
FIRST AMENDMENT — REGULATION OF ONLINE SPEECH —
NINTH CIRCUIT FINDS FIRST AMENDMENT VIOLATION IN
SCHOOL DISTRICT OFFICIALS’ BLOCKING OF PARENTS ON
SOCIAL MEDIA. — *Garnier v. O’Connor-Ratcliff*, 41 F.4th 1158 (9th
Cir. 2022).

The Internet has changed the way we speak — and also the way we disrupt speech. In the ever-protean First Amendment jurisprudence of free speech, novel forms of communication on the web should make us proceed with caution as we pour new wine into old skins. Recently, in *Garnier v. O’Connor-Ratcliff*,¹ the Ninth Circuit held that the First Amendment restricts the ability of public officials to block private individuals on social media. While the court reached the right outcome, its analysis of relevant government interest hinged too heavily on analogizing to situations arising in physical fora. Such reasoning leaves little room for acknowledging unprecedented speech disruptions enabled by social media. Shifting attention toward the “felicity conditions” of speech — conditions that must be satisfied in order for speech to achieve its intended effect — may facilitate a more precise understanding of speech disruption.

Michelle O’Connor-Ratcliff and T.J. Zane (“Trustees”) were elected to the Poway Unified School District (PUSD) Board of Trustees in November 2014.² While running for election, they each created public Facebook pages to promote their campaigns.³ They continued to use the pages to announce PUSD-related information and solicit public opinion about the Board’s decisions after assuming office.⁴ In 2016, O’Connor-Ratcliff created a public Twitter page for similar uses.⁵ These social media spaces, which were identified as official pages of government officials, allowed members of the public to reply to original posts made by the Trustees in the form of comments or to register non-verbal reactions.⁶

Christopher and Kimberly Garnier, parents of children in PUSD schools, began leaving comments on these pages sometime in 2015.⁷ The Garniers had been active critics of the Board for years, participating in public meetings and emailing the Board to express concerns about race relations in the District and the financial misconduct of PUSD’s superintendent at the time.⁸ Frustrated by the PUSD’s lack of response, the Garniers posted lengthy and repetitive comments on the Trustees’

¹ 41 F.4th 1158 (9th Cir. 2022).

² *Id.* at 1163.

³ *Id.*

⁴ *Id.* at 1164–65.

⁵ *Id.* at 1163.

⁶ *Id.* at 1163–64.

⁷ *Id.* at 1165–66.

⁸ *Id.*

Facebook and Twitter pages.⁹ For example, Christopher Garnier left nearly identical comments on 42 separate posts on O'Connor-Ratcliff's Facebook page; he also posted 226 identical replies to her Twitter page within ten minutes.¹⁰ The Trustees initially responded by deleting or hiding these comments individually; around October 2017, they blocked the Garniers from their social media pages.¹¹ Later, the Trustees used the "word filter" function on their Facebook pages to prevent comments containing designated words from being posted.¹² The broad list of filtered words practically disabled any viewer from posting new comments.¹³

The Garniers then filed suit in federal court under 42 U.S.C. § 1983, claiming that the Trustees could not block them on social media consistent with the First Amendment.¹⁴ After a bench trial, Judge Benitez found for the Garniers.¹⁵ He held that the Trustees acted under color of state law and that their social media pages were designated public fora.¹⁶ Judge Benitez found that the *initial* blocking of the Garniers served the substantial government interest of "facilitat[ing] transparency in government" and "promoting online interaction with constituents through social media."¹⁷ In his view, the blocking also constituted a narrowly tailored content-neutral regulation because it was based on the repetitive nature of the Garniers' comments, rather than on their criticism of the Board.¹⁸ However, he concluded that the *continued* blocking of the Garniers for the next three years was no longer narrowly tailored to the transparency interest.¹⁹ He cautioned that the defendants may legitimately reblock the Garniers should they repeat their "repetitive and largely unreasonable behavior"²⁰ and that the defendants may also adopt content-neutral rules of decorum.²¹ Both parties appealed.²²

⁹ *Id.* at 1166.

¹⁰ *Id.*

¹¹ *Id.*

¹² *Id.*

¹³ *Id.*

¹⁴ *Garnier v. O'Connor-Ratcliff*, 513 F. Supp. 3d 1229, 1232 (S.D. Cal. 2021). Judge Whelan, to whom the case was first assigned, dismissed the Garniers' claim for damages as barred by qualified immunity but allowed their claims for declaratory and injunctive relief to proceed to trial. *See Garnier v. Poway Unified Sch. Dist.*, No. 17-CV-2215, 2019 WL 4736208, at *5 (S.D. Cal. Sept. 26, 2019). The case was then reassigned to Judge Benitez. *See Garnier*, 513 F. Supp. 3d at 1232.

¹⁵ *See Garnier*, 513 F. Supp. 3d at 1233.

¹⁶ *Id.*

¹⁷ *Id.* at 1252.

¹⁸ *Id.* at 1248–49.

¹⁹ *Id.* at 1251.

²⁰ *Id.*

²¹ *Id.* at 1252.

²² *Garnier*, 41 F.4th at 1167.

The Ninth Circuit affirmed.²³ Writing for a unanimous panel, Judge Berzon²⁴ ruled that the Trustees “violate[d] the First Amendment by creating a publicly accessible social media page related to [their] official duties and then blocking certain members of the public from that page.”²⁵ Noting that a successful § 1983 claim requires state action, she chose the “nexus test” — which asks whether there exists “such a close nexus between the State and the challenged action that the seemingly private behavior may be fairly treated as that of the State itself” — as the appropriate test in the instant case.²⁶ She applied the three-pronged test announced in *Naffe v. Frey*²⁷ to demonstrate that the Trustees’ “use of their social media pages qualifie[d] as state action under § 1983.”²⁸ First, the Trustees “purport[ed] . . . to act in the performance of [their] official duties,” as evidenced by their self-identification as government officials and the chief use of the social media pages to announce PUSD-related information.²⁹ Second, the significant number of followers on the pages and the Trustees’ active solicitation of “constituent input about official PUSD matters”³⁰ illustrated “the purpose and effect of influencing the behavior of others.”³¹ Third, the informative function of the social media pages related meaningfully to the Trustees’ “governmental status” and “to the performance of [their] duties.”³²

Next, the court held that the Trustees “violated the First Amendment when they blocked the Garniers from their social media pages.”³³ Judge Berzon analyzed the issue through the lens of the public forum doctrine, which scrutinizes speech regulation based on the category of forum regulated.³⁴ Focusing on the pages’ open access to the public and the initial lack of any content regulation policy, she found the Facebook pages prior to the implementation of the word filter and O’Connor-Ratcliff’s Twitter page to be *designated* public fora — in which restrictions must be both content neutral and “narrowly tailored to serve a significant

²³ *Id.* at 1163.

²⁴ Judge Berzon was joined by Judges Tallman and Friedland.

²⁵ *Garnier*, 41 F.4th at 1163. In addition, Judge Berzon rejected on two bases the Trustees’ argument that — given their effective elimination of comments on the Facebook pages via the word-filter function — the case was moot. *Id.* at 1168. First, she noted that the public could still offer “non-verbal feedback” such as emoticons and that the Garniers could thus seek “effective relief.” *Id.* Second, she stressed that “voluntary cessation of allegedly unlawful activity ordinarily does not moot a case.” *Id.* Here, the Trustees failed to meet the burden of demonstrating that they would not “remove the word filters from their Facebook pages and again open those pages for verbal comments from the public.” *Id.*

²⁶ *Id.* at 1169 (quoting *Kirtley v. Rainey*, 326 F.3d 1088, 1094–95 (9th Cir. 2003)).

²⁷ 789 F.3d 1030 (9th Cir. 2015).

²⁸ *Garnier*, 41 F.4th at 1171.

²⁹ *Id.* (alterations in original) (quoting *Anderson v. Warner*, 451 F.3d 1063, 1069 (9th Cir. 2006)).

³⁰ *Id.*

³¹ *Id.* (quoting *Naffe*, 789 F.3d at 1037).

³² *Id.* (alteration in original) (quoting *Naffe*, 789 F.3d at 1037).

³³ *Id.* at 1177.

³⁴ *Id.* at 1177–78.

government interest.”³⁵ However, once the Trustees disabled comments via the word filter, she found that the pages turned into *limited* public fora, in which all reasonable viewpoint-neutral restrictions are permissible.³⁶ According to Judge Berzon, whether the initial blocking of the Garniers was content neutral posed “a close question.”³⁷ Even if content neutral, however, the blocking served no significant government interest because the technical features of the Facebook and Twitter pages minimized the extent of disruption caused by the repetitive comments, either by trimming lengthy comments or limiting their visibility.³⁸ Judge Berzon cited the holding from *Norse v. City of Santa Cruz*³⁹ that a significant government interest in forum maintenance required a showing of “actual disruption” on the forum, which does not encompass “constructive disruption, technical disruption, virtual disruption, *nunc pro tunc* disruption, or imaginary disruption.”⁴⁰ The visual clutter of the Garniers’ comments fell short of the regulation-worthy disruption that may be caused by an unruly speaker in a physical city hall meeting.⁴¹

Neither was the blocking narrowly tailored, given that the Trustees could have pursued alternative solutions, such as deleting individual comments or establishing rules of decorum.⁴² In addition, the continued blocking of the Garniers after the use of the word filter exceeded the bounds of permissible speech restrictions in a limited public forum.⁴³ By effectively eliminating all comments, the word filter rendered the continued ban superfluous and, therefore, unreasonable.⁴⁴

While the Ninth Circuit reached the correct outcome, its analytic approach is limited to recognizing forms of speech disruption in physical spaces. The features of social media that distinguish it from traditional fora should require a recalibration of the philosophical assumptions underlying speech. The court’s treatment of the significant-government-interest prong in its forum analysis overfocused on analogizing to disruptions of speech in physical fora and failed to consider novel forms

³⁵ *Id.* at 1178 (quoting *Ward v. Rock Against Racism*, 491 U.S. 781, 791 (1989)); *see id.* at 1178–79.

³⁶ *Id.* at 1178–79.

³⁷ *Id.* at 1179.

³⁸ *Id.* at 1181.

³⁹ 629 F.3d 966 (9th Cir. 2010).

⁴⁰ *Garnier*, 41 F.4th at 1181–82 (quoting *Norse*, 629 F.3d at 976).

⁴¹ *Id.* at 1181.

⁴² *Id.* at 1182.

⁴³ *Id.* at 1182–83.

⁴⁴ *Id.* at 1183. Judge Berzon also affirmed the grant of qualified immunity to the defendants regarding the damages claims, explaining that the novel issue of public members being ejected from a public official’s social media did not implicate a clearly defined right. *Id.* at 1184. The Trustees have since filed a petition for certiorari in the Supreme Court. *See* Petition for a Writ of Certiorari, *Garnier* (No. 22-324).

of “actual disruption” that social media enables. A more precise understanding of speech disruption might pay attention to speech that thwarts the conditions necessary for other speech to be successful.

Judge Berzon’s attempt to transplant existing free speech rules onto social media is symptomatic of a general tendency to treat the virtual forum as just another species of the physical forum. In *Packingham v. North Carolina*,⁴⁵ the Supreme Court characterized social media as the contemporary version of the quintessential public forum for exchanging views, such as “a street or a park.”⁴⁶ Lower courts have interpreted this language as establishing the presumption that “social media is entitled to the same First Amendment protections as other forms of media.”⁴⁷

This inclination toward doctrinal consistency should not, however, ignore the caveat that certain forms of speech disruption are unique to cyberspace. The suggested equivalency between physical and virtual fora focuses on the nature of activities — *what things people do* — and not on the mechanism of communication — *how people do things*. While *Packingham* astutely observed that platforms such as Facebook, LinkedIn, and Twitter allow members of the public to “debate religion[,] . . . advertise for employees, . . . [and] petition their elected representatives,”⁴⁸ the issue in that case — the constitutionality of a law banning sex offenders from even viewing, rather than posting on, social media⁴⁹ — did not require noting that communication on these platforms occurs asynchronously or that algorithms may influence page accessibility.⁵⁰

Similarity in the nature of communicative activities does not entail similarity in the modes of disruption to which those activities are vulnerable. The visual-clutter effect contemplated by *Garnier* itself is telling. While it may be tempting to treat this effect as a virtual analog of the unruly speaker in a city hall meeting, as the court did, visual clutter constitutes a different kind of disturbance altogether from auditory bombardment. For example, comments on a webpage remain in space,

⁴⁵ 137 S. Ct. 1730 (2017).

⁴⁶ *Id.* at 1735; *see also* *Reno v. ACLU*, 521 U.S. 844, 868 (1997) (describing cyberspace as “vast democratic forums of the Internet”).

⁴⁷ *Knight First Amend. Inst. at Columbia Univ. v. Trump*, 928 F.3d 226, 237 (2d Cir. 2019), *vacated sub nom. Biden v. Knight First Amend. Inst. at Columbia Univ.*, 141 S. Ct. 1220 (2021) (mem.); *see also* *United States v. Eaglin*, 913 F.3d 88, 96 (2d Cir. 2019) (acknowledging a “First Amendment right to access the Internet”); *United States v. Ellis*, 984 F.3d 1092, 1105 (4th Cir. 2021) (“[A]n Internet ban implicates fundamental rights . . .”).

⁴⁸ *Packingham*, 137 S. Ct. at 1735.

⁴⁹ *See id.* at 1733.

⁵⁰ *See, e.g.*, Jan L. Jacobowitz, *Lawyers Beware: You Are What You Post — The Case for Integrating Cultural Competence, Legal Ethics, and Social Media*, 17 SMU SCI. & TECH. L. REV. 541, 563–64 (2014) (“Online communication is considered to be asynchronous communication, meaning that there is a time gap between when a message is sent and received.”); THE SOCIAL DILEMMA at 56:04 (Exposure Labs 2020) (explaining how a Facebook user’s feed depends on Facebook’s customization algorithm).

whereas spoken words dissipate.⁵¹ In a matter of a few minutes, a single commenter can create lasting disruption; the city hall interrupter can sustain only a temporary disturbance. *Garnier* neglected to explore such differences, apparently deeming the meaning of “actual disruption” self-evident even in the virtual context.⁵²

Sketching the criteria for actual disruption, in fact, deserves more attention. The task requires moving past imperfect analogies to the brick-and-mortar forum. The fundamental question of how speech gets disrupted can be recast to ask how speech fails, which in turn demands an understanding of how speech succeeds. Philosopher of language J.L. Austin’s seminal speech act theory can aid this enterprise. Challenging the traditional view that speech primarily states propositions, Austin observed that we *do* many other things with words — promising, commanding, betting, and so forth.⁵³ This view of speech as the performance of speech acts enables a more precise assessment of the success or failure of speech by examining a speech act’s felicity conditions — conditions that must be satisfied in order for a speech act to achieve its intended effect.⁵⁴ For example, the speech act of commanding requires, among other things, the commander’s authority as a felicity condition, in much the same way that the act of signing a contract must be performed by the correct signer to take effect. A low-ranking soldier uttering the word “Charge!” at a captain fails to command.

The dearth of discussion on what qualifies as actual disruption — in both cyberspace *and* physical space⁵⁵ — is bound to cause thorny issues as online speech outruns comparisons to traditional speech. Given this uncertainty, Austin’s framework offers an analytic schema worth exploring: any speech act that thwarts the felicity condition of someone else’s speech act can plausibly be described as actually disrupting, or even silencing, that speech act. The prima facie plausibility of this schema lies in its explanatory capacity to account for traditional speech disruptions. One obvious way in which the city hall interrupter disrupts is by preventing other speakers from being heard. After all, most communicative statements share the felicity condition that the audience hear them; these speech acts must “secure uptake.”⁵⁶

This flexible conception of speech disruption would invite courts and scholars to contemplate the unique felicity conditions of speech acts on

⁵¹ See Jill I. Goldenziel & Manal Cheema, *The New Fighting Words?: How U.S. Law Hampers the Fight Against Information Warfare*, 22 U. PA. J. CONST. L. 81, 101 (2019).

⁵² See *Garnier*, 41 F.4th at 1181–82 (briefly mentioning *Norse*’s “actual disruption” standard without elaborating).

⁵³ See J.L. AUSTIN, HOW TO DO THINGS WITH WORDS 5–6 (1962).

⁵⁴ See *id.* at 45.

⁵⁵ See *Norse v. City of Santa Cruz*, 629 F.3d 966, 976 (9th Cir. 2010) (omitting any discussion of what counts as actual disruption in a physical forum).

⁵⁶ AUSTIN, *supra* note 53, at 138; see *id.* at 115 (“I cannot be said to have warned an audience unless it hears what I say and takes what I say in a certain sense.”).

social media. *Garnier* hints at one such condition by entertaining the defendants' argument that the repetitive comments had "a net effect of slightly pushing down" other posts.⁵⁷ The Trustees' speech acts of informing and announcing, as well as other viewers' speech acts of opining via comments, all depend on the felicity condition that their text be visible. In other words, the security of uptake on social media depends on visibility, which may be impaired by repetitious comments. Of course, the act of *posting* a comment does not depend on visibility; a mouse click accomplishes the posting, whether that comment reaches an audience or not. The natural assumption, however, is that the Trustees' social media pages differ from, say, a YouTube comment section, in that they were meant to facilitate dialogue rather than merely allow unilateral self-expression, whether the speakers are the Trustees announcing news or parents criticizing the Trustees.

Social media is also especially vulnerable to trolling, spamming, and disinformation. Trolls, people "who post[] deliberately inflammatory messages online,"⁵⁸ take advantage of the unsupervised, anonymous, and asynchronous nature of online forum discussions. While a physical forum is better equipped to handle such comments — by concurrent feedback from other participants, for example — an online forum lacks similar measures. By inciting irrelevant discussions, trolling not only dilutes the substance of the intended discussion hosted by a forum but also risks creating the perception that all comments on a forum are insincere.⁵⁹ An observer will have difficulty distinguishing trolls from sincere participants and thereby attribute "a generalized gross insincerity" to the forum.⁶⁰ Hence, trolling attacks the felicity condition that speakers on a forum be trusted with some minimal level of sincerity. More invidious means of speech disruption might capitalize on social media algorithms. For instance, spam comments can cause Instagram's algorithm to recognize a page as "spammy" and decrease the likelihood of the page appearing to new viewers.⁶¹

⁵⁷ *Garnier*, 41 F.4th at 1180.

⁵⁸ *Troll*, COLLINS, <https://www.collinsdictionary.com/us/dictionary/english/troll> [<https://perma.cc/5QVT-9262>].

⁵⁹ Professor Jason Stanley makes a similar argument about certain types of political speech. See Jason Stanley, Opinion, *The Ways of Silencing*, N.Y. TIMES (June 25, 2011, 9:12 AM), <https://archive.nytimes.com/opinionator.blogs.nytimes.com/2011/06/25/the-ways-of-silencing> [<https://perma.cc/7CWS-V8J8>]. According to Stanley, one purpose of conspiracy theories against former President Obama was to silence him by "undermin[ing] the public's trust in him, so that nothing he says can be taken at face value." *Id.*

⁶⁰ *Id.*

⁶¹ Steph Bechard, *Using Hidden Words to Avoid Instagram Spam Comments*, CRYSTAL MEDIA (May 16, 2022), <https://www.crystalmediaco.com/using-hidden-words-to-avoid-instagram-spam-comments> [<https://perma.cc/63WU-3PAQ>]; see also Anjali Kandey, *How to Prevent Spam Comments from Razing Your Brand Reputation*, STATUSBREW (Apr. 11, 2021), <https://statusbrew.com/insights/prevent-spam-from-banishing-brand-reputation> [<https://perma.cc/ZQ6V-R9LQ>].

Still more problematic are disinformation campaigns, which pose grave threats in virtual space: “[Social media] does not possess the filters and vetting systems of traditional news media to process what is true and what is false. Thus, the platforms enable false information to spread widely and quickly.”⁶² Given the current First Amendment doctrine’s general tolerance of false speech,⁶³ the speech act framework might equip courts with an alternative justification for regulating some types of disinformation by considering its disruptive effect on free speech. When fake news targets a specific individual, for example, it can undermine her credibility, disabling her from securing the minimal amount of trust necessary to perform basic speech acts.⁶⁴ In these cases, the problem of disinformation can be construed not as a conflict between liberty and truth, but as “a conflict between liberty and liberty.”⁶⁵

These examples may raise concerns about aggressive government regulation of speech. However, the analysis of actual disruption goes only to the prong of significant government interest and leaves intact the other prong of narrow tailoring, which requires that the government exhaust “easily available alternative modes of regulation.”⁶⁶ Indeed, *Garnier*’s analysis that the Trustees could have “delet[ed] only repetitive comments rather than blocking the Garniers entirely” would preserve the holding in favor of the Garniers, even had the court acknowledged a significant government interest in forum maintenance.⁶⁷

Today, social media and other cyber platforms are increasingly the dominant venues of public discourse. The characterization of these platforms as the “modern public square,”⁶⁸ when taken too literally, risks missing the unique attributes of online speech that render it vulnerable to new forms of disruption. Going forward, courts should be prepared to recalibrate the philosophical assumptions underlying speech so as to entertain a broader array of speech disruptions.

⁶² Goldenziel & Cheema, *supra* note 51, at 101 (footnote omitted). The risk and reach of disinformation have become more prominent in recent politics. *See id.*

⁶³ *See, e.g.,* United States v. Alvarez, 567 U.S. 709, 719 (2012) (refusing to exempt false statements from First Amendment protection).

⁶⁴ *See* Stanley, *supra* note 59; *see also* JASON STANLEY, HOW FASCISM WORKS: THE POLITICS OF US AND THEM 58 (2018) (describing the function of conspiracy theories as “rais[ing] general suspicion about the credibility and the decency of their targets”).

⁶⁵ Rae Langton, *Speech Acts and Unspeakable Acts*, 22 PHIL. & PUB. AFFS. 293, 299 (1993).

⁶⁶ Santa Monica Food Not Bombs v. City of Santa Monica, 450 F.3d 1022, 1041 (9th Cir. 2006).

⁶⁷ *Garnier*, 41 F.4th at 1182.

⁶⁸ Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017).