Picketing the Virtual Storefront: Content Moderation
and Political Criticism of Businesses on Yelp

BEN MEDEIROS[1]
State University of New York at Plattsburgh, USA

This article examines incidents in which business owners incur criticism on the consumer review platform Yelp based on political ideology. I analyze two case studies from the summer of 2018 by considering the sentiments expressed in the review texts, the application of Yelp's relevant policies, and the tactical adaptations of reviewers. The case studies evince a normative conflict over how the platform should treat viral criticism of this sort. While Yelp clearly cannot truly function as a laissez-faire public forum, its moderation criteria can be gamed, and its efforts evidently exclude a range of sentiments that some users find meaningful. The article provides an in-depth exploration of a platform that has received somewhat less attention in the growing literature on the role of private intermediaries in shaping what kinds of speech attain visibility in the digital public sphere.

Keywords: platforms, content moderation, free speech, Yelp, commodity activism

This article examines the phenomenon that *Wired* magazine recently called "the weaponization of Yelp" (Matsakis, 2018)—or incidents in which the online consumer review platform Yelp becomes a site for viral condemnation of individuals (in their capacities as business proprietors) in response to some public political statement or perceived social transgression. Such incidents have received news coverage since at least 2011, when a bridal shop owner refused to serve a same-sex couple (Kuehn, 2017, p. 205). Readers might also recall news reporting on a 2012 incident in which a pizza restaurant owner was flooded with negative criticism in response to a picture of him hugging then president Barack Obama or the infamous Minnesota dentist whose practice temporarily closed following outcry over his 2015 killing of "Cecil the lion" on a hunting trip in Zimbabwe. Such incidents prompt questions about the goals of commenters and their choice of the Yelp platform specifically as well as questions about Yelp's response. What does the phenomenon ultimately tell us about the ways that private digital intermediaries govern public expression?

I explore these questions through two more recent case studies. In June 2018, a Virginia restaurant came under fire from Trump supporters because it refused service to Trump press secretary Sarah Huckabee Sanders. Only days later, the proprietor of a construction company in Woodland Hills, California, received criticism on Yelp after a photograph circulated on social media depicting her wearing the iconic "Make America Great Again" hat in an exchange with a teenage boy at a city council hearing on immigration issues. These two case studies were chosen because they represent paradigmatic instances of business criticism based on political ideology and because the media coverage and activity of reviewers could be followed as they unfolded during the summer of 2018. The article considers both the tactical behavior of the different entities involved (business, reviewers, and platform) as well as the content of the review texts and Yelp's relevant policies.

While previous scholarship on Yelp has assessed its ideological significance and affordances for "commodity activism," these case studies suggest that Yelp also provides a window into the ways that the digital age has compounded perennial concerns in legal scholarship about the privatization of public venues for expression. I argue that the combination of review content and tactical behavior on display suggests a normative conflict over the platform's role in the contemporary digital ecosystem of information. Yelp weaponization embodies an adaptation to the increasing prominence of privately owned venues for expression. The cases ultimately illustrate that while Yelp clearly cannot (and need not) function as a content-neutral, laissez-faire public forum, its efforts at moderation evidently exclude a range of sentiments that some users find meaningful to express on the Yelp platform, and its moderation criteria can be gamed. The article concludes by suggesting that if the platform can adjust its moderation to better represent the range of opinions being expressed about a business (including those about the ideology of the proprietors), it might help discourage the kind of rule-gaming and belligerence that reviewers displayed in the case studies.

The article provides an in-depth exploration of a popular platform that has nonetheless received somewhat less attention in the growing literature on the role of private intermediaries in shaping what kinds of speech attain visibility in the digital public sphere. It adds to the research on content moderation exemplified by Sarah Myers West (2018) that focuses on "understanding content moderation as a relational process that incorporates both company systems and resources and user labor" (p. 4368). In doing so, it responds to Tarleton Gillespie's (2018) general call to "move th[e] fundamental reconsideration [of social media platforms] forward" by "examin[ing] the moderation apparatus that has been built over the past decade: the policies of content moderation, the sociotechnical mechanisms for its enforcement, the business expectations it must serve, the justifications articulated to support it" (pp. 12–13).

Literature Review

The use of Yelp to express politically resonant commentary about businesses has been theorized by Kathleen Kuehn (2017) as a modality of what Sarah Banet-Weiser and Roopali Mukherjee (2012) labeled "commodity activism"—a term that encompasses "the vexed and contradictory means by which individuals and communities have marshaled the ideological and cultural frameworks of consumption to challenge, support, and reimagine the political and social dynamics of power" (p. 3). Kuehn's (2017)

analysis of the commodity activism that is enacted through Yelp identifies two frames that recur in Yelp reviews: socially responsible shopping—such as "addressing issues around production, ecology, labor practices" (p. 214)—and "the celebration of localism as a means of re-directing economic flows" (p. 215).

Banet-Weiser and Mukherjee (2012) argue that commodity activism is grounded in an ethos of neoliberal individualism, and Kuehn (2017) likewise describes how the dominant invocations of commodity activism on Yelp "focu[s] on individual as opposed to collective interests, expressed in everyday decisions and practices such as what, where and how one consumes" (p. 210). In presenting Yelp as a "review democracy" in its promotional media, moreover, the platform "effectively reconfigures democratic participation within a market based, economized notion of citizenship" (Kuehn, 2013, p. 607). In the final analysis, Kuehn (2017) concludes that "reviews are not so much written to collectively mobilize other locals around pressing social issues, democratic rights and responsibilities as spaces for expressing lifestyle politics" (p. 220).

Especially important for studying the weaponization of Yelp phenomenon is Kuehn's (2017) finding that some reviewers reinforce the discourse of neoliberal individualism through their normative rejection of politics as relevant to the consumer experience. These users position political commentary as antithetical to the "generally shared . . . goal of warning, deterring and/or encouraging consumers to spend their money at particular places" (p. 212), and on these grounds, they sometimes flag reviews based on perceived violations of Yelp's terms of service (p. 214). This article does not challenge Kuehn's characterization of commodity activism on Yelp per se. Instead, it builds on her observations to situate Yelp's policing of political speech within the ongoing conversation in freedom of expression scholarship about the governance of privately owned spaces for expression (both digital and physical).

Much freedom of expression scholarship and jurisprudence considers the proper parameters for regulating (political) expression in different venues. First Amendment jurisprudence generally prohibits viewpoint-based regulation of speech in "traditional public forums," or a forum (such as a street or park) "that has 'by long tradition or by government fiat' been 'devoted to assembly and debate,'" as well as in "designated" public forums, or "public property which the state has opened for use by the public as a place for expressive activity" (Lidsky, 2011, pp. 1981–1984; for the criteria used to distinguish different forums, see *Perry Education Ass'n v. Perry Local Educators' Ass'n*, 1983).

Some scholars have explored how the architecture of the forums where people tend to congregate influences the kinds of ideas to which they are exposed. In an article examining state legal protections for speech in shopping malls, for instance, legal scholar Jennifer Coffin (2000) called malls "the modern replacement for the traditional town square" because they afford public congregation in addition to shopping (p. 615). Yet they are technically privately owned spaces that are free to restrict expression as they see fit (with some state law exceptions), and thus Coffin laments "the cultural evolution in which our public spaces have become private fortresses, 'protected' from political speech in the interests of providing 'safe' and unmolested shopping experiences for consumers" (p. 615). The U.S. Supreme Court has recognized that boycotting businesses to express disagreement with the views or practices of their proprietors represents "essential political speech lying at the core of the First Amendment" (*NAACP v. Claiborne Hardware*, 1982, p. 915). An increase in consumer-oriented space

sanitized of political messages would thus represent the diminished circulation of this kind of "essential political speech."

Relatedly, these "private fortresses" of consumption threaten the likelihood of exposure to what Cass Sunstein (2007) calls "unplanned encounters." Such encounters with unanticipated or even annoying ideas "are central to democracy" because "in a democracy deserving the name, people often come across views and topics that they have not specifically selected" (p. 5). Concern about the decline of viewpoint-neutral venues for expression and the decline of unplanned encounters is particularly relevant to the Internet. Indeed, Sunstein's book from which the concept is taken was directly critical of the increasing cloistering of online discourse into "information silos." Media scholar Molly Sauter (2014) frames the problem in terms of governance, asserting that "the overwhelmingly privatized nature of the internet is a challenge to the practice of activism online" (p. 3). Specifically, "there is no 'street' on the internet" where critiques can be voiced in a more traditionally accepted manner (p. 4). Such a characterization is echoed in other general formulations of the dominance of platforms in the digital public sphere. According to Tarleton Gillespie (2018), we have effectively "handed over to private companies the power to set and enforce the boundaries of appropriate public speech for us" (p. 197). Kate Klonick (2018) has dubbed platforms "the new governors" given the "power and scope these private platforms wield through their moderation systems and . . . their role in democratic culture" (p. 1663).

Like in malls, U.S. law generally protects platforms' discretion to determine what does and does not belong on their sites without risking liability for what is posted (with exceptions for content such as copyrighted material and child pornography). For instance, pursuant to Section 230 of the Communications Decency Act, a platform like Yelp cannot be compelled to remove reviews that it does not believe violate its terms of service even when a court has found them defamatory. The Supreme Court of California recently affirmed that this is true of Yelp in *Hassell v. Bird* (2018). Likewise, in *Demetriades v. Yelp*, a California trial court "confirm[ed] that Yelp isn't legally liable for filtering users' reviews as it sees fit" (Goldman, 2013, para. 1). Correspondingly, Yelp is free to strike any speech from its platform that it deems undesirable regardless of whether that speech otherwise enjoys constitutional protection from government censorship, and recent lawsuits arguing that platforms are full-fledged state actors bound by the First Amendment have proven unsuccessful (see, e.g., *Prager University v. Google*, 2018). This is not entirely unfortunate. As Klonick (2018) notes, treating them as true state actors who must comply with the First Amendment would likely preclude their moderation prerogative to such a degree that it "would also likely create an internet nobody wants" (p. 1659). Nonetheless, while the use of Yelp for different variants of commodity activism might indeed be steeped in the logic of neoliberal individualism, it also appears in an environment marked by the increasing partitioning of commerce and politics and the decline of traditional venues for viewpoint-neutral expression in the first place.

Given their sovereign role in the digital speech landscape and their relative freedom from First Amendment constraint, platforms face the enormous task of deciding how exactly to moderate the speech of their users. These processes have been explored extensively by Klonick (2018), Gillespie (2018), Sarah T. Roberts (2019), and Sarah Myers West (2018). Though Klonick's (2018) article focuses on YouTube and Facebook, it details how platforms have generally shifted from a moderation approach based in standards (general and sometimes subjective norms such as "don't drive too fast") to a system of precise rules that

try to define the boundary between permitted and prohibited conduct (i.e., exact speed limits; pp. 1631–1632). Such a shift came in response to the challenges of moderating content on an increasingly larger scale (pp. 1632–1633). The shift toward rules-based moderation at scale in some sense necessitates less responsiveness to unique user demands. While "the principles formalized in [both] rules and standards are rooted in the social norms and values of a community," rules are "more distant from the [community] norms they are based on and 'do not depend on ongoing dialogue to gain dimension or consent'" (p. 1632). Nonetheless, Klonick notes that in the platforms' constant revisions of the internal guidelines for moderation, they seem to be "attempting, in large part, to rapidly reflect the norms and expectations of users" (p. 1649).

Gillespie (2018) also emphasizes the difficulty of delineating precise guidelines for what does and does not belong on a platform while staying responsive to their protean uses. He formulates this in terms of the tension between "the aims of independent content providers who want their work to appear on a public forum and the platform's own economic and political imperative to survive and flourish" (p. 19). Just as Klonick suggested that platforms remain mindful of the sometimes dissenting norms communicated by users, Gillespie also asserts that power goes both ways in the moderation equation: "While platforms structure user activity, users also have power over platforms—maybe less so as mere individuals or groups, but more in the aggregate, the slow, unrelenting shifts in what people seem to want to do" (p. 23).

West (2018) and Freberg and Weed (2018) also provide important foundational analyses of the dissonances that arise between user and platform. West's (2018) research on user perceptions of content moderation across different platforms indicates that some users "expressed the belief that, regardless of the reason for the takedown, social media companies *should* be held to the same legal requirements as the U.S. government under the First Amendment" (p. 4373, emphasis). To vindicate these feelings and to overcome the opacity of moderation decisions, she describes how users sometimes "s[eek] out a variety of inventive mechanisms to assert their agency, turning to various parts of the platform in order to break through the systems of automation" (p. 4379). While West's respondents primarily discussed moderation on Facebook, Twitter, and Instagram, the weaponization of Yelp provides another opportunity to observe such negotiations of rules and agency in action in the context of a particular platform. Likewise, while Freberg and Weed (2018) suggest that "monitoring and moderation of reviews may lead to a perception of censorship to arise" (p. 72)—in their case study of a restaurant excoriated on Yelp for not disclosing the use of Popeyes chicken in its dishes—they do not examine how such perceptions are articulated by users or how they stem from Yelp's endeavors to police the boundary between political and consumer sentiments in reviews.

There is therefore additional room to analyze Yelp's guidelines and their operationalization in the 2018 weaponization case studies to observe how tensions regarding the boundaries of acceptable speech are negotiated in practice. How does Yelp moderate content, and how does this accord or conflict with the ways in which different factions of users use the platform in cases of weaponization?

Methodological Approach

The remaining sections analyze Yelp's architecture and policies for moderating speech and then trace how these factors are manifest in the two weaponization case studies. The cases were observed in a similar but not identical manner. Review text as well as metadata (star rating, date, and user profile attributes) were manually entered into a spreadsheet at intervals of one to five days in July and August 2018. By observing the changes to the page in this ongoing manner, the research attempted to account for the downgrading or outright removal of individual reviews in as much detail as possible. In addition, this approach allowed for the original text of some reviews to be recorded before they were removed for violation of the terms of service (which erases the text of a review and leaves only the user, date, and star rating), thus offering some additional insight into precisely what language might cause reviews to be filtered. Effort was made to track any changes in the treatment of each review by Yelp's filter (i.e., from "recommended" to "removed" or to "not recommended"), though given the intervals of data collection, the exact time when such changes occurred was sometimes not captured with precision (Yelp pages do not indicate when a review's status is changed, only when it was originally posted). This procedure continued until the composition of filtered, unfiltered, and new reviews on each of the pages in question appeared to have reached an equilibrium—that is, they had not changed substantially for several days.

In the case of RC Design (the firm targeted for its proprietor's behavior at an immigration rally), the Yelp page was monitored starting shortly after the incident occurred on June 28, 2018. The text of 83 reviews was collected, with the majority of these (71) coming from users who had previously written reviews of other businesses. In the case of the Red Hen restaurant (the restaurant that refused service to Trump administration press secretary Sarah Huckabee Sanders), the changes in the Yelp page were monitored in a similar manner starting several weeks after the incident, when the researcher became aware of it. The text of 40 reviews was collected, and all but two of these reviews came from users who had left other reviews on the site. Though some reviewers reposted their filtered reviews (often many times), the full text of several hundred filtered reviews was not captured for the Red Hen case because these reviews had already been demoted to the removed section by the time the research began.

Finally, a Yelp spokesperson provided answers to two rounds of questions via e-mail in July 2018. The questions concerned both Yelp's general ethos of content moderation and mechanisms for enforcing its content guidelines as well as its approach in the particular cases analyzed here (though the latter questions were not addressed in the spokesperson's answers).

Yelp Platform: Design, Filter, and Content Guidelines

Founded in 2004, Yelp hosts approximately 163 million reviews, and its press materials claim this "mak[es] Yelp the leading local guide for real word-of-mouth on everything from boutiques and mechanics to restaurants and dentists" (Yelp, n.d.d). This perception of the platform's prominence is shared to at least some degree. Describing Yelp's centrality in the information landscape of the Web, an article on the *PC World* website states, "After Google, Twitter, and Facebook, Yelp is arguably the Internet company next in line to have its name converted into a verb" (Null, 2012, para. 1). The site's guidelines stipulate a preference for reviews of consumer-facing businesses and service providers that have physical locations

with some exceptions (Yelp, n.d.e). While the review page for an individual business can be submitted by users of the site, a business may claim the page to keep its vital information updated or add other locations, track visitor metrics for the page, and respond to reviews (Yelp, n.d.c). Yelp also sells advertising space to businesses, which appear in organic Yelp search results for particular service or product terms and even on competitor pages. Buying Yelp advertising also affords the business an upgraded profile with options that include expanded promotional features (such as deals and videos) or the removal of competitor ads from a business's page (Yelp, n.d.a). Despite perennial speculation, Yelp vehemently insists that businesses that purchase advertising do not enjoy any kind of preferential treatment of their reviews (Yelp, n.d.f). According to Kuehn (2013), the Yelp platform has "redesigned its architectures of participation in accordance with the site's commercial objectives," which has "reshap[ed] the site's evaluative model toward one arguably more business-centric than user-centric" (p. 617).

Content on Yelp is managed through the site's distinct user interface and approach to content moderation. Yelp's central reputation metric for businesses is the star rating, which ranges from one to five. A Harvard Business School study on Yelp has found that differences in star ratings can have a significant impact on the success of a business, with a one-star increase generating a 9% increase in revenue (Luca & Zervas, 2010). According to several explanations featured on the website, Yelp's filter polices reviews in several ways. First, its proprietary algorithm automatically screens out some reviews that it determines are not likely to be useful to consumers. According to an explanatory video, these are from "users we don't know much about, or that might be fakes, or unhelpful rants or raves" (Yelp, 2010). A Yelp spokesperson (personal communication, 2018) added "biased" and "solicited" to this list of attributes that might cause a review to be filtered, though the spokesperson did not elaborate on how such determinations are made. In terms of approaches to governance, Yelp's content guidelines are thus defined according to both more subjective standards (no unhelpful rants) as well as those that attempt to approximate rules that are more exact (no fake reviews).

While the recommendation software is opaque by design (Yelp, in fact, claims that this helps prevent businesses from gaming it), Yelp's explanatory pages also allude to criteria such as reviewer metadata (such as number of previous reviews, location, and presence of a profile picture), star rating (because research indicates that extreme reviews are more likely to get filtered), and the language used in the review itself as central factors in the filter's judgments about review quality (Yelp, 2010). Data from a 2013 *Marketing Land* study indicate that 70% of reviews written by a user with no other reviews were filtered, as were 41% of those submitted by a user lacking a profile picture (McGee, 2013). Likewise, the Harvard study cited above (Luca & Zervas, 2010) indicates that one- and five-star reviews were significantly more likely to be filtered than less extreme reviews.

Filtered reviews are sequestered to a separate "not recommended" page; their text is preserved, but they do not factor into the aggregate star rating. This is, by nature, an imperfect process, but as Yelp sees it, it is better for "reviews that reflect perfectly legitimate experiences [to be] sometimes unrecommended out by the recommendation software's algorithmic processes" than to "b[e] a laissez-faire review site that people stop using" (Yelp, 2010). In fact, Yelp regards itself as a leader in the practice of online content moderation. As the spokesperson explained:

We have been dealing with this type of issue since our inception over a decade ago. Unlike other tech platforms that may be discovering the need for policing content in the wake of the 2016 election, this is something we've been doing for a long time, as everyone from small businesses to online pranksters have attempted to use our platform for purposes beyond review content. (personal communication, 2018)

One approach to policing content that Yelp added to the site in September 2015 is the "active cleanup alert" that is placed on the home Yelp page of a business that has received a large spike in review activity because of media scrutiny (Slegg, 2015). As the alert in its current form informs readers, "This business recently made waves in the news, which often means that people come to this page to post their views on the news." The message then directs users to the site's separate discussion forum, Yelp Talk, and warns prospective reviewers that their reviews might be removed "as part of the cleanup process" (screen capture available).

Yelp also preserves a separate page of reviews that have been "removed for violating Yelp's terms of service." The text of these reviews is not visible, though the profile of the reviewer is. In addition to the algorithmic filter, Yelp employs a "user support team" that reviews content that has been flagged by consumers and businesses because it might violate Yelp's terms of service. Yelp's content guidelines delineate a particular vision for the type of information that the company thinks will be most useful for consumers. As the guidelines explain: "The best reviews are passionate and personal. . . . We want to hear about your firsthand consumer experience[;] feel free to air your opinions, but don't exaggerate or misrepresent your experience" (Yelp, n.d.b). In addition to personal experience and factuality, the guidelines stipulate that "reviews aren't the place for rants about a business's employment practices, political ideologies, extraordinary circumstances, or other matters that don't address the core of the consumer experience" (Yelp, n.d.b).

The characterization of reviews that comment on political matters as "rants and raves" is strikingly consistent across Yelp's published guidelines and explanations, and it was also invoked verbatim by the spokesperson in explaining the review criteria (in addition to the allusion to "pranksters" who have long necessitated content moderation on the site). The choice of word *rant* is itself perhaps significant in expressing a kind of disdain for this content. As the cases discussed subsequently will demonstrate, it is certainly possible to refer to a business owner's ideology without the review being tantamount to a rant. Regardless, given the overarching descriptions of the ideal review in the content guidelines, it seems clear that Yelp has constructed a schema for content moderation that grounds its normative standard of review content usefulness in a particular conception of customer experience that does not include discussion of politics. In a more rule-based sense, furthermore, it offers a tangible but somewhat inexact list of criteria that are used to judge whether reviews are likely to come from real customers.

Case Studies

The discussion of the case studies will proceed in the following manner. After a brief summary of the inciting events and media coverage, the evolution of the content of the Yelp page is discussed in terms of both the critical sentiments expressed by reviewers as well as their tactical and rhetorical engagement

with the Yelp filter. The analysis concludes with a consideration of how this kind of criticism has a lasting impact—both on its target as well as in the propagation of an enduring topical and normative message. Generating this comprehensive portrait of the spread of the story, the substance of the reviews and behavior of the reviewers, and the possible impact allows us to understand how the weaponization of Yelp both constitutes an intractable moderation problem for the platform and reflects a particular normative conception of the role of private intermediaries in the digitally mediated public forum.

In June 2018, a photograph circulated on social media platforms of a woman in a red "Make America Great Again" hat who appeared to be screaming at a teenage immigration activist. The woman was soon outed as Roslyn La Liberte, the proprietor of RC Associates, a contractor in Woodland Hills, California. Starting on June 28, hundreds of tweets mentioned her and her business by name; one that was retweeted repeatedly implored readers to "boycott RC Design" (caps omitted). On June 29, MSNBC anchor Joy Reid retweeted one of the most ubiquitous denunciations of La Liberte. Conservative websites denounced Reid for "lead[ing] a Twitter mob in false allegations," as several news outlets had subsequently reported that when interviewed, the boy in the photo had contended that the exchange had in fact been civil (Taylor, 2018). The Yelp page was likewise inundated with one-star reviews starting on June 28, and Yelp placed its active cleanup alert message on the page on June 29.

Overwhelmingly, the negative reviews written in response to the viral photograph decried the racism that posters assumed from La Liberte's political stance on immigration. While the subsequent reporting suggests that the picture is a misleading representation of the exchange and Reid was eventually sued for defamation in October 2018 because of her amplification of the misleading tweet, La Liberte visually appears to be screaming at the teenage counterprotester, which undoubtedly made it easier for those already unsympathetic to her beliefs to assume she is a hostile zealot. Of the 83 reviews whose text was collected for this project, nearly one-third explicitly allude to hate or racism as part of their criticism; others criticize what they saw as the apparent incivility of La Liberte's engagement with the boy.

Several of these reviews directly advance a counternorm to Yelp's characterization of a "good" review (and to the users cited in Kuehn's study of commodity activism on Yelp): Such sociopolitical perceptions are indeed directly relevant to consumer decisions. One user declared that she "wouldn't ever hire someone, or their company, [who] seem[s] to condone this behavior" (Jodie C.), and another stipulated that because of the incident, she "wouldn't be able to trust [RC's] workers in my home!" (Judy M.). While not referring to the incident, a third stated, "This company unfortunately has hired someone that I do not see myself wanting to do business with" (Danielle G.).

Others reviewers conceived of punishment in the economic marketplace as the appropriate means of signaling their social disapproval. One reviewer formulated this argument comprehensively:

> Hateful actions deserve consequences, including to her bottom line, which the Trump administration was promising to grow with his policies. Let's prove that this harmful thinking and rhetoric will have the opposite effect and hit her where it hurts—her wallet. (Katherine O.)

In addition to again implying that political sentiments are indeed relevant to otherwise commercial decisions, calling for others to boycott the company because of La Liberte's remarks works to advance a norm of tolerance and respect through the realm of consumer information.

Given Yelp's professed intolerance for rants inspired by media controversy, it is perhaps not surprising that all but one of the reviews quoted above were removed from the recommended page by July 17, 2018 (several weeks after the incident). Despite many of the reviews originating in Southern California (and thus theoretically more likely to have been written by customers), their references to the incident were probably too direct. The review that merely referenced the fact that the company had "hired someone" was still recommended as of late August 2018. Its author lists a location of San Antonio, Texas, but perhaps the reference to undesirable personnel was oblique enough that it appeared to reflect authentic customer experience. (The Yelp spokesperson contacted did not comment on questions about individual cases.)

Many other reviews were less explicit in criticizing the political beliefs of the business in justifying their one-star rating. A number of these remained recommended as of August 22, 2018. For instance, one from June 29 by a reviewer in Palo Alto, California, described being "really unsatisfied with the customer service I received from I believe it was Rosalyn" (Ignacio B.). This reviewer further explained that he "hope[s] they hire better people in the future, they are losing customers." Such a review nominally refers to customer service, yet it focuses on La Liberte as the source of the problem. Simply put, given the timing and the focus on La Liberte, the likelihood that this review reflects the experience of a potential customer who sought construction services seems low.

Other reviews that have remained recommended display a similar approach. One from July 11 by a reviewer in Los Angeles states, "They have no idea what they're doing. From ownership on down, this company is unethical and will overpromise and under-deliver" (Artin P.). Another from June 30 alleges that the business had "no morals or respect at all" (JJ G.) without elaborating on any reasons for this conclusion. A reviewer whose initial review had been removed for violating the terms of service (it referred to La Liberte as "spew[ing] out her vile, hateful rhetoric at a 14-year-old") wrote a second that remained recommended; this review only cryptically claimed, "They are representing themselves something other than they are" (Giorgio T.). Again, given the direct reference to the incident in the first review, it seems unlikely that the second reflects experience as a customer. Rather, these reviewers have simply written reviews that evade removal by staying under a threshold level of explicitness in referring to the media incident.

Other reviews that managed to stay recommended throughout the page cleanup went further to describe scenarios involving their purported customer interactions with the business. "I recently had some work done on a house remodel and these people were not on time and were rude when they finally showed up," wrote one reviewer (Andres F.) from Las Vegas on June 29. Another from June 30 describes a dissatisfying experience having his grandmother's house remodeled in great detail, down to "'granite' counter tops [that] were fake, cracking, and warped" (Jae D.). The reviewer mentions that the company "also referred to her as 'you people,' referring to her Haitian accent and skin color." While the review could

theoretically describe a real experience, the timing and the inclusion of a comment related to racial prejudice again could well indicate a response to La Liberte's public behavior in disguise.

The volume of reviews greatly diminished after mid-July (about two weeks after the incident). Nearly 300 reviews from the initial flurry of scrutiny were eventually removed for violating the terms of service, but the last review was removed on July 3. On the other hand, at least 11 that were initially not recommended or removed for terms violations were subsequently reclassified as recommended by late July; nearly all reference the kind of vague deficiencies of character, craftsmanship, and work ethic cited in the examples above. Curiously enough, the business neglected to claim the Yelp page, though doing so would have afforded it the opportunity to respond to reviews (especially, perhaps, to those that might have fabricated customer experiences). Roughly two months after the late June incident, the 29 recommended reviews for the business still had an average rating of one star. All but two of them were left between June 28 and June 30, during the peak of the response to the viral photograph.

While the situation seems to have stabilized, the business's page is still dominated by response to the incident. Yelp clearly removed the vast majority of the reviews that referred directly to the incident, but the remaining recommended reviews seem to have largely succeeded in compromising a significant part of the company's online presence. Further, Yelp's moderation efforts in this case appear to have treated a set of reviews that simply offered opinion commentary on a media incident more harshly than some (such as the review referencing "fake granite countertops") that could well contain false and materially consequent statements of fact about the business's construction acumen. Thus, the Yelp page clearly functions in this case as a symbolically significant means of signaling social disapproval that is intertwined with (rather than distinct from) commercial considerations. Likewise, the criticism in this case appears to exploit a genuinely vexing moderation challenge for Yelp, suggesting a limit to the site's ability to enforce rules that dichotomize political and commercial content.

The Red Hen restaurant in Lexington, Virginia, made headlines of major professional outlets immediately when it refused service to Sarah Huckabee Sanders on June 22, 2018. As Huckabee Sanders (2018) confirmed in a tweet on June 23, "Last night I was told by the owner of Red Hen in Lexington, VA to leave because I work for @POTUS and I politely left." Conservative activists swiftly called for a boycott of the restaurant. In a comment subsequently cited in numerous news reports, Turning Point USA leader Charlie Kirk (2018) declared that the backlash represented "the market at work" and argued that "if you refuse [to serve someone], you must accept the consequences." According to *Fortune*, the restaurant had 15,000 (largely negative) reviews by June 25 (Morris, 2018).

Reviewers echoed Kirk's general sentiment but also expressed reservations about the idea of rejecting customers based on political leanings. The reviews consistently evince a normative ethos in which businesses should expect to be punished by consumers when they stray into the realm of politics. "The Red Hen deserves to go out of business. . . . We can't have arrogant owners booting people out simply because she doesn't like their politics," argued Johnny H. In a similar vein, Adam R. argued, "When you disrespect customers you should be recognized for it." As with the RC case, many commenters decried what they saw as the fundamental incivility of the restaurant's treatment of Sanders (regardless of Sanders' own account of

the event, which seemed to indicate that the confrontation was not particularly contentious). As one review stated,

> I have not been to your restaurant but the way you treated her speaks volume. i am a nurse and i provide care to everyone regardless of skin color, religious belief (even those i completely disagree and abhor), gender (& identity), political persuasion, nationality, economic status, profession (or the lack of), and so on. (Emay B.)

Others described the proprietor as "bigoted" (e.g., John O.), "hate-filled" (D.K.), or "pompous, intolerant, and hypocritical" (C.W.). Some reviewers indulged in more hyperbolic rhetoric, with one decrying the owner's "discrimination" based on political party; another commenter (evidently without irony) compared Sanders' treatment to "Nazi Germany and the Jews who were thrown out of public establishments" (Tom S.).

While these sentiments are premised on a normative rejection of the idea that politics and consumption should overlap, the very act of leaving the reviews of course reinforces a normative ethos among these reviewers that consumer decisions are, in fact, political. Thus, it is not surprising that others expressed incredulity about Yelp's efforts to cleanse the page of reviews expressing reaction to the incident with Sanders. "This policy, to not serve certain people, may seem odd but should be disclosed up front. No judgement, let the consumer decide," wrote a reviewer from New York. The Red Hen incident thus complements the RC Associates incident in evincing a segment of users who reject Yelp's neat dichotomy between politics and consumption.

Compared with the RC Associates case, however, the reviewers who flocked to the Red Hen page in response to the incident appear to have been far more persistent (or at least belligerent) in their efforts to evade the filter. As with RC Associates, reviews that explicitly mentioned the incident with Sanders were quickly removed. Several reviewers repeatedly reposted their reviews over July and August 2018—sometimes upward of 10 times. For some of these reviewers, Yelp's policy made little sense: "Yelp keeps removing one star reviews, which violates the very spirit of the app" (Jeremy S.). Filtering reviews does, of course, technically distort the true scope of public sentiment about the company; Yelp believes that doing so is actually advantageous because it curates the visible information in a way that readers will (ideally) find useful. The comment by Jeremy S. suggests a different conception of the spirit of the platform in which the star rating aggregation is expected to reflect the complete volume of different sentiments that have been expressed about a business.

To others, the filtering was tantamount to bias on the part of the platform. After excoriating Yelp for being a "political POS," one review argued, "Yelp needs to get a grip on its opinion of we the people who make this site what it is" (Bo N.). For another, the filtering meant that he would "have to decide if I want to continue to value ANY of the content on this website since it is clearly controlling the type of reviews that are being posted" (Jesse C.). Such a comment not only expresses discontent with the filter as applied in this case; it suggests that the content moderation itself created a kind of disinformation effect in which the ability to distinguish legitimate from illegitimate reviews of the business was compromised. Many others were explicit about how many times the reviews had been removed and vowed to continue posting them. Several

of these (e.g., Kyle W.) were still being reposted as verbatim copies of reviews initially left in late June and early July and temporarily appearing on the recommended page over two months after the incident.

Because of the cleanup notice, the persistent repeat reviewers, and a handful of new ones, the page of recommended reviews for the Red Hen still bears traces of the incident. At the same time, the kind of one- and two-star reviews from faraway posters that might well represent fabricated customer experiences have not remained to the degree that they did in the RC case. Further, the restaurant had a much more established history of positive feedback at the time of the incident. While a GoFundMe page was set up to aid the business immediately after the incident, the restaurant was back to an average star rating of 4.5 based on 106 recommended reviews as of August 25, 2018.

Conclusion: Political Speech and Content Moderation on Yelp

The review texts and the actions of the reviewers in both cases illustrate how Yelp attempts to police the boundary between politics and consumption through its moderation policies and the ways in which some users of the platform contest this normative dichotomy—both substantively and tactically. First, many reviewers in each case advocated for an understanding of commerce that included considerations of political ideology and cast Yelp as a meaningful venue to advocate for social norms such as tolerance, diversity, or civility. Regardless of whether they coalesce into a collective struggle, such postings represent a contribution to the cultural contestation over the meaning of these terms and thus contribute to the kind of unplanned encounters with ideas that scholars like Sunstein see as vital to democracy.

Relatedly, the reviews represent a novel means of disrupting the gears of both commerce and the consumer information-seeking experience. In the short term, attacking the pages creates a headache for the business proprietors that Yelp's content moderation does not fully rectify. In the longer term, the fact that the pages still bear the trace of the incidents months later means that they act as a kind of digital media archive that cuts against the trend toward sanitized commercial spaces (whether physical or virtual).

If a significant faction of users clearly feels that the platform has assumed a responsibility to reflect some kind of comprehensive portrait of information about a business (indeed, Yelp has positioned the platform as a central arbiter of consumer information), then how should this color our judgment of the platform's approach? One of the company's preferred solutions—to sequester political discussion on separate Yelp Talk pages—is perhaps a productive alternative, yet it also seems analogous to the often-maligned "free speech zones" in physical protest situations (see Bovard, 2003). While the Yelp Talk pages technically afford another platform for the kind of criticism that Yelp endeavors to filter from the main business pages, the reviewers are not simply seeking somewhere to discuss the businesses; they are deliberately targeting the pages as one might picket outside a business's physical location. Rather than pedantically reminding users that they have no right to post Yelp reviews, we should also acknowledge what the sentiment signals about normative contemporary understandings of platforms as a kind of virtual storefront in these kinds of cases.

At the same time, a fully laissez-faire approach to content moderation leads to the tragedy of the commons, where a user captures the benefit of using a resource (i.e., the liberty to post a review) while the cost (i.e., the clutter of many reviews) is distributed. "[The] tragic result of each person thinking this way . . . is ruin of the commons, and thus of everyone using it" (Myers, 2000). As Gillespie (2018) asserts regarding platforms categorically, "There is no platform that does not impose rules to some degree. Not to do so would simply be untenable" (p. 5). Yelp must therefore reconcile the tragedy of the commons risks with the demand for the platform to host political condemnation of business.

Though it has no legal obligation to do so, Yelp might nonetheless orient itself toward this challenge in the future by considering the spirit of the town meeting framework famously outlined by free speech theorist Alexander Meiklejohn: "Rather than a laissez-faire marketplace . . . the open-but-orderly process of debate represented by a democratically conducted town-hall meeting provides the most effective manner for 'protecting the common needs of all the members of the body politic'" (Kerr, 2007, p. 65). Informed decisions can only be "achieved by allowing citizens access to the full range of ideas in public discussion" (p. 65).

Meiklejohn would agree with Yelp's efforts to keep the platform from being overrun with one redundant point of view (e.g., political condemnation of a business owner). Yet his emphasis on representing the range of discussion would also include those voices that condemn a business proprietor's words or actions as part of their consumer commentary. To its credit, the active cleanup alert message perhaps helps keep redundant political criticism from dominating pages while also preserving a trace of the relevant controversy. Readers who are especially curious can discover what happened through other information sources and come to their own decisions about its significance. Then the platform can perhaps continue to evolve with the norms of its use by considering additional creative ways to preserve some record of the volume and vehemence that a business owner elicited in cases of weaponization while also more consistently identifying and discouraging reviews that simply add false information to the narrative by gaming the platform's consumer experience content criteria.

References

Banet-Weiser, S., & Mukherjee, R. (2012). Introduction. In S. Banet-Weiser & R. Mukherjee (Eds.), *Commodity activism: Cultural resistance in neoliberal times* (pp. 1–22). New York, NY: New York University Press.

Bovard, J. (2003, December 15). Free-speech zone: The administration quarantines dissent. *The American Conservative*. Retrieved from https://www.theamericanconservative.com/articles/free-speech-zone/

Coffin, J. (2000). The united mall of America: Free speech, state constitutions, and the growing fortress of private property. *University of Michigan Journal of Legal Reform*, *33*(4), 615–649.

Freberg, K., & Weed, A. (2018). Yelp reviewers call fowl for restaurant's menu switcheroo. In A. Laskin (Ed.), *Social, mobile, and emerging media around the world* (pp. 69–82). Lanham, MD: Lexington.

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

Goldman, E. (2013, February 6). Yelp defeats legal challenge to its user review filter. *Forbes*. Retrieved from https://www.forbes.com/sites/ericgoldman/2013/02/06/yelp-defeats-legal-challenge-to-its-user-review-filter/#43f373b442da

Hassell v. Bird, 247 Cal. App. 4th 1336 (2018).

Kerr, R. (2007). Justifying corporate speech regulation through a town-meeting understanding of the marketplace of ideas. *Communication Monographs*, *9*(2), 57–113.

Kirk, C. (2018, June 24). [Tweet]. Retrieved from https://twitter.com/charliekirk11/status/1011040773712867329

Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, *131*, 1598–1670.

Kuehn, K. (2013). "There's got to be a review democracy": Communicative capitalism, neoliberal citizenship and the politics of participation on the consumer evaluation website yelp.com. *International Journal of Communication*, *7*, 607–625.

Kuehn, K. (2017). Brand local: Consumer evaluations as commodity activism on Yelp.com. *Journal of Consumer Culture*, *17*(2), 205–224.

Lidsky, L. (2011). Public forum 2.0. *Boston University Law Review*, *91*, 1975–2028.

Luca, M., & Zervas, G. (2010). *Fake it till you make it: Reputation, competition, and Yelp review fraud* (Harvard Business School NOM Unit Working Paper No. 14-006). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2293164

Matsakis, L. (2018, June 26). The Red Hen and the weaponization of Yelp. *Wired*. Retrieved from https://www.wired.com/story/red-hen-trump-and-weaponization-of-yelp/

McGee, M. (2013, July 30). Yelp more likely to filter extreme reviews and new accounts. *Marketing Land.* Retrieved from http://marketingland.com/yelp-more-likely-to-filter-extreme-reviews-new-accounts-study-53622

Morris, D. (2018, June 23). Trump press secretary Sarah Sanders kicked out of Virginia restaurant. *Fortune*. Retrieved from http://fortune.com/2018/06/23/trump-press-secretary-sarah-sanders-kicked-out-red-hen-virginia-restaurant/

Myers, G. (2000). Thoughts on Hardin's tragedy of the commons. Retrieved from http://faculty.wwu.edu/gmyers/esssa/Hardin.html

NAACP v. Claiborne Hardware, 458 U.S. 886 (1982).

Null, C. (2012, February 12). Yelp alternatives: Which user review services matter? *PC World*. Retrieved from https://www.pcworld.com/article/249770/beyond_yelp_which_user_review_services_matter_.html

Perry Education Association v. Perry Local Educators' Association, 460 U.S. 37 (1983).

Prager University v. Google, 2018 WL 1471939 (N.D. Cal. March 26, 2018).

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media.* New Haven, CT: Yale University Press.

Sanders, S.H. (2018, June 23). [Tweet]. Retrieved from https://twitter.com/PressSec/status/1010536237457924096

Sauter, M. (2014). *The coming swarm: DDOS actions, hacktivism, and civil disobedience on the Internet.* New York, NY: Bloomsbury.

Slegg, J. (2015, September 17). Yelp adds new active cleanup alerts for local businesses in the news. *The SEM Post*. Retrieved from http://www.thesempost.com/yelp-adds-new-active-cleanup-alerts-for-local-businesses-in-the-news/

Sunstein, C. R. (2007). *Republic.com 2.0.* Princeton, NJ: Princeton University Press.

Taylor, K. (2018, July 1). Horrible! Conservative woman falsely accused of calling kid a "dirty Mexican"—Joy Reid leads Twitter mob in false allegations. *Gateway Pundit*. Retrieved from https://www.thegatewaypundit.com/2018/07/woman-in-viral-photo-put-on-twitter-mob-blast-by-teen-activist-and-joy-reid-was-falsely-accused-by-them-of-calling-boy-dirty-mexican/

West, S. M. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, *20*(11), 4366–4383.

Yelp. (2010, March 18). *Yelp's recommendation software explained.* Retrieved from https://www.yelpblog.com/2010/03/yelp-review-filter-explained

Yelp. (n.d.a). *Advertising with Yelp.* Retrieved from https://biz.yelp.com/video/yelp_intro_ads

Yelp. (n.d.b). *Content guidelines*. Retrieved from https://www.yelp.com/guidelines

Yelp. (n.d.c). *Find and claim your Yelp business page*. Retrieved from https://biz.yelp.com/?utm_source
    =how_to_claim&utm_medium=yelp_support&utm_content=search_biz_link&utm_campaign
    =claim_business

Yelp. (n.d.d). *Press*. Retrieved from https://www.yelp.com/press

Yelp (n.d.e). *What are the guidelines for adding a business to Yelp?* Retrieved from https://www.yelp-
    support.com/article/What-are-the-guidelines-for-adding-a-business-to-Yelp?l=en_US

Yelp. (n.d.f). *Yelp does not extort local businesses or manipulate ratings*. Retrieved from
    https://www.yelp.com/extortion