.g., how people are talking about them) and different signals of virality (e.g., retweet metrics vs. trending pages). If a violent extremist group emerges, we may see the name of the extremist group trending as a hashtag on Twitter, along with viral posts supporting extremism (e.g., "Now is the time for violence!"; Bail et al., 2018; Bakshy et al., 2015; Barnidge, 2017). In these contexts, greater perceptions of virality may also make threats seem especially influential and dangerous (J. W. Kim, 2018). At the same time, we may also see viral posts from allies condemning or "dunking" on the threatening group, which could provide some assurance that others' are already addressing threats. However, even viral posts condemning threats can still make those threats more salient (Berger et al., 2010; Chen & Berger, 2013). All signals of virality may not function identically, but we do expect them to generally shape our perceptions of what content is important—amplifying the power of potential threats.

The impact of virality also depends upon people's personal fears. Historically, media stories about *Dungeons & Dragons* and "mugging" would likely not have amplified feelings of danger for people with no fears of satanism or street crime. Even so, there is still considerable flexibility in the types of events that can make us feel that society is in danger. Natural disasters and pandemics are often not initially moralized, but they can still destabilize society (Gelfand & Lorente, 2021) and cause our minds to search for immoral agents to blame (Schein & Gray, 2018). Many potential threats that trigger moral panics seem mild or ambiguous at first (Garland, 2008). For example, drag shows have recently emerged as a focal societal threat among the political right in the United States (Carless, 2023), but drag shows have long existed without being labeled a major societal threat. At the same time, drag shows only transformed into a societal threat among people with specific concerns (e.g., fears of sexual minorities destroying society). While our minds can quickly envision how different types of threats could endanger society—especially once everyone around us starts talking about them—people's subjective beliefs also limit what can be potentially threatening.

Feelings of Danger

After social amplification turns potential threats into dangerous societal threats (Kasperson et al., 1988; Renn, 2011), people experience a mix of emotions and perceptions that we call feelings of danger. Sociological research on moral panics describes this key psychological experience as a heightened sense of concern, characterized by increasing perceptions that our institutions are failing to contain threats. This undermines the apparent safety and stability of society, causing people to feel that they need to do something about the threat (Hier, 2008; Rohloff, 2011). Integrating this work with research from psychology, we argue that the core psychological experience during moral panics is increasing perceptions of danger and instability in society (Cook et al., 2018; Duckitt & Fisher, 2003; Rucker et al., 2004). These perceptions are typically accompanied by emotions like fear and distress (Richards et al., 2014; Shaffer & Duckitt, 2013; Steimer, 2002; Zoellner et al., 2014), which are highly aversive and motivate us to restore safety and stability to society.

People have multiple strategies for coping with negative emotions (Gross, 1998), including those triggered by societal threats. Sometimes people reappraise the threat (e.g., by rationalizing societal injustice; Solak et al., 2021); other times they distract themselves from events that are difficult to reappraise (e.g., disengaging from politics; Mehta et al., 2020). But for societal threats that evoke intense emotions and feel widely impactful, reappraising or distancing oneself from the threat may not feel like an effective option (Spring et al., 2018). In the face of socially amplified threats, people may opt to fight back (Ford & Feinberg, 2020). But fights against societal threats are often not physical fights; they are fights to punish moral deviants and restore stability to society (Darley & Pittman, 2003; Gelfand & Lorente, 2021). While it is difficult to run from threats on social media, expressing outrage to fight back against them involves relatively little effort and risk (Brady et al., 2020).

Moral Punishment via Outrage Expression

Humans evolved multiple strategies for responding to threats—including fighting or fleeing—but we also evolved a strategy for combating social threats: outrage expression. Morality is an old toolset for regulating harmful behaviors (Cosmides et al., 2018), and people are strongly motivated to punish the moral deviants who seem to threaten society (Darley & Pittman, 2003; Kahneman & Sunstein, 2005). People punish moral deviants in multiple ways, including direct hostility (Cook et al., 2018; Rudert & Speckert, 2023), relational aggression (Fernandes et al., 2017), and ostracism (Hales et al., 2017). Outrage expression is another tool for moral punishment that can directly tarnish a target's reputation (Brady et al., 2020; DeScioli & Kurzban, 2013) and motivate onlookers to deprive them of resources (Gamez-Djokic & Molden, 2016; Henderson & Schnall, 2021). In the face of looming threats, people may deploy moral outrage to punish moral deviants, alleviate distress, and restore a sense of security and stability within society.

Social media creates opportunities to express outrage against moral deviants. Historically, condemning moral deviants—from violent criminals to *Dungeons & Dragons* players—carried risks of retaliation (Nelissen, 2008). And recruiting allies to engage in collective action required effort and coordination. On social media, physical distance between users and anonymity can reduce the risks of retaliation (Brady et al., 2020), and signals of virality may help users identify dangerous threats to simultaneously mobilize against. In short, social media may amplify dangerous threats (via signals of virality), but they also reduce the costs of deploying outrage to address those threats. This may cause users to respond to threats on social media not by running from or reinterpreting them but by deploying outrage against those responsible.

The Theoretical Context of the Social Amplification Model of Moral Panics

Our model complements and expands existing work on social media outrage. First, previous work focuses on how outrage may cause virality (Brady et al., 2020), but our model argues for the complementary pathway: that virality also causes outrage (the direct path from Steps 2 to 4). These two causal directions may reflect a dynamic feedback loop between virality and outrage: viral threats may amplify feelings of danger and trigger outrage expression (the paths emphasized in our model), but outraged posts may also cause those same posts to spread more widely, further amplifying the threat. Within this potential cycle of outrage, we focus on testing the specific claims of the social amplification model of moral panics: the pathway from potential threats to signals of virality to feelings of danger to outrage expression.

Our model also complements existing models that highlight other motivations of outrage expression and design features of social media (Brady et al., 2020). Brady et al.'s motivation, attention, and design model posits that moral content is prevalent on social media because people have group-based identity motivations to share moral content, because moral content captures our attention, and because the design of social media amplifies our biases for moral content. Some elements of this model dovetail with our own. The motivation, attention, and design model posits that social media algorithms interact with our biases to promote outrage-inducing content, which helps explain how social media exposes users to potential threats (Step 1 in our model). We also agree that people on social media often express outrage to maintain a positive group image (e.g., by derogating competing outgroups) in the face of group identity threats (e.g., when Democrats see conservative media criticizing their ingroup). But our model emphasizes threats to the rules, stability, and safety of society (Darley & Pittman, 2003; Rohloff, 2011; Rucker et al., 2004), which motivates us to punish harm-doers and restore security and stability. We also highlight different design features of social media—signals of virality—that amplify these threats.

Last, our model emphasizes how feelings of danger help drive moral outrage expressions (Steps 3 to 4). Recent work emphasizes the extrinsic benefits of expressing moral outrage: how it reaps social rewards by attracting likes and shares (Brady et al., 2021) and how it signals one's virtues to others (Grubbs et al., 2019; Jordan & Rand, 2019). These benefits are well documented, but here we reemphasize the link between moral outrage and internal feelings of danger. This emphasis is supported by work revealing that perceptions of (Duckitt & Fisher, 2003; Rucker et al., 2004) and fears of societal threats (Henderson & Schnall, 2021; Murray et al., 2019) are key subjective experiences that drive us to punish others and defend society. These perspectives are compatible—just as any social phenomenon can be jointly shaped by intrinsic motivations (e.g., helping behavior driven by genuine feelings of empathy) and extrinsic benefits (e.g., helping behavior driven to earn prestige). Expressing outrage serves multiple functions. Here, we argue that a portion of digital outrage stems from feelings of danger and threat.

Study Aims and Overview

Eight studies—including naturalistic studies of social media and experiments—test the key paths in the social amplification model of moral panics in the context of social media (specifically Twitter/X). Naturalistic studies (Studies 1–3) first examine whether highly viral content about three potential threats (i.e., climate change, immigration, and COVID-19) attracts higher proportions of outraged replies than less viral content. Additional analyses with tight controls isolate the impact of virality on outrage—testing whether tweets from the same types of users, using the same language, to discuss the same topics evoke more outrage expression when highly viral. Controlled experiments (Studies 4–8) then test whether identical content evokes greater feelings of danger and intentions to express outrage when viral.

These studies examine virality primarily via the metrics of "shares" on social media—a relatively novel form of social information that has been shown to influence social judgments (Calabrese & Zhang, 2019; J. W. Kim, 2018; J. Kim, 2021; Lee-Won et al., 2016, 2017). In Studies 1–3, we measured these share metrics using metadata obtained from the Twitter application programming interface. In Studies 4–8, we showed participants a series of posts and manipulated the share metrics in all studies (we also displayed metrics of likes in Studies 5–8, adjusting them proportionally to our manipulation of share metrics). We expected these explicit metrics to have strong effects on subjective perceptions of virality. Our experiments also tested this idea, examining whether these metrics caused content to seem more widely shared, influential, and important. In Supplemental Materials, we also report a correlational study, which explores how often people report checking virality metrics and whether people who check virality metrics perceive more danger in the world (potentially from virality causing danger to feel like it is more noteworthy or impactful).

Across these studies, we also explore the generalizability and specificity of our model. We test the effects of virality on potential threats related to the environment, pandemics, politics, prejudice, and animal abuse. We test whether virality primarily amplifies threats relevant to users' subjective fears and concerns (e.g., immigration among conservatives and climate change among liberals). We also test whether the effect of virality on outrage is especially strong for posts from political opponents, which should consistently pose potential threats. Relatedly, we investigate how much the effects of virality reflect ideological conflict (e.g., liberals and conservatives reacting to the spread of ideas from political opponents) and whether virality also amplifies a wide range of dangerous behaviors and issues. In sum, these studies test the paths in our model, examine the effects of virality in multiple contexts, and explore the roots of digital outrage in feelings of danger.

We maximized statistical power across these studies by incorporating large social media data sets and ensuring each of our experiments had a minimum of 100 participants in each condition. We also conducted sensitivity analyses for each of our experiments to identify the smallest effect sizes our samples were capable of detecting with 80% power. For studies that were analyzed with mixed-effects models, we used simulation methods (script available on our open science framework page), and for Studies 7–8, sensitivity analyses were conducted in *G*Power*. Our simulations for Studies 4–6 identified the smallest standardized betas that our samples could detect with 80% power, simulating the effects of virality upon outrage. The smallest detectable effects were as follows: Study 4: $\beta = .22$, Study 5: $\beta = .18$, and Study 6: $\beta = .22$. For Studies 7 and 8, the smallest detectable effect was $d = .25$. In sum, sensitivity analyses confirmed that the samples collected in our experiments were adequately powered to detect small effect sizes.

Data for all studies and R code for analyses are available on our open science framework page (https://osf.io/wr69q/; Puryear et al., 2023).

Naturalistic Studies: Viral Tweets Receive Greater Proportions of Outraged Replies

One key prediction of the model is that signals of virality, combined with potential threats, fuel outrage expression (i.e., Steps 1 and 2 to Step 4 in Figure 1). It is well established that outrage expression leads to virality (Brady et al., 2020), but we expected the opposite to also be true: more viral tweets would evoke more outrage, even when controlling for multiple factors, including the language used, the status of the author, and the specific topic being discussed.

Measuring how much people feel society is in danger is difficult with naturalistic data, and so we do not directly test the mediating path through feelings of danger in these naturalistic studies. Still, we can measure people's ideology (Barberá, 2020), and certain topics should be more likely to cause feelings of danger and outrage expression among conservatives than liberals. We tested this idea in two ways: first, we thought posts about climate change and immigration should, on average, be more relevant to the fears of liberals and conservatives, respectively. Second, since partisans in the United States increasingly believe that political opponents hold extreme and harmful opinions (Lees & Cikara, 2021; Moore-Berg et al., 2020), we also expected posts about threats from political opponents (vs. allies) to be relevant to active Twitter users. These two contexts allowed us to examine posts that should contain more relevant threats to users in our sample, providing an indirect test of the idea that virality predicts outrage because it amplifies threats.

In sum, Studies 1–3 tested whether virality in tweets predicted outrage in replies (Steps 2–4 in our model) and whether this relationship was stronger among threats that should be especially concerning to users, and we explored these effects within posts about specific, real-world events.

Data

We tested our model in studies of three different issues on Twitter—climate change (Study 1; N = 97,088), immigration (Study 2; N = 43,531), and COVID-19 (Study 3; N = 96,611). We wanted to collect discussions about threats that should be especially concerning to liberals (i.e., climate change) and to conservatives (i.e., immigration). We also wanted to examine discussions about a more novel threat that was not a long-standing political issue at the time of data collection (i.e., COVID-19, which was collected in April 2020). Each of these data sets comprised tweets and replies to those tweets. Tweets IDs for each data set were obtained from repositories of tweets related to each issue, and then the full data for each tweet was obtained via the Twitter application programming interface (see Supplemental Materials for additional details). We measured virality using retweet counts obtained from Twitter metadata.

We measured moral outrage on a binary scale (0 = no moral outrage; 1 = moral outrage) using a validated machine learning classifier (Brady et al., 2021). This classifier was trained on 26,000 tweets and annotated based on three criteria:

> A person can be viewed as expressing moral outrage if (a) they have feelings in response to a perceived violation of their personal morals; (b) their feelings comprise emotions such as anger, disgust, and contempt; and (c) the feelings are associated with specific reactions, including blaming people/events/things, holding them responsible, or wanting to punish them. (Brady et al., 2021, p. 3)

We conducted Studies 1–3 at different times, first collecting and analyzing the climate change corpus. We then wanted to examine a topic that we thought would be more threatening to conservatives (i.e., immigration) and a more novel threat (i.e., COVID-19). To simplify the presentation of results, we have grouped our Results section by our research questions/hypotheses (rather than reporting Studies 1–3 sequentially).

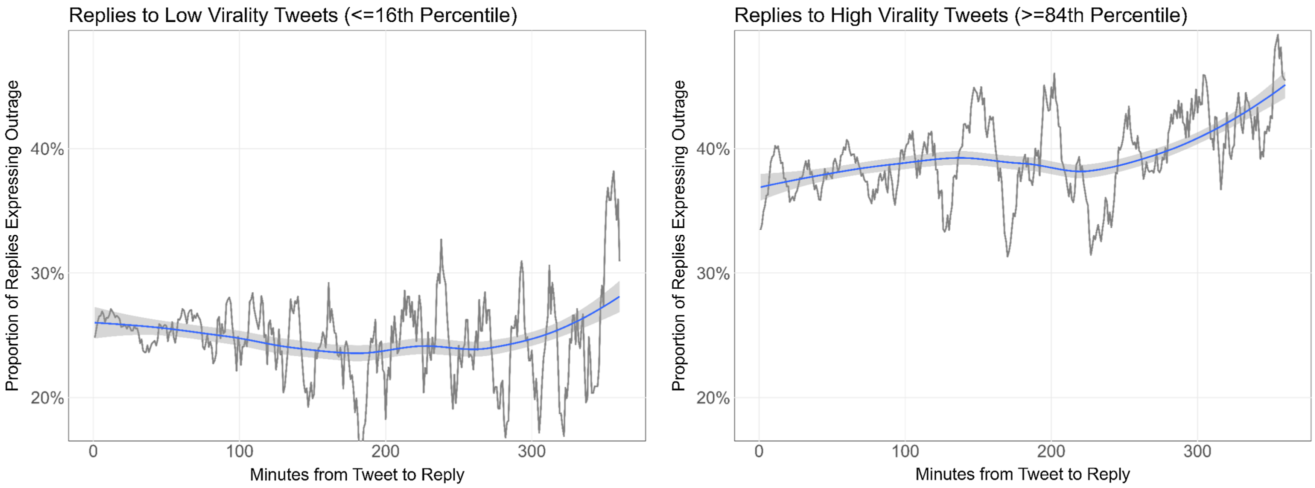
Results

The overall pattern of the data can be seen in Figure 2, which combines all data from Studies 1–3. As expected, highly viral tweets (i.e., the 84th percentile or above in number of shares received; right panel) received higher proportions of outraged replies than low virality tweets (i.e., the 16th percentile or below; left panel) in the first 6 hr since tweets were posted. This also revealed that the proportion of replies that express outrage gradually increased over time for high virality tweets (as they “go viral”), which is consistent with the idea that more viral tweets are more likely to be targeted with outrage.

We now present statistical tests of the model’s predictions while controlling for (among other things) specific tweet content. Supplemental Materials include additional robustness analyses. One of our goals was to address the concern that perhaps some content is

Figure 2

Average Proportion of Replies That Express Outrage in Response to Low and High Virality Tweets (Combining Studies 1–3)



Note. Proportions of replies expressing outrage in the first 6 hr after tweets are posted. We calculated these proportions every minute, and then applied two smoothing methods for visualization. The gray line depicts the 10-min moving average; the blue line is a loess curve. See the online article for the color version of this figure.

both more likely to go viral and to elicit more outraged replies. Although these analyses cannot perfectly account for all alternative explanations—which is why we conducted several experiments—they provide converging support for our predictions.

Virality Predicts Outrage

A series of linear-mixed effects logistic regressions tested whether virality predicts outrage expression on social media while controlling for other causes of outrage in replies. These controls included (a) moral contagion (i.e., the possibility that replies mirror outrage expressed by viral tweets; Brady et al., 2020) by accounting for outrage expressed by the original tweet (i.e., the target of replies), (b) follower-size effects (i.e., the possibility that users are more willing to express outrage toward powerful users; Sawaoka & Monin, 2018) by accounting for the following size of tweet authors, (c) the overall linguistic style of the tweet by controlling for the four summary variables from Linguistic Inquiry and Word Count 2022 (Tone, Authenticity, Clout, and Analytic), and (d) political ideology (using ideology estimates obtained in 2018; Barberá, 2020) by controlling both for ideology of users and the ideological difference between tweets and replies. To account for the fact that some replies targeted the same tweets or were authored by the same users, we included random intercepts for the tweet ID and for the ID of the reply author. Both retweet counts and follower counts were log transformed prior to analysis, although similar results are obtained when analyses are conducted on untransformed data (see Supplemental Materials for additional details and analyses). Models were fit using the glmer function from the lme4 package in R.

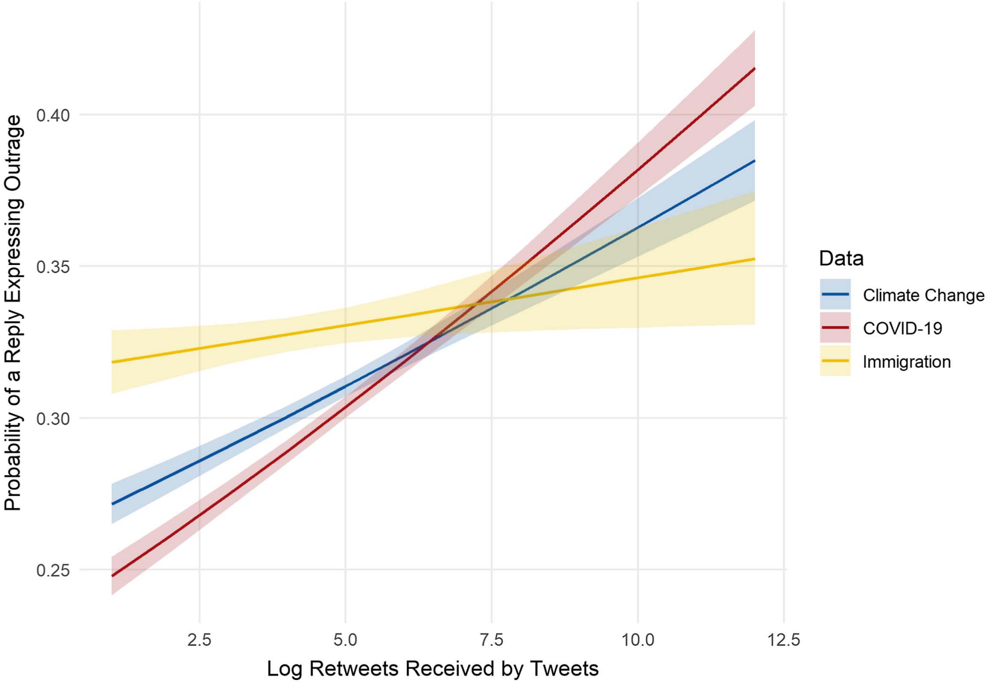
As predicted, virality (among tweets) predicted the probability of replies expressing outrage across all three issues, including climate change: odds ratio (*OR*) = 1.05, 95% CI [1.04, 1.06], z = 12.17, $p < .001$; immigration: *OR* = 1.01, 95% CI [1.00, 1.03], z = 2.43, $p = .025$; and COVID-19: *OR* = 1.07, 95% CI [1.06, 1.08], z = 19.39, $p < .001$. For every one-unit increase in log retweets, the odds of an individual reply expressing outrage increased by 5%, 1%, and 7% in the climate change, immigration, and COVID-19 corpora, respectively. Across all three issues, highly viral tweets evoked more outrage while controlling for other causes of outrage expression (see Figure 3 for predicted values from statistical models). This relationship was small in our immigration corpus, but our sample of social media users contained more liberals who likely find the topic of immigration less threatening on average. This is consistent with the idea that virality primarily amplifies threats users' are concerned about.

Virality Predicts Outrage Within Specific Topics

We next tested whether virality predicted outrage within tweets about even more specific topics. Each of our corpora contained tweets about more specific events, and we wanted to replicate the relationship between virality and outrage while controlling for the fact that some of the tweets in our data set may have been about more controversial events (which could have caused them to be more viral and more outrage-inducing). We identified these topics using biterm topic models (which are especially effective at discovering coherent topics in short social media texts; see Method section for details; Yan et al., 2013) and classified tweets based on the most probable

Figure 3

The 10 Topics With the Largest Estimated Ideology × Retweet Interactions, Studies 1–3



Note. See the online article for the color version of this figure.

topic predicted by our model. This enabled us to test whether more viral topics generated more outrage and, more importantly, whether virality predicted outrage within tweets about the same specific topics (still controlling for the language used by the tweet and the following size of the author).

The fine-grained topics extracted from this model enabled us to test whether highly viral (vs. less viral) tweets about the same specific events received greater proportions of replies that expressed outrage replies. For example, the immigration data set contained a broad topic discussing immigration and health care ($n = 552$) and a judge in Hawaii who overruled efforts to ban travel from majority Muslim countries to the United States ($n = 1{,}814$). Similarly, the COVID-19 data set contained larger topics discussing symptoms ($n = 1{,}001$) and death rates ($n = 2{,}949$) but also contained specific events like an outbreak on U.S.S. Roosevelt ($n = 294$). We fit mixed models—containing the same controls for language in tweets and characteristics of the tweet author as before—including random slopes for virality, intercepts for the topic each tweet was about, intercepts for the tweet ID, and intercepts for the reply author. Models included both the average log-transformed retweets for each specific topic and the topic-centered log retweets for each individual tweet. Since the random slopes in our model accounted for the different corpora used in our three studies, we combined all three data sets in these analyses. This analysis revealed that topics with more viral tweets contained greater proportions of replies that expressed outrage, $OR = 1.13$, 95% CI [1.09, 1.17], $z = 7.00$, $p < .001$. More importantly, results confirmed that tweets within these topics—even when they use similar language to discuss the same specific events—evoked more outrage when they were more viral, $OR = 1.05$, 95% CI [1.04, 1.06], $z = 14.86$, $p < .001$.[1]

Virality Predicts Outrage More Strongly for Threats Users Care About

Our model predicts that virality produces outrage by amplifying threats on social media, but societal threats are inherently subjective. And we thought some tweets in our data would be more threat-relevant to users: tweets about certain issues (i.e., climate change for liberals and immigration for conservatives) and tweets from political opponents—whose ideas and behaviors threaten to destroy society in the eyes of many U.S. partisans. Consistent with our predictions, virality within the climate change corpus was a weaker predictor of outrage among conservatives: $OR = 1.03$, 95% CI [1.02, 1.04], compared to liberals: $OR = 1.06$, 95% CI [1.05, 1.07]. This pattern was reversed for immigration, with virality failing to predict outrage among liberals: $OR = 1.01$, 95% CI [.99, 1.02] but predicting outrage among conservatives: $OR = 1.03$, 95% CI [1.01, 1.05]. Combining the data from both corpora revealed a significant three-way interaction between virality, ideology of the replier, and discussion topic: $OR = 1.02$, 95% CI [1.01, 1.03], suggesting that these patterns differed significantly from one another.

Users should also be especially concerned about tweets from political opponents. For tweets from political opponents, we thought increases in virality would predict an especially large increase in outrage expression. Consistent with this idea, virality was a stronger predictor of outrage in cross-partisan interactions, climate change: $OR = 1.06$, 95% CI [1.05, 1.07]; immigration: $OR = 1.03$, 95% CI [1.02, 1.05]; COVID-19: $OR = 1.08$, 95% CI [1.07, 1.09], than in copartisan interactions, climate change: $OR = 1.04$, 95% CI [1.03, 1.05], immigration: $OR = 1.00$, 95% CI [.98, 1.01]; COVID-19: $OR = 1.06$, 95% CI [1.06, 1.07], test of interaction for climate change: $OR = 1.01$, 95% CI [1.00, 1.01]; immigration: $OR = 1.02$, 95% CI [1.01, 1.02]; COVID-19: $OR = 1.01$, 95% CI [1.00, 1.01]. In sum, increased virality predicted stronger increases in outrage among tweets about topics (i.e., climate change/immigration for liberals/conservatives) and tweets from users (i.e., political opponents) that should be more potentially threatening. These results provide indirect evidence that virality predicts outrage by amplifying potential threats.

We also thought highly viral tweets about certain, more specific events might be more likely to trigger outrage among conservatives versus liberals (and vice versa). We fit another mixed model, including the interaction between virality and the ideology of the reply author, allowing the slope of this interaction to vary across topics from our topic model. We then identified topics where increases in virality corresponded with especially large increases in outrage among liberals versus conservatives. As our model would predict, the top five topics with stronger effects among liberals were related to events that should be more threatening to liberals (see Figure 4), such as U.S. President Trump withdrawing from the Paris Climate Agreement, early discussion of symptoms of COVID-19, and businesses reopening during COVID-19. By contrast, the topics with stronger effects of virality among conservatives were rarer, and the estimated interaction effects were smaller. Among the topics that did have stronger effects among conservatives, the two with the largest interaction effects were both about immigration. One topic was about protests against U.S. President Trump's immigration ban, and another focused on a judge ruling that would allow immigrants from Muslim majorities to enter the United States. Two of the remaining topics focused on threats that affected regions where more conservatives live (a hurricane and harmful algae in the American South), but results suggested that virality had relatively minor effects on posts about these topics. Overall, these exploratory findings were mostly consistent with the idea that tweets about relevant threats should be more easily amplified by virality.[2]

Contagion Effects

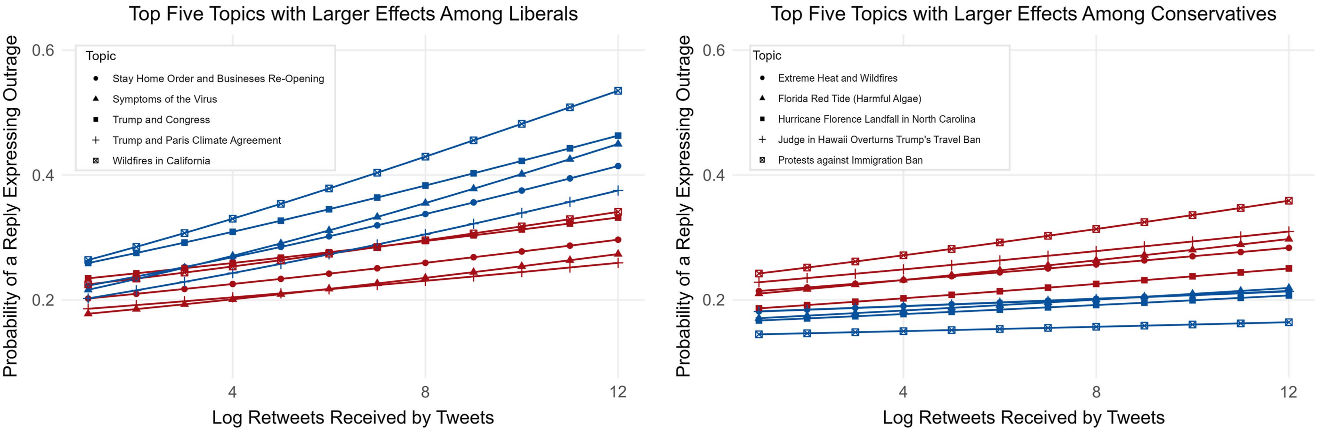
Our model proposes that virality makes potentially threatening tweets appear more threatening, leading to greater proportions of replies that expressed outrage. This effect should exist separate

[1] Our topic models extracted some mixed topics with no clear theme, as is almost always the case in topic modeling (Yan et al., 2013). Our primary analyses included tweets that belonged to these topics, but we also tested whether our effects were robust to an analysis that excluded tweets in these mixed topics. Virality predicted outrage just as strongly in this smaller data set ($N = 154{,}271$), $OR = 1.05$, 95% CI [1.05, 1.06], $z = 13.51$, $p < .001$.

[2] We leveraged information about users' ideology to test our prediction that virality would have stronger effects on tweets about threats that are especially concerning to users. But there are also likely to be threats that are especially concerning to all users, regardless of their ideology. We also explored the topics where increases in virality predicted the largest increases in outrage overall. Perhaps unsurprisingly, topics about COVID-19—a novel and highly salient threat at the time our data were obtained (April 2020)—consistently had the largest effects of virality. We also found that topics about immoral behavior (based on crowd-sourced ratings of our topic labels) yielded consistent effects of virality, but not to the same degree as topics related to COVID-19. We describe these results in detail in the Supplemental Materials.

Figure 4

The 10 Topics With the Largest Estimated Ideology × Retweet Interactions, Studies 1–3



Note. See the online article for the color version of this figure.

from the contagion of moral emotions from tweets to replies. To account for this, each of our models controlled for moral outrage expressed in tweets, but we were still interested in whether virality could also increase contagion effects. Namely, virality may cause posts expressing outrage to appear more normative (J. W. Kim, 2018; Lee-Won et al., 2016), making replies more likely to echo outrage expressed in tweets. However, in contrast with this idea, the relationship between outrage expressed in tweets and outrage expressed in replies was stronger for low-virality interactions in both the immigration, test of interaction: $OR = .97$, 95% CI [.95, 1.00], $z = -2.39$, $p = .017$; and COVID-19 corpora, test of interaction: $OR = .98$, 95% CI [.96, .99], $z = -3.72$, $p < .001$. No such interaction effect emerged in the climate change corpus, $OR = 1.00$, 95% CI [.99, 1.02], $z = .47$, $p = .640$. Of course, moral contagion between tweets and replies could be less pronounced for a variety of reasons, and viral tweets expressing outrage could still make outrage feel more normal, increasing outrage expression in other posts. Nevertheless, this pattern suggests it is unlikely that viral tweets attract outraged replies because of emotional contagion.

Discussion: Naturalistic Twitter Studies

Across three societal threats (climate change, immigration, and COVID-19; Studies 1–3), one of the most common signals of virality on social media—public metrics of retweets—predicted outrage expression. Highly viral tweets received higher proportions of replies that expressed outrage. This pattern grew over the first 6 hr of a tweet's life span—the crucial period of time when tweets "go viral" (or not; Mathews et al., 2017), and these higher proportions of outrage in replies were robust to multiple approaches to controlling for the content of tweets. Analyses exploring nearly 100 more specific topics revealed that more viral topics received greater proportions of replies that expressed outrage, and tweets within those topics—from users of similar status, using similar language—evoked more outrage when they were viral. Our model proposed that this relationship is due to virality, making the same potentially threatening content appear more threatening and important. Consistent with this idea, we found weaker effects on topics that should be less potentially threatening on average (i.e., no panic-inducing behaviors emerge from threats that do not seem real), and we found larger effects among tweets from political opponents (i.e., political opponents' behaviors and ideas are more threatening than allies'). Combined, these results provide initial, naturalistic evidence that signals of virality on social media may produce moral outrage (Steps 2 to 4 of the model are shown in Figure 1). Of course, any naturalistic data come with caveats: large-scale online data cannot directly measure psychological processes and cannot completely rule out third variables even with statistical controls—which is why we manipulated virality in experiments.

Experiments: Does Virality Cause Feelings of Danger and Outrage Expression?

Five experiments tested the social amplification model of moral panics by showing participants tweets that should be potentially threatening to them (Step 1), manipulating how viral they were (Step 2), and then measuring feelings of danger (Step 3) and behavioral intentions to express outrage in response (Step 4). These experiments further tested the boundary conditions of our model and clarified how virality amplifies threats. We examined a broader range of threats beyond hot-button issues including the threat of political opponents (Study 4), the expression of prejudice (Studies 5 and 6), and a novel dangerous trend that we invented (Study 7). We again tested whether virality had stronger effects on threats that users were especially concerned about (Study 6) and also teased apart the threat of tweets themselves from the issues connected to those tweets (Study 4). Last, in Study 8, we examined multiple types of outrage expression and tested whether people also respond to viral threats with other types of replies besides outrage. Studies 5–8 were preregistered.

Study 4: Viral Posts From Political Opponents (and Allies)

Our first experiment tested whether increasing the perceived virality of posts from political opponents would cause feelings of danger and intentions to express outrage. We also examined the effects of virality on posts from political allies, hoping to clarify results from our naturalistic studies. In our social media data, we found that viral (vs. less viral) tweets from political allies also evoked outrage (though this relationship was not as strong as it was for tweets from opponents). This could simply reflect the fact that people are often alarmed by posts from ideological allies too. However, it is also possible that virality amplified whatever threats political allies were discussing in their posts (e.g., a warning of environmental destruction may feel more important when it is shared widely). To test this, we showed participants posts from both political opponents and allies, manipulating their virality.

Method

Participants (see Table 1 for summary of participant information, the number of tweets rated, types of tweets rated, and example tweets for all experiments) each rated two tweets from political opponents and two from political allies in random order. Liberal tweets discussed issues like police violence and threats to the environment. Conservative tweets discussed issues like critical race theory and crises at the U.S. border. Following previous research (Calabrese & Zhang, 2019; J. W. Kim, 2018; Lee-Won et al., 2017), we manipulated perceptions of virality via the metrics accompanying each tweet. Highly viral tweets received between 4 K and 15 K retweets; nonviral tweets received zero to five retweets.

Our primary tests focused on reactions to seeing opponents' messages spread. To measure these reactions, we asked participants about the danger posed by the tweet (Step 3 in Figure 1) and their intentions to express outrage in response (Step 4). To measure danger, we asked participants how much "the above tweet is dangerous" and "the above tweet poses a serious threat" (two items, α = .96). To measure intentions to express outrage, we asked participants how likely they would be to write a post "expressing outrage toward this tweet" and "condemning this tweet" (two items, α = .95)

To test whether signals of virality (i.e., greater virality metrics) could also amplify threats mentioned in tweets from allies, we also asked participants about their feelings toward the issues discussed by tweets (separate from measuring feelings toward the tweets). To measure danger, we also asked participants to respond to two items following the stem "The issue raised by this tweet … ," including "is a dangerous issue" and "is an issue that poses a serious threat to this country" (α = .93). Similarly, to measure intentions to express outrage about the issues discussed by each tweet, we asked participants to indicate how likely they would be to "write a tweet expressing outrage over the issue discussed" and "write a tweet condemning someone related to the issue" (α = .93).

Study 4 also contained supplemental measures. We focused primarily on measures of perceived danger, but feelings of danger also include emotional experiences, like fear and distress. So, we measured additional emotional experiences. We also asked about participants' moral convictions (two items, e.g., "The issue discussed by the tweet is connected to my fundamental beliefs about right and wrong"), the perceived importance of issues (two items, e.g., "How important is the issue raised by the above tweet?"), subjective feelings of outrage (using items adapted from Tetlock et al., 2000),[3] and a manipulation check comprising a single item, asking "How much has the above tweet been shared across Twitter?"

After each tweet, participants responded to measures in the following order: the manipulation check, importance of the issue discussed by the tweet, how much they agreed with the tweet, feelings of danger about the issue, feelings of danger about the tweet, intentions to express outrage about the issue and toward the tweet, emotional reactions, and moral convictions about the issue.

All measures used 7-point Likert scales. Analyses used multilevel models with random intercepts for participant and for stimuli (i.e., which tweets they saw from our pool of tweets). Results for supplemental measures are reported in Supplemental Materials.

Results

We first tested whether our manipulation of virality metrics successfully increased perceptions of virality. Looking at our manipulation check, the high virality (vs. low virality) tweets successfully made participants think the tweets were more widely shared, $b = 3.94$, $t(951) = 76.53$, $p < .001$. As expected, changes in virality metrics strongly affected perceived virality.

We next tested the model outlined in Figure 1: the effect of high (vs. low) virality on danger posed by tweets from opponents (Steps 2–3), on intentions to express outrage (Steps 2–4), and the relationship between danger and intentions to express outrage (Steps 3–4). Supporting our predictions, virality increased ratings of danger posed by tweets from political opponents, $b = .46$, $t(264) = 4.05$, $p < .001$, 95% CI [.24, .68]; intentions to express outrage, $b = .37$, $t(264) = 3.46$, $p < .001$, 95% CI [.16, .58]; and ratings of danger predicted intentions to express outrage, $b = .51$, $t(532) = 15.20$, $p < .001$. We also conducted formal tests of mediation using the JSmediation package in R to estimate the within-subjects indirect effects with Monte Carlo methods (Mackinnon et al., 2004). This approach revealed a significant indirect effect of virality on intentions to express outrage through danger, indirect effect = .22, 95% CI [.10, .35], and the direct effect was nonsignificant, direct effect = .14, $t(265) = .137$. These results supported our model, revealing that when users witnessed opponents' posts spread, it made them appear more dangerous, which then predicted increases in intentions to express outrage.

We also explored whether virality could amplify threats discussed by tweets that users agreed with. Our naturalistic studies found that viral (vs. less viral) tweets received greater proportions of replies that expressed outrage, even from allies. This could simply reflect the fact that political infighting is also common (i.e., Seeing posts from allies whom we disagree with "go viral" may feel threatening). However, we also predicted that virality makes the content of posts seem more noteworthy, which could amplify threats even if we agree with the post (i.e., If people are sharing a post about a threat, then we may infer that the threat is important). Supporting this idea, high virality tweets also made the issues discussed by tweets seem more important, $b = .78$, $t(947) = 8.49$, $p < .001$, and the perceived virality of tweets predicted these ratings of importance, $b = .21$, $t(1084) = 10.23$, $p < .001$. As theorized, perceptions of virality were closely related to the perceived importance of the tweet.

Since virality caused tweets to generally seem more important, we also thought that higher virality metrics could amplify whatever threats allies are discussing and evoke outrage about the threat. We found partial support for this idea. Virality amplified general

[3] We predicted that virality would also increase subjective feelings of outrage in our preregistered hypotheses. We found support for this hypothesis, though effects of virality on subjective feelings of outrage were not as consistent as the effects on our key indicators for feelings of danger. We report these analyses in Supplemental Materials.

Table 1
Potential Threats Examined in Each Experiment and Example Stimuli

Study (Demographic)	Tweets rated (Number at low and high virality)	Example tweet text
Study 4 (N = 319; M_{age} = 39.91, 152 women, 162 men, three other gender, two no response)	Political opponents (one low virality and one high virality) and Political allies (one low virality and one high virality)	(Example of liberal tweet) “Police violence shows no sign of slowing down in the US. Every month we see another video of officers abusing their power. They’re on camera and they’re doing it more than ever.”
Study 5 (N = 102; M_{age} = 38.89, 50 women, 52 men)	Prejudiced (e.g., anti-Black, anti-Muslim, sexist) (two low virality, two medium virality, and two high virality)	“Blacks and Whites just don’t go together. People just need to admit that we’re fundamentally different, and we never should’ve let them into our schools”
Study 6 (N = 189; M_{age} = 38.11, 91 women, 97 men, one no response)	Annoying (see right) (one low virality and one high virality) and Prejudiced (from Study 6) (one low virality and one high virality)	“Yaayy, my dad FINALLY bought me the car I’ve been bugging him about for months for my birthday!!! I was honestly starting to get tired of asking him about it lol. Don’t accept anything but the best! Haha”
Study 7 (N = 396; M_{age} = 38.27, 193 women, 199 men, four other gender)	Novel harmful behavior (one low virality or one high virality)	“I got him dizzy as hell with this laser pointer. Wait till he starts crashing into everything at the end haha #dizzydogging”
Study 8 (N = 493; M_{age} = 45.38, 239 women, 249 men, five other gender)	Harmful behavior (one low virality or one high virality)	“Huffing gasoline is actually legit. Y’all gotta try this shit. Hit me up after school. Bye bye brain cells”

Note. Participants were collected from Mechanical Turk via CloudResearch, requiring workers be located in the United States and have a Human Intelligence Task approval rate of 95%. Participants in Studies 4–7 also had to have a Twitter account. Studies 4–6 were within subjects; Studies 7 and 8 were between subjects. In Supplemental Materials, we report an additional experiment that finds virality increases intentions to “speak up” in response to tweets from political opponents. This experiment used different tweet materials but did not measure feelings of danger or intentions to express outrage.

feelings of danger about issues,[4] b = .43, t(944) = 4.25, p < .001, regardless of whether the post was from a political ally or opponent, test of interaction: b = −.15, t(941) = −.78, p = .434. These feelings of danger also predicted intentions to express outrage about the issue (while controlling for the virality manipulation), b = .30, t(1099) = 18.48, p < .001. However, higher virality did not significantly increase intentions to express outrage about the issues discussed by tweets, b = .11, t(947) = 1.73, p = .085, and this effect was also not moderated by whether the post was from a political ally or opponent, test of interaction: b = .07, t(942) = .51, p = .613. One possible reason for this is that seeing viral posts from allies made participants feel like the threat was important (increasing feelings of danger) but also assured participants that the threat was already being addressed by others (reducing the urgency to express outrage). This is also consistent with the weaker effects of virality among copartisan interactions in Studies 1–3.

Discussion

This study confirmed our prediction that increases in virality metrics would increase perceived virality and that these increases in perceived virality would make posts appear more noteworthy. We also found support for key steps in our model: when posts from political opponents appeared to be spreading, they appeared more dangerous and evoked more intentions to express outrage, and danger mediated this effect upon intentions to express outrage.

These results also helped clarify how virality may amplify threats from posts we agree with: it causes the issues discussed within posts to feel more dangerous. These feelings of danger also predicted intentions to express outrage about the issue, suggesting that they may translate into outrage expression. However, in the present study, the overall effect of virality did not significantly increase intentions to express outrage about issues. There are multiple possible explanations for this. Perhaps seeing highly viral posts from allies both amplifies the threats they discuss while also providing some reassurance that the threat is being addressed. People could also be less likely to express outrage without a clear target (i.e., Our items here asked about intentions to express outrage about the issue or toward “someone”). Nevertheless, these results, along with the patterns we observed in our naturalistic studies, reveal that virality can amplify threats in multiple ways. In our remaining experiments, we replicate and extend the context in which we have observed the strongest effects thus far: the viral spread of content that poses potential threats.

Study 5: Viral Prejudice

Watching the ideas of our political adversaries “go viral” is threatening and evokes outrage, but we do not expect these effects of virality to be limited to partisan competition. We next tested whether virality would also amplify threats from a more general harmful

[4] We also examined the effects of virality on feelings of danger and intentions to express outrage toward tweets from allies. Here, whether the tweet was authored by a political ally or opponent did moderate feelings of danger from the tweet, b = .40, t(943) = 2.60, p = .009, and intentions to express outrage toward the tweet, b = .39, t(943) = 2.95, p = .003. Unsurprisingly, when tweets from allies were more viral, the tweets themselves did not seem any more dangerous, b = .01, t(943) = 1.01, p = .311, nor did participants report increased intentions to express outrage against allied tweet b = .01, t(943) = 0.06, p = .953. In other words, virality caused tweets from opponents to seem more dangerous (see results in the main text) but did not cause tweets from allies to seem more dangerous (though virality did cause the issues mentioned by allies to seem more dangerous).

behavior on social media: expressions of prejudice. This study also tested whether participants were sensitive to the degree of virality, an important question given that real-world virality is continuous. We tested whether posts that received only hundreds of retweets would amplify danger and outrage to the same degree as posts that received thousands of retweets.

Method

Participants (N = 102; M_{age} = 38.89, 50 women, 52 men) each rated—in random order—two high virality tweets, two low virality tweets, and two moderate virality tweets. We manipulated high and low virality the same as we did in Study 4. The moderate virality tweets received 200–400 retweets. We used the same measures of danger and intentions to express outrage as in Study 4. To check whether our manipulation successfully affected perceived virality, we asked participants how many people they thought saw the tweet (two items, e.g., "How many people are likely to see the above tweet?"). Study 5 also included additional measures of emotional reactions, subjective feelings of outrage, and a more general measure of participants' intentions to speak up (two items, e.g., "If you saw this on Twitter, would you feel the need to speak up?"; the results for these additional measures are reported in Supplemental Materials). After seeing each tweet, participants completed the manipulation check, measure of feelings of danger, emotional reactions to tweets, intentions to reply to the tweet, and then intentions to reply with outrage.

Results

Our manipulation of virality metrics again increased perceived virality. These results also suggested that participants were sensitive to the degree of virality. The effect of the high (vs. low) virality, b = 4.31, t(508) = 56.32, $p < .001$, was larger than the effect of moderate (vs. low) virality, b = 2.07, t(508) = 27.09, $p < .001$. This difference between the high and moderate virality tweets was significant, b = 2.24, t(504) = 29.31, $p < .001$.

We first tested our model by comparing the high virality (vs. low virality) conditions. Supporting our predictions, virality increased ratings of danger posed by prejudiced behavior, b = .87, t(504) = 7.20, $p < .001$, 95% CI [.63, 1.11]; intentions to express outrage toward the prejudiced users, b = .44, t(504) = 3.58, $p < .001$, 95% CI [.20, .68]; and ratings of danger predicted intentions to express outrage, b = .42, t(550) = 11.21, $p < .001$, 95% CI [.35, .50]. Formal tests of mediation again revealed a significant indirect effect of virality on intentions to express outrage through danger, indirect effect = .54, 95% CI [.32, .79], and the direct effect was nonsignificant, direct effect = −.06, t(99) = .38, p = .702. These results show that virality also amplifies threats from—and outrage in reply to—a common harmful behavior on social media: expressions of prejudice.

To test whether intentions to express outrage were sensitive to the degree of virality, we also examined tweets with hundreds (instead of thousands) of retweets. Compared to low virality tweets, these moderately viral tweets had a small but nonsignificant effect on intentions to express outrage, b = .23, t(504) = 1.91, p = .057, 95% CI [−.01, .48]. The difference between moderately viral and highly viral tweets was also nonsignificant, b = .21, t(503) = 1.69, p = .093, 95% CI [−.03, .45]. These results provide suggestive evidence that participants' intentions to express outrage may depend not just on the presence of but also on the degree of virality.

Discussion

This study replicated and extended the results from Study 4, revealing that virality amplified the perceived threat of a prejudiced tweet. We also found these effects to be more robust at higher levels of virality (i.e., thousands of retweets vs. hundreds).

Study 6: Potential Threats Versus Annoying Behavior

Virality can make any content look more noteworthy and influential, but we also thought this would primarily increase feelings of danger and outrage expression for content that is connected to threats that people are concerned about (just as fears of Satanism spreading across society primarily affected people who feared the devil). Study 6 included tweets that we thought would be less likely to be perceived as potential threats (Step 1 in our model), making them less likely to be affected by virality (Step 2). We compared these to the prejudiced tweets from Study 5. Study 6 also used a new measure of feelings of danger, designed to better capture feelings of instability in society.

Method

Participants (N = 189; M_{age} = 38.11, 91 women, 97 men, one no response) each rated—in random order—two high virality tweets and two low virality tweets using the same manipulation from Study 4. Participants saw two tweets (one of the high virality and one of the low virality tweets) expressing prejudice (taken from Study 5) and two tweets that were merely annoying (like spoiled teenagers; see Table 1 for example materials). We used the same measures of intentions to express outrage as in Studies 4 and 5, but used a different measure of feelings of danger adapted from Altemeyer (1988; three items, e.g., "The above Tweet makes me feel that any day now chaos and anarchy could erupt around us"), which we thought may better capture generalized feelings of societal danger. We used the same manipulation check from Study 5. We again included additional measures of emotional reactions, subjective feelings of outrage, and the same moral conviction items used in Study 4 (see Supplemental Materials for results). After seeing each tweet, participants completed manipulation checks, items measuring moral convictions, emotional reactions to tweets, feelings of danger, and intentions to reply with outrage.

Results

As in Studies 4 and 5, our manipulation successfully increased perceived virality. The high (vs. low) virality condition significantly increased perceived virality, b = 4.27, t(560) = 85.79, $p < .001$.

Study 6 again used mixed-effect models with random intercepts for participant and stimuli ID, while also including an interaction between virality (high vs. low) and tweet type (prejudiced vs. annoying). Categorical variables were effect-coded (i.e., low virality = −.5, high virality = .5). The results of this model revealed a main effect of virality on feelings of danger, b = .29, t(560) = 3.71, $p < .001$, 95% CI [.14, 45]; and intentions to express outrage, b = .19, t(560) = 2.18, p = .030, 95% CI [.02, 36]. To our surprise, we observed no interaction between virality and tweet type for feelings of danger, $b = .09$, $t(560) = .56$, $p = .574$, 95% CI [−.22, .40]; or for intentions to express outrage, $b = .11$, $t(561) = .62$, $p = .536$, 95% CI [−.24, .45]. This potentially suggested that virality may cause even annoying content to feel dangerous and deserving of outrage. We again found that feelings of danger predicted intentions to express outrage, $b = .32$, $t(541) = 8.93$, $p < .001$, 95% CI [.25 .40], and formal tests of mediation found that the main effect of virality on intentions to express outrage was mediated by feelings of danger, indirect effect = .09, 95% CI [.04, .15]; direct effect = .10, $t(186) = 1.73$, $p = .085$.

Next, we further explored whether virality truly had the same effects for tweets expected to be more concerning (i.e., prejudiced tweets) versus less concerning (i.e., annoying tweets) to participants in Study 6. We did predict that virality could make any content look more noteworthy and influential. It is also true that our minds are flexible in what they can perceive to be societal threats (e.g., If today's youth think that being a spoiled brat is acceptable, what might happen to the future of society?). Still, we did not expect virality to cause feelings of danger and outrage expression as strongly for annoying behavior as it did prejudice. So, we next explored whether this unexpected pattern could be due to the fact that some participants did not find prejudice to be a potential threat.

We conducted an exploratory test of a three-way interaction between virality, threat type, and participant ideology (which was measured on a 1–7 scale from very liberal to very conservative in the demographic sections of our experiments). Since liberals are more concerned about prejudice toward the groups that were targeted in our tweets (i.e., Muslims, women, and Black People; see Neal, 2017), we thought that we may observe our predicted results among liberals. This three-way interaction was significant for feelings of danger, $b = -.20$, $t(558) = -2.33$, $p = .020$. Among liberals, virality increased feelings of danger for prejudiced content, $b = .60$, $t(557) = 3.81$, $p < .001$, but not for annoying behavior, $b = .14$, $t(557) = .92$, $p = .358$. Among conservatives, virality did not increase feelings of danger for prejudiced content, $b = .08$, $t(557) = .50$, $p = .618$, but did for annoying behavior, $b = .35$, $t(557) = 2.28$, $p = .023$. For intentions to express outrage, the three-way interaction was not significant, $b = -.13$, $t(560) = -1.33$, $p = .185$, but the simple effects followed a similar pattern: among liberals, virality increased outrage expression for prejudiced, $b = .37$, $t(558) = 2.15$, $p = .032$, but not annoying content, $b = .03$, $t(558) = .20$, $p = .843$. Though these additional analyses were exploratory, they demonstrate how subjective beliefs and concerns may place boundary conditions on the effects of virality. Among people who are especially concerned about threats from prejudice in the United States (liberals vs. conservatives), virality had significantly stronger effects on feelings of danger for prejudiced tweets. This is consistent with our predictions of how virality should affect different types of threats.[5]

Discussion

Study 6 replicated tests of steps 2–4 in our model, using a measure of feelings of danger that better captured rising feelings of threat and instability in society. We also tested the interaction between potential threats and virality (Steps 1 and 2), which proposed that virality should be more likely to amplify threats that people actually care about. We expected virality to consistently amplify threats from prejudiced tweets compared to annoying tweets. But this pattern was only true for liberals, and it reversed for conservatives. These results highlight how threats from overtly harmful behaviors—like expressions of racial and religious prejudice—are perceived subjectively. As research has long demonstrated (Kasperson et al., 1988), social amplification is a complex process, shaped both by the characteristics of the threat and of the perceiver.

Study 7: Creating a Novel Threat

People's pre-existing beliefs appear to affect how people respond to spreading threats—including ideas from political opponents and expressions of prejudice. However, would people react similarly to a potentially harmful behavior in the absence of pre-existing beliefs? We next tested whether virality could amplify a novel threat that we invented—a fake social media trend called "#dizzydogging." This posed a "purer" test of our model, free from the influence of pre-existing ideological biases. This study also aimed to replicate the results from Studies 4–6 using a between-subjects design.

Method

Participants ($N = 396$; $M_{\text{age}} = 38.27$, 193 women, 199 men, four other gender) were randomly assigned to either the low virality or high virality condition. Both conditions showed participants a headline from a Buzzfeed article about a potentially harmful social media trend, which we then either depicted as being shared widely (the high virality condition) or not (the low virality condition). This headline was embedded within five other real headlines from BuzzFeed, and participants rated how likely they would be to click on each. The target headline discussed a potential trend on social media that involved dog owners posting videos of themselves making their dogs dizzy and sometimes hurting themselves.

Participants then evaluated a distractor tweet about one of the Buzzfeed articles they had just seen (i.e., a photo of researchers involved in the Human Genome Project). They then saw a tweet that read, "I got him dizzy as hell with this laser pointer. Wait till he starts crashing into everything at the end haha. #dizzydogging." The post also contained an image of a paused video with a dog spinning and chasing a laser pointer. In the low virality condition, the tweet had one retweet and one like. In the high virality condition, the tweet had "4,367" retweets and "16.2K" likes (the formatting of these values matched how Twitter formats numbers over 10,000).

Study 7 used the same measures as Study 6 (except for using the manipulation check from Study 4, "How much has the above tweet been shared across Twitter?"). After each tweet, participants completed manipulation checks, then rated feelings of danger, intention to reply with outrage, emotional reactions, and moral convictions.

Results

Because Study 7 was a between-subjects design, the results reported are from linear regression (as opposed to the mixed-effects models used in our previous experiments).

[5] Since Study 5 contained the same prejudiced tweets used in Study 6 (plus three additional prejudiced tweets), we also retrospectively examined whether ideology moderated these results. These results followed a similar pattern as Study 6, as the effects of virality on intentions to express outrage were significant for liberals ($b = .61$, $p < .001$) but not conservatives ($b = .27$, $p = .123$). But the overall interaction was nonsignificant ($b = -.10$, $p = .161$).

Results showed that our virality manipulation again increased perceptions of virality, high virality: M = 5.99, SD = 1.12; low virality: M = 1.83, SD = .75; b = 4.16, t(676) = 43.64, p < .001. We also found support for Steps 2–4 in our model. The high virality condition (M = 2.67, SD = 1.80) increased feelings of danger compared to the low virality condition, M = 2.17, SD = 1.55; b = .50, t(381) = 2.96, p = .003, 95% CI [.17, .83]; increased intentions to express outrage, high virality: M = 2.87, SD = 2.08; low virality: M = 2.27, SD = 1.73; b = .61, t(373) = 3.15, p = .002, 95% CI [.23, .98], and feelings of danger predicted intentions to express outrage, b = .62, t(392) = 12.96, p < .001, 95% CI [.53, .72]. Formal tests of mediation again showed feelings of danger mediated the effect upon intentions to express outrage, indirect effect = .31, 95% CI [.10, .53]; direct effect = .29, t(392) = 1.77, p = .078.

We also predicted that #dizzydogging should pose a novel threat that has not been politicized. We thought that neither liberals nor conservatives should be more concerned about the threat of #dizzydogging. Thus, political ideology should not moderate the effects of virality (as it did with tweets expressing prejudice). Supporting this prediction, ideology did not moderate the effects of virality on feelings of danger, b = .04, t(387) = .39, p = .695, 95% CI [−.15, .22] or intentions to express outrage, b = .07, t(386) = .69, p = .491, 95% CI [−.14, .29].

Discussion

Study 7 found support for our model using a novel threat and a between-subjects design. Study 7 also confirmed our predictions that the effects of virality should not be limited to political conflict (e.g., liberals reacting to conservatives' posts spreading). We replicated the effects of virality on feelings of danger and intentions to express outrage for a novel threat that should not be associated with pre-existing ideological biases. Indeed, participants' ideology did not moderate the effects of virality.

Study 8: Comparing Different Responses to Viral Threats

Virality amplifies outrage expression, but outrage is only one reaction to potential threats. In our final experiment, we measured a larger range of possible replies (including worried and sad replies, as well as two types of outrage expression) and tested whether outraged replies were specifically caused by viral threats. We also included new measures of perceived virality to further explore how participants interpret virality metrics.

Method

As in Study 7, this study used a between-subjects design. Participants (N = 493; M_{age} = 45.38, 239 women, 249 men, five other gender) were randomly assigned to either a high or low virality condition, where they evaluated a single social media post. In both conditions, participants read, "Below is a recent post on X (formerly known as Twitter). Look at it carefully and answer the questions that follow." The post was authored by someone trying to convince students to huff gasoline, which we chose since harmful adolescent behavior is a common source of moral panic. The post read: "Huffing gasoline is actually legit. Y'all gotta try this shit. Hit me up after school. Bye bye brain cells." We manipulated virality in the same fashion as Study 7, using the same metrics but with the labels altered to reflect recent changes on Twitter/X (i.e., formerly the word "Retweets" appeared next to the number of shares, but now this metric is called reposts "Reposts"; "Quote Tweets" was changed to "Quotes").

After rating the post, participants completed new measures of perceptions of virality. Most of our studies showed multiple virality metrics for each post, manipulating metrics of retweets and likes proportional to one another. But these metrics could be interpreted in different ways, and thinking a post is widely liked could have different effects than thinking a post is influencing people. To understand how each of these perceptions contributes to our effect, we asked participants, "How much do you think this post was widely liked?" and "How much do you think this post influences people?" (anchors: 1—not at all to 7—very much).

Participants then completed the same measure of feelings of danger from Studies 6 and 7 and new measures of reply intentions. To measure general intentions to respond to the post, we asked, "If you saw this on Twitter/X, how likely would you be to post something about it?" (from 1—not at all likely to 7—very likely). We then asked, "Say you were to reply to the above post. How similar are each of the following to what you would say?" (anchors 1—not at all similar to what I would say, 7—very similar to what I would say). Below this question were five possible replies. Two reply options were negative but nonoutraged. One expressed sadness (e.g., "Seeing this just makes me sad") and one expressed worry (e.g., "This is really concerning"). We included two different ways that people may express outrage, both of which expressed emotional condemnation. One type of outrage focused more on a blanket condemnation of the entire situation (e.g., "What the hell?! This is terrible"), and the other focused on condemning the character of the tweeter (e.g., "You seriously disgust me"). We refer to the former type of outrage expression as "outrage (overall)" and the latter as "outrage (personal)." We also included a positive reply option (e.g., "This is amazing!"). Participants rated one of each type of reply in random order.

To help ensure that the results were not affected by the wording of each reply option, participants saw one of three different variations at random. We also conducted a small pilot study to make sure each reply actually expressed sadness, worry, and outrage as we expected. We showed participants in the pilot study the same tweets used in this study, along with each of the reply options from our total pool (three options for each of the five types). We asked, "The person who wrote this reply felt …," followed by four items: "Sad," "Worried," "Outraged," and "Happy" (rated from 1—not at all to 7—very much). As expected, participants rated both types of outraged replies significantly higher on outrage than the other types of replies (see Table 2 for results). Similarly, the worried reply options scored significantly higher on "worried" than the other reply options, and the sad options scored significantly higher than all other reply options on "sad."

Results

We again tested our full model: whether virality caused feelings of danger, whether virality caused outrage expression, and whether feelings of danger predicted outrage expression. In contrast to our previous experiments, we tested the effect of condition on intentions to express outrage by first examining general intentions to respond and then examining the type of reply that participants' would write

Table 2
Pilot Study Results and Example Reply Options From Study 8

	"The person who wrote this reply felt …"			
	Outraged	Worried	Sad	Happy
Reply type	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Outraged overall: "What the hell?! This is terrible"	5.93 (1.40)$_{a}$	4.78 (1.89)$_{a}$	4.10 (1.96)$_{a}$	1.41 (1.12)$_{a}$
Outraged personal: "You seriously disgust me"	6.04 (1.47)$_{a}$	3.70 (1.96)$_{b}$	3.35 (1.93)$_{b}$	1.39 (1.13)$_{a}$
Worried: "This is really concerning"	3.70 (1.82)$_{b}$	6.18 (1.26)$_{c}$	4.67 (1.77)$_{c}$	1.42 (1.12)$_{a}$
Sad: "Seeing this just makes me sad"	3.40 (1.86)$_{c}$	5.52 (1.65)$_{d}$	5.86 (1.52)$_{d}$	1.49 (1.23)$_{a}$
Positive: "This is amazing!"	1.53 (1.29)$_{d}$	1.51 (1.27)$_{e}$	1.48 (1.23)$_{e}$	5.84 (1.57)$_{b}$

Note. The leftmost column contains one example of three variations of each reply type. Participants in this pilot study rated all 15 of the reply options, and the means in this table are the average of all three variations for each reply type. Participants in Study 8 were shown one reply option of each type at random. Means within the same column that do not share a subscript letter in common are significantly different from one another at $p < .05$.

(if they were to write a reply).[6] This enabled us to separate general intentions to respond to the threat in some way from the specific ways that people may reply to threats on social media. Because participants saw one of three possible variations of the reply options (i.e., We wrote three variations of each reply type), all analyses of reply intentions included a random intercept for the variant that participants saw. Effects on other outcome variables were estimated in simple linear regression.

We again found that higher virality metrics significantly increased feelings of danger, $b = .73$, $t(490) = 4.33$, $p < .001$, 95% CI [.39, 1.06]. We also found that virality increased general intentions to respond to the post, $b = .29$, $t(479) = 2.26$, $p = .024$, 95% CI [.04, .55], and feelings of danger predicted these intentions to respond (while controlling for the virality manipulation), $b = .14$, $t(477)$, $p <$ .001, 95% CI [.08, 21]. Virality caused feelings of danger and made participants feel more compelled to respond to the post in some way. We next examined whether these intentions to respond manifested as outrage specifically.

Participants in the high virality (vs. low virality) condition were significantly more likely to say that the outraged (overall) replies reflected how they would respond to the post, $b = .47$, $t(487) = 2.38$, $p = .018$, 95% CI [.08, .85]. Higher virality metrics also significantly increased sad replies, $b = .42$, $t(488) = 2.23$, $p = .026$, 95% CI [.05, .78], suggesting that people may not respond to viral threats exclusively with outrage. The effects on worried, $b = .30$, $t(489) = 1.61$, $p = .108$, 95% CI [−.06, .67], and positive replies, $b = .01$, $t(487) = .133$, $p = .894$, 95% CI [−.16, .18], were not significant. Virality also did not significantly increase the outraged (personal) replies, $b = .04$, $t(490) = .19$, $p = .846$, 95% CI [−32, .39]. This finding suggests that—despite increasing people's willingness to express outrage about the overall situation—people were more reluctant to express outrage that attacked the post author personally. This may be explained by the target of these replies: an adolescent. The phenomenon of moral typecasting suggests that people are reluctant to personally blame those who evoke sympathy (Gray & Wegner, 2009). In this case, people appear to feel bad for the author of the post but nevertheless express outrage at the overall situation in line with their feelings of threat.

Results from this study confirm that virality specifically causes intentions to express outrage, but they also suggest that people may feel and express multiple different emotions in the face of viral threats. Examining the relationship between feelings of danger and each type of reply supported this idea. Feelings of danger predicted each type of reply, outrage (overall): $b = .54$, $t(489) = 11.45$, $p < .001$; outrage (personal): $b = .53$, $t(489) = 12.19$, $p < .001$; worried: $b = .53$, $t(487) = 12.03$, $p < .001$; and sad: $b = .50$, $t(487) = 11.15$, $p < .001$, except for positive replies, $b = −.004$, $t(487) = −.18$, $p = .858$. Though feelings of danger help explain outrage on social media, they may also help explain other negative, emotional reactions to threats on social media as well.

Similar to Study 7, we also did not expect teenagers huffing gasoline to be a politicized issue. Supporting this prediction, ideology did not moderate the effects of virality on feelings of danger, $b = −.08$, $t(479) = −.90$, $p = .367$, 95% CI [−.27, .10]; outraged replies, $b = −.09$, $t(480) = −.85$, $p = .396$, 95% CI [−.31, .12]; or outraged (personal) replies, $b = −.12$, $t(479) = −1.20$, $p = .230$, 95% CI [−.32, .08]. The findings from this study provide further evidence that virality amplifies threats and outrage outside of the context of partisan conflicts.

Last, we further explored how virality metrics affected different possible perceptions of virality, including how much they caused people to think the post was widely liked and influential. We were especially interested in whether one of these perceptions was better at explaining the effects of virality on outrage. Our virality manipulation significantly increased both perceived liking, $b = 3.80$, $t(486) = 33.79$, $p < .001$, 95% CI [3.58, 4.02], and perceived influence, $b = 2.53$, $t(485) = 19.88$, $p < .001$, 95% CI [2.28, 2.78]. These two perceptions were also highly correlated ($r = .79$). However, perceived influence appeared to be a stronger predictor of outraged (nonpersonal) replies, $b = .25$, $t(483) = 4.84$, $p < .001$, 95% CI [.15, .35], than perceived liking, $b = .13$, $t(485) = 2.99$, $p = .003$, 95% CI [.04, .21] (in two separate regression models). And perceived liking did not significantly predict outrage (personal) replies, $b = .06$, $t(484) = 1.62$, $p = .106$, 95% CI [−.01, .14], but perceived influence did, $b = .18$, $t(483) = 3.75$, $p < .001$, 95% CI [.09, .27] (again, in separate regression models). This may suggest that virality metrics

[6] Our preregistered analyses originally planned to combine the measures for our two types of outraged replies and the two types of negative, nonoutraged replies. The analyses reported here—which examine the effects on each of the five reply types individually—was listed as an additional analysis in our preregistration. However, since results revealed participants responded differently to these reply types, we decided to report the analyses for each reply type separately.

drive outrage by making harmful posts appear to be influencing people. If this is true, then this could have implications for the design of virality metrics: metrics that are better at making content look influential and impactful (vs. widely liked) may be the most powerful drivers of outrage.

Discussion

This study found further support for our model and better revealed how people respond to viral threats. Perceived virality caused people to express outrage about the behavior, but not outrage that personally attacked the tweet author. Further, virality also increased expressions of sadness about the harmful post. These findings could be partly explained by the specific post we used, which suggested the target was someone still in school. Perhaps in political contexts, we would see less sadness in replies and more personal attacks. Nevertheless, the post we examined in this study resembles many real-world moral panics over adolescent drug use (or other behaviors like eating Tide Pods). These findings support our prediction that viral threats cause outrage while also providing an important reminder that outrage is not the only way people may respond to seeing harmful behavior spread and influence others (see Figures 5–7 for summaries of results from experiments).

Discussion: Experiments

Five experiments provided further evidence that virality on social media amplifies perceptions of threat and evokes outrage expression. Virality caused intentions to express outrage in replies to potentially threatening tweets (Steps 2–4), caused feelings of danger (Steps 2–3), and feelings of danger consistently predicted intentions to express outrage (Steps 3–4). This supports the idea that outrage in reply to viral messages at least partly stems from feelings of danger, which virality amplifies. Results from Study 4 also suggested viral tweets about threats—whether from allies or from opponents—can increase feelings of danger, but users' outrage typically targets specific (ideologically opposed) users. Study 8 replicated the effect of virality on intentions to express outrage, even when people had multiple negative reply options. However, Study 8 also found an increase in sad replies, suggesting that viral threats may trigger multiple types of reactions.

Results also provided insights into the types of threats affected by virality. We found that the effects of virality generalized beyond political contexts (Studies 7 and 8)—helping to rule out the possibility that people were expressing outrage merely to derogate political opponents. Exploratory analyses from Study 6 also supported our prediction that virality would have stronger effects on threats that are most concerning to people. Just as we found stronger effects of virality for liberals among climate change topics in our naturalistic data, Study 6 revealed stronger effects for tweets about prejudice (toward Muslims, women, and Black people) among liberals (issues that are especially concerning to liberals in the United States).

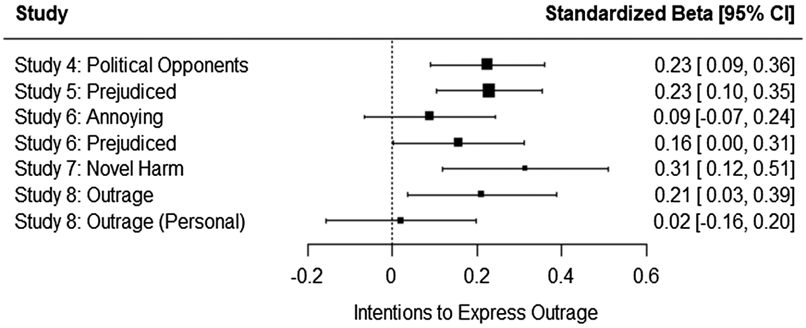
General Discussion

Moral panics seem pervasive on social media, but there is a lack of consensus on what moral panics are, what drives them, and whether they help explain outpourings of outrage on social media (Crockett, 2017; Garland, 2008; Hier, 2008; Ungar, 2001). Synthesizing work from sociology and cognitive science, we distill moral panics to four elements—potential threats, social amplification, feelings of danger, and moral punishment. We then describe how social media accelerates moral panic by providing abundant potential threats (Step 1), providing signals of virality that tell us threats are important and influential (Step 2), causing people to feel society is in danger (Step 3), and reducing the risks and effort necessary to address these threats by deploying moral outrage against those responsible (Step 4).

Eight studies—including naturistic studies on Twitter and experiments—support the role of virality in amplifying threats and suggest that digital outrage partly takes root in feelings of danger. Three studies of real-world Twitter discussions of societal threats found that more viral tweets received replies that were more likely to express outrage (the path from Step 2 to 4 in our model). Five experiments then found that identical content evoked more feelings of danger and intentions to express outrage when the

Figure 5

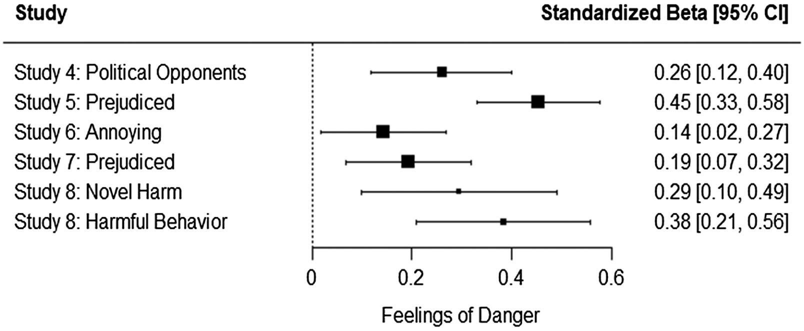
Effects of Virality on Intentions to Express Outrage



Note. We standardized regression coefficients to aid comparisons across studies, but results in text report unstandardized coefficients. The label for each study in the left column represents the type of threat used in that study, except for Study 8, which reflects the two types of outrage expression examined in that study.

Figure 6

Effects of Virality on Feelings of Danger

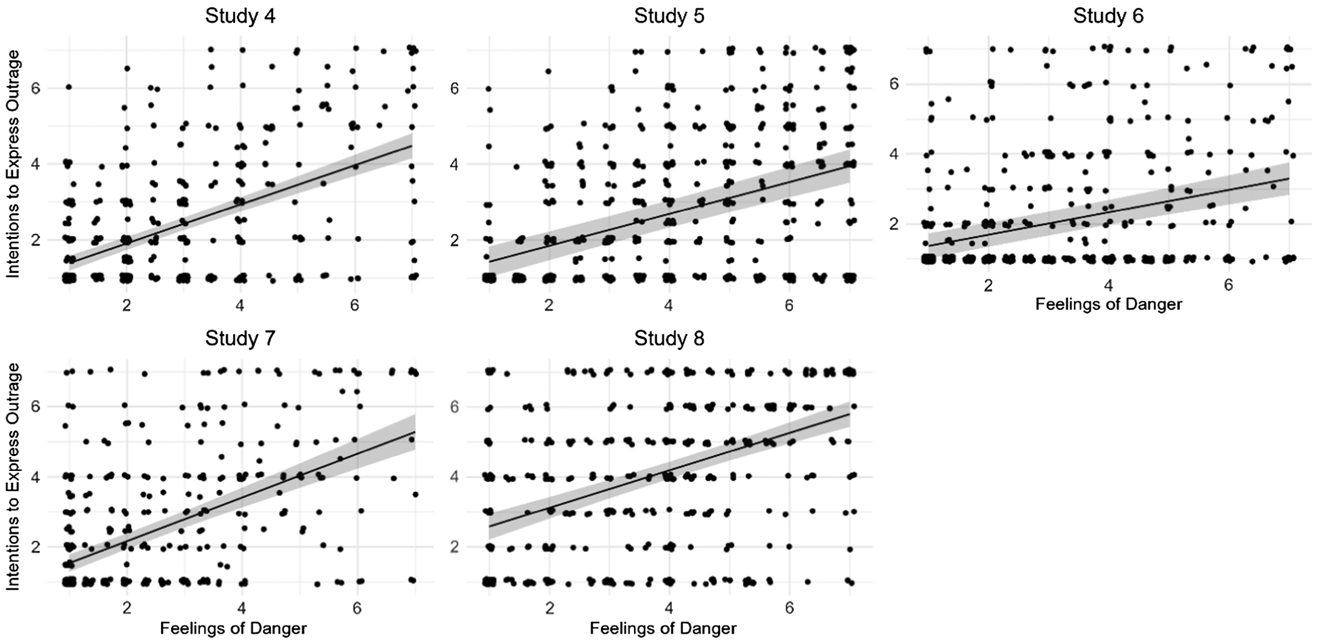


content was portrayed as highly viral (vs. not viral). We also found support for our prediction that virality would be especially likely to amplify threats that are relevant to users' concerns. In our naturalistic studies, more viral tweets about threats that should be especially concerning to users (e.g., climate change to liberals) corresponded with especially large increases in outrage among replies. In Study 6, viral tweets expressing prejudice toward minorities caused feelings of danger, specifically among liberals. Together, these findings provide converging evidence that signals of virality on social media amplify threats and that outrage expression on social media is partly fueled by feelings of danger.

The present work provides insight into moral panics and the nature of social media outrage. The term "moral panic" is sometimes used dismissively to characterize societal reactions as overblown or unreasonable (Garland, 2008), leading some to suggest that the term is more of a political tool than a valid scientific concept (Waddington, 1986). However, semantic debates over the term "moral panic" can obscure an important sociopsychological phenomenon: when societies focus their attention and concern on a specific threat, causing eruptions of outrage to punish the moral deviants who propagate those threats. This phenomenon is neither necessarily rational nor irrational. Panicking and expressing outrage to stop moral threats is no more an irrational overreaction than panicking and running away from physical threats. To explain this phenomenon, we have developed a model that synthesizes research on human threat detection (Blanchard et al., 2011), social amplification (Kasperson et al., 1988; Renn, 2011), and the functions of moral punishment (Darley & Pittman, 2003; Rucker et al., 2004). The present work reveals how feelings of danger—across a wide array of real and imagined threats—drive expressions of moral outrage, suggesting that the eruptions of outrage on social media at least partly reflect how societies have long adapted to emerging threats.

Figure 7

Feelings of Danger Predict Outrage Expression Across Experiments



Implications of Viewing Digital Outrage Through the Lens of Moral Panic

If digital outrage reflects genuine distress and panic, then it may also reflect underlying costs to well-being (e.g., distress; Twenge et al., 2019) and increasing punitiveness in society (Gelfand & Lorente, 2021). We did conduct one preliminary, correlational study of the potential costs of moral panics on well-being (see Supplemental Study 2). This study found (*N* = 299; 122 women; 173 men; four other gender) that social media use predicted greater feelings of danger, perceptions of extremist groups in society, and past outrage expression on social media, but only among participants who also said they paid attention to social media trends and virality metrics. Moreover, this combination of social media use and virality checking also predicted general distress. In other words, the people who report expressing outrage not only report greater perceptions of danger and extremism in society, they are also significantly more distressed. The results from this preliminary study—and the other studies reported here—suggest that these feelings of distress and danger are partly fueled by viral threats on social media.

Potential effects of moral panics on punitive attitudes and well-being have important implications for society, but what percentage of digital outrage reflects moral panic? Inferring the motives behind each expression of outrage on social media is impossible, especially since outrage expression stems from multiple motivations. Further, moral panics are a dynamic process, and a wide range of behaviors on social media can help threats spread widely (Kasperson et al., 1988). Each expression of moral outrage—regardless of users' motivation—may also contribute to moral panic by helping threatening content spread (Brady et al., 2020). Precisely identifying which or how much behavior on social media is moral panic may not be possible, but this is not our goal. Instead, the present work contributes to a larger literature identifying the multiple ways that the design of social media can fuel outpourings of outrage.

Future work should also examine how moral panics are self-perpetuating on social media. Given that outrage also helps content achieve higher virality (Brady et al., 2020; Heath et al., 2001), moral panics may be one piece of a larger cycle on social media in which (a) virality signals that something is a dangerous threat, (b) users respond with outrage to mitigate danger and, (c) outrage then reinforces the signal that some behavior or idea must be dangerous, which then fuels further virality. The present results, paired with previous research, provide evidence for the psychological mechanisms that drive both halves of this cycle of virality and outrage.

Statement of Limitations

While we found support for our model using diverse methods and materials, every set of studies comes with limitations (see Table 3 for summaries). We primarily focused on one signal of virality on social media—the explicit metrics of shares and likes that pervade many platforms. Future work may examine other features of social media that signal virality (e.g., trending pages and exposing users to multiple posts about the same topic). To provide causal evidence, we used multiple methods to control for potential confounds in our naturalistic studies (see Supplemental Materials for additional results) and conducted controlled experiments. But future work may strengthen this evidence further with longitudinal analyses in naturalistic data and by manipulating the mediator variable in our model (feelings of danger, which were measured in the present studies). Samples in our experiments were from the United States,

Table 3
Summary of Limitations

Limitation	Summary
Operationalization	Virality: We focused primarily on metrics of shares because they are common signals of virality on social media. But social media contains other signals of virality, like trending tabs or the volume of posts that users encounter. Future work should examine whether—and how much—additional signals of virality fuel moral panic.
Causal inference	Effect of virality: Naturalistic studies provided suggestive but limited evidence that virality causes outraged replies in the real world (using multiple approaches to control for the content of tweets). Experiments provided additional evidence that virality causally amplifies feelings of danger and outrage expression. Future work may use additional methods (e.g., longitudinal analyses) to test the causal effects of virality in naturalistic settings.
	Effects of feelings of danger: Experiments suggested that feelings of danger mediated the causal effect of virality upon outrage expression. These studies were limited by merely measuring the mediator. However, existing work supports a causal effect of danger upon outrage (e.g., Duckitt & Fisher, 2003; Rucker et al., 2004), and we found that feelings of danger were a more consistent mediator than an alternative mechanism—subjective feelings of outrage (see Supplemental Materials).
Generalizability	Samples: Experiments relied on samples from Mechanical Turk in the United States, collected via CloudResearch. This limits our ability to know whether people outside of the United States interpret signals of virality differently or think outrage expression is an effective way to respond to threats. Future work may test these effects in different cultures.
	Social media platforms: We focused primarily on Twitter/X. We expect our model to generalize to other platforms depending on how much they expose users to threats (Step 1 in our model), provide signals of virality (Step 2), and facilitate outrage expression (Step 4). Future work should test whether the present effects generalize to other platforms as our model would predict.
	Context of outrage expression: We focused mostly upon outrage in direct reply to content on social media. This was a straightforward approach to testing how virality affected responses to posts. However, people can express outrage in other contexts (e.g., someone may see a post and then create a separate post later in time). Future work may explore whether expressing outrage in different contexts feels more or less effective at mitigating threats.

and future work should examine whether outrage expression is a common response to threat on social media across cultures. The present methods focus primarily on Twitter/X, though our model outlines how we expect the present effects to generalize (i.e., platforms that combine potential threats with signals of virality). Last, we focused mostly on outrage expression in direct response to other users, but there are multiple ways to post content on social media outside of direct replies. Future work should examine whether expressing outrage in some contexts feels more effective at mitigating threats.

Conclusion

Much research on social media has examined the type of language that helps content go viral, but the type of content that dominates social media may also have important effects on society. We outline a model of one such outcome of virality on social media—moral panic. Social media helps threatening content spread widely, and it provides explicit signals of the viral spread of this content. Information about virality on social media appeals to our minds, which prioritize detecting and interpreting the scope of threats around us. Further, social media provides frequent opportunities to immediately express outrage toward people who are deemed responsible for threats, further facilitating outpourings of outrage. If this enables moral panics to occur more frequently and to repeatedly evoke feelings that society is in danger, then this may have implications for potential shifts in well-being, the frequency of unrest, and support for punitive policies. We hope the present work spurs future research that pursues these questions.

References

Aguirre, B. E. (2005). Emergency evacuations, panic, and social psychology. *Psychiatry: Interpersonal and Biological Processes*, *68*(2), 121–129. https://doi.org/10.1521/psyc.2005.68.2.121

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. Jossey-Bass.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barberá, P. (2020). Social media, echo chambers, and political polarization. In J. A. Tucker & N. Persily (Eds.), *Social media and democracy: The state of the field, prospects for reform* (pp. 34–55). Cambridge University Press. https://doi.org/10.1017/9781108890960.004

Barnidge, M. (2017). Exposure to political disagreement in social media versus face-to-face and anonymous online settings. *Political Communication*, *34*(2), 302–321. https://doi.org/10.1080/10584609.2016.1235639

BBC News. (2014, April 11). *The great 1980s dungeons & dragons panic*. https://www.bbc.com/news/magazine-26328105

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

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, *29*(5), 815–827. https://doi.org/10.1287/mksc.1090.0557

Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience and Biobehavioral Reviews*, *35*(4), 991–998. https://doi.org/10.1016/j.neubiorev.2010.10.016

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., Doan, T. N., & Crockett, M. (2021). *How social learning amplifies moral outrage expression in online social networks*. PsyArXiv. https://doi.org/10.31234/osf.io/gf7t5

Calabrese, C., & Zhang, J. (2019). Inferring norms from numbers: Boomerang effects of online virality metrics on normative perceptions and behavioral intention. *Telematics and Informatics*, *45*, Article 101279. https://doi.org/10.1016/j.tele.2019.101279

Carless, W. (2023, February 1). *Ban drag shows? Analysis finds bills targeting them in 8 states*. https://www.usatoday.com/story/news/nation/2023/02/01/ban-drag-shows-analysis-finds-bills-targeting-them-8-states/11150196002/

Chen, Z., & Berger, J. (2013). When, why, and how controversy causes conversation. *Journal of Consumer Research*, *40*(3), 580–593. https://doi.org/10.1086/671465

Chung, M. (2017). Not just numbers: The role of social media metrics in online news evaluations. *Computers in Human Behavior*, *75*, 949–957. https://doi.org/10.1016/j.chb.2017.06.022

Cohen, S. (1972). *Folk devils and moral panics*. Routledge. https://doi.org/10.4324/9780203828250

Cook, C. L., Li, Y. J., Newell, S. M., Cottrell, C. A., & Neel, R. (2018). The world is a scary place: Individual differences in belief in a dangerous world predict specific intergroup prejudices. *Group Processes & Intergroup Relations*, *21*(4), 584–596. https://doi.org/10.1177/1368430216670024

Cosmides, L., Guzmán, R., & Tooby, J. (2018). The evolution of moral cognition. In A. Zimmerman, K. Jones, & M. Timmons (Eds.), *The routledge handbook of moral epistemology* (pp. 174–228). Routledge. https://doi.org/10.4324/9781315719696-10

Costello, M., Hawdon, J., Ratliff, T., & Grantham, T. (2016). Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior*, *63*, 311–320. https://doi.org/10.1016/j.chb.2016.05.033

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*(4), 324–336. https://doi.org/10.1207/S15327957PSPR0704_05

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*(2), 477–496. https://doi.org/10.1037/a0029065

Duckitt, J., & Fisher, K. (2003). The impact of social threat on worldview and ideological attitudes. *Political Psychology*, *24*(1), 199–222. https://doi.org/10.1111/0162-895X.00322

Fernandes, S., Kapoor, H., & Karandikar, S. (2017). Do we gossip for moral reasons? The intersection of moral foundations and gossip. *Basic and Applied Social Psychology*, *39*(4), 218–230. https://doi.org/10.1080/01973533.2017.1336713

Ford, B. Q., & Feinberg, M. (2020). Coping with politics: The benefits and costs of emotion regulation. *Current Opinion in Behavioral Sciences*, *34*, 123–128. https://doi.org/10.1016/j.cobeha.2020.02.014

Gamez-Djokic, M., & Molden, D. (2016). Beyond affective influences on deontological moral judgment: The role of motivations for prevention in the moral condemnation of harm. *Personality and Social Psychology Bulletin*, *42*(11), 1522–1537. https://doi.org/10.1177/0146167216665094

Garland, D. (2008). On the concept of moral panic. *Crime, Media, Culture*, *4*(1), 9–30. https://doi.org/10.1177/1741659007087270

Gelfand, M. J., & Lorente, R. (2021). Threat, tightness, and the evolutionary appeal of populist leaders. In J. P. Forgas, W. D. Crano, & K. Fiedler (Eds.), *The psychology of populism* (1st ed., pp. 276–294). Routledge. https://doi.org/10.4324/9781003057680-18

Goode, E., & Ben-Yehuda, N. (1994). Moral panics: Culture, politics, and social construction. *Annual Review of Sociology*, *20*(1), 149–171. https://doi.org/10.1146/annurev.so.20.080194.001053

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520. https://doi.org/10.1037/a0013748

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PLOS ONE*, *14*(10), Article e0223749. https://doi.org/10.1371/journal.pone.0223749

Hales, A., Ren, D., & Williams, K. (2017). Protect, correct, and eject: Ostracism as a social influence tool. In S. G. Harkins, K. D. Williams, & J. M. Burger (Eds.), *The Oxford handbook of social influence* (pp. 205–217). Oxford University Press.

Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: The case of urban legends. *Journal of Personality and Social Psychology*, *81*(6), 1028–1041. https://doi.org/10.1037/0022-3514.81.6.1028

Henderson, R. K., & Schnall, S. (2021). Social threat indirectly increases moral condemnation via thwarting fundamental social needs. *Scientific Reports*, *11*(1), Article 21709. https://doi.org/10.1038/s41598-021-00752-2

Hier, S. (2008). Thinking beyond moral panic: Risk, responsibility, and the politics of moralization. *Theoretical Criminology*, *12*(2), 173–190. https://doi.org/10.1177/1362480608089239

Jordan, J. J., & Rand, D. G. (2019). 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

Kahneman, D., & Sunstein, C. R. (2005). Cognitive psychology of moral intuitions. In J. P. Changeux, A. R. Damasio, W. Singer, & Y. Christen (Eds.), *Neurobiology of human values. Research and perspectives in neurosciences* (Vol. 15, pp. 91–105). Springer. https://doi.org/10.1007/3-540-29803-7_8

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X., & Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*, *8*(2), 177–187. https://doi.org/10.1111/j.1539-6924.1988.tb01168.x

Kim, J. (2021). The meaning of numbers: Effect of social media engagement metrics in risk communication. *Communication Studies*, *72*(2), 195–213. https://doi.org/10.1080/10510974.2020.1819842

Kim, J. W. (2018). They liked and shared: Effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Computers in Human Behavior*, *84*, 153–161. https://doi.org/10.1016/j.chb.2018.01.030

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1822), Article 20200143. https://doi.org/10.1098/rstb.2020.0143

Lee-Won, R. J., Abo, M. M., Na, K., & White, T. N. (2016). More than numbers: Effects of social media virality metrics on intention to help unknown others in the context of bone marrow donation. *Cyberpsychology, Behavior, and Social Networking*, *19*(6), 404–411. https://doi.org/10.1089/cyber.2016.0080

Lee-Won, R. J., Na, K., & Coduto, K. D. (2017). The effects of social media virality metrics, message framing, and perceived susceptibility on cancer screening intention: The mediating role of fear. *Telematics and Informatics*, *34*(8), 1387–1397. https://doi.org/10.1016/j.tele.2017.06.002

LoBue, V., & Rakison, D. H. (2013). What we fear most: A developmental advantage for threat-relevant stimuli. *Developmental Review*, *33*(4), 285–303. https://doi.org/10.1016/j.dr.2013.07.005

Mackinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

Mathews, P., Mitchell, L., Nguyen, G., & Bean, N. (2017). *The nature and origin of heavy tails in retweet activity* [Conference session]. Proceedings of the 26th International Conference on World Wide Web Companion, Republic and Canton of Geneva, Switzerland. https://doi.org/10.1145/3041021.3053903

Mehta, A., Formanowicz, M., Uusberg, A., Uusberg, H., Gross, J. J., & Suri, G. (2020). The regulation of recurrent negative emotion in the aftermath of a lost election. *Cognition and Emotion*, *34*(4), 848–857. https://doi.org/10.1080/02699931.2019.1682970

Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(26), 14864–14872. https://doi.org/10.1073/pnas.2001263117

Murray, D. R., Kerry, N., & Gervais, W. M. (2019). On disease and deontology: Multiple tests of the influence of disease threat on moral vigilance. *Social Psychological and Personality Science*, *10*(1), 44–52. https://doi.org/10.1177/1948550617733518

Neal, S. (2017). *Views of racism as a major problem increase sharply, especially among democrats*. Pew Research Center. https://www.pewresearch.org/short-reads/2017/08/29/views-of-racism-as-a-major-problem-increase-sharply-especially-among-democrats/

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, *29*(4), 242–248. https://doi.org/10.1016/j.evolhumbehav.2008.01.001

Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience and Biobehavioral Reviews*, *35*(4), 1042–1051. https://doi.org/10.1016/j.neubiorev.2010.08.011

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. In M. N. Warehime (Ed.), *Soul of society: A focus on the lives of children & youth* (Vol. 18, pp. 253–273). Emerald Group Publishing Limited. https://doi.org/10.1108/S1537-466120140000018021

Puryear, C., Vandello, J. A., & Gray, K. (2023). *Moral panics on social media are fueled by signals of virality*. https://osf.io/wr69q/

Renn, O. (2011). The social amplification/attenuation of risk framework: Application to climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(2), 154–169. https://doi.org/10.1002/wcc.99

Richards, H. J., Benson, V., Donnelly, N., & Hadwin, J. A. (2014). Exploring the function of selective attention and hypervigilance for threat in anxiety. *Clinical Psychology Review*, *34*(1), 1–13. https://doi.org/10.1016/j.cpr.2013.10.006

Rohloff, A. (2011). Extending the concept of moral panic: Elias, climate change and civilization. *Sociology*, *45*(4), 634–649. https://doi.org/10.1177/0038038511406597

Rucker, D. D., Polifroni, M., Tetlock, P. E., & Scott, A. L. (2004). On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin*, *30*(6), 673–684. https://doi.org/10.1177/0146167203262849

Rudert, S. C., & Speckert, K. (2023). You shouldn't have shut them out: Justice sensitivity and norm adherence affect moral reactions to observed ostracism. *Personality and Individual Differences*, *201*, Article 111929. https://doi.org/10.1016/j.paid.2022.111929

Sawaoka, T., & Monin, B. (2018). The paradox of viral outrage. *Psychological Science*, *29*(10), 1665–1678. https://doi.org/10.1177/0956797618780658

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Shaffer, B., & Duckitt, J. (2013). The dimensional structure of people's fears, threats, and concerns and their relationship with right-wing authoritarianism and social dominance orientation. *International Journal of Psychology*, *48*(1), 6–17. https://doi.org/10.1080/00207594.2012.696651

Solak, N., Tamir, M., Sümer, N., Jost, J. T., & Halperin, E. (2021). Expressive suppression as an obstacle to social change: Linking system justification, emotion regulation, and collective action. *Motivation and Emotion*, *45*(5), 661–682. https://doi.org/10.1007/s11031-021-09883-5

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

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

Steimer, T. (2002). The biology of fear- and anxiety-related behaviors. *Dialogues in Clinical Neuroscience*, *4*(3), 231–249. https://doi.org/10.31887/DCNS.2002.4.3/tsteimer

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Turner, J., Sicha, C., Waldman, K., Hess, A., Paskin, W., Bouie, J., Woodruff, B., Weissmann, J., Goldman, A., Ford, P., & Kois, D. (2014, December 17). 2014: The year of outrage. *Slate Magazine*. https://www.slate.com/articles/life/culturebox/2014/12/the_year_of_outrage_2014_everything_you_were_angry_about_on_social_media.html

Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, *128*(3), 185–199. https://doi.org/10.1037/abn0000410

Ungar, S. (2001). Moral panic versus the risk society: The implications of the changing sites of social anxiety. *The British Journal of Sociology*, *52*(2), 271–291. https://doi.org/10.1080/00071310120044980

Waddington, P. A. J. (1986). Mugging as a moral panic: A question of proportion. *The British Journal of Sociology*, *37*(2), 245–259. https://doi.org/10.2307/590356

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). *A biterm topic model for short texts* [Conference session]. Proceedings of the 22nd International Conference on World Wide Web, New York, NY, United States. https://doi.org/10.1145/2488388.2488514

Zoellner, L. A., Pruitt, L. D., Farach, F. J., & Jun, J. J. (2014). Understanding heterogeneity in PTSD: Fear, dysphoria, and distress. *Depression and Anxiety*, *31*(2), 97–106. https://doi.org/10.1002/da.22133

Received November 14, 2022
Revision received December 4, 2023
Accepted December 18, 2023 ■