White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/170127/

Version: Accepted Version

Article:
Christodoulides, G, Gerrath, MHEE and Siamagka, NT (2021) Don't be rude! The effect of content moderation on consumer-brand forgiveness. Psychology and Marketing, 38 (10). pp. 1686-1699. ISSN: 0742-6046

https://doi.org/10.1002/mar.21458

© 2021 Wiley Periodicals LLC. This is the peer reviewed version of the following article: Christodoulides, G, Gerrath, MHEE and Siamagka, NT (2021) Don't be rude! The effect of content moderation on consumer-brand forgiveness. Psychology and Marketing, 38 (10). pp. 1686-1699. ISSN 0742-6046, which has been published in final form at https://doi.org/10.1002/mar.21458. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Don't be rude! The effect of content moderation on consumer-brand forgiveness

GEORGE CHRISTODOULIDES*
American University of Sharjah

MAXIMILIAN H. E. E. GERRATH
University of Leeds

NIKOLETTA THEOFANIA SIAMAGKA
Aristotle University of Thessaloniki

George Christodoulides is the Chalhoub Group Professor of Luxury Brand Management, American University of Sharjah, University City PO Box 26666 Sharjah, UAE (gchristodoulides@aus.edu).

Maximilian H. E. E. Gerrath is a Lecturer (Assistant Professor) of Marketing, Marketing Division, Leeds University Business School, University of Leeds, Maurice Keyworth Building, LS2 9JT, Leeds, UK (m.gerrath@leeds.ac.uk).

Nikoletta Theofania Siamagka is an Assistant Professor of Marketing, Aristotle University of Thessaloniki (AUTH), Department of Economics, Aristotle University Campus, 54124, Thessaloniki, Greece (nsiamagka@econ.auth.gr).

*Please address correspondence concerning this article to George Christodoulides (gchristodoulides@aus.edu). All authors have contributed equally and are listed in alphabetical order.

SHORT RUNNING TITLE

Content Moderation and Forgiveness

KEYWORDS

negative word of mouth, content moderation, complaint management, free speech, emotions, consumer-brand forgiveness

CONFLICT OF INTEREST, ETHICS AND INTEGRITY

The authors confirm they have no conflict of interest to declare. Authors also confirm that this article adheres to the ethical guidelines specified in the APA Code of Conduct as well as the authors' national ethics guidelines. The authors confirm that the article submitted, to the knowledge of all authors, has not been published elsewhere previously and is not under consideration for publication elsewhere. This submission is approved by all authors and by the responsible authorities where the authors carried out the work. If accepted for publication, this article will not be published elsewhere including electronically in the same form, in English, or in any other language, without the written consent of the copyright-holder.

Don’t be rude! The effect of content moderation on consumer-brand forgiveness

Abstract

While it is a popular belief that venting helps unload frustrations about negative (customer) experiences, its effects on consumers’ emotional states and consumer-brand forgiveness (CBF) remain to be explored. Given that a lot of customer complaints are made online, brands seem ambivalent about managing these public complaints without violating consumers’ right to free speech. In two experiments, we find that writing a customer complaint increases negative emotions regarding an incident. Moreover, brands can mitigate this negative effect by asking consumers to moderate their speech—for example, with the help of a content moderation policy. Specifically, if brands impose restrictions on consumers’ freedom to express their frustrations, essentially asking them for self-censorship, anger levels in their language decrease and CBF increases. We demonstrate that this effect is stronger for consumers with strong self-brand connection.

KEYWORDS

Negative word of mouth, Content moderation, Complaint management, Free speech, Emotions, Consumer-brand forgiveness

1. INTRODUCTION

Even good brands may at some point fail to deliver an experience that is in line with customer expectations (Aaker, Fournier, & Brasel, 2004; Khamitov, Grégoire, & Suri, 2020). In such cases, and given today's hyperconnected world, companies and brands are exposed to the impact of negative online word of mouth (WOM) and customer complaints; this has often been referred to as "social media call-out culture" (Sprout Social, 2020; Suciu, 2020). Specifically, many customers, empowered by technology, turn to social media to vent their frustrations, register their complaints, and seek retribution for the perceived wrongdoings of brands (Berger, 2014; Grégoire, Salle, & Tripp, 2015). According to a study conducted by Sprout Social (2017), 81% of consumers believe that social media has increased brand accountability. Moreover, a recent study by WhoIsHostingThis.com (2019) reveals that 51% of users had called out a brand on social media before. Specifically, platforms such as Twitter and Facebook are perceived as optimal media for customers to voice their concerns and/or to actively provide feedback to brands (WhoIsHostingThis.com, 2019). Indeed, firms often proactively use their social media accounts to receive and address customer service inquiries and complaints.

Social media complaints clearly present both opportunities and threats for brands. From a customer relationship and retention perspective, brands need to manage customers' frustrations by attempting to mitigate their negative emotions and ultimately convince them to forgive the offending brand. At the same time, complaints on social media are publicly visible to other consumers (Schaefers & Schamari, 2016). As such, they can be seen as negative WOM that potentially threatens the reputation of the firm.

In their efforts to manage customers' frustrations and protect their own reputation on social media, brands often face a conundrum, as their attempts to moderate consumers' negative WOM

may be perceived as an attack on freedom of expression. For example, brands may introduce policies that enable them to moderate the content of (negative) online reviews or complaints posted on their social media accounts (e.g., their official Facebook page). In more extreme cases, brands have been reported to censor or even completely remove legitimate consumer online reviews from the public sphere (Shaw & Coker, 2012; Trustpilot, 2020).

While there is a sizeable stream of literature on brand response to consumer complaints on social media (e.g., Johnen & Schnittka, 2019) and subsequent recovery strategies (e.g., Hogreve, Bilstein, & Hoerner, 2019), particularly in services, the focus has been on strategies deployed *after* a customer complaint. The current research goes beyond this extant work by examining how the management of negative customer emotions may start even *before* complaints appear on social media, thus activating the consumer-brand forgiveness (CBF) process at an earlier point in time. Specifically, this paper contributes to research by examining how content moderation reduces negative emotional responses from consumers (as reflected by sentiment analysis of written complaints) and increases CBF.

We begin by reviewing relevant theory on WOM and free speech, the psychology of venting negative emotions, and CBF. We then develop hypotheses about the impact of content moderation policies on CBF, the mediating effect of emotional tone, and the moderating effect of self-brand connection. Next, we outline the scale development procedures we followed to arrive at a multidimensional measure of CBF and report the findings of two experimental studies. We conclude by discussing the implications of our findings for theory and practice and by identifying avenues for further research.

2. THEORETICAL BACKGROUND AND HYPOTHESES DEVELOPMENT

2.1. Interpersonal forgiveness

Interpersonal forgiveness has been studied extensively in the psychology domain (e.g., Barcaccia et al., 2020; Burnette, McCullough, Van Tongeren, & Davis, 2012; Fehr, Gelfand, & Nag, 2010). It is characterized as a therapeutic process (Freedman & Zarifkar, 2016) that involves forgoing vengeful behavior (e.g., McCullough et al., 1998) in which individuals who are hurt or violated engage in a series of processes to "replace destructive response with constructive behavior to achieve reconciliation with others" (Yao, Chen, Yu, & Sang, 2017, p. 585). Despite diverse conceptualizations in the literature (Freedman, Enright, & Knutson, 2005; Riek & Mania, 2012), interpersonal forgiveness has been recognized as a multidimensional construct, consisting of *affective*, *cognitive*, and *behavioral* facets (e.g., Subkoviak et al., 1995; Takaku, 2001). These facets reflect the following three changes that victims experience during the process of forgiveness: the generation of positive and benevolent attributions, the reduction of negative affect (e.g., anger, resentment) and development of positive emotions (e.g., sympathy, compassion), and finally the increased likelihood of positive behavior toward the offender/transgressor (Takaku, 2001).

Interpersonal forgiveness has been linked to several positive outcomes, including improved health and quality of life (e.g., Bono & McCullough, 2006; Currier, Drescher, Holland, Lisman, & Foy, 2016), well-being (e.g., Chen, Harris, Worthington, & VanderWeele, 2019; Karremans, Van Lange, Ouwerkerk, & Kluwer, 2003; Yao et al., 2017), cooperation (e.g., Clarke & Rhodes, 2020; Komorita, Hilty, & Parks, 1991), relationship satisfaction, commitment, and trust (e.g., Fahimdanesh, Noferesti, & Tavakol, 2020; McCullough et al., 1998). In contrast, less forgiving individuals tend to experience higher levels of negative emotions, such as anxiety, anger, and

bitterness (e.g., Gençoğlu, Şahin, & Topkaya, 2018; Little, Simmons, & Nelson, 2007; Takaku, 2001).

2.2. Consumer-brand forgiveness (CBF)

Similar to interpersonal forgiveness, where the precursor is an interpersonal offence or wrongdoing, consumer-brand forgiveness (CBF) can be triggered after a perceived negative incident with a brand. Specifically, CBF can be viewed as a relationship constructive mechanism, whereby consumers engage in a process to overcome post-transgression negative emotions and maintain a constructive relationship with the brand (Chung & Beverland, 2006). As Joireman, Grégoire, and Tripp (2016, p. 76) note, most definitions of forgiveness in consumer research have the following in common: "reduced anger and obsession with the offender or offense, a willingness to foreswear revenge, and enhanced compassion and generosity toward the offender." Similar to interpersonal forgiveness, CBF suggests that consumers refrain from vengeful behavior and release the brand from its debt (Schnebelen & Bruhn, 2018; Xie & Peng, 2009). CBF is critically important to businesses that have offended or failed consumers. Numerous studies have explored consumer outcomes and have linked forgiveness to WOM (e.g., Casidy & Shin, 2015; Harrison-Walker, 2019; Tsarenko & Tojib, 2012), switching behavior (e.g., Tsarenko & Tojib, 2012), relationship satisfaction (Muhammad & Gul-E-Rana, 2020), trust (Schoorman, Mayer, & Davis, 2007; Xie & Peng, 2009), repatronage behavior, and reconciliation (e.g., Harrison-Walker, 2019).

In contrast to forgiveness in the psychology domain, the literature on CBF is much less developed, and the construct has only recently received increased interest among scholars (Casidy & Shin, 2015). Interestingly, the dimensionality of consumer forgiveness diverges from that of interpersonal forgiveness in that it emphasizes only the behavioral component of the construct. As

Xie and Peng (2009, p. 578) argue, CBF captures "consumers' willingness to give up retaliation, alienation, and other destructive behaviors, and to respond in constructive ways after an organizational violation of trust and the related recovery efforts." This definition highlights the behavioral component of forgiveness but fails to capture the cognitive and emotional aspects inherent to the concept (McCullough, Worthington, & Rachal, 1997). Thus, the operationalization of consumer forgiveness has been limited to unidimensional scales and, in some cases, to single-item measures (e.g., Sinha & Lu, 2016). In other cases, scholars have inherited measures from the psychology domain (e.g., Casidy & Shin, 2015; Harrison-Walker, 2019; Wei & Ran, 2019; Xie & Peng, 2009), which fail to capture the idiosyncrasies of forgiveness in a consumer-brand relationship context.

2.3. Venting, consumers' emotional states, and CBF

Negative emotions, such as anger, can be experienced as well as expressed. People often feel the need to communicate their negative feelings to others (Canary, Spitzberg, & Semic, 1998; Rimé, Mesquita, Boca, & Philippot, 1991). The suppression or overexpression of anger may lead to serious mental and physical health issues (e.g., Leonhardt, Lysaker, Vohs, James, & Davis, 2018; Pennebaker, Kiecolt-Glaser, & Glaser, 1988). Because people have this need to communicate negative emotional experiences, a popular lay hypothesis is that verbalizing and sharing negative experiences with others may give the affected person a certain feeling of relief. This behavior is often referred to as "venting," "letting off steam," or "looking for catharsis" (Rimé, 2009). According to Zech's (1999) estimates, approximately 80% of people believe that venting helps. Research in social psychology, however, does not confirm this theory in a general sense (Nils & Rimé, 2012; Pennebaker, Zech, & Rimé, 2001). For example, Bushman (2002) shows that people given the opportunity to hit a punching bag while thinking of someone who has angered them were

angrier than people who did not have this opportunity. Similarly, Dalebroux, Goldstein, and Winner (2008) contend that encouraging positive emotions is a more effective strategy than venting when it comes to short-term mood repair.

According to Nils and Rimé (2012), the conclusion that sharing negative emotions, or venting, does not lead to positive outcomes is not surprising. The authors argue that venting may compel people to reaccess memories, reappraise a situation in the same way as before, and ultimately reactivate the original negative experience. Thus, prior literature concludes that the outcomes of venting may depend on several factors, such as the way in which a negative experience is shared or the audience with whom it is shared (Lepore, Fernandez-Berrocal, Ragan, & Ramos, 2004; Parlamis, 2012). Parlamis (2012) suggests that venting is a strategy that depends on a person's interaction with others—specifically, the mere act of venting does not regulate emotions. However, people reported feeling better if venting was succeeded by feedback from others (e.g., the offender)—especially if offenders attributed their actions to uncontrollable causes.

In a consumption context, consumers often share their negative brand experiences with others (Berger, 2014). This act of venting takes often place online in the form of customer complaints, negative product reviews, or other types of negative WOM (Grewal, Stephen & Bart, 2019; Nam, Baker, Ahmad, & Goo, 2020; Nyer, 1997; Wetzer, Zeelenberg, & Pieters, 2007; Zeelenberg & Pieters, 2004). Consumer venting is mostly targeted at an offending brand and does not necessarily generate an immediate response from the brand. Thus, the mere act of venting, in the form of a customer complaint, may be presumed to drive consumers to revisit this negative customer experience, which may worsen their negative emotional state (e.g., higher anger levels). We expect this negative emotional state to ultimately drive consumers to be less forgiving toward the brand.

Indeed, this negative link between venting and forgiveness can be explained through the process of rumination, whereby people mull the incident over in their minds (Caprara, 1986). This process results in lower levels of forgiveness (e.g., Barber, Maltby, & Macaskill, 2005; McCullough, 2000; McCullough et al., 1998), as ruminating is associated with higher levels of avoidance and revenge (McCullough, 2000). It is, therefore, rational to argue that the mere act of venting might instigate the process of rumination, leading to lower levels of forgiveness. Accordingly, we hypothesize the following:

H1: *Venting by means of writing a customer complaint is negatively associated with CBF.*

2.4. Negative WOM, freedom of speech, and content moderation

Consumers rely on the experiences of others when making purchase decisions (Babić Rosario, Sotgiu, De Valck, & Bijmolt, 2016). Negativity bias suggests that people are more strongly influenced by negative than by positive entities (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). Laczniak, DeCarlo, and Ramaswami (2001) show that negative WOM can determine consumers' brand evaluations. Specifically, consumers may attribute negative WOM to the product's poor quality or the communicator's poor judgment—which ultimately determines whether negative WOM affects brand evaluations. Chevalier and Mayzlin (2006) establish a link between negative WOM and product sales. Furthermore, the authors find that consumers rely more strongly on a review's textual content than on its summary statistics (e.g., star rating). Babić Rosario et al. (2016) build on this to show that negative reviews are particularly damaging if there is high variability in review valence. In addition, Hornik, Satchi, Cesareo, and Pastore (2015) show that people share negative WOM more quickly and elaborately than positive WOM.

Thus, brand managers are looking for ways to manage and improve the way consumers talk about their brands online (Berger, 2014; Grewal, Stephen, & Coleman, 2019; Reimer & Benkenstein, 2016). One way to manage negative WOM is to impose restrictions on consumers' ability to freely express their feelings about a brand; this is often accomplished with a content moderation policy. In the spirit of Gillespie (2018) and Gibson (2019), we define content moderation as a brand's active attempt to monitor and censor consumer-generated content. This process includes establishing, communicating, and reinforcing community guidelines. Despite the importance of moderating the content of negative WOM for brands, marketing research needs to develop a deeper understanding of this practice. Most research in marketing focuses on consumers' individual differences or biases that drive them to self-moderate (Antonetti & Maklan, 2018; Khamitov et al., 2020), while few marketing papers examine content moderation as an intervention by the brand.

For example, Madio and Quinn (2020) examine the impact of content moderation on ad pricing in the context of brand safety, while Shaw and Coker (2012) demonstrate that deleting or censoring negative WOM decreases a brand's credibility and trustworthiness. Despite this, Dekay (2012) finds that large corporations often censor or ignore critical feedback from consumers. Similarly, Pantano and Corvello (2013) show that removing consumer posts is a commonly employed technique by more experienced brands with many followers. Finally, in a linguistic analysis of postings about sociopolitical issues on Reddit's message board, Gibson (2019) finds that content moderation policies may influence the levels of anger and negativity in posts. Specifically, Gibson (2019) argues that "safe space" (vs. "free speech") content moderation policies may lead to self-censorship and affect language use.

Thus, we posit that content moderation affects how consumers express their emotions in written customer complaints and ultimately influences whether they forgive offending brands. Specifically, we postulate that content moderation activates a certain level of self-censorship, which results in a more positive emotional tone in the written customer complaints. This self-censoring effect on consumers' freedom to express their emotions causes a lower level of reactivation of the negative emotions experienced due to reduced rumination on the negative customer experience (Nils & Rimé, 2012), thus leading to lower salience levels of negative emotions. Finally, due to the more positive emotional state, consumers may, therefore, be more willing to forgive the brand that wronged them.

H2a: *The emotional tone of a customer complaint is more positive if a content moderation policy is in place than if there is no content moderation policy.*

H2b: *Consumers are more forgiving of a brand if a content moderation policy is in place than if there is no content moderation policy.*

H3: *The relationship between content moderation and CBF is mediated by the emotional tone of the customer complaint.*

2.5. Self-brand connection

Brands are a part of consumers' identities (Belk, 1989, 2013; Stokburger-Sauer, Ratneshwar, & Sen, 2012). As consumers pursue self-enhancement and self-verification goals, they establish and maintain strong self-brand connections (Escalas & Bettman, 2003). Thus, despite the need to share negative experiences with others, many consumers are hesitant to share their true emotions and opinions in a public setting. This is because posting about products and brands online may be considered an act of identity signaling, which is accompanied by some degree of social risk (Berger & Heath, 2007; De Angelis, Bonezzi, Peluso, Rucker, & Costabile, 2012; Grewal, Stephen, & Coleman, 2019; Palmeira, Spassova, & Quoidbach, 2020). Hence, consumers are generally more

hesitant to ruminate about negative aspects of a brand, or spread negative WOM about brands that are a part of their own identity, as this would constitute a threat to their self-concept (Cheng, White, & Chaplin, 2012; De Angelis et al., 2012; Lisjak, Lee, & Gardner, 2012).

Similarly, consumers who maintain strong emotional bonds with brands are usually more likely to recommend these brands to others and spread positive WOM (Batra, Ahuvia, & Bagozzi, 2012; Carroll & Ahuvia, 2006; Park, MacInnis, Priester, Eisingerich, & Iacobucci, 2010). Strong and positive brand relationships can also act as safety cushions that drive consumers to react more positively to negative brand information and experiences (Ahluwalia, Burnkrant, & Unnava, 2000; Khamitov et al., 2020). In addition, one may assume that highly attached consumers are more willing to obey content moderation policies of a brand. In contrast, a consumer with a weak self-brand connection prior to an incident may be less likely to follow the content moderation policies that a brand imposes. Thus, we hypothesize the following:

H4: *A consumer's self-brand connection moderates the effect of content moderation on the emotional tone of customer complaints (such that low self-brand connection mitigates the positive effect of content moderation on emotional tone).*

3. OVERVIEW OF STUDIES

To examine how content moderation influences CBF, we conducted two experiments preceded by qualitative interviews and two survey studies to develop and validate a new measurement scale of CBF. We deemed this necessary given that none of the existing measures of consumer forgiveness treats the construct as multidimensional, thus failing to capture its complexity and nuanced manifestations. The purpose of experimental Study 1 is twofold. First, we explore whether the mere act of freely expressing one's frustration about a negative customer experience decreases CBF. Second, Study 1 examines whether content moderation imposed by brands on consumers enhances consumers' forgiveness levels following a negative experience with a given brand. In

Study 2, we test the robustness of our results based on actual negative customer experiences while analyzing the mediating role of consumers' emotional states reflected in their written complaints. In Study 2, we also propose self-brand connection as a moderator. We provide an overview of the relationships we aim to examine in Figure 1.

Insert Figure 1 Here

4. SCALE DEVELOPMENT: CBF

Due to the absence of a multidimensional scale to measure CBF that captures different facets of forgiveness in line with the literature from interpersonal psychology, we developed a scale, following established scaling procedures (Netemeyer, Bearden, & Sharma, 2003). We first conducted in-depth interviews to better understand the concept of CBF and to identify pertinent dimensions. Qualitative research was necessary in this case as the existing literature offers different definitions of consumer forgiveness and the research in this area has been relatively limited. We then used the main interview findings and the relevant literature on interpersonal forgiveness to develop the initial item pool and conducted two survey studies to develop and validate the CBF scale.

4.1. Qualitative stage

We adopted a grounded theory approach in which we drew on prior knowledge "while keeping a fresh and open mind to new concepts as they emerge from the data" (Goulding, 2005, p. 296). We interviewed 16 consumers face-to-face and asked them to recall an incidence with a brand that let them down and to describe, inter alia, their emotions, thoughts, and behaviors vis-à-vis the brand.

On average, interviews lasted 40 minutes each. We recorded and transcribed all interviews verbatim.

We reviewed the interview transcripts and made notes with regard to the process that consumers followed after having a negative experience with a brand. The total list of notes constituted our codebook. We categorized codes based on their content, and the main components of CBF emerged through this process. The results supported a three-dimensional structure of consumer forgiveness, encompassing affective, cognitive, and behavioral facets.

The affective component is captured by feelings of betrayal, disappointment, and loss of faith in the brand (e.g., Interviewee 13 claimed that she felt "*disappointed […], almost betrayed*"). The cognitive component encapsulates consumer evaluations and thoughts following negative experiences with a brand (e.g., Interviewee 4 indicated that he does not buy the brand anymore "*because of my perception of the brand now and the actual products they sell*"). Finally, the behavioral facet encompasses consumers' behavioral intentions, suggesting that unforgiveness is associated with switching behavior (e.g., Interviewee 5 claimed that if she has options, she would "*go somewhere else rather than them*"). Thus, in line with the extant literature on forgiveness and our qualitative findings, we define CBF as the consumer's cognitive, affective, and behavioral response to a brand's perceived wrongdoing, with the aim of maintaining a constructive relationship with the brand.

4.2. Quantitative stage

Next, we embarked on two survey rounds to develop the scale (calibration sample) and then validate it (validation sample). The initial item pool involved items drawn from the qualitative stage as well as the relevant literature from psychology and consisted of 41 items (11 items for

cognitive, 16 items for affective, and 14 items for behavioral forgiveness). In line with existing research (McCullough et al., 1998), in both surveys, respondents were instructed to recall an incident with a brand that let them down. Based on this incident, they were then asked to answer a series of questions, including those measuring CBF, as well as demographic and nomological network variables. For both rounds, we collected data in the United Kingdom using an online consumer panel. In total, we received 257 fully completed questionnaires from our calibration sample and 301 from our validation sample.

We performed a confirmatory factor analysis on the calibration sample to test the dimensionality of the scale and develop the scale. Through an iterative process that involved deleting items on the basis of modification indices and standardized residuals, we reduced the initial item pool and obtained an acceptable fit ($\chi^2(24) = 63.974$, comparative fit index [CFI] = .98, normed fit index [NFI] = .97, root mean square error of approximation [RMSEA] = .08). The resultant scale consisted of three dimensions (affective, cognitive, and behavioral) and nine items in total (three per dimension). Internal consistency, composite reliabilities, and average variances extracted (AVEs) were all within acceptable levels (Bagozzi & Yi, 1988; Fornell & Larcker, 1981): cognitive ($\alpha = .92$, CR = .83, AVE: = .80), affective ($\alpha = .85$, CR = .85, AVE = .66), and behavioral ($\alpha = .93$, CR = .94, AVE = .83). Internal consistency reliability for the CBF scale was acceptable ($\alpha = .93$). Using Fornell and Larcker's (1981) criteria, we established convergent and discriminant validity. More specifically, all AVEs of the three dimensions were greater than .50, indicating convergent validity, and the square root of each dimension's AVE was greater than the correlations between the dimensions, supporting discriminant validity. The correlations between the dimensions range from .55 to .82, and the square root of the AVEs range from .81 to .91. The items and item loadings appear in Table 1.

Following the development of the CBF scale, the second stage involved more validation tests on a fresh sample. We performed CFA again, this time on the validation sample (n = 301). The fit indices were satisfactory ($\chi^2(24) = 69.759$, CFI = .98, NFI = .97, RMSEA = .08), supporting the three-dimensional structure of the construct. Internal consistency and composite reliabilities as well as AVEs were within acceptable levels: cognitive (α = .91, CR = .91, AVE: = .78), affective, (α =.82, CR = .83, AVE = .62), and behavioral (α = .91, CR = .91, AVE = .77). The CBF scale was reliable (α = .91), and the results support both discriminant and convergent validities (Fornell & Larcker, 1981). The correlations between the dimensions ranged from .47 to .80, and the square root of the AVEs ranged from .87 to .93. We assessed criterion-related validity by establishing the relationship between CBF and two outcome variables from the literature: brand love (measured with a six-item scale adapted from Bagozzi, Batra, and Ahuvia [2017]) and relationship commitment (measured with a three-item scale adapted from Morgan and Hunt [1994]). We performed regression analysis using the two aforementioned variables as the dependent variables. The results confirm that CBF is positively associated with brand love (β = .38, $p < .01$) and relationship commitment (β = .77, $p < .01$).

Insert Table 1 Here

5. STUDY 1

We designed Study 1 to test our prediction that freely expressed frustrations about a negative customer experience with a brand (i.e., lack of a moderation policy) decreases consumers' willingness to forgive a brand. Moreover, we examine if a brand's content moderation policies mitigate this effect by increasing consumers' brand forgiveness levels.

5.1. Method

146 American Amazon Mechanical Turk (MTurk) users (M_{age} = 38 years; 52.7% male, 46.6% female, 0.7% other) participated in a three-condition (no content moderation vs. content moderation vs. control) between-subjects experiment. First, participants had to pass an attention check, in line with Oppenheimer, Meyvis, and Davidenko's (2009) recommendations. Then, we asked all participants to read a scenario that described a negative customer experience with the fictitious brand "Compobooks." Specifically, the scenario instructed participants to imagine that they did not receive the correct laptop model they had ordered online and that it took the brand more than a month to replace the device with the correct version of the laptop. After this reading task, we randomly allocated participants to one of the three experimental conditions (we provide an overview of the manipulations in Appendix A).

In the content moderation condition, we asked participants to imagine that they decided to write a complaint on the brand's official message board. Moreover, we asked participants to read the sign-up agreement for the message board, which stated that "aggressive language and threats will not be tolerated" and that "posts that violate this agreement will be removed." Finally, we asked the participants to draft a post for the message board.

In the no-content moderation condition, participants were also asked to imagine writing a complaint on the brand's official message board. In contrast to the content moderation condition, however, the no-content moderation condition did not include any sign-up agreement or content moderation policy. Finally, we asked participants to write a post for the message board.

In the control condition, we provided participants with a filler writing task. Instead of writing a customer complaint on the brand's message board, as in the other two conditions, we asked participants to describe a picture of an American town.

Following the writing task, we asked participants to indicate their level of forgiveness on our newly developed nine-item seven-point CBF Likert scale (α = .89). We also captured participant demographics. Finally, we administered a manipulation check by asking participants how free they felt in performing the writing task using a seven-point scale (i.e., "not at all free" vs. "extremely free").

5.2. Results

In terms of manipulation checks, we found a significant difference across all three conditions ($F(2, 143) = 3.438$, $p < .05$). As we predicted, a Tukey HSD post-hoc test revealed a significant difference in perceived freedom between the content moderation ($M = 4.52$, $SD = 1.75$) and no-content moderation ($M = 5.39$, $SD = 1.73$; $p < .05$) conditions, while the control condition did not differ significantly with either of the other two conditions ($M = 4.85$, $SD = 1.43$).

After running the manipulation check, we performed a one-way analysis of variance (ANOVA) with the manipulated factor (no content moderation vs. content moderation vs. control) as the independent variable (IV) and CBF as the dependent variable (DV). Again, we found a significant main effect of the manipulations on CBF ($F(2, 143) = 3.914$, $p < .05$). As we predicted in H1, a Tukey HSD post-hoc test revealed a significant difference between the no-content moderation ($M = 2.81$, $SD = 0.95$) and the control ($M = 3.40$, $SD = 1.40$; $p < .05$) conditions. Moreover, in line with H2b, we found a significant difference between the no-content moderation ($M = 2.81$, $SD = 0.95$) and content moderation ($M = 3.39$, $SD = 1.04$; $p < .05$) conditions. When running the ANOVAs for each of the three dimensions of CBF individually, we found similar interaction patterns as on aggregate level. While the manipulation (no content moderation vs. content moderation vs. control) generated a significant main effect for affective CBF ($F(2, 143) = 3.565$, $p < .05$), the effects for cognitive CBF ($F(2, 143) = 2.849$, $p = .06$) and behavioral CBF

($F(2, 143) = 2.823$, $p = .06$) only reached marginal significance. This may perhaps be unsurprising due to the fictitious nature of the brand used in the scenarios. Participants may be able to form affective reactions in relation to a fictitious brand based on a scenario. However, it may be harder for participants to develop complex cognitive thoughts (e.g., deeming a brand unworthy) or to generate behavioral intentions (e.g., avoiding a brand) for brands that they are not familiar with. Thus, in Study 2 we aim to build on those findings by examining the effects of content moderation in the context of real brand incidences.

Insert Figure 2 Here

5.3. Discussion

As predicted, we find that the mere act of freely expressing frustration with a brand decreases consumers' levels of forgiveness. Indeed, consumers who were not restricted by any content moderation showed lower CBF levels than those who did not write a customer complaint at all. Furthermore, those who were given less freedom to express their discontent—due to a content moderation policy—showed higher levels of CBF than those who were free to express their anger without restriction. Thus, Study 1 supports our prediction that the act of complaining about a negative customer experience decreases consumers' CBF. However, the findings of Study 1 also indicate that brands might mitigate this negative effect of venting on CBF by imposing content restrictions on consumers' freedom to express their frustration.

6. STUDY 2

Study 2 builds on Study 1 in several ways. First, instead of using a fictitious brand and scenario-based design, we asked consumers to recall an actual negative experience with a *real* brand. Second, we tested our predictions in another context—Facebook. Third, the real nature of the

customer complaints in Study 2 allows us to analyze the emotional tone portrayed in the Facebook posts. Finally, we introduce the self-brand connection as a potential moderator of the effect.

6.1. Method

In Study 2, after passing attention checks (Oppenheimer et al., 2009), we randomly assigned 442 American MTurk respondents (M_{age} = 41 years; 50.7% male, 49.1% female, 0.2% other) to an experiment with one between-subjects factor (no content moderation vs. content moderation) and one measured factor (prior self-brand connection).

First, we asked participants to think of an incident with a brand that let them down or disappointed them. Moreover, we instructed participants to think of an incident that affected them personally when they purchased or used the brand themselves. We then asked participants to take a moment to think about the incident and write down the name of the brand that wronged them.

Second, to manipulate content moderation, we asked participants to imagine that they visited the brand's Facebook page to reach out and complain about the incident. Afterward, we provided a list of guidelines to follow while drafting the Facebook post. In the content moderation condition, the guidelines stated that participants should "not write any disrespectful or defamatory content" and that they should moderate their language by not using "any foul language." To balance the length of the instructions in the no-content moderation condition, participants were told that "there are no restrictions in terms of content in the posting" and that they could use their "regular writing style." We provide a detailed overview of both manipulations in Appendix B.

Third, we asked the participants to complete a questionnaire. We measured CBF using our newly developed nine-item seven-point Likert scale (α = .91). Furthermore, we measured participants' self-brand connection prior to the incident with an adaptation of Escalas and Bettman's (2003) scale (α = .96). The measurement instructs participants to indicate their prior

self-brand connection on items such as "I felt a personal connection to this brand" on a six-item seven-point Likert scale. Afterward, we collected participants' demographics. In addition, participants answered several manipulation check questions. As in the previous study, we asked participants to indicate their perceived freedom regarding the writing task on a seven-point scale (i.e., "not at all free" vs. "extremely free"). Finally, we measured the severity of the reported cases using a single-item seven-point scale (i.e., "not at all severe" vs. "extremely severe").

6.2. Results

An analysis of the manipulation check revealed that participants perceived higher levels of freedom in the no-content moderation condition (M = 6.38, SD = 1.03) than in the content moderation condition (M = 4.91, SD = 1.65; $F(1, 440) = 125.437$, $p < .01$). Moreover, an ANOVA did not reveal any significant differences between the severity of the incidences reported in either condition (M_{NoMod} = 4.44, SD_{NoMod} = 1.66 vs. M_{Mod} = 4.40, SD_{Mod} = 1.45; $F(1, 440) = 0.062$, p = .80). Thus, participants reported equally severe incidents in both conditions. Similarly, and as intended, participants' self-brand connection prior to the incident was not affected by the manipulations (M_{NoMod} = 4.58, SD_{NoMod} = 1.64 vs. M_{Mod} = 4.48, SD_{Mod} = 1.55; $F(1, 440) = 0.441$, p = .51). In addition to the manipulation checks, we also examined whether consumers assessed the content moderation policies as a restriction of their free speech. Indeed, in comparison to the participants in the no content moderation condition, participants in the content moderation condition agreed more strongly with the notion that the brand's policy restricted their free speech (M_{NoMod} = 2.58, SD_{NoMod} = 1.49 vs. M_{Mod} = 3.72, SD_{Mod} = 1.76; $F(1, 440) = 53.981$, $p < .01$). Similarly, participants in the content moderation condition indicated to feel more negative about the brand due to the content moderation policy than their counterparts in the no content moderation

condition (M_{NoMod} = 2.96, SD_{NoMod} = 1.66 vs. M_{Mod} = 3.29, SD_{Mod} = 1.77; $F(1, 440) = 3.998$, $p < .05$).

The real nature of the incidents that participants reported in this study allowed us to analyze the emotional tone of Facebook posts with an application called Linguistic Inquiry and Word Count (LIWC). We chose LIWC due to its wide use in prior marketing research and its effectiveness in the domain of sentiment analysis—especially in the context of negative emotional events (Berger et al., 2020; Cohn, Mehl, & Pennebaker, 2004; Pennebaker, Boyd, Jordan, & Blackburn, 2015). LIWC uses a dictionary-based approach to count the number of words related to certain linguistic categories (e.g., "happy" relates to positive emotions, while "hurt" relates to negative emotions). The emotional tone variable in LIWC describes both the positivity and the negativity in text to calculate a score that ranges from 0 to 100, with higher scores indicating higher levels of positivity. Cohn et al. (2004) provide a detailed description of how LIWC's emotional tone scores are calculated. We provide sample posts that scored high and low on the emotional tone variable in Appendix C.

As we predict in H2a, an ANOVA with the manipulated factor of no content moderation versus content moderation as the IV and the LIWC emotional tone scores as the DV showed that participants in the content moderation condition (M = 36.06, SD = 29.90) wrote their Facebook post using a more positive emotional tone than participants in the no-content moderation condition (M = 28.33, SD = 27.29; $F(1, 440) = 8.055$, $p < .01$).

We ran another ANOVA to determine whether the content moderation manipulations affected CBF levels. Indeed, the results provide further evidence in support of H2b, as participants showed higher levels of CBF in the content moderation condition than in the no content moderation condition (M_{NoMod} = 2.97, SD_{NoMod} = 1.41 vs. M_{Mod} = 3.40, SD_{Mod} = 1.39; $F(1, 440) = 10.634$, $p <$

.01). In the context of customer complaints in relation to real incidents (i.e., Study 2), we found significant differences between the no-content moderation and content moderation conditions for all three subdimensions of CBF (Affective CBF: M_{NoMod} = 2.77, SD_{NoMod} = 1.49 vs. M_{Mod} = 3.23, SD_{Mod} = 1.48; $F(1, 440) = 10.405$, $p < .01$; Cognitive CBF: M_{NoMod} = 3.10, SD_{NoMod} = 1.47 vs. M_{Mod} = 3.60, SD_{Mod} = 1.46; $F(1, 440) = 12.873$, $p < .01$; Behavioral CBF: M_{NoMod} = 3.03, SD_{NoMod} = 1.59 vs. M_{Mod} = 3.38, SD_{Mod} = 1.64; $F(1, 440) = 5.168$, $p < .05$).

To test the indirect effects of content moderation on CBF via emotional tone, as well as the moderation effect of self-brand connection, we ran a moderated mediation using PROCESS model 8 (Hayes, 2017). In the model, we included no content moderation versus content moderation as the IV (0,1 coded dichotomous variable), self-brand connection as the moderator (MOD), emotional tone as the mediator (MED), and CBF as the DV. The analysis revealed a significant relationship between the interaction of IV × MOD and the MED (a_3 *path* = 3.50, $t(438) = 2.06$, $p < .05$). A floodlight analysis revealed that the IV (content moderation) significantly predicted the MED (emotional tone) if the MOD (prior self-brand connection) was below a value of $J - N$ = 3.90, which was true in 69.23% of cases. Moreover, the MED (emotional tone) was positively associated with the DV (CBF) (*b path* = .01, $t(437) = 2.91$, $p < .01$). The conditional indirect effects of content moderation on CBF showed that emotional tone acts as a mediator for cases in which consumers' selves were moderately or strongly connected to the brand (*indirect effect* $_{\text{MEAN MOD}}$ = .05, CI = .01 to .11; *indirect effect* $_{+1\text{SD MOD}}$ = .09, CI = .02 to .18), while no significant indirect effect could be established for consumers with a low prior self-brand connection (*indirect effect* $_{-1\text{SD MOD}}$ = .02, confidence interval [CI] = -.03 to .07). This provides conditional support for H3 while accounting for the predictions in H4. Moreover, the direct effect for low self-brand connection was significant (*direct effect* $_{-1\text{SD MOD}}$ = .31, CI = –.06 to .68), while the direct effects

for moderate and high self-brand connection were significant (*direct effect* $_{\text{MEAN MOD}}$ = .39, CI = .13 to .65; *direct effect* $_{+1\text{SD MOD}}$ = .47, CI = .10 to .85). Finally, the index of moderated mediation was significant, thus providing support for H4 (*index mod med* = .02, CI = .00 to .05). For an overview of the moderated mediation model, see Figure 3. Additional analyses with the three types of CBF as DV revealed similar results and significant indices moderated mediation (*index mod* $med_{Affective\ CBF}$ = .02, CI = .00 to .05; *index mod* $med_{Cognitive\ CBF}$ = .02, CI = .00 to .06; *index mod* $med_{Behavioral\ CBF}$ = .03, CI = .00 to .06).

Insert Figure 3 Here

6.3. Discussion

In Study 2, we find that restricting consumers' freedom to express their frustration about a brand by moderating the content they post on a brand's Facebook page may decrease the negativity in written customer complaints. Furthermore, we find that the emotional tone contained in the text ultimately influences CBF. Finally, this effect is particularly strong for consumers who felt strongly connected to the brand before the incident, while those who had a low self-brand connection appear to be less influenced by content moderation measures.

7. CONCLUSION

In this paper, we show that the mere act of writing a customer complaint increases consumers' negative emotions, as consumers reaccess and ruminate on the negative experience they had with a brand. Moreover, we demonstrate that brands are capable of mitigating this negative effect by asking consumers to moderate their speech, for example, by means of a content moderation policy. Specifically, we demonstrate that when brands impose restrictions on consumers' freedom to

express their frustrations—in essence, asking them to self-censor—emotions detected in the language of the corresponding complaint are more positive, ultimately generating higher levels of CBF. Finally, we determine that this effect is stronger for consumers with strong prior self-brand connection.

7.1 Theoretical contribution and managerial implications

First, we contribute to the current venting and emotion regulation literature by providing evidence that a consumer's act of venting about negative brand experiences does not resolve their emotional state or result in catharsis (Pennebaker et al., 2001; Rimé, 2009). Prior literature on WOM has identified the need to vent as a driver of negative WOM (Grewal, Stephen & Bart, 2019; Nam et al., 2020; Nyer, 1997; Wetzer et al., 2007; Zeelenberg & Pieters, 2004) and speculates that venting may be beneficial for regulating negative emotions about a brand transgression (Berger, 2014). However, our findings indicate that in a consumption context, particularly those characterized by high levels of consumer empowerment, venting triggers the process of rumination, leading to increased negative emotions and ultimately to lower levels of forgiveness toward the offending brand (McCullough, 2000; McCullough et al., 1998; Pennebaker et al., 2001; Rimé, 2009). Furthermore, consumers allowed to vent freely without any restrictions (e.g., due to the lack of a content moderation policy) wrote the most negative reviews, which, in our specific context, were visible to others. This suggests a double jeopardy for brands; not only was the corresponding complainant less forgiving toward the offending brand, but also this (more) negative review could be seen by others.

Second, we contribute to the existing WOM literature (Berger, 2014; Grewal, Stephen & Bart, 2019; Reimer & Benkenstein, 2016) by examining how a content moderation policy imposed

by brands affects consumers. We built on extant free speech literature (Gibson, 2019; Gillespie, 2018) by showing that content moderation can, indeed, be an effective tool for brands to generate more favorable WOM. Specifically, we show that consumers self-censor their content according to moderation policies, which is reflected in a more positive emotional tone in written complaints and ultimately in higher levels of CBF. Thus, while previous research shows that deleting negative WOM can damage brand credibility (e.g., Shaw & Coker, 2012), we demonstrate that content moderation policies may help brands mitigate negative comments in their digital marketing channels while acting as short-term mood repair tool, ultimately resulting in higher levels of forgiveness.

Moreover, we add to the current literature on content moderation policies and WOM from a signaling perspective (De Angelis et al., 2012; Gibson, 2019; Grewal, Stephen & Coleman, 2019; Palmeira et al., 2020) by demonstrating that consumers with a stronger self-brand connection are particularly prone to adapt their emotional tone and forgiveness levels according to content moderation policies. In particular, consumers who are more strongly connected to a brand may be more willing to obey their brand's content moderation policy, be less prone to disclose their true emotions in public and are therefore more likely to self-censor their free speech.

Third, we contribute to the emerging stream of literature on consumer forgiveness by developing and validating a new scale to measure CBF. This represents the first systematic effort to conceptualize and operationalize the construct as multidimensional, in line with the interpersonal psychology literature. The resultant scale involves three dimensions and nine items that capture cognitive, affective, and behavioral aspects of CBF. By applying the scale, we also contribute to the literature on factors that enable or hinder forgiveness by examining CBF in a context in which it has not previously been examined, namely, consumers' online complaints.

Specifically, we demonstrate how the management of negative consumer emotions might begin even earlier than the submission of a complaint on social media, thus triggering the CBF process.

In addition to the theoretical contributions, we provide actionable recommendations for businesses. While many companies have designed and adopted sophisticated complaint-handling procedures in social media channels, these are commonly activated *after* a consumer complaint has been submitted. We advise brands to adopt a proactive approach to complaint management, one that starts before a complaint is submitted and involves a content moderation policy in all digital communication channels—especially those visible to others—in the form of self-censorship. Prior research (Shaw & Coker, 2012) has discouraged brands from engaging in forms of content moderation that are more radical and controversial, such as removing or censoring consumer comments. However, our findings demonstrate that merely instructing consumers to apply self-censorship when writing their complaints can be an effective tool in complaint management. The benefits for businesses of content moderation policies, specifically self-censorship, are twofold. On the one hand, consumers will write their customer complaints in a more positive emotional tone, which ultimately benefits a brand's image. On the other hand, consumers may be more willing to forgive the brand for its wrongdoing, which might result in greater relationship satisfaction (Muhammad & Gul-E-Rana, 2020), trust (Schoorman et al., 2007; Xie & Peng, 2009), and repatronage behavior (e.g., Harrison-Walker, 2019). Finally, we find evidence to suggest that such content moderation policies are particularly effective in managing negative WOM of customers with pre-existing and moderate to strong brand relationships. Thus, managers of brands with a loyal following may be more likely to find that content moderation impacts positively on their consumers' WOM and forgiveness following a negative incident.

7.2 Limitations and further research

There are some limitations of this research that need to be addressed and warrant further attention. First, our study focused on negative experiences that consumers personally experienced with their purchased products or services. According to Dutta and Pullig (2011), brand crises can be categorized into two main types: performance-related or values-related crises. Thus, while our study focused on examining the effects of content moderation on CBF in the context of performance-related issues, it would be interesting to test whether the assumptions hold for values-related brand crises. From a theoretical point of view, one may assume that values-based issues, which do not directly involve the product but rather social or ethical issues embodied or reflected in the brand, might activate a more emotional rather than rational evaluation of a brand's transgression. Thus, it would be interesting to examine whether content moderation has a larger or smaller impact on the emotional states of consumers' written complaints in the context of values-related issues.

Second, and also related to values, brands may impose content moderation policies that are based on their own communicated values. While our paper focuses on content moderation in relation to language, brands may also restrict consumers' freedom of speech by engaging in content moderation due to ideological concerns. In a polarized society, brands have become much more aware of their power to promote certain (often controversial) social and political issues (Mukherjee & Althuizen, 2020). Future studies could examine the effects of different motivations behind content moderation and how consumers with similar or opposing value systems receive this.

Third, while we find some evidence that language-focused content moderation policies are perceived to infringe on consumers' free speech, we also observe that most consumers obey those policies in a form of self-censorship. Future research may lay more emphasis on examining the

negative outcomes of content moderation. For example, strong beliefs related to free speech may cause content moderation to negatively affect CBF. Moreover, content moderation may take different forms, ranging from weaker interventions, such as calls for self-censorship (the focus of the present research) to stronger forms of censorship, such as the proactive deletion of comments. We used only one form of content moderation, whereby participants were instructed to self-censor their communication with the brand that failed them, thus representing a weaker intervention. Scholars may examine whether more invasive and restrictive types of content moderation (e.g., deletion of negative product reviews) may cause consumer backlash and ultimately harm the brand. Moreover, further research could investigate how consumers adapt their complaint-writing behavior and emotional tone if the deletion of comments is salient. While prior research has found that content deletion decreases the credibility of a brand (Shaw & Coker, 2012), it remains to be investigated how more extreme forms of free speech suppression affect consumers' emotional states and, ultimately, CBF.

Finally, our research builds on venting theory which suggests that rumination is the underlying mechanism for the increase in anger levels due to venting (Caprara, 1986; Nils & Rimé, 2012; Rimé, 2009). However, our current study did not aim to test this theory empirically due to the difficulty in measuring state rumination in reaction to an event based on self-reporting in a reliable manner. Instead, our paper measured the anger levels contained in the customer complaints with the help of sentiment analysis (i.e. LIWC). Future work could use advanced neurological measurements (e.g., Kelley, Hortensius, & Harmon-Jones, 2013) to examine whether brain regions related to rumination, such as the prefrontal cortex, may indeed be more active if consumers vent about negative brand experiences. Alternatively, one may measure the physiological reactions in

relation to rumination—for example, in the form of blood pressure recovery (e.g., Gerin, Davidson, Christenfeld, Goyal, & Schwartz, 2006).

References

Aaker, J. L., Fournier, S., & Brasel, S. A. (2004). When good brands do bad. *Journal of Consumer Research*, *31*(1), 1–16.

Ahluwalia, R., Burnkrant, R. E., & Unnava, H. R. (2000). Consumer response to negative publicity: The moderating role of commitment. *Journal of Marketing Research*, 37(2), 203-214.

Antonetti, P., & Maklan, S. (2018). Identity bias in negative word of mouth following irresponsible corporate behavior: A research model and moderating effects. *Journal of Business Ethics*, *149*(4), 1005–1023.

Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, *53*(3), 297–318.

Bagozzi, R. P., Batra, R., & Ahuvia, A. (2017). Brand love: Development and validation of a practical scale. *Marketing Letters*, *28*(1), 1–14.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of The Academy of Marketing Science*, 16(1), 74-94.

Barber, L., Maltby, J., & Macaskill, A. (2005). Angry memories and thoughts of revenge: The relationship between forgiveness and anger rumination. *Personality and Individual Differences*, *39*(2), 253–262.

Barcaccia, B., Salvati, M., Pallini, S., Baiocco, R., Curcio, G., Mancini, F., & Vecchio, G. M. (2020). Interpersonal forgiveness and adolescent depression. The mediational role of self-reassurance and self-criticism. *Journal of Child and Family Studies*, *29*(2), 462–470.

Batra, R., Ahuvia, A., & Bagozzi, R. P. (2012). Brand love. Journal of Marketing, 76(2), 1-16.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Belk, R. W. (1989). Extended self and extending paradigmatic perspective. *Journal of Consumer Research*, 16(1), 129-132.

Belk, R. W. (2013). Extended self in a digital world. *Journal of Consumer Research*, 40(3), 477-500.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, *24*(4), 586–607.

Berger, J., & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, *34*(2), 121–134.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25.

Bono, G., & McCullough, M. E. (2006). Positive responses to benefit and harm: Bringing forgiveness and gratitude into cognitive psychotherapy. *Journal of Cognitive Psychotherapy*, *20*(2), 147–158.

Burnette, J. L., McCullough, M. E., Van Tongeren, D. R., & Davis, D. E. (2012). Forgiveness results from integrating information about relationship value and exploitation risk. *Personality and Social Psychology Bulletin*, *38*(3), 345–356.

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, *28*(6), 724–731.

Canary, D. J., Spitzberg, B. H., & Semic, B. A. (1998). The experience and expression of anger in interpersonal settings. In *Handbook of communication and emotion* (pp. 189–213). New York: Academic Press.

Caprara, G. V. (1986). Indicators of aggression: The dissipation-rumination scale. *Personality and Individual Differences*, *7*(6), 763–769.

Carroll, B. A., & Ahuvia, A. C. (2006). Some antecedents and outcomes of brand love. *Marketing Letters*, 17(2), 79-89.

Casidy, R., & Shin, H. (2015). The effects of harm directions and service recovery strategies on customer forgiveness and negative word-of-mouth intentions. *Journal of Retailing and Consumer Services*, *27*, 103–112.

Chen, Y., Harris, S. K., Worthington Jr., E. L., & VanderWeele, T. J. (2019). Religiously or spiritually-motivated forgiveness and subsequent health and well-being among young adults: An outcome-wide analysis. *Journal of Positive Psychology*, *14*(5), 649–658.

Cheng, S. Y., White, T. B., & Chaplin, L. N. (2012). The effects of self-brand connections on responses to brand failure: A new look at the consumer–brand relationship. *Journal of Consumer Psychology*, 22(2), 280-288.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Chung, E., & Beverland, M. (2006). An exploration of consumer forgiveness following marketer transgressions. In C. Pechmann and L. Price (Eds.), *NA – Advances in consumer research* (Vol. 33, pp. 98–99). Duluth, MN: Association for Consumer Research.

Clarke, S. L., & Rhodes, E. S. (2020). Entrepreneurial apologies: The mediating role of forgiveness on future cooperation. *Journal of Business Venturing Insights*, *13*, e00147.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693.

Currier, J. M., Drescher, K. D., Holland, J. M., Lisman, R., & Foy, D. W. (2016). Spirituality, forgiveness, and quality of life: Testing a mediational model with military veterans with PTSD. *The International Journal for the Psychology of Religion*, *26*(2), 167–179.

Dalebroux, A., Goldstein, T. R., & Winner, E. (2008). Short-term mood repair through art-making: Positive emotion is more effective than venting. *Motivation and Emotion*, *32*(4), 288–295.

De Angelis, M., Bonezzi, A., Peluso, A. M., Rucker, D. D., & Costabile, M. (2012). On braggarts and gossips: A self-enhancement account of word-of-mouth generation and transmission. *Journal of Marketing Research*, 49(4), 551-563.

Dekay, S. H. (2012). How large companies react to negative Facebook comments. *Corporate Communications: An International Journal*, *17*(3), 289–299.

Dutta, S., & Pullig, C. (2011). Effectiveness of corporate responses to brand crises: The role of crisis type and response strategies. *Journal of Business Research*, *64*, 1281–1287.

Escalas, J. E., & Bettman, J. R. (2003). You are what they eat: The influence of reference groups on consumers' connections to brands. *Journal of Consumer Psychology*, 13(3), 339-348.

Fahimdanesh, F., Noferesti, A., & Tavakol, K. (2020). Self-compassion and forgiveness: Major predictors of marital satisfaction in young couples. *American Journal of Family Therapy*, *48*(3), 221–234.Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin*, *136*(5), 894–914.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics.

Freedman, S., Enright, R. D., & Knutson, J. (2005). A progress report on the process model of forgiveness. In E. Worthington Jr. (Ed.), *Handbook of forgiveness* (pp. 393–406). New York: Routledge.

Freedman, S., & Zarifkar, T. (2016). The psychology of interpersonal forgiveness and guidelines for forgiveness therapy: What therapists need to know to help their clients forgive. *Spirituality in Clinical Practice*, *3*(1), 45–58.

Gençoğlu, C., Şahin, E., & Topkaya, N. (2018). General self-efficacy and forgiveness of self, others, and situations as predictors of depression, anxiety, and stress in university students. *Educational Sciences: Theory & Practice*, *18*(3), 605–626.

Gerin, W., Davidson, K. W., Christenfeld, N. J., Goyal, T., & Schwartz, J. E. (2006). The role of angry rumination and distraction in blood pressure recovery from emotional arousal. *Psychosomatic Medicine*, 68(1), 64-72.

Gibson, A. (2019). Free speech and safe spaces: How moderation policies shape online discussion spaces. *Social Media + Society*, *5*(1), https://doi.org/10.1177/2056305119832588.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Goulding, C. (2005). Grounded theory, ethnography and phenomenology: A comparative analysis of three qualitative strategies for marketing research. *European Journal of Marketing*, *39*(3/4), 294–308

Grégoire, Y., Salle, A., & Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons*, *58*(2), 173–182.

Grewal, L., Stephen, A. T., & Bart, Y. (2019). Managing Online Venting: the Impact of Temporal Distance on the Perception of Negative Online Reviews. *ACR North American Advances*, 47,

19-23.

Grewal, L., Stephen, A. T., & Coleman, N. V. (2019). When posting about products on social media backfires: The negative effects of consumer identity signaling on product interest. *Journal of Marketing Research*, *56*(2), 197–210.

Harrison-Walker, L. J. (2019). The critical role of customer forgiveness in successful service recovery. *Journal of Business Research*, *95*, 376–391.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*: New York: Guilford Press.

Hogreve, J., Bilstein, N., & Hoerner, K. (2019). Service recovery on stage: Effects of social media recovery on virtually present others. *Journal of Service Research*, *22*(4), 421–439.

Hornik, J., Satchi, R. S., Cesareo, L., & Pastore, A. (2015). Information dissemination via electronic word-of-mouth: Good news travels fast, bad news travels faster!. *Computers in Human Behavior*, *45*, 273–280.

Johnen, M., & Schnittka, O. (2019). When pushing back is good: The effectiveness of brand responses to social media complaints. *Journal of the Academy of Marketing Science*, *47*(5), 858–878.

Joireman, J., Grégoire, Y., & Tripp, T. M. (2016). Customer forgiveness following service failures. *Current Opinion in Psychology*, *10*, 76–82.

Karremans, J. C., Van Lange, P. A., Ouwerkerk, J. W., & Kluwer, E. S. (2003). When forgiving enhances psychological well-being: The role of interpersonal commitment. *Journal of Personality and Social Psychology*, *84*(5), 1011–10216.

Kelley, N. J., Hortensius, R., & Harmon-Jones, E. (2013). When anger leads to rumination: Induction of relative right frontal cortical activity with transcranial direct current stimulation increases anger-related rumination. *Psychological Science*, 24(4), 475-481.

Khamitov, M., Grégoire, Y., & Suri, A. (2020). A systematic review of brand transgression, service failure recovery and product-harm crisis: integration and guiding insights. *Journal of the Academy of Marketing Science*, *48*, 519–542.

Komorita, S. S., Hilty, J. A., & Parks, C. D. (1991). Reciprocity and cooperation in social dilemmas. *Journal of Conflict Resolution*, *35*(3), 494–518.

Laczniak, R. N., DeCarlo, T. E., & Ramaswami, S. N. (2001). Consumers' responses to negative word-of-mouth communication: An attribution theory perspective. *Journal of Consumer Psychology*, *11*(1), 57–73.

Leonhardt, B. L., Lysaker, P. H., Vohs, J. L., James, A. V., & Davis, L. W. (2018). The experience and expression of anger in posttraumatic stress disorder: The relationship with metacognition. *Journal of Mental Health*, *27*(5), 432–437.

Lepore, S. J., Fernandez-Berrocal, P., Ragan, J., & Ramos, N. (2004). It's not that bad: Social challenges to emotional disclosure enhance adjustment to stress. *Anxiety, Stress & Coping*, *17*(4), 341–361.

Lisjak, M., Lee, A. Y., & Gardner, W. L. (2012). When a threat to the brand is a threat to the self: The importance of brand identification and implicit self-esteem in predicting defensiveness. *Personality and Social Psychology Bulletin*, 38(9), 1120-1132.

Little, L. M., Simmons, B. L., & Nelson, D. L. (2007). Health among leaders: Positive and negative affect, engagement and burnout, forgiveness and revenge. *Journal of Management Studies*, *44*(2), 243–260.

Madio, L., & Quinn, M. (2020). *User-generated content, strategic moderation, and advertising*. Available at SSRN 3551103.

McCullough, M. E. (2000). Forgiveness as human strength: Theory, measurement, and links to well-being. *Journal of Social and Clinical Psychology*, *19*(1), 43–55.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington Jr., E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, *75*(6), 1586–1603.

McCullough, M. E., Worthington Jr., E. L., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, 73(2), 321–336.

Morgan, R. M., & Hunt, S. D. (1994). The commitment–trust theory of relationship marketing. *Journal of Marketing*, *58*(3), 20–38.

Muhammad, L., & Gul-E-Rana (2020). Mediating role of customer forgiveness between perceived justice and satisfaction. *Journal of Retailing and Consumer Services*, *52*, DOI: 10.1016/j.jretconser.2019.101886.

Mukherjee, S., & Althuizen, N. (2020). Brand activism: Does courting controversy help or hurt a brand? *International Journal of Research in Marketing*, https://doi.org/10.1016/j.i Wei, H., & Ran, Y. (2019). Male versus female: How the gender of apologizers influences consumer forgiveness. *Journal of Business Ethics*, 154(2), 371-387.

Nam, K., Baker, J., Ahmad, N., & Goo, J. (2020). Determinants of writing positive and negative electronic word-of-mouth: Empirical evidence for two types of expectation confirmation. *Decision Support Systems*, *129*, https://doi.org/10.1016/j.dss.2019.113168.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications.

Nils, F., & Rimé, B. (2012). Beyond the myth of venting: Social sharing modes determine the benefits of emotional disclosure. *European Journal of Social Psychology*, *42*(6), 672–681.

Nyer, P. U. (1997). A study of the relationships between cognitive appraisals and consumption emotions. *Journal of the Academy of Marketing Science*, *25*(4), 296–304.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872.

Palmeira, M., Spassova, G., & Quoidbach, J. (2020). You're not yelping your case: The unexpected social consequences of word of mouth. *European Journal of Marketing*, *54*(2), 419–447.

Pantano, E., & Corvello, V. (2013). The impact of experience on companies' reactions to negative comments on social networks. *Journal of Direct, Data and Digital Marketing Practice*, *14*(3), 214–223.

Park, C. W., MacInnis, D. J., Priester, J., Eisingerich, A. B., & Iacobucci, D. (2010). Brand attachment and brand attitude strength: Conceptual and empirical differentiation of two critical brand equity drivers. *Journal of Marketing*, 74(6), 1-17.

Parlamis, J. D. (2012). Venting as emotion regulation: The influence of venting responses and respondent identity on anger and emotional tone. *International Journal of Conflict Management*, *23*(1), 77–96.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin

Pennebaker, J. W., Kiecolt-Glaser, J. K., & Glaser, R. (1988). Disclosure of traumas and immune function: Health implications for psychotherapy. *Journal of Consulting and Clinical Psychology*, *56*, 239–245.

Pennebaker, J. W., Zech, E., & Rimé, B. (2001). Disclosing and sharing emotion: Psychological, social, and health consequences. In *Handbook of bereavement research: Consequences, coping, and care*. (pp. 517–543). Washington, DC: American Psychological Association.

Reimer, T., & Benkenstein, M. (2016). When good WOM hurts and bad WOM gains: The effect of untrustworthy online reviews. *Journal of Business Research*, *69*(12), 5993–6001.

Riek, B. M., & Mania, E. W. (2012). The antecedents and consequences of interpersonal forgiveness: A meta-analytic review. *Personal Relationships*, *19*, 304–325.

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review*, *1*(1), 60–85.

Rimé, B., Mesquita, B., Boca, S., & Philippot, P. (1991). Beyond the emotional event: Six studies on the social sharing of emotion. *Cognition & Emotion*, *5*(5-6), 435–465.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320.

Schaefers, T., & Schamari, J. (2016). Service recovery via social media: The social influence effects of virtual presence. *Journal of Service Research*, *19*(2), 192–208.

Schnebelen, S., & Bruhn, M. (2018). An appraisal framework of the determinants and consequences of brand happiness. *Psychology & Marketing*, *35*(2), 101–119.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344–354.

Shaw, V., & Coker, B. (2012). Keeping negative Facebook comments leads to more trust in your brand. In *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Sinha, J., & Lu, F. C. (2016). "I" value justice, but "we" value relationships: Self-construal effects on post-transgression consumer forgiveness. *Journal of Consumer Psychology*, *26*(2), 265–274.

Sprout Social (2017). The Sprout Social Index, Edition XII: Call-out culture. Retrieved from https://sproutsocial.com/insights/data/q3-2017/.

Sprout Social (2020). Call out culture. Retrieved from https://sproutsocial.com/glossary/call-out-culture/.

Stokburger-Sauer, N., Ratneshwar, S., & Sen, S. (2012). Drivers of consumer–brand identification. *International Journal of Research in Marketing*, 29(4), 406-418.

Subkoviak, M. J., Enright, R. D., Wu, C. R., Gassin, E. A., Freedman, S., Olson, L. M., & Sarinopoulos, I. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence*, *18*(6), 641–655.

Suciu, P. (2020). Social media's "callout culture" continues to improve customer service. Retrieved from https://www.forbes.com/sites/petersuciu/2020/01/08/social-medias-callout-culture-continues-to-improve-customer-service/#287670dc6d99.

Takaku, S. (2001). The effects of apology and perspective taking on interpersonal forgiveness: A dissonance-attribution model of interpersonal forgiveness. *Journal of Social Psychology*, *141*(4), 494–508.

Trustpilot (2020). The critical role of reviews in Internet trust. Available at https://business.trustpilot.com/guides-reports/build-trusted-brand/the-critical-role-of-reviews-in-internet-trust (accessed 22 June 22, 2020).

Tsarenko, Y., & Tojib, D. (2012). The role of personality characteristics and service failure severity in consumer forgiveness and service outcomes. *Journal of Marketing Management*, *28*(9-10), 1217–1239.

Wei, H., & Ran, Y. (2019). Male versus female: How the gender of apologizers influences consumer forgiveness. *Journal of Business Ethics*, *154*(2), 371–387.

Wetzer, I. M., Zeelenberg, M., & Pieters, R. (2007). "Never eat in that restaurant, I did!": Exploring why people engage in negative word-of-mouth communication. *Psychology & Marketing*, *24*(8), 661–680.

WhoIsHostingThis.com. (2019). Complaining customers & calling out businesses online (survey results). Retrieved from https://www.whoishostingthis.com/blog/2019/12/19/business-callout/.

Xie, Y., & Peng, S. (2009). How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness. *Psychology & Marketing*, *26*(7), 572–589.

Yao, S., Chen, J., Yu, X., & Sang, J. (2017). Mediator roles of interpersonal forgiveness and self-forgiveness between self-esteem and subjective well-being. *Current Psychology*, *36*(3), 585–592.

Zech, E. (1999). Is it really helpful to verbalise ones emotions? *Gedrag & Gezondheid: Tijdschrift voor Psychologie en Gezondheid*, *27*(1-2), 42–47.

Zeelenberg, M., & Pieters, R. (2004). Beyond valence in customer dissatisfaction: A review and new findings on behavioral responses to regret and disappointment in failed services. *Journal of Business Research*, *57*(4), 445–455.

Table

TABLE 1 CFA results for calibration (C) and validation (V) samples

	Factor loadings		*t*-Values	
	Calibration	**Validation**	**Calibration**	**Validation**
Cognitive α = .92 (C), .91 (V)				
I think the brand should get what it deserves.	.84	.79	20.115	18.884
I wish that others could see that this brand is unworthy.	.93	.94	*	*
I disapprove of this brand.	.91	.92	24.413	26.812
Affective α = .85 (C), .82 (V)				
I feel sympathetic towards this brand.	.82	.83	14.021	12.971
I have compassion for the brand, which wronged me.	.84	.79	*	*
I feel as if I have restored my faith in this brand.	.77	.74	13.209	12.138
Behavioral α = .93 (C), .91 (V)				
I avoid using this brand.	.97	.91	*	*
I do not consider this brand anymore when evaluating alternatives.	.94	.90	32.717	23.197
I am less likely to try this brand again.	.82	.83	20.710	19.678

*Parameters fixed to 1.

Figures

FIGURE 1 Conceptual model

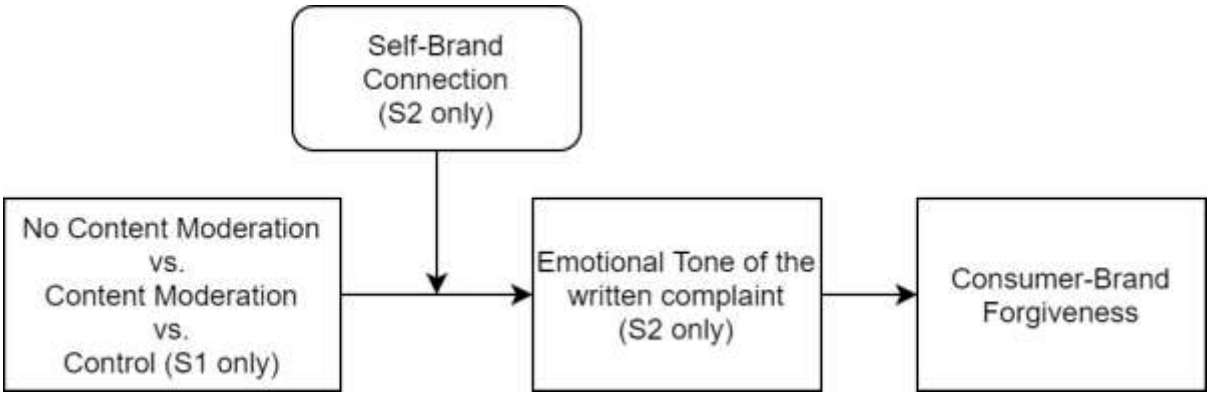


FIGURE 2 CBF levels of Study 1

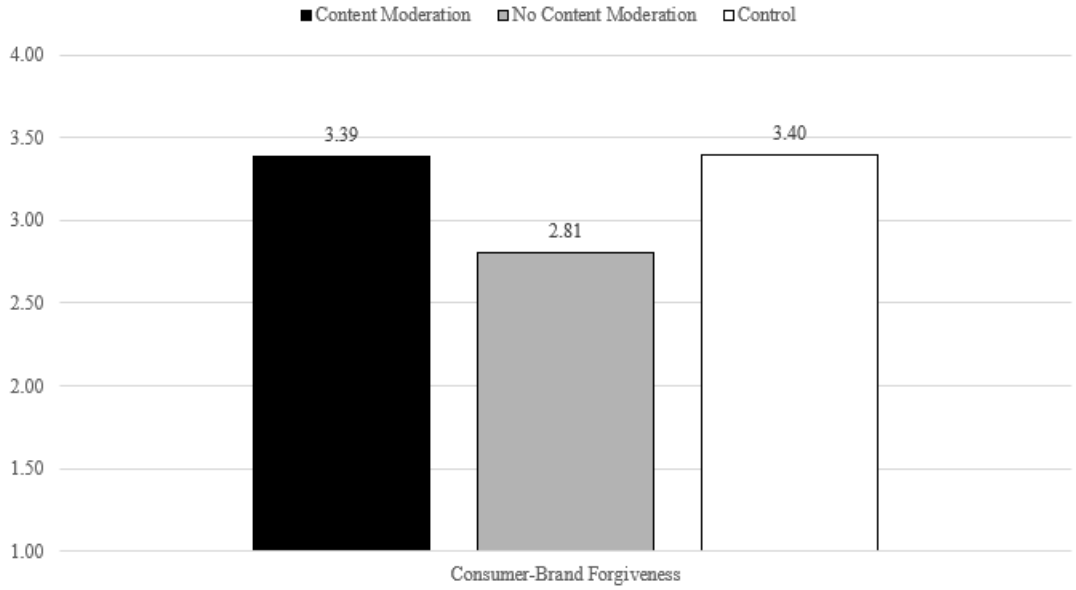
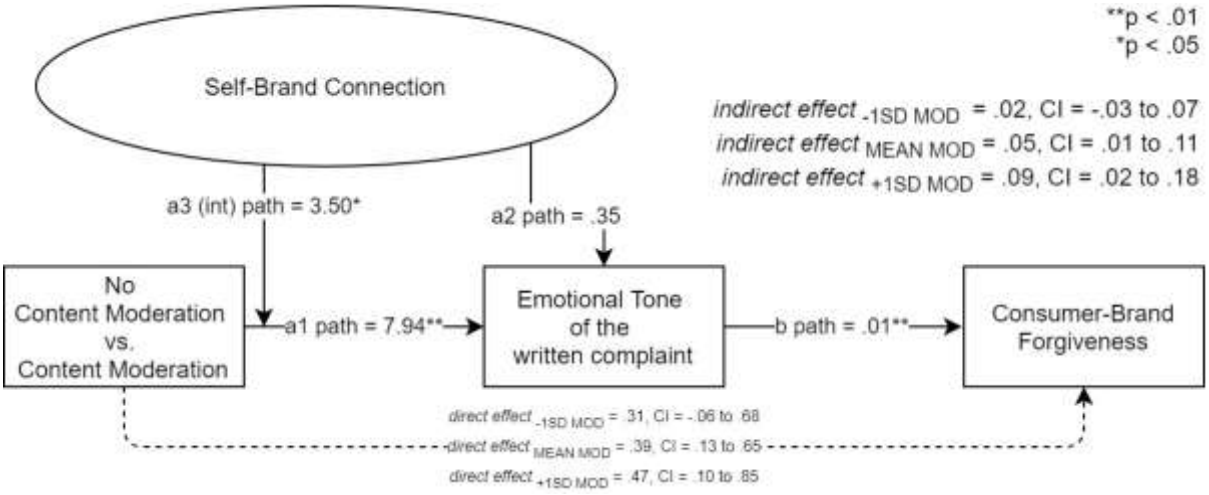


FIGURE 3 Moderated mediation model of Study 2



Appendix

Appendix A: Study 1 Manipulations

Condition: Content Moderation

Now, please imagine that you decide to inform the brand (Compobooks) about your negative customer experience. For this purpose, you are signing up on Compobooks' customer service message board. Among other things, the sign-up page of the message board states:

Sign-up agreement:

- Aggressive language and threats will not be tolerated.
- Be respectful of the customer service team and others.
- Do not make personal attacks on customer service representatives or other community members, including arguments, bullying, prejudiced remarks and defamatory or false comments.
- Posts that violate this agreement will be removed.

Condition: No Content Moderation

Now, please imagine that you decide to inform the brand (Compobooks) about your negative customer experience. For this purpose, you are signing up on Compobooks' customer service message board.

Condition: Control Group

In the following, you will be shown a picture.

Please look at the picture thoroughly and carefully.

Afterwards, you will be asked to describe the picture in text.

Appendix B: Study 2 Manipulations

Condition: Content Moderation

- Please do not confront the brand by using any foul language or with a bad temper.

  Please use moderate language for this task.
- Please do not write any disrespectful or defamatory content.

Condition: No Content Moderation

- You are completely free in the way you would like to confront the brand about their wrongdoings.
- Please use your regular writing style.
- There are no restrictions in terms of content in the posting.

Appendix C: Sample posts and corresponding LIWC scores

Positive post example / LIWC emotional tone score: 97.07

"I recently placed an order for my vitamins which I rely on for energy, focus, and sleep. The expected delivery date was for yesterday, and I am still waiting. Whatever happened to two-day delivery? I get that with Covid, people are ordering more, but your company is the richest in the world. Surely you can fix your issues."

Negative post example / LIWC emotional tone score: 4.02

"I got an order from X last week. I ordered a new charger for my iPhone but they sent me a totally different one meant for Android that I can't even use! I contacted customer service and they were surprisingly rude to me and would not even allow me to return this item for a replacement! I was very upset and angry. I basically wasted $30 on something that I cannot even use. This will definitely make me reconsider buying from X again in the future for products like this! :("

Note: The name of the company was anonymized and replaced with X