RESEARCH ARTICLE OPEN ACCESS

Beyond Moral Outrage: The Role of the Ingroup in Online Condemnations

Jeff D. Rotman[1] (iD) | Virginia Weber[2] | Andrew W. Perkins[3] (iD) | Americus Reed II[4]

[1]Deakin Business School, Deakin University, Melbourne, Australia | [2]Bissett School of Business, Mount Royal University, Canada | [3]College of Business, Washington State University, Carson, USA | [4]Wharton School of the University of Pennsylvania, USA

Correspondence: Jeff D. Rotman (jeff.rotman@deakin.edu.au)

ABSTRACT

Although expressions of condemnation are pervasive on social media and often shape the narratives of traditional news outlets, brands continue to struggle with predicting and managing these responses effectively. To address this challenge, it is crucial to first understand the underlying reasons for such behavior. While existing research argues that moral outrage alone is sufficient to drive online condemnations, we present a more nuanced approach as to when and why this behavior occurs. Across six studies (including Supplementary Appendix A; $N = 1285$) we argue and find that the perception that there is a like-minded ingroup to whom one can signal is critical for promulgating online condemnations. Notably, this audience must be perceived to feel similarly outraged as the condemner (studies 1a-c), and the condemner must be able to signal to this audience through a public post (vs. merely expressing their condemnation privately with no audience; study 2). Engaging in such condemnations predicts subsequent actual consumption choices (study 3), with substantive implications for marketers.

1 | Introduction

Eleven years ago, in December of 2013, Justine Sacco posted the following tweet to her one-hundred-and-seventy followers: *"Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!"* By the time she landed in Africa, Twitter users around the world had blasted Justine's tweet as racist, callous and uncaring. Critics demanded that she be fired from her job and some even called for her death. These reactions provide one of the first examples of massive online outrage and condemnation (Ronson 2016; Crockett 2017). In the decade since, little has changed, with cases like these remaining frequent and consequential (Brady et al. 2020).

From a business perspective, the ease and speed with which consumers can respond to business activities and marketing communications creates a field full of potential landmines for organizations. For example, when Bud Light advertised the slogan "*The perfect beer for removing 'No' from your vocabulary*

for the night" as part of its 'Up For Whatever' campaign, it was immediately condemned for supporting sexual assault, and subsequent independent brand analyses determined it had weakened the Bud Light brand (BrandIndex 2015). Eight years later, Bud Light was faced with similar backlash when it ran a marketing campaign with trans activist Dylan Mulvaney, leading to a $27 billion loss in market cap (O'Halloran 2023).

As Bud Light demonstrates, even experienced brands can struggle to properly predict or manage these responses (Brady and Crockett 2019; Kermani et al. 2024; Spring et al. 2018). Despite this, there is little research that examines individuals' underlying motivations for engaging in these public condemnations (Rothschild and Keefer 2017), with recent calls for more empirical work in this area (Brady et al. 2020).

A *condemnation* is an expression of moral disapproval against an action, behavior, or person deemed to have violated important normative moral standards or principles (Mooijman and

Van Dijk 2015; Rozin et al. 1999; Brady et al. 2020). For example, in response to the 'Up For Whatever' campaign, consumers expressed their moral disapproval across social media by condemning the brand for undercutting antirape initiatives (Glenza 2015). In line with research in this area, we use the term *moral* in the context of moral pluralism, which acknowledges the legitimacy of different varieties of moral violations and may include violations of fairness, loyalty, or purity (Schein and Gray 2018) or other "cultural differences about what is perceived as "right" and "wrong." (Brady et al. 2020, p. 979).

Understanding these condemnations requires first understanding the underlying reasons for why this behavior manifests. On the one hand, a condemnation may be simply the result of feelings of outrage, as outrage alone has been argued to motivate individuals to shame, gossip, and punish others (Crockett 2017). On the other hand, it is possible that these condemnations serve a social-signaling function for impression management, often referred to as 'virtue-signaling' (Shariatmadadri 2016). Such virtue-signaling is theorized to be a performative expression of moral goodness for status-based reasons rather than a genuine expression of one's morality or values (Westra 2021; Dilevko 2019). Ultimately, understanding why a consumer expresses condemnation is critical in knowing how to predict when a consumer will express condemnation and how to respond if they do. In this paper, we argue that these public expressions of condemnation are motivated by genuine outrage but are strategic in their consideration of the signal value of such expressions, making them more likely to occur when there is a like-minded ingroup for the condemner to signal to.

This study makes several contributions to the literature. First, we extend existing research on online condemnations (Brady et al. 2020; Rothschild and Keefer 2017) to show that although expressing online condemnation is predicated on personal feelings of outrage, this expression of condemnation is moderated by the perception that one's ingroup will be similarly offended (studies 1a-c). This extends prior research arguing that moral actions are undertaken for reputational reasons (Jordan and Rand 2020) by demonstrating the important moderating role shared ingroup perception plays in determining when and why individuals will engage in public and potentially socially risky behaviors such as condemnation.

Second, we demonstrate an important boundary to expressing online condemnations. Specifically, because of the need for a like-minded ingroup to signal to when making a condemnation, these condemnations manifest predominately when the opportunity to condemn a moral violation is public rather than private (study 2). This provides nuance to research on virtue-signaling – which is characterized as 'hypocritical' and 'ethically-dubious' (Dilevko 2019) – by demonstrating that although expressing condemnation online is motivated, in part, by social-signaling reasons, it is nonetheless also founded in outrage.

Third, the present research builds on work that has previously found that online behavior does not generally predict accompanying 'offline' consumption behavior (Kristofferson et al. 2014; Barnett et al. 2020; Chou et al. 2020). However, we demonstrate that online condemnation does indeed predict individuals' consequential choices in the real world (study 3).

In the following section, we draw from research on moral outrage and the social functions of condemnation to motivate our framework (see Table 1 for an overview of empirical moral outrage research), followed by five studies to support our theorizing.

2 | Conceptual Development

2.1 | Moral Outrage and Online Condemnations

Why do individuals express online condemnations? A core motivation is because the individual is experiencing *moral outrage*, defined as the anger produced by the perception that a moral standard has been violated (Batson et al. 2007; O'Mara et al. 2011). Such outrage is argued to not be dependent on others' beliefs or values but rather reflects one's own personal emotional and cognitive appraisals of anger, contempt, and disgust (Rozin et al. 1999; Salerno and Peter-Hagene 2013).

Moral outrage is a potent motivator to action (Crockett 2017). Outrage predicts perceptions of punishment deservingness (Darley et al. 2000), desire for more severe punishments (Carlsmith et al. 2002), confidence in a guilty murder verdict (Salerno and Peter-Hagene 2013), negative word of mouth following corporate social irresponsibility (Antonetti and Maklan 2016) and condemnation (Konishi et al. 2017). In this vein, moral outrage on its own, irrespective of other's beliefs and values, may be a sufficient and uniquely predictive motive to explain when and why individuals express condemnations against those that are perceived to violate a moral standard. We argue, however, that moral outrage alone may often not be enough and that social motivations will impact the likelihood of expressing such a condemnation. Several researchers have used terms like 'expressing moral outrage' or 'expressing moral emotions' as synonymous with condemnation (e.g., Crockett 2017; Brady and Crockett 2019; Brady et al. 2020), in that these authors are referring to the expression of moral disapproval (e.g., condemnation). For example, Brady et al. (2020) use the example of condemning a hypothetical senator 'John Doe' by writing 'John Doe is the worst', or 'John Doe needs to resign', to illustrate their conceptualization of expressing moral emotion or outrage. However, whether the expression of moral outrage is the sole product of the experience of moral outrage remains an open question.

Thus, central to the current research is to what degree, and in what contexts, do online criticisms genuinely reflect moral emotions, versus being primarily "virtue-signaling," where people express outrage to appear morally superior or gain social approval? For example, someone can claim to find a behavior disgusting without actually feeling disgust, just as people may feel moral outrage but choose not to express it on social media. Treating public expressions of moral outrage as direct evidence of underlying emotions limits our understanding of when and why people engage in online condemnation. Conflating the expression with the underlying emotion also obscures whether such condemnations reflection authentic emotional reactions or strategic social signaling, making it harder to determine when moral emotions truly drive condemnations and influence consumer purchasing behavior.

TABLE 1 | Overview of Empirical Moral Outrage Research from the Past Fifteen Years (2010-2025).

Reference	Independent variable(s)	Dependent variable(s)	Context	Key findings
Antonetti and Maklan (2016)	Attributions of blame, perceived transgression severity	Moral outrage	Irresponsible corporate actions	Perceiving transgressions as more severe and attributing blame to the perpetrator increases the amount of moral outrage experienced.
Bastian et al. (2013)	Moral outrage and dehumanization	Perceived harm and severity of punishment	Punishment of criminal offenses	Moral outrage and dehumanization independently predicted the severity of punishment assigned to offenders.
Brady and Van Bavel (2025)	Moral-emotional language in political social media posts; partisan social identity	Sharing moral-emotional political posts	Political social media posts	The inclusion of moral-emotional language in political social media posts increases the intention to share those posts, for those with a partisan political identity.
Brady et al. (2021)	Moral outrage, social feedback	Online expressions of moral outrage	Online secondary (Twitter data), political tweets/transgressions	When online posts expressing moral outrage received more positive social feedback, users were more likely to express outrage in future posts.
Brady et al. (2020)	Moral outrage of posters/authors	Perceptions of moral outrage from observers, group polarization	Online (Twitter)	Individuals overperceive the amount of moral outrage felt by others who post moral content on social media.
Brown et al. (2022)	Display of moral outrage	Attraction, long-term mate desirability, trustworthiness	Online dating	Participants (especially women), perceived prospective mates who engaged in outrage-signaling displays as more desirable for long-term relationships.
Crockett (2017)	Immoral acts	Moral outrage	Online and offline	Online immoral acts evoked more moral outrage than immoral acts seen in person.
Effron and Brady (2025)	Frequency/repetition of learning about a transgression from different sources	Moral judgements	Various transgressions from organizations	Affective and cognitive processes both underpin how likely an individual is to condemn a transgression when news of that transgression is widespread/viral.
Forbes (2023)	Power (e.g. powerful perpetrator)	Moral outrage (anger and disgust), support for punishment, moral evaluation (and various mediators).	Various (e.g. bullying, sexual comments, spreading false rumors, sexual assault)	Powerful people who commit transgressions elicit greater moral outrage, as well as lower moral evaluation and greater support for punishment.
Ginther et al. (2022)	Moral outrage, intent, and harm	Punishment	Scenario depicting an individual harming another person	Moral outrage is the product of harm and intent and results in heightened punishment intentions.

(Continues)

TABLE 1 | (Continued)

Reference	Independent variable(s)	Dependent variable(s)	Context	Key findings
Grappi et al. (2013)	Moral outrage (CAD), Social virtues (e.g. equality, justice)	Negative word of mouth, protest behaviors	Various vignettes of company irresponsible actions.	Transgressions result in greater moral outrage (CAD), which results in greater negative WOM and protesting, particularly for those who hold stronger social virtues.
Henderson and Schnall (2021)	Exclusion, social needs threat	Condemnation	60 Moral Foundations Vignette (12 violations for each foundation)	Exclusion increased social needs-threat and subsequent moral condemnation. Threatened social standing resulted in greater moral condemnation.
Hofmann et al. (2014)	Moral and immoral acts, religiosity	Moral emotions (proud/grateful (positive) and contempt/disgust (outrage)	Momentary assessment of daily experiences	Experiencing an immoral (moral) act results in greater outrage (positive) emotions that moral (immoral) acts. Religious people experience greater moral emotions.
Jordan and Rand (2020)	Third-party condemnation and pro-sociality (sharing a resource)	Outrage, punishment	Third-party condemnation game	Individuals express more moral outrage and greater punishment when they cannot signal their trustworthiness through direct pro-sociality.
Kerman et al. (2024)	Expressions of outrage	Self-brand connection, purchase intentions	Social marketing campaigns	When consumers express outrage against a brand for endorsing a social value in a marketing campaign, individuals who witness the outrage will strengthen their connection with the brand if they personally hold the social value being condemned.
Leach et al. (2025)	Expressions of moral outrage	Liking and reposting of posts (virality), petition signatures	X (formerly Twitter)	Moral outrage led to an increased virality, but resulted in fewer signatures when controlling for virality.
Lindenmeier et al. (2012)	Moral outrage, moral inequity, disconfirmation of moral norms	Boycott intentions	Company spying on its employees	Unethical corporate conduct led to consumer moral outrage which subsequently led to boycott intentions.
O'Mara et al. (2011)	Victim identity (self vs. stranger); fair versus unfair treatment	Moral judgement, anger, retribution intentions	Fair or unfair exclusion from an experience	When the victim of an unfair treatment (exclusion) is the self rather than a stranger, individuals experience more anger and greater retribution intention, calling into question whether certain moral/fairness transgressions lead to genuine moral

(Continues)

33

TABLE 1 | (Continued)

Reference	Independent variable(s)	Dependent variable(s)	Context	Key findings
				outrage or merely anger for harm done to the self.
Romani et al. (2013)	Corporate misconduct, moral outrage emotions (anger, disgust, and contempt)	Constructive and destructive punitive actions	Corporate wrongdoing (child slave labor and an oil spill)	Corporate misconduct leads to anger and contempt. Anger leads to constructive punitive actions, while contempt leads to destructive punitive actions. Disgust was not a significant unique predictor of these actions.
Rothschild and Keefer (2017)	Guilt, moral outrage (and various moderators)	Moral outrage, (support for) punishment, guilt	Corporate transgressions (labor exploitation and environmental destruction)	Guilt about being complicit in wrongdoing predicted greater moral outrage at a third-party offender, which in turn increased support for punitive actions against the offender.
Salerno and Peter-Hagene (2013)	Moral outrage (requiring both anger and disgust)	Confidence of a guilty verdict	Evidence (modified) from an actual murder case	Moral outrage resulted in greater verdict confidence.
Salerno and Slepian (2022)	Moral perceptions (morally wrong vs. not wrong)	Punishment (revealing another's immoral secret)	Personal and leaked secrets	Individuals are more likely to reveal others' private secrets when these secrets violate their moral values to punish them, as mediated by moral outrage.
Sawaoka and Monin (2018)	Online expressions of outrage (comments, condemnations)	Impression of comments	Various online contexts and transgressions	A greater number of comments resulted in poorer evaluation of the condemner. Large number of condemnations can elicit sympathy and reduce condemnation.
Sawaoka and Monin (2020)	Witnessing a low versus high volume of condemnations of a transgression.	Moral outrage/offense; sympathy	Various transgressions as social media posts.	Witnessing a high volume of condemnations (i.e., viral outrage) in response to a transgression induces both sympathy and perceptions of outrage being normative, with implications for engaging in moral condemnation.
Shah et al. (2020)	Perceived public condemnation	Moral outrage, avoidance intentions	Users of a riding service in Pakistan	Perceived public condemnation was correlated with avoidance and moral outrage.

(Continues)

TABLE 1 | (Continued)

Reference	Independent variable(s)	Dependent variable(s)	Context	Key findings
Xie and Bagozzi (2019)	Corporate social irresponsibility, social cognitions (e.g. moral identity, empathy)	Moral emotions (contempt and anger), Negative WOM, complaining, boycotting	Offshore shipping corporate social irresponsibility	Corporate transgressions (moderated by various social cognitions) result in negative contempt, anger, and overall negative attitudes, subsequently leading to negative word of mouth, complaining behaviors, and boycotting.
Current Research	Moral outrage, personal offense, other-offense, public likeminded ingroup	Condemnation, product choice	Various online contexts and transgressions	Condemnation as well as changes in product choice/preference is a function of both personal-offense (moral outrage) and other-offense perceptions. Condemnation is amplified by the presence of a likeminded ingroup to witness the condemnation.

2.2 | The Importance of a Like-Minded Ingroup

Moral rules and principles tend to be associated with groups (e.g., one's nation, religion, or political party) and morality is based in part on a duty to follow those rules and principles Nasa et al. 2025). Group identities are often defined by shared moral norms, which are essential for cooperation in large groups (Ellemers and van den Bos 2012) and individuals who present a discrepant moral attitude are often viewed negatively and shamed by other group members (Skitka et al. 2005; Tetlock et al. 2000). The social nature of morality means that there are interpersonal factors that determine how individuals respond to those that are perceived to violate a moral standard (Skitka et al. 2004).

Thus, drawing on this prior research, we propose that individuals are more likely to publicly condemn a perceived violation of moral standards when they believe their feelings are shared by like-minded others. Further, positive feedback on one's online expressions of outrage increases the likelihood of continued expressions of outrage (Brady et al. 2021). However, condemning others can be risky, as it may provoke physical or social retaliation (Jordan and Rand 2020; Balafoutas et al. 2014). Therefore, we argue that moral outrage alone does not fully drive public condemnation; rather, the opportunity to signal to a like-minded ingroup amplifies this behavior. A supportive, like-minded ingroup provides both motivation and a sense of safety for expressing outrage. As a result, we expect that individuals who perceive a "sympathetic audience" (Crockett 2017, p. 770; see also Brady et al. 2017) in the form of a like-minded ingroup will be more inclined to engage in condemnation.

H1. *Outraged individuals are more likely to engage in online condemnation against a perceived moral violation when they believe their ingroup shares their outrage.*

Further, we argue that not only the *existence* of a like-minded ingroup, but the *presence* of this audience is important for determining when and why individuals engage in online condemnations. More specifically, we predict that individuals will be more likely to condemn moral transgressors when the opportunity to do so is public instead of private (i.e., not witnessed by the ingroup). A large body of research has demonstrated that a primary motivation for engaging in condemnations and other third-party punishments is to signal trustworthiness or other reputational effects (Raihani and Bshary 2015; Jordan et al. 2016). Relatedly, empirical research has found that status-seeking, general prestige seeking, and dominance striving are key predictors for individuals to engage in political social media posting (Tosi and Warmke 2016; Grubbs et al. 2019). However, the idea that condemnations serve as reputational signaling implies that the condemnation must be public. That is, the reputational benefits of condemnation are contingent on the presence of an audience who will evaluate and approve of the condemner's actions.

Thus, we propose that the likelihood of condemnation is moderated by whether the condemnation occurs in a public or private context. When the opportunity to condemn is private, it is unlikely to be observed by others and thus lacks the reputational motivation. Supporting this notion, research has demonstrated that people are more likely to punish when in the presence of an audience (Kurzban et al. 2007) and are more

likely to buy green products to signal their status or altruism in public compared to private settings (Griskevicius et al. 2010). This distinction further underscores the strategic nature of the condemnation and the importance of interpersonal factors in determining how individuals respond to those that are perceived to violate a moral standard (Skitka et al. 2004): condemnations serve not only as expressions of moral outrage but also as a way to signal moral commitment and build social capital within a like-minded community. As such, the public visibility of moral behavior is not merely a backdrop but a key determinant of whether individuals choose to engage in condemnation at all. This is because once an individual is outraged or offended, the likelihood of expressing that outrage via condemnation relies on it having a positive ingroup signaling value (Brady et al. 2020), thus requiring the potential for an audience to witness it. As noted, general prestige seeking and dominance striving are key predictors for individuals to engage in political social media posting (Tosi and Warmke 2016). Building from this prior work, we hypothesize:

H2. *Individuals experiencing moral outrage are more likely to condemn a moral violation in a public online setting compared to a private one.*

2.3 | Downstream Consequences on Real-World Behaviors

While online condemnations are most easily witnessed in online social media spaces, we argue that engaging in this behavior has a tangible impact on individuals' actions in the real world. Although there are proponents of collective activism online (Spring et al. 2018), it has also been described as mere 'slacktivism' – a phenomenon whereby individuals engage in a public token act of support for a cause but not necessarily an accompanying willingness to engage in actual meaningful behavior (Kristofferson et al. 2014). Moreover, other researchers have suggested that because consumers are fragmented, lack formal leadership, and are generally unmotivated, condemnations and other calls for change are regularly dismissed (Barnett et al. 2020; Leonel et al. 2024). In contrast, we argue that engaging in online condemnations has a positive relationship to real-world actions, such that the financially costless 'token' of public condemnation will predict meaningful real-world consumer choices, including purchasing behavior, and that firms should pay attention to these actions.

Specifically, we anticipate that individuals experiencing moral outrage who condemn a perceived moral transgressor will be less likely to select products from that transgressor, even when foregoing those options is more costly. This is because, in contrast to many forms of (online) pro-social activism, online condemnations are grounded in outrage and identity, which are the product of and imbued with their moral and personal values. Consistent with this, research on slacktivism finds that when such public tokens of support are aligned with individuals' values, their subsequent meaningful support is increased (Study 4; Kristofferson et al. 2014).

Indeed, research has long demonstrated that individuals strive to align their behavior with their values to maintain

authenticity (Schmader and Sedikides 2018) and cognitive consistency (Festiger 1957). For example, committing to engage in environmentally sustainable behaviors results in greater likelihood of engaging in those behaviors (Bodur et al. 2015). Moreover, across domains ranging from singing over the phone to volunteering, condom use, or making a donation; advocating for specific issues (Stone and Fernandez 2008) or predicting how you will behave (Spangenberg and Greenwald, 1999) facilitates greater alignment.

Thus, given that these online condemnations reflect a violation of one's values, resulting in moral outrage, we anticipate that individuals will make choices in a manner consistent with their condemnations.

H3. *Individuals who publicly condemn a target online for a perceived moral violation are less likely to subsequently purchase products from that target.*

3 | Studies Overview

We test our predictions across five studies. We first provide evidence for the role that social signaling plays in moderating when feelings of outrage will be expressed as online condemnations. We do so by first measuring perceptions of ingroup-outrage (H1; studies 1a-c), and then by providing private versus public opportunities for condemnation (H2; study 2). Our final studies describe the consequences of expressing online outrage by examining subsequent real-world choices (H3; study 3). For a conceptual overview of the studies, see Figure 1.

For each study, sample sizes were determined a priori with no intervening data analysis. We report all manipulations, measures, and exclusions in these studies, with the details of all data, stimuli, and measures available at (https://osf.io/2qm9b/?view_only=7158714b637d4a438cc6d0bc0c744657). Data cleaning for each study was done via an attention check asking participants to identify the focal tweet they had seen, with any additional cleaning reported in individual study procedures (see Supplementary Appendix B for sampling details). For a summary of descriptive statistics across all studies, see Supplementary Appendix C.

4 | Study 1a

Study 1a provides an initial test of our theorizing by showing that individuals who are outraged and believe a like-minded ingroup shares their outrage in response to an offensive tweet are more likely to engage in an online condemnation of the transgressor.

4.1 | Method and Procedure

All participants initially viewed an offensive social media message (hereafter called 'the Tweet') that stated, "*PRESIDENT BARACK OBAMA*' - rearrange the letters: '*AN ARAB BACKED IMPOSTER*'." As an attention check here and in all subsequent
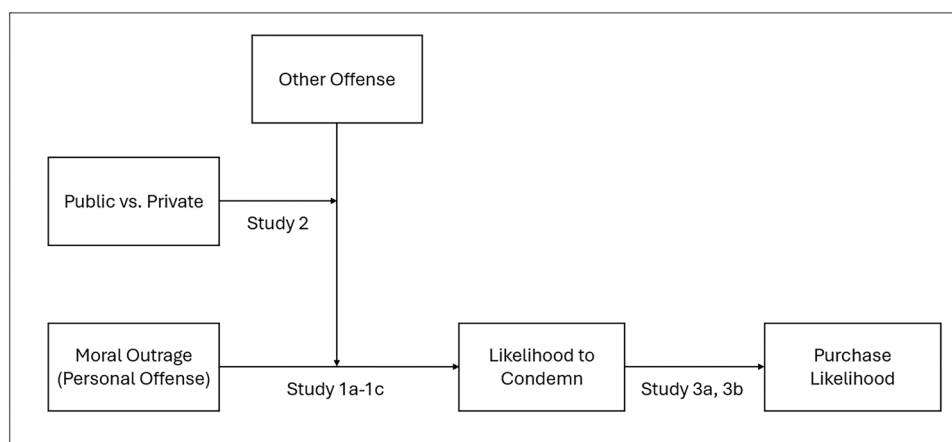
FIGURE 1 | Conceptual framework.

studies (unless otherwise specified), participants were later asked to identify the tweet they saw.

As our independent variable, participants then reported the extent to which they experienced *Moral Outrage* (Batson et al. 2007) using nine outrage adjectives (*irritated, angry, upset, annoyed, offended, outraged, mad, perturbed, and frustrated*, $\alpha = 0.96$) on 6-point Likert scales (1 = Not at all to 6 = Extremely). Interspersed within the emotions were seven filler emotions (*surprised, uninterested, frantic, tense, strange, calm, and pity*, see Supplementary Appendix D for a factor analysis). The order of emotions was randomly presented for each participant. As the moderator, participants indicated whether others in their own social media streams would find the Tweet offensive (*Other Offense;* 1 = Strongly Disagree to 7 = Strongly Agree).

As our dependent variable, we assessed how likely participants would be to condemn the Tweet on their own social media (adapted from Taylor et al. 2012; 3 items; $\alpha = 0.99$; *Likelihood To Condemn*; e.g., "Assuming you saw this on a social media platform you use, how likely would you be to condemn this via social media" 1 = Very Unlikely, 7 = Very Likely). The order of the dependent variable and the two independent variables were counterbalanced and there was no effect of counterbalancing (all p's > 0.16).

Finally, participants responded to covariate measures and basic demographics (see OSF for all included measures). These included a single item measuring *Political Orientation* (1 = Very Liberal to 7 = Very Conservative), two measures asking participants how often they post and how often they post things that are political in nature on social media ($r = 0.63$, $p < 0.001$), and whether they had seen the tweet presented previously (Yes/No). Across all studies controlling for or excluding people who had seen the tweet did not affect the results and thus all participants are included in the analyses.

For this study (but not subsequent studies), we also included single-items measuring whether their desire to post was based on the relevance to others (*relevance*), whether their desire to post was based on if they thought it could make a difference (*efficacy*), to rule these out as alternate motivations (*vs.* moral

outrage and social signaling) for individuals to post in response to potentially offensive tweets.

4.2 | Results

As anticipated, the interaction between *Moral Outrage* and *Other Offense* on *Likelihood To Condemn* was significant (F (1,291) = 5.67, $p = 0.018$, $\eta_p^2 = 0.019$). We assessed the relative effects using a floodlight analysis in PROCESS (Hayes 2017; Model 1, 5000 bootstraps) with *Outrage* as the independent variable and *Other Offense* as the moderator (see Figure 2). As expected, moral outrage predicted posting behavior only when other offense was relatively high (Johnson–Neyman significance region > 4.05).

While relevance and efficacy predicted *Likelihood to Condemn* ($r = 0.388$ and $r = 0.354$, $ps < 0.001$), controlling for them did not affect the interaction results ($p = 0.031$). The correlation between *Moral Outrage* and *Other Offense* was modest, ($r = 0.46$, $p < 0.001$; Cohen 1992). Finally, there was no main or interacting effects of counterbalancing (see Supplementary Appendix E for test for equivalence), political orientation, seeing the tweet or of the filler emotions (all p's > 0.164).

5 | Study 1b

A potential concern with study 1a is that we measured outrage using emotional self-report items, where drawing awareness to one's emotions, mood, or affect can subsequently impact behavior (e.g., Raghunathan and Corfman 2006). Although the counterbalancing provides evidence that directly measuring emotions did not impact our results, study 1b further rules this alternative explanation out by utilizing a single-item measure of personal offense in place of outrage, similar to our measure of ingroup (other) offense. To ensure that personal offense and moral outrage reflect the same construct (and that other-offense does not), a pre-test ($N = 98$) was conducted measuring personal offense, other offense, and moral outrage towards an offensive tweet. A factor analysis (see Supplementary Appendix F) revealed that all moral outrage items along with the personal offense item loaded onto one factor (Eigenvalue = 6.70), while
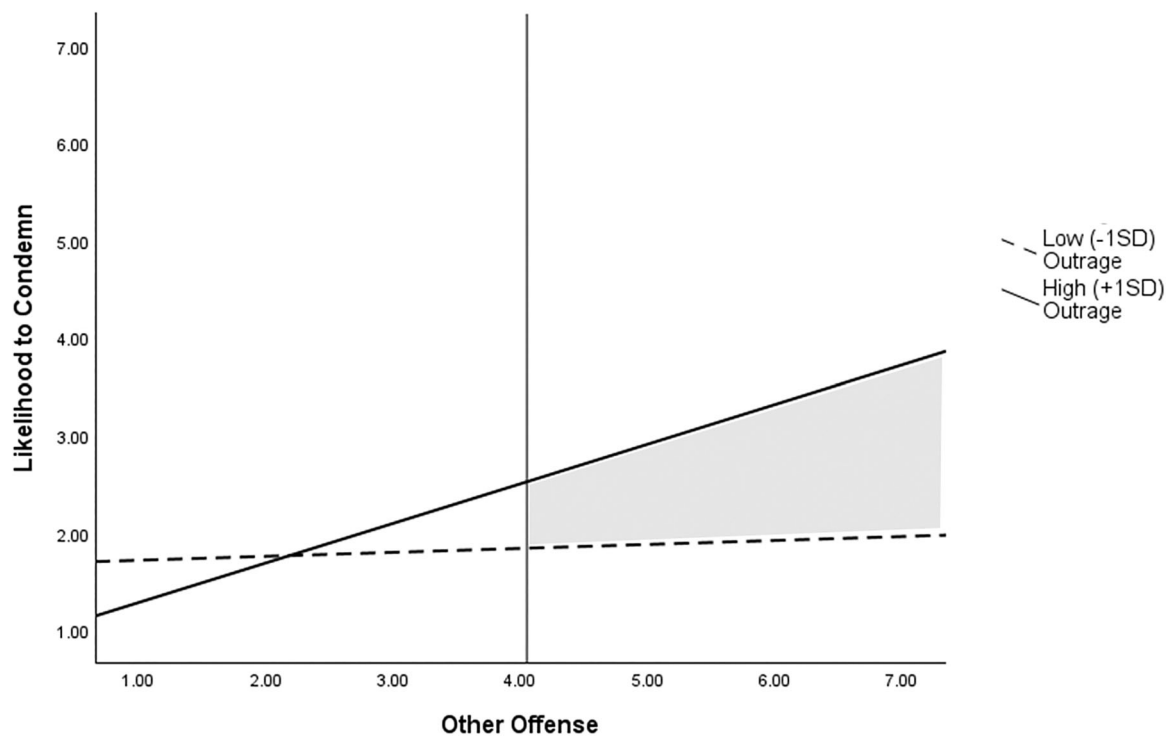
FIGURE 2 | Likelihood to post a condemnation to an offensive tweet is greatest when there is both high moral outrage (+1 SD) and high (> 4.05) perceptions of others' offensiveness.

other offense loaded onto a separate factor (Eigenvalue = 1.29). Study 1b also utilized a different tweet to demonstrate that our results are not idiosyncratic to our stimuli and are robust across various potential triggers of outrage.

5.1 | Method and Procedure

The method for study 1b was similar to 1a, however, instead all participants initially saw the tweet "*Going to Africa, Hope I don't AIDS... Haha, just kidding, I'm white.*" Participants then reported whether they themselves found the initial tweet offensive on a single-item scale (*Personal Offense*; 1 = Strongly Disagree to 7 = Strongly Agree), and whether others in their own social media streams would find the Tweet offensive, (*Other Offense;* 1 = Strongly Disagree to 7 = Strongly Agree). Once again, our dependent measure was *Likelihood to Condemn* (3 questions, $\alpha = 0.98$).

Participants finally responded to the demographics and covariates as in study 1a for *Political Orientation*, *Typical Political Posting Behavior*, and whether they had seen the Tweet presented previously (Yes/No).

5.2 | Results

As in study 1a, the interaction between *Personal Offense* and *Other Offense* on *Likelihood To Condemn* was significant (F (1,191) = 7.19, $p = 0.008$, $\eta_p^2 = 0.036$). Controlling for *Typical Political Posting Behavior*, *Political Orientation*, and whether participants had seen the tweet did not affect the interaction results ($p = 0.015$), nor did they moderate the interaction (all

$p's > 0.317$). Further, the correlation between personal offense and other offense ($r = 0.23$, $p = 0.001$) was again quite modest (Cohen 1992)

Using the PROCESS macro (Hayes 2017; Model 1, 5000 bootstraps) and consistent with study 1a, the interaction utilizing *Personal Offense* as a moderator revealed the same pattern of interaction, with *Other Offense* significantly predicting posting only when *Personal Offense* was relatively high, (Johnson-Neyman significance region > 4.84), but not when *Personal Offense* was relatively low (Johnson-Neyman significance region < 4.84). This replicates our finding from study 1a while demonstrating our results are robust to differences in stimuli.

6 | Study 1c

An important limitation of the prior studies is that although personal outrage/offense and perceived ingroup offense showed only modest correlations, they may not be fully orthogonal. Indeed, given that people share similar morals and values with their ingroup, we would expect that the two constructs should typically be correlated (and we found that they are). Nevertheless, our framework relies on individuals increasing their desire to condemn transgressors in the presence of ingroup agreement; if personal outrage and other offense always co-occur, one or the other should always be sufficient for expressions of condemnation to occur. However, despite individuals generally sharing morals with their ingroup (Ellemers and van den Bos 2012), this is not always the case; individuals do not share the exact same moral attitudes or values with their ingroup on every issue (Rozin et al. 1999; Salerno and Peter-Hagene 2013). To this end, study 1c manipulates personal

offense and perceptions of ingroup offense independently to demonstrate their unique impacts on the likelihood to condemn transgressors.

Further, a second goal of this study was to provide greater external validity to our design. In particular, individuals viewing social media posts often do not do so in isolation, but instead are perusing multiple posts at a time, subsequently choosing whether and how to respond to each. Thus, in this study, we allow participants to view and respond to multiple tweets.

6.1 | Method and Procedure

Participants were assigned to a 2 *Self-Offensive* (Offensive to Self *vs.* Not Offensive to Self) x 2 *Other-Offensive* (Perceived Offensive to Others *vs.* Perceived Not Offensive to Others), repeated-measures design. In each condition, participants saw a list of ten tweets informed from research on offensive and hate speech and cross referenced with the virality of the tweets (Burnap and Williams 2016). Participants were told, "People often tweet things that some people find offensive while others do not. Please select one of the following tweets, that you find offensive (that you do not find offensive); however, think that, in general, your friends would find offensive (not find offensive)". All participants selected a tweet for all 4 conditions; the order of these conditions was randomly assigned to preclude potential order effects.

As a manipulation check, each time participants selected a tweet they were taken to a new page, shown the tweet, and asked to rate their agreement with two statements: "I personally find this tweet offensive" and "I think most of my friends would find the tweet offensive" (1 = Strongly Disagree to 7 = Strongly Disagree). For the main dependent variable, participants were asked the same three questions regarding their likelihood to post for each tweet they selected as in prior studies, this time anchored as Likert scales instead of semantic differentials

(1 = Strongly Disagree to 7 = Strongly Disagree; average $\alpha = 0.96$). Finally, participants provided demographic information and covariates similar to previous studies, including *Political Orientation* (measured on a six-point scale to avoid neutral responses, following Kray et al. 2017), *Typical Political Posting Behavior*, and whether they had previously seen each tweet.

6.2 | Results

6.2.1 | Manipulation Checks

Participants rated tweets in the *Offensive to Self* conditions ($M = 5.57$, SD = 1.16) as more personally offensive than in the *Not Offensive to Self* conditions ($M = 2.76$, SD = 1.39; t (200) = 22.72, $p < 0.001$, $d = 1.60$). Similarly, participants rated tweets in the *Offensive to Others* ($M = 5.27$, SD = 1.17) conditions as more offensive to others than in the *Not Offensive to Others* conditions ($M = 3.18$, SD = 1.32; $t(200) = 16.79$, $p < 0.001$, $d = 1.19$). There were no significant interactions between the offensiveness conditions *(Personal x Others)* on participants' feelings of personal offense ($p = 0.53$) or their perceptions of other offense ($p = 0.29$), indicating that there is not an amplifying effect on perceptions of offensiveness (to themselves or others).

6.2.2 | Primary Results

A repeated-measures analysis assessing *Offensive to Self* and *Offensive to Others* revealed a significant interaction ($F(1, 199) = 15.11$, $p < 0.001$, $\eta_p^2 = 0.071$). As expected, participants were significantly more likely to condemn when a tweet was both offensive to self and offensive to others ($M = 4.54$, SD = 1.87), relative to when a tweet was offensive to self but not others ($M = 3.84$, SD = 1.82, $p < 0.001$), offensive to others but not oneself ($M = 2.68$, SD = 1.60), and neither offensive to self or others ($M = 2.47$, SD = 1.54, $p < 0.001$; see Figure 3). Neither
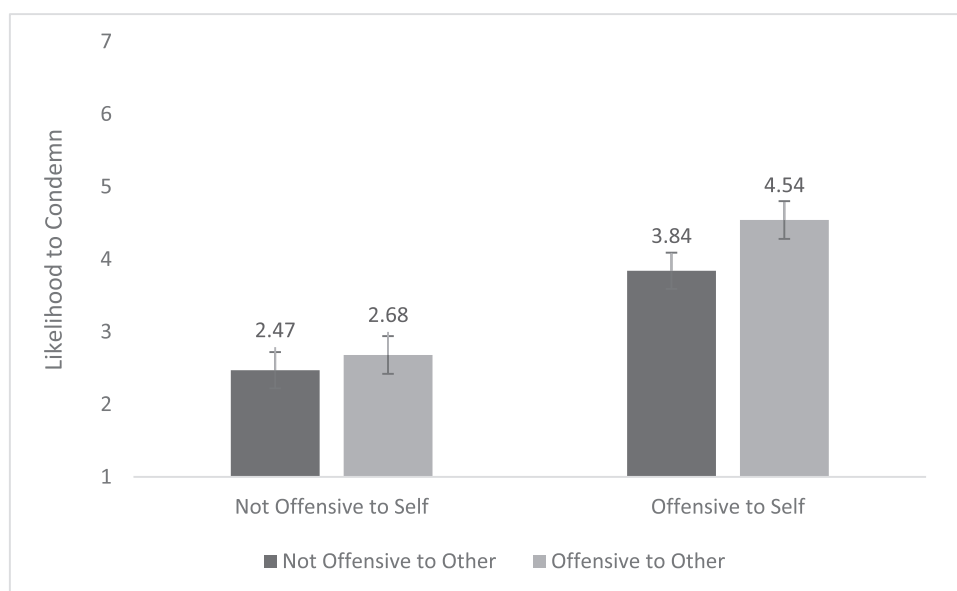
FIGURE 3 | Likelihood to condemn is strongest when the selected tweet is perceived as both offensive to self and offensive to others in one's ingroup.

political orientation nor posting frequency impacted the results (all p's > 0.366).

6.3 | Discussion of Studies 1a – 1c

The results of our initial studies provide converging evidence for our theoretical framework. Specifically, participants were more likely to condemn an offensive social media message when they believed a like-minded ingroup would also find the message offensive. These effects are robust across different stimuli and outrage-provoking messages, do not arise due to measurement effects for emotion (cf. Raghunathan and Corfman 2006), hold even when participants are exposed to multiple potentially outrage-inducing messages, and occur when personal and ingroup offense are manipulated orthogonally. Taken together, these experiments build on traditional moral outrage and identity signaling research by showing how ingroup offense moderates the likelihood of personal outrage leading to public condemnations.

7 | Study 2

Study 2 is designed to test our prediction that individuals are more likely to engage in condemnation of a transgressor when the opportunity to do is public versus when it is private (i.e., when the condemnation is witnessed only by the perceived transgressor and not by others). We argue that this will occur because personal outrage alone is not always sufficient for individuals to engage in condemnation, and that the social signaling value of the condemnation is a key factor motivating individuals to take the potentially risky step of publicly condemning a perceived moral violation (Crockett 2017; Haidt 2001; Skitka et al. 2005; Tetlock et al. 2000).

7.1 | Method and Procedure

This study was conducted in October 2016 amidst a highly polarizing United States election between Donald Trump and Hillary Clinton. In additional to cleaning based on an attention check recognizing the topic of the article at the end of the study, we restricted our analysis only to individuals who planned to vote for either Hillary Clinton or Donald Trump, leaving a final sample of 220 (144 Clinton supporters and 76 Trump supporters).

Participants were assigned to a 2 (*Support: Hillary Clinton, Donald Trump*) x 2 (*Article Content: Hillary Clinton, Donald Trump*) x 2 (*Response: Private vs. Public*). Participants were first asked who they intended to support in the upcoming election. Next, participants read one of two randomly assigned articles. Both articles were identical but implicated either Hillary Clinton or Donald Trump as the source for the offensive statements in the article, implying that people on welfare were "deadbeats" (See OSF link above). We posited that participants would find the content of the article offensive regardless of who they supported or who the article was said to be quoting, whereas their perception of ingroup agreement with their condemnation

would be a function of a match between *Support* and *Article Content*, such that participants would perceive their ingroup to agree with the condemnation only if the offensive content was said by the opposing political candidate. It should be noted that the article was developed from an actual article which linked Hillary Clinton to that statement.

After reading the article, participants were asked 'how likely would you be to send a (public) private tweet to Hillary Clinton (Donald Trump)? (*Likelihood to Condemn;* Taylor et al. 2012; $\alpha = 0.98$). Participants in the private condition were told that only the recipient would be able to see it, whereas in the public conditions, participants were told that everyone following you would be able to see it. Finally, basic demographic data was collected, including whether they used Twitter, whether they had read the article before, and due to the public/private manipulation, we also asked participants to report how many followers they had on the social media platform but none of these variables affected the results.

7.2 | Results

A three-way ANOVA (*Support, Article Content*, and *Response: Private vs. Public*) indicated a significant three-way interaction ($F(1,212) = 4.40$, $p = 0.018$, $\eta_p^2 = 0.026$) and did not interact with whether people utilized Twitter or not ($p = 0.16$). As anticipated, assessing the two-way interactions for each of the *Response: Private versus Public* conditions revealed a significant effect in the *Public* condition ($F(1,98) = 7.94$, $p = 0.006$, $\eta_p^2 = 0.075$), but not in the *Private* condition ($F(1,114) = 0.35$, $p = 0.56$, $\eta_p^2 = 0.003$).

Decomposing the interaction for the *Public* condition revealed that, in line with our predictions, participants were more likely to post a public tweet when there was a mismatch between *Support* and *Article Content* (see Figure 4). In particular, in the *Article Content: Hillary Clinton* condition, Trump supporters were more likely to send a public tweet than Clinton supporters ($p < 0.05$) whereas in the *Article Content: Donald Trump*, Clinton supporters were marginally more likely to send a public tweet compared to Trump supporters ($p = 0.06$; See Supplementary Appendix G for means/SDs).

7.3 | Discussion

This study provides evidence for the signaling role of online condemnations. Individuals were more likely to condemn an opposing candidate's statement when the opportunity to do so was public, but not when it was private. Specifically, when the opportunity to condemn a moral transgressor was private, individuals were equally likely to condemn regardless of whether their ingroup was likely to agree with the condemnation (i.e., if the article content matched their ingroup's political norms). However, when the opportunity to condemn was public, individuals were more likely to condemn a moral transgressor when their ingroup was likely to agree with the condemnation (i.e., when the article content matched their ingroup's political norms), underscoring that the expression of outrage is contingent on its public signaling value within the ingroup. This has
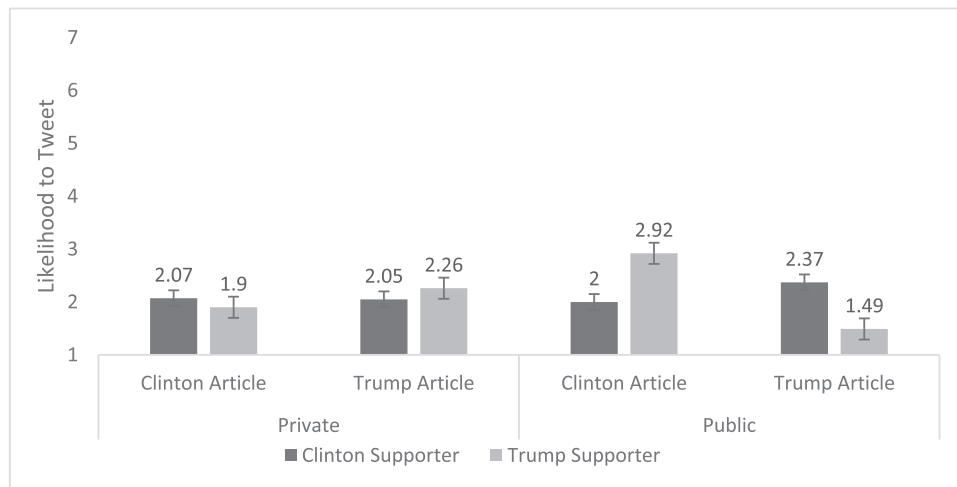
FIGURE 4 | Likelihood to tweet is greater against an opposing political candidate, but only when the post is public.

implications for how social media platforms are designed to promulgate expressions of moral outrage (Brady et al. 2020).

8 | Study 3

Study 3 examines the downstream effects of one's initial condemnation on a subsequent consequential behavior. In particular, we examine whether engaging in an online condemnation predicts subsequent choice to support or not support the moral transgressor. To test our assertion that engaging in condemnations influences offline behavior, we use a consequential choice task for a real gift card to assess the likelihood of choosing a branded product where a brand's social media (McDonald's Twitter account) has written an offensive tweet, and predict that an increased likelihood to condemn will predict decreased subsequent likelihood to purchase from the brand.

8.1 | Method and Procedure

This study uses a tweet which is likely to be more offensive to conservatives, deriding former President Donald Trump, and employs a real consequential choice for a real brand to provide a stronger test for our proposed effects. In particular, participants had the opportunity to enter a genuine lottery with the potential to receive a real eGift-card from the brand they selected as the dependent measure in this study.

Participants were randomly assigned to a 2 (*Offensive vs. Non-offensive*) x Continuous (*Political Orientation*) design. All participants saw a real[1] Tweet from the brand McDonald's (a QSR restaurant). Participants in the offensive condition saw a tweet that read "*@realDonaldTrump You are actually a disgusting excuse of a President and we would love to have @BarackObama back, also you have tiny hands,*" whereas participants in the non-offensive condition saw a tweet that read: "*Innovations from this revolutionary restaurant will start to be rolled out in new McDonald's restaurants built in the UK from this year. The best is yet to come.*"

Following the manipulation, participants indicated to which they would be likely post a response condemning the tweet (*Likelihood to Condemn*), along with the extent they found personally and whether they believed the tweet would be offensive to others. Participants were then given the option to select a $25 gift certificate from one of four fast-food brands (McDonald's, Burger King, Arby's, and Wendy's; coded as McDonald's = 1 and all others = 2), and told (truthfully) that one participant would actually receive the gift certificate of their choice, making this decision consequential. Next, participants were asked about their *Political Orientation* (1 = Very Liberal to 7 = Very Conservative), demographics, and typical posting behavior (identical to previous studies). We also asked participants whether they believed that it was a real tweet, their attitude towards the fast-food restaurants, and whether they took the survey seriously; none of these responses impacted our results and are not discussed further.

8.2 | Results

8.2.1 | Manipulation Check

Participants rated the offensive post as more personally offensive and perceived it to be more offensive to others ($M_{offense} = 4.06$, SD = 2.26 and $M_{other} = 4.33$, SD = 1.85) compared to the non-offensive post ($M_{offense} = 1.32$, SD = 0.91 and $M_{other} = 1.54$, SD = 1.21), (*Personal Offense* $t(193) = -.11.04$, $p < 0.001$, $d = 1.73$ and *Other Offense* $t(193) = -12.27$, $p < 0.001$, $d = 1.58$). However, these were both qualified by a significant interaction with *Political Orientation* ($F_{personaloffense}(1,192) = 18.25$, $p < 0.001$ and $F_{otheroffense}(1,192) = 17.65$, $p < 0.001$). Moreover, consistent with our assumption that the tweet would be more offensive to conservatives, there was a significant correlation between *Offensiveness* and *Political Orientation* (r = 0.25, $p < 0.001$).

8.2.2 | Condemnation

We observed a significant interaction on *Likelihood to Condemn* ($F(1,192) = 10.41$, $p < 0.001$, $\eta_p^2 = 0.052$) even while controlling

for typical posting behavior ($F(1,191) = 13.09$, $p < 0.001$, $\eta_p^2 = 0.062$). In particular, there was no effect of *Tweet Offensiveness* on *Likelihood to Condemn* among those that were relatively more liberal but *Tweet Offensiveness* impacted *Likelihood to Condemn* among those that were relatively more conservative (Johnson-Neyman significance region > 2.23). In addition, consistent with previous studies, the personal-other offense interaction also predicted *Likelihood to Condemn* ($F(1,190) = 5.55$, $p = 0.019$, $\eta_p^2 = 0.028$).

8.2.3 | Downstream Choice

A binary logistic regression revealed the same significant pattern of interaction ($B = -.849$, S.E. $= 0.311$, Wald $\chi^2 = 7.47$, $p = 0.006$), with those that were more conservative (Johnson-Neyman significance region > 3.53; see Figure 5) more likely to choose a non-McDonald's gift-certificate, and with *Likelihood to Condemn* marginally predicting downstream choice ($r = 0.129$, $p = 0.07$). This effect was consistent when controlling for typical posting ($B = 0.047$, SE $= 0.02$, $p = 0.02$). Moreover, although the directionality reversed for liberal consumers, there was no significant difference between the conditions, even for the most liberal consumers.

8.3 | Discussion

In support of our prediction that engaging in online condemnations has downstream consequences on actual, consequential choice (H3), study 3 found that individuals who were more likely to publicly condemn an offensive tweet from a brand were subsequently less likely to choose a product from that brand. In a conceptual replication (see Supplementary Appendix A), we also

found that this effect held across the political spectrum, and occurred despite such choices being disadvantageous to the individual (i.e., foregoing a desirable cost-savings). These results demonstrate that online condemnation behavior has substantive implications for real-world consumer choice.

9 | General Discussion

In this manuscript we provide a framework to better understand when and why individuals condemn other persons, brands, or corporations they deem offensive or outrageous on social media. Building on the morality and identity signaling literatures (Brady and Crockett 2019; Jordan and Rand 2020; Reed et al. 2012), we show that public displays of outrage on social media are not solely driven by moral outrage but occur more often in contexts where individuals perceive their ingroup will agree and be similarly outraged. We discuss the theoretical and managerial implications of these findings below.

9.1 | Theoretical Contributions

This study contributes to the literatures on moral outrage and social signaling by revealing that individuals' condemnation often functions as a way to signal ingroup membership, especially in public contexts (Brady et al. 2017; Brady et al. 2020). While prior work suggests that outrage alone is sufficient for condemnation (Darley et al. 2000), we demonstrate that signaling to one's ingroup plays a significant role in determining when condemnation occurs. This insight extends recent findings that condemnation can be a reputational tool used to signal trustworthiness. Whereas past research found that these effects occurred whether or not others witnessed the moral acts, our
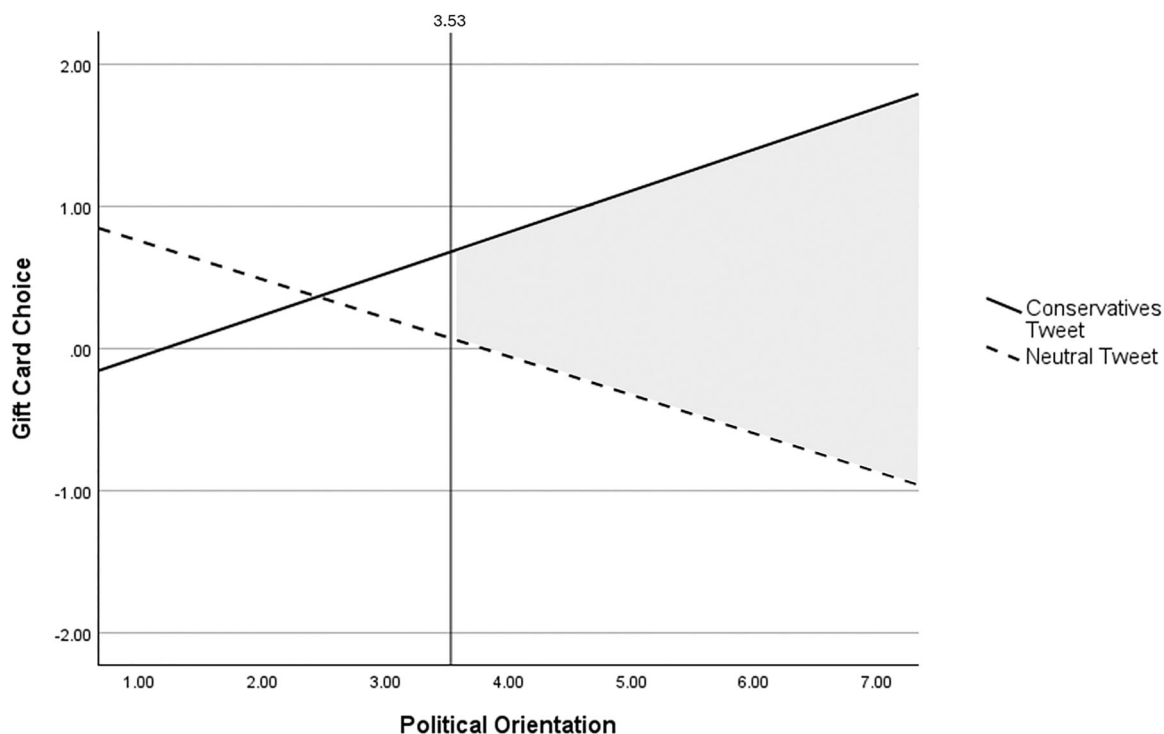

FIGURE 5 | Giftcard selection as a function of tweet condition and relative ingroup (political) identification.

research diverges from these previous findings by showing that online condemnations are specifically tailored to an audience of like-minded individuals, thus functioning differently than private moral acts (Taylor et al. 2012).

Relatedly, this study helps disentangle condemnation (e.g. expressing moral outrage) from the experience of moral outrage, which has often been conflated in prior work. Our findings suggest that condemnations may not always stem from the experience of moral outrage but are, at the very least, amplified when in the presence of a like-minded ingroup. Moreover, these findings provide explanatory insight into the related research of online hate. Online hate focuses on condemnation of a specific group or identity, rather than disagreement with a particular action or statement (Tong 2024). Consistent with the current research, Walther and Rice (2025), have similarly noted that "why people express and propagate hate online has been understudied relative to [documenting the pervasiveness and describing the contents of online hate]" (p.10). While online hate research has argued for the importance of both individual attitudes and social signaling to a like-minded ingroup (Walther 2024), the present research would suggest that the occurrence of both is required. Indeed, this speaks to why online hate predominantly occurs on websites or forums explicitly created for ideologically aligned users (Bliuc et al. 2018).

Finally, in contrast to some research that suggests that online moral discourse lacks behavioral consequences, we find that expressions of condemnation predict tangible consumer choices, such as product avoidance. These findings link online moral signaling to downstream market behaviors, suggesting that online outrage can have meaningful economic implications. Together, our work advances multiple literatures by revealing when condemnation is more than just expression, and when it becomes action.

9.2 | Practical Implications

Integrating the perspectives on whether online condemnation stems from genuine moral outrage or virtue-signaling is essential for brands to navigate today's volatile online environment. For example, as we mentioned earlier, Bud Light faced backlash after launching a campaign with transgender influencer Dylan Mulvaney. While some decried the backlash as 'fake outrage' (Parkel 2023), this study suggests that it was not only genuine but was also accompanied by, at the very least, the perception that others felt similarly. This study underscores the importance of responding to high-profile situations quickly, as 'trending' situations will provide further evidence that like-minded people are offended, further proliferating the backlash.

Our findings also have substantive implications for managerial decision-making, as our results suggest that brands need to pay attention and respond to online outrage. Indeed, we find that engaging in online condemnation impacts individuals' real-world choices, such as choosing to avoid products associated with moral transgressors, even when those choices are personally disadvantageous. The present work highlights a need to respond to backlash quicky to curtail outrage and condemnation when this arises. However, brands should consider whether it the outrage is coming from a very vocal but small and insular segment or whether it reflects a wider demographic. Particularly when the outrage is in response to a purposeful marketing decision, brands need to consider whether they ought to apologize and reverse or double-down on their message. A prime example that illustrates this phenomenon is Nike's 2018 advertising campaign featuring Colin Kaepernick. The backlash Nike faced after launching this campaign stemmed from outrage over Kaepernick's protests. While this outrage led to some customers abandoning the Nike brand, Nike strategically emphasized its commitment to social justice, staying consistent with its values while navigating shifting consumer expectations. This highlights the importance for brands to remain true to their values, appealing to loyal customers without over-committing resources to managing performative outrage.

9.3 | Limitations and Future Research Directions

One limitation of the present research is that all studies were conducted as experiments. Future research may seek to explore these effects using different methodologies, including qualitative methods such as interviews to derive rich insights as to how people feel when they see an online post they deem as condemnation-worthy and what they feel inspires them to respond. Moreover, future research may benefit from examining expressions of online outrage in a longitudinal fashion to measure its impact on longer term real-world consumption decisions. Although our study 3 measured real, substantive choice for an actual gift card, it would be of interest to explore how long these effects hold for and what type of consumption decisions in the real world are impacted.

Future research should explore also the interaction between morality and identity signaling, particularly in online spaces. For instance, under what circumstances might individuals prefer private condemnation, particularly when the transgressor is an ingroup member or personal acquaintance? Additionally, more work is needed to understand the downstream effects of online outrage, such as whether engaging in online condemnation might act as a moral license, leading to less prosocial behavior, or if it motivates virtue-signaling through other means, such as conspicuous donations.

Furthermore, it would be valuable to examine how these signaling acts influence brand attitudes, identification, and other marketing outcomes (Kermani et al. 2024). The interpersonal outcomes of brand condemnation, such as group polarization, should also be examined. Ultimately, integrating both viewpoints—moral outrage and virtue-signaling—provides a richer understanding of how individuals use morality to manage their social media presence, navigate consumer choices, and interact with brands. This nuanced approach enables brands to craft more effective, authentic responses, build long-term relationships with their audience, and manage public relations crises with greater precision.

10 | Conclusion

The current research examines the ever-abundant phenomenon of online condemnations as a way to punish perceived moral

transgressors. We highlight that these condemnations have an important signaling function and occur predominately in public and in view of those who are likely to be similarly offended by the perceived transgression. In doing so, we provide crucial nuance for understanding how online condemnation has real-world consequences for consumer behavior and brand management.

Ethics Statement

The study was supported by internal research funds from the authors. None of the authors' professional or financial affiliations can be perceived to have biased the research. The human participants gave informed consent to participate voluntarily in the studies after the general nature of the studies had been explained to them. The studies were approved by Washington State University's Institutional Review Board and were conducted in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Data Availability Statement

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/2qm9b/?view_only=7158714b637d4a438cc6d0bc0c744657.

Endnotes

[1] While both were real tweets from the McDonald's Corp Twitter account, the offensive tweet was the product of a hack, thus participants in the offensive condition were told that at the end of the experiment.

References

Antonetti, P., and S. Maklan. 2016. "An Extended Model of Moral Outrage at Corporate Social Irresponsibility." *Journal of Business Ethics* 135, no. 3: 429–444.

Balafoutas, L., K. Grechenig, and N. Nikiforakis. 2014. "Third-Party Punishment and Counter-Punishment in One-Shot Interactions." *Economics Letters* 122, no. 2: 308–310.

Barnett, M. L., H. Irene, and B. W. Husted. 2020. "The Rise and Stall of Stakeholder Influence: How the Digital Age Limits Social Control." *Academy of Management Perspectives* 34, no. 1: 48–64.

Bastian, B., T. F. Denson, and N. Haslam. 2013. "The Roles of Dehumanization and Moral Outrage in Retributive Justice." *PLoS One* 8, no. 4: e61842.

Batson, C. D., C. L. Kennedy, L. A. Nord, et al. 2007. "Anger At Unfairness: Is It Moral Outrage?" *European Journal of Social Psychology* 37, no. 6: 1272–1285.

Bliuc, A. M., N. Faulkner, A. Jakubowicz, and C. McGarty. 2018. "Online Networks of Racial Hate: A Systematic Review of 10 Years of Research on Cyber-Racism." *Computers in Human Behavior* 87: 75–86.

Bodur, H. O., K. M. Duval, and B. Grohmann. 2015. "Will You Purchase Environmentally Friendly Products? Using Prediction Requests to Increase Choice of Sustainable Products." *Journal of Business Ethics* 129, no. 1: 59–75.

Brady, W. J., and J. J. Van Bavel. 2025. "Social Identity Shapes Antecedents and Functional Outcomes of Moral Emotion Expression." *Journal of Experimental Psychology: General* 154: 1505–1522.

Brady, W. J., and M. J. Crockett. 2019. "How Effective Is Online Outrage?" *Trends in Cognitive Sciences* 23, no. 2: 79–80.

Brady, W. J., M. J. Crockett, and J. J. Van Bavel. 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, no. 4: 978–1010.

Brady, W. J., K. McLoughlin, T. N. Doan, and M. J. Crockett. 2021. "How Social Learning Amplifies Moral Outrage Expression in Online Social Networks." *Science Advances* 7, no. 33: eabe5641.

Brady, W. J., J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114, no. 28: 7313–7318.

BrandIndex. 2015. *Millennials Not Always* #upforwhatever. https://www.forbes.com/sites/brandindex/2015/08/05/millennials-not-always-upforwhatever-3/.

Brown, M., L. A. Keefer, D. F. Sacco, and F. L. Brown. 2022. "Demonstrate Values: Behavioral Displays of Moral Outrage as a Cue to Long-Term Mate Potential." *Emotion (Washington, D.C.)* 22, no. 6: 1239–1254.

Burnap, P., and M. L. Williams. 2016. "Us and Them: Identifying Cyber Hate on Twitter Across Multiple Protected Characteristics." *EPJ Data Science* 5: 11.

Carlsmith, K. M., J. M. Darley, and P. H. Robinson. 2002. "Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment." *Journal of Personality and Social Psychology* 83, no. 2: 284–299.

Chou, E. Y., D. Y. Hsu, and E. Hernon. 2020. "From Slacktivism to Activism: Improving the Commitment Power of E-Pledges for Prosocial Causes." *PLoS One* 15, no. 4: e0231314.

Cohen, J. 1992. "A Power Primer." *Psychological Bulletin* 112: 155–159.

Crockett, M. J. 2017. "Moral Outrage in the Digital Age." *Nature Human Behaviour* 1, no. 11: 769–771.

Darley, J. M., K. M. Carlsmith, and P. H. Robinson. 2000. "Incapacitation and Just Deserts as Motives for Punishment." *Law and Human Behavior* 24, no. 6: 659–683.

Dilevko, J. 2019. "Virtue-Signaling and Inclusivity." *Journal of Information Ethics* 28, no. 2: 11–15.

Effron, D. A., and W. J. Brady. 2025. "Affective and Cognitive Underpinnings of Moral Condemnation When News of Transgressions Goes Viral." *Journal of Experimental Psychology: General* 154: 2553–2568.

Ellemers, N., and K. van den Bos. 2012. "Morality in Groups: on the Social-Regulatory Functions of Right and Wrong." *Social and Personality Psychology Compass* 6, no. 12: 878–889.

Festiger, L. 1957. *A Theory of Cognitive Dissonance*.

Forbes, R. C. 2023. *When the Mighty Fall: Moral Outrage Toward Powerful Perpetrators*. Canada: University of Toronto.

Ginther, M. R., L. E. S. Hartsough, and R. Marois. 2022. "Moral Outrage Drives the Interaction of Harm and Culpable Intent in Third-Party Punishment Decisions." *Emotion (Washington, D.C.)* 22, no. 4: 795–804.

Glenza, J. (2015), *Bud Light Sorry for 'Removing no From Your Vocabulary for the Night' Label*, The Guardian, https://www.theguardian.com/business/2015/apr/29/bud-light-slogan-no-means-no, retrieved April 3, 2025.

Grappi, S., S. Romani, and R. P. Bagozzi. 2013. "Consumer Response to Corporate Irresponsible Behavior: Moral Emotions and Virtues." *Journal of Business Research* 66, no. 10: 1814–1821.

Griskevicius, V., J. M. Tybur, and B. Van den Bergh. 2010. "Going Green to be Seen: Status, Reputation, and Conspicuous Conservation." *Journal of Personality and Social Psychology* 98, no. 3: 392–404.

Grubbs, J. B., B. Warmke, J. Tosi, A. S. James, and W. K. Campbell. 2019. "Moral Grandstanding in Public Discourse: Status-Seeking

Motives as a Potential Explanatory Mechanism in Predicting Conflict." *PLoS One* 14, no. 10: e0223749.

Haidt, J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108, no. 4: 814–834.

Hayes, A. F. 2017. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Publications.

Henderson, R. K., and S. Schnall. 2021. "Social Threat Indirectly Increases Moral Condemnation via Thwarting Fundamental Social Needs." *Scientific Reports* 11, no. 1: 21709.

Hofmann, W., D. C. Wisneski, M. J. Brandt, and L. J. Skitka. 2014. "Morality in Everyday Life." *Science* 345, no. 6202: 1340–1343.

Jordan, J. J., M. Hoffman, P. Bloom, and D. G. Rand. 2016. "Third-Party Punishment as a Costly Signal of Trustworthiness." *Nature* 530, no. 7591: 473–476.

Jordan, J. J., and D. G. Rand. 2020. "Signaling When No One Is Watching: A Reputation Heuristics Account of Outrage and Punishment in One-Shot Anonymous Interactions." *Journal of Personality and Social Psychology* 118, no. 1: 57–88.

Kermani, M. S., T. J. Noseworthy, and P. R. Darke. 2024. "Getting Political: The Value-Protective Effects of Expressed Outgroup Outrage on Self-Brand Connection." *Journal of Consumer Psychology* 34, no. 3: 385–405.

Konishi, N., T. Oe, H. Shimizu, K. Tanaka, and Y. Ohtsubo. 2017. "Perceived Shared Condemnation Intensifies Punitive Moral Emotions." *Scientific Reports* 7, no. 1: 7289.

Kray, L. J., L. Howland, A. G. Russell, and L. M. Jackman. 2017. "The Effects of Implicit Gender Role Theories on Gender System Justification: Fixed Beliefs Strengthen Masculinity to Preserve the Status Quo." *Journal of Personality and Social Psychology* 112, no. 1: 98.

Kristofferson, K., K. White, and J. Peloza. 2014. "The Nature of Slacktivism: How the Social Observability of an Initial Act of Token Support Affects Subsequent Prosocial Action." *Journal of Consumer Research* 40, no. 6: 1149–1166.

Kurzban, R., P. DeScioli, and E. O'Brien. 2007. "Audience Effects on Moralistic Punishment." *Evolution and Human Behavior* 28, no. 2: 75–84.

Leach, S., M. Formanowicz, J. Nikadon, and A. Cichocka. 2025. "Moral Outrage Predicts the Virality of Petitions for Change on Social Media, but Not the Number of Signatures They Receive." *Social Psychological and Personality Science*. In press. https://journals.sagepub.com/doi/10.1177/19485506251335373.

Leonel, R., K. Rehbein, M. Westermann-Behaylo, and E. Perrault. 2024. "Firms' Response to Slacktivism: When and Why Are E-Petitions Effective?" *Journal of Management Studies* 61, no. 7: 3148–3183.

Lindenmeier, J., C. Schleer, and D. Pricl. 2012. "Consumer Outrage: Emotional Reactions to Unethical Corporate Behavior." *Journal of Business Research* 65, no. 9: 1364–1373.

Mooijman, M., and W. W. Van Dijk. 2015. "The Self in Moral Judgement: How Self-Affirmation Affects the Moral Condemnation of Harmless Sexual Taboo Violations." *Cognition and emotion* 29, no. 7: 1326–1334.

Nasa, J., J. D. Rotman, K. R. Mercurio, M. G. Staton, and A. Vocino. 2025. "The Moral Compass of Identity: Ethical Predispositions Predict the Importance Consumers Ascribe to Their Group and Individual Identities." *Journal of the Association for Consumer Research* 10, no. 1: 11–23.

O'Mara, E. M., L. E. Jackson, C. D. Batson, and L. Gaertner. 2011. "Will Moral Outrage Stand Up?: Distinguishing Among Emotional Reactions to a Moral Violation." *European Journal of Social Psychology* 41, no. 2: 173–179.

O'Halloran, S. (2023, May 31). Bud Light Parent Anheuser-Busch Sees $27 Billion Gone, Shares Near Bear Market. FOXBusiness, https://www.foxbusiness.com/markets/bud-light-parent-anheuser-busch-sees-27-billion-gone-shares-near-bear-market.

Parkel. 2023. *John Oliver Condemns Bud Light's Response Following Dylan Mulvaney Backlash*. https://www.independent.co.uk/arts-entertainment/tv/news/john-oliver-bud-light-dylan-mulvaney-b2325781.html.

Raghunathan, R., and K. Corfman. 2006. "Is Happiness Shared Doubled and Sadness Shared Halved? Social Influence on Enjoyment of Hedonic Experiences." *Journal of Marketing Research* 43, no. 3: 386–394.

Raihani, N. J., and R. Bshary. 2015. "The Reputation of Punishers." *Trends in Ecology & Evolution* 30, no. 2: 98–103.

Reed, A., M. R. Forehand, S. Puntoni, and L. Warlop. 2012. "Identity-Based Consumer Behavior." *International Journal of Research in Marketing* 29, no. 4: 310–321.

Romani, S., S. Grappi, and R. P. Bagozzi. 2013. "My Anger Is Your Gain, My Contempt Your Loss: Explaining Consumer Responses to Corporate Wrongdoing." *Psychology & Marketing* 30, no. 12: 1029–1042.

Ronson, J. 2016. So You've Been Publicly Shamed: *Riverhead Books*. Hardcover.

Rothschild, Z. K., and L. A. Keefer. 2017. "A Cleansing Fire: Moral Outrage Alleviates Guilt and Buffers Threats to One's Moral Identity." *Motivation and Emotion* 41, no. 2: 209–229.

Rozin, P., L. Lowery, S. Imada, and J. Haidt. 1999. "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)." *Journal of Personality and Social Psychology* 76, no. 4: 574–586.

Salerno, J. M., and L. C. Peter-Hagene. 2013. "The Interactive Effect of Anger and Disgust on Moral Outrage and Judgments." *Psychological Science* 24, no. 10: 2069–2078.

Salerno, J. M., and M. L. Slepian. 2022. "Morality, Punishment, and Revealing Other People's Secrets." *Journal of Personality and Social Psychology* 122, no. 4: 606–633.

Sawaoka, T., and B. Monin. 2018. "The Paradox of Viral Outrage." *Psychological Science* 29, no. 10: 1665–1678.

Sawaoka, T., and B. Monin. 2020. "Outraged but Sympathetic: Ambivalent Emotions Limit the Influence of Viral Outrage." *Social Psychological and Personality Science* 11, no. 4: 499–512.

Schein, C., and K. Gray. 2018. "The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm." *Personality and Social Psychology Review* 22, no. 1: 32–70.

Schmader, T., and C. Sedikides. 2018. "State Authenticity as Fit to Environment: The Implications of Social Identity for Fit, Authenticity, and Self-Segregation." *Personality and Social Psychology Review* 22, no. 3: 228–259.

Shah, Z., J. Chu, S. Qaisar, Z. Hassan, and U. Ghani. 2020. "Perceived Public Condemnation and Avoidance Intentions: The Mediating Role of Moral Outrage." *Journal of Public Affairs* 20, no. 1: e2027.

Shariatmadari, D. (2016). 'Virtue-Signalling' – The Putdown That Has Passed Its Sell-by Date. *The Guardian*. https://www.theguardian.com/commentisfree/2016/jan/20/virtue-signalling-putdown-passed-sell-by-date.

Skitka, L. J., C. W. Bauman, and E. Mullen. 2004. "Political Tolerance and Coming to Psychological Closure Following the September 11, 2001, Terrorist Attacks: An Integrative Approach." *Personality and Social Psychology Bulletin* 30, no. 6: 743–756.

Skitka, L. J., C. W. Bauman, and E. G. Sargis. 2005. "Moral Conviction: Another Contributor to Attitude Strength or Something More?" *Journal of Personality and Social Psychology* 88, no. 6: 895–917.

Spring, V. L., C. D. Cameron, and M. Cikara. 2018. "The Upside of Outrage." *Trends in Cognitive Sciences* 22, no. 12: 1067–1069.

Stone, J., and N. C. Fernandez. 2008. "To Practice What we Preach: The Use of Hypocrisy and Cognitive Dissonance to Motivate Behavior Change." *Social and Personality Psychology Compass* 2, no. 2: 1024–1051.

Taylor, D. G., D. Strutton, and K. Thompson. 2012. "Self-Enhancement as a Motivation for Sharing Online Advertising." *Journal of Interactive Advertising* 12, no. 2: 13–28.

Tetlock, P. E., O. V. Kristel, S. B. Elson, M. C. Green, and J. S. Lerner. 2000. "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals." *Journal of Personality and Social Psychology* 78, no. 5: 853–870.

Tong, S. T. 2024. "Foundations, Definitions, and Directions in Online Hate Research." In *Social Processes of Online Hate*, 37–72. Routledge.

Tosi, J., and B. Warmke. 2016. "Moral Grandstanding." *Philosophy & Public Affairs* 44, no. 3: 197–217.

Walther, J. B. 2024. "The Effects of Social Approval Signals on the Production of Online Hate: A Theoretical Explication." *Communication Research*. In press. https://journals.sagepub.com/doi/10.1177/00936502241278944.

Walther, J. B., and R. E. Rice. 2025. *Social Processes of Online Hate*, 336. Taylor & Francis.

Westra, E. 2021. "Virtue Signaling and Moral Progress." *Philosophy & Public Affairs* 49, no. 2: 156–178.

Xie, C., and R. P. Bagozzi. 2019. "Consumer Responses to Corporate Social Irresponsibility: The Role of Moral Emotions, Evaluations, and Social Cognitions." *Psychology & Marketing* 36, no. 6: 565–586.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.
Supplementary Appendix - Final.