

Harvard Journal of Law & Technology
Volume 37, Digest Fall 2023

**ASSESSING ONLINE TRUE THREATS AND THEIR IMPACTS:
THE NEW STANDARD OF *COUNTERMAN V. COLORADO***

*Jacob Kovacs-Goodman** & *Alice Schoenauer Sebag***

TABLE OF CONTENTS

I. INTRODUCTION.....	1
II. TRUE THREATS JURISPRUDENCE.....	3
III. A NEW RECKLESSNESS STANDARD	5
A. <i>Oral Argument</i>	6
B. <i>The Counterman Opinions</i>	7
IV. ONLINE THREATS: NEW NORMS FOR A NEW CONTEXT?	9
A. <i>The Offline Impacts of Online Hate</i>	10
B. <i>Tailoring Contextual Analyses to the Online Setting</i>	13
V. A RECKLESSNESS STANDARD IN PRACTICE.....	15
VI. CONCLUSION	18

I. INTRODUCTION

In *Counterman v. Colorado*,¹ the United States Supreme Court held that the minimum mental state (mens rea) required to trigger the “true threats” exception to First Amendment speech protections is recklessness.² Though the majority opinion never mentioned the Internet, the decision will shape the way individuals interact online.

A Colorado court convicted Counterman of stalking for sending thousands of unsolicited Facebook messages to the complaining

* Jacob Kovacs-Goodman, J.D., is a Redstone Public Service Venture Fellow at Harvard Law School. He is a graduate of Harvard Law School.

** Alice Schoenauer Sebag, Ph.D., is a member of the technical staff in ML Safety at Cohere, Inc. She was previously a machine learning engineer in Trust & Safety at Twitter (now X), and chief data scientist at the General Inspectorate of Finance (French Ministry for the Economy and Finance). She completed her Ph.D. at Mines Paris PSL (France) after graduating from the École Polytechnique and University Paris I – Pantheon – Sorbonne (France). This work was conducted prior to joining Cohere.

The authors would like to thank Geoffrey Négier for his precious support and help proofreading and improving the Note, as well as the editors for their thoughtful comments and feedback.

1. 143 S. Ct. 2106 (2023).

2. *Id.* at 2111–12.

witness over the course of two years.³ Whenever she blocked his account, Counterman would make a new account and resume contact, including with provocative messages such as: “Five years on Facebook. Only a couple physical sightings,” and “[y]ou’re not being good for human relations. Die. Don’t need you.”⁴ Counterman appealed, challenging his conviction. On appeal, Counterman framed the issue as a violation of his constitutional right to freedom of speech, but the Colorado Court of Appeals affirmed Counterman’s conviction.⁵ The appellate court used an objective test of whether a reasonable person would have found the messages threatening to conclude that Counterman’s messages constituted true threats and were therefore exempt from First Amendment protection.⁶ The Supreme Court granted review to decide whether this objective test for true threats was permissible, or whether the test for true threats must show that the speaker subjectively intended to threaten.⁷

Lower courts have been split regarding the mental state sufficient for a true threat, with some using an objective test and others requiring subjective intent tied to the threat.⁸ This Note examines the legal context and potential impact of the *Counterman* decision, with emphasis on the specific qualities of online speech. In Part II, we review true threats jurisprudence at both the federal and state levels. Part III details how, during oral argument and in the *Counterman* decision itself, Justices probed a recklessness standard, breaking with the two mainstream standards courts have used, and conflating two distinct crimes in the process. Part IV analyzes threatening and hateful speech in the online environment. Part V considers how a recklessness standard, which allows for conviction when a speaker consciously disregarded a substantial risk that the content of their speech was threatening, compares favorably against the two previous prevailing

3. See *id.* at 2112–13; see also Paul Cassell, *Intervening to Stop “True Threats” from Delusional Stalkers and Devious Stalkers*, VOLOKH CONSPIRACY (Apr. 19, 2023), <https://reason.com/volokh/2023/04/19/intervening-to-stop-true-threats-from-delusional-stalkers-and-devious-stalkers/> [https://perma.cc/VBN4-S4Q8].

4. People v. Counterman, 497 P.3d 1039, 1044 (Colo. App. 2021).

5. *Id.*

6. “In the absence of additional guidance from the U.S. Supreme Court, we decline today to say that a speaker’s subjective intent to threaten is necessary for a statement to constitute a true threat for First Amendment purposes.” *Id.* (citing People in Int. of R.D., 464 P.3d 717 (Colo. 2020)). The Colorado Supreme Court denied certiorari in 2022. *Counterman v. People*, No. 17CA1465, 2022 WL 1086644 (Colo. Apr. 11, 2022); see also Dan Schweitzer, *Supreme Court Report: Counterman v. Colorado*, 22-138, NAT’L ASS’N ATT’YS GEN. (Jan. 30, 2023), <https://www.naag.org/attorney-general-journal/supreme-court-report-counterman-v-colorado-22-138/> [https://perma.cc/2AWL-T7SJ].

7. *Counterman v. Colorado*, 143 S. Ct. 2106, 2116 (2023).

8. *Id.* at 2113 (“Courts are divided about (1) whether the First Amendment requires proof of a defendant’s subjective mindset in true-threats cases, and (2) if so, what *mens rea* standard is sufficient. We therefore granted certiorari.”).

tests. The Note concludes by sketching a multistakeholder path towards a healthier online “marketplace of ideas.”

II. TRUE THREATS JURISPRUDENCE

The First Amendment guarantees that “Congress shall make no law . . . abridging the freedom of speech.”⁹ The Supreme Court has articulated several exceptions for categories of speech that do not enjoy First Amendment protection.¹⁰ One such category is “true threats.”¹¹ The concept is best introduced through a discussion of how courts analyze potential true threats statements and the evolution of the approach.

Identifying true threats involves a contextual analysis. In *Watts v. United States*,¹² a young Vietnam War protestor proclaimed to a small crowd, “If they ever make me carry a rifle the first man I want to get in my sights is [President] L. B. J.”¹³ The Court held that the protestor had not issued a true threat against the President due to the context of the speech and its conditional nature.¹⁴ The Court’s contextual analysis weighed the laughter of the petitioner and his audience in the moment.¹⁵ The Court in *Watts* withheld making a determination about the “willfulness” of the threats it considered.¹⁶

In 2003, *Virginia v. Black*¹⁷ defined true threats as “statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals.”¹⁸ In *Black*, the Court considered two consolidated cases from Virginia: one in which the Ku Klux Klan burned a cross at a rally and another in which two men burned a cross

9. U.S. CONST. amend. I.

10. See, e.g., *Chaplinsky v. New Hampshire*, 315 U.S. 568, 571–72 (1942) (“There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem.”).

11. These categories include incitement, obscenity, defamation, fighting words, and commercial speech. See ERWIN CHEMERINSKY, THE FIRST AMENDMENT (2d ed. 2021) (Chapter 3).

12. 394 U.S. 705 (1969).

13. *Id.* at 706.

14. *Id.*

15. *See id.* at 707 (highlighting that counsel “stressed the fact that petitioner’s statement was made during a political debate, that it was expressly made conditional upon an event-induction into the Armed Forces—which petitioner vowed would never occur, and that both petitioner and the crowd laughed after the statement was made.”).

16. *Id.* at 707–08 (concluding, only with respect to the statute at issue, that “the Court of Appeals differed over whether or not the ‘willfulness’ requirement of the statute implied that a defendant must have intended to carry out his ‘threat.’ Some early cases found the willfulness requirement met if the speaker voluntarily uttered the charged words with ‘an apparent determination to carry them into execution’ Perhaps this interpretation is correct, although we have grave doubts about it.” (internal citations omitted)).

17. 538 U.S. 343 (2003).

18. *Id.* at 359.

in their Black neighbor's yard.¹⁹ A Virginia statute that prohibited cross burning specified that the act of burning crosses was itself "prima facie evidence of an intent to intimidate a person or group."²⁰ However, Justice O'Connor, writing for the plurality, found the statute unconstitutionally overbroad given the "prima facie" language it employed to criminalize cross burning.²¹ The Court held that because people might burn a cross without an intent to intimidate, for instance as a political statement, the statute impermissibly chilled protected speech.²² The *Black* plurality nevertheless would have considered cross burning with the intent to threaten as a true threat not protected by the First Amendment.²³ By contrast, Justice Thomas, writing in dissent, found "the majority err[ed] in imputing an expressive component to the activity in question."²⁴

Lower courts began to split over what level of intent was required under *Black*, with the Seventh, Ninth, and Tenth Circuits mandating that the threatener subjectively intend to threaten, while the remaining courts adopted an objective standard.²⁵ In 2015, commentators anticipated the Supreme Court would use *Elonis v. United States*²⁶ both to clarify the mens rea requirement for true threats and to evaluate First Amendment exceptions for the digital age.²⁷

In *Elonis*, the petitioner, after his wife left him, changed his name on Facebook "to a rap-style nom de plume, 'Tone Dougie,'" and then posted about his wife using "graphically violent language and imagery . . . interspersed with disclaimers that the lyrics were 'fictitious.'"²⁸ Chief Justice Roberts's opinion for the Court was

19. *Id.* at 348–50.

20. *Id.* at 347.

21. *Id.* at 363–64 (finding statute unconstitutional where it stated, "burning of a cross shall be prima facie evidence of an intent to intimidate a person or group of persons" (quoting VA. CODE ANN. § 18.2-423 (1996))).

22. *Id.* at 365–66 ("[S]ometimes the cross burning is a statement of ideology, a symbol of group solidarity. It is a ritual used at Klan gatherings, and it is used to represent the Klan itself.").

23. See *id.* at 362–63 ("The First Amendment permits Virginia to outlaw cross burnings done with the intent to intimidate because burning a cross is particularly virulent form of intimidation.") (citing R.A.V. v. St. Paul, 505 U.S. 377, 391 (1992)).

24. *Id.* at 388 (Thomas, J., dissenting); see also JEREMY WALDRON, THE HARM IN HATE SPEECH 34–39 (2012) (describing cross burning, and more generally, the most harmful manifestations of hate speech as problematic because (1) they are not speech *per se*, but rather tangible and "permanent or semipermanent part[s] of the visible environment in which our lives, and the lives of members of vulnerable minorities, have to be lived" and (2) they are incitement to hatred).

25. Jing Xun Quek, *Elonis v. United States: The Next Twelve Years*, 31 BERKELEY TECH. L.J. 1109, 1110 (2016).

26. 575 U.S. 723 (2015).

27. See, e.g., P. Brooks Fuller, *Evaluating Intent in True Threats Cases: The Importance of Context in Analyzing Threatening Internet Messages*, 37 HASTINGS COMM'CN & ENT. L.J. 37 (2015) (surveying scholarly previews of *Elonis*).

28. *Elonis*, 575 U.S. at 727.

narrow, reaching neither the precise mens rea standard nor the online context of the threats. Acknowledging that the Court has “long been reluctant to infer . . . a negligence standard,” Roberts “decline[d] to determine “whether recklessness suffices for liability.”²⁹ The Chief Justice held only that, as far as the federal threats statute in question was concerned, a mens rea beyond negligence is required.³⁰

In opinions that foreshadowed *Counterman*, Justice Alito, in concurrence, and Justice Thomas, in dissent, both advocated for the Court to define recklessness or general intent as sufficient for true threats.³¹ Since the Court narrowed the *Elonis* ruling solely to interpreting the federal statute at issue in that case, lower courts continued to disagree over the intent requirement for true threats.³²

III. A NEW RECKLESSNESS STANDARD

Once again, in *Counterman*, the Court homed in on the question of whether the government must prove specific intent in a true threats case. This Part provides additional context from oral argument, during which the Justices weighed the viability of a recklessness test for true threats, introducing a standard not typically encountered in First Amendment jurisprudence.³³ It then analyzes the competing perspectives that emerged in the *Counterman* opinions.

Counterman was a bad case for a true threats ruling. Professors Evelyn Douek and Genevieve Lakier argued that stalking is a separate crime from true threats, emphasizing *Counterman* was convicted of the former and not the latter, and raised the alarm in advance of the decision that the Supreme Court would significantly weaken stalking laws if it introduced true threats analysis to the stalking context.³⁴ We agree with

29. *Id.* at 741.

30. *See generally* Note, *Elonis v. United States*, 129 HARV. L. REV. 331 (2015).

31. *Elonis*, 575 U.S. at 745 (Alito, J., concurring); *id.* at 750 (Thomas, J., dissenting).

32. Prior to the *Counterman* decision, recklessness was the least common mens rea standard for true threats. Only three states have criminal threats statutes with a specified mens rea of recklessness: Kansas, Georgia, and Connecticut. *See* KAN. STAT. ANN. §21-5415(a)(1) (2018); GA. CODE ANN. § 16-11-37(a) (2010); CONN. GEN. STAT. § 53a-61aa (2012). In all three states, the state supreme courts recently reviewed true threats cases under their respective state statutes. The Connecticut Supreme Court ruled an objective standard is permissible. *State v. Taupier*, 193 A.3d 1, 8 (Conn. 2018). The Georgia Supreme Court rejected the argument that a recklessness standard is comparable to negligence because recklessness sufficiently examines a defendant’s state of mind. *Major v. State*, 800 S.E.2d 348, 349–350 (Ga. 2017). The Kansas Supreme Court found the recklessness standard unconstitutional. *State v. Boettger*, 450 P.3d 805, 807 (Kan. 2019). The U.S. Supreme Court declined to grant certiorari, but Justice Thomas dissented, wishing to resolve the split between Kansas, Connecticut, and Georgia on the grounds that a historical analysis demonstrates that no subjective intent should be required in true threats cases. *See generally* *Kansas v. Boettger*, 140 S. Ct. 1956 (2020) (Thomas, J., dissenting from denial of petition for writ of certiorari).

33. Courts either use subjective intent or an objective standard, as detailed *supra* Part II.

34. Evelyn Douek & Genevieve Lakier, *The Supreme Court Seems Poised to Decide an Imaginary Case*, ATLANTIC (Apr. 26, 2023),

this assessment, and had hoped that the Court would explicitly exclude stalking crimes from the true threats ambit and instead treat stalking as conduct that only sometimes involves speech.³⁵

The subjective intent of a stalker, which is often untethered from reality, can be lethal; more than half of female homicide victims reported being stalked before they were murdered by their stalker.³⁶ Professors Douek and Lakier note that Justice Gorsuch, during oral arguments, wondered aloud why Colorado pursued a true threats conviction and not a stalking conviction, when in fact the reverse was the case; Colorado had charged Counterman with stalking.³⁷ This shows that, at least during oral argument, the Court was conflating the two crimes in a way that could prove exceedingly harmful. Although Justice Sotomayor argued in concurrence that stalking should remain distinct from true threats, the *Counterman* majority did not draw a precise line between the two crimes.³⁸

A. Oral Argument

The oral arguments offered clues that the Court was eyeing a recklessness standard even though neither party advocated for it in briefing. The petitioner only advocated for an intent requirement, and Colorado argued for an objective standard.³⁹ Nevertheless, Justices Kavanaugh, Alito, and Thomas all asked questions about recklessness during oral argument.⁴⁰ Justice Kavanaugh voiced his approval for a recklessness standard that would leave “plenty of room . . . to make sure threats are captured.”⁴¹ Justices Alito and Thomas echoed

<https://www.theatlantic.com/ideas/archive/2023/04/supreme-court-social-media-stalking-case-colorado/673849/> [<https://perma.cc/NH5R-CDKY>]; *see* Brief of First Amendment Scholars as Amici Curiae in Support of Respondent at 2, *Counterman v. Colorado*, 143 S. Ct. 2106 (2023) (No. 22-138), 2023 WL 2816039 (describing how threats statutes “forbid speech because of its communicative content alone” while “laws that prohibit stalking . . . prohibit a pattern of conduct, [only] sometimes involving speech[.]”).

35. *See* Brief of First Amendment Scholars, *supra* note 34, at 2.

36. Judith M. McFarlane et al., *Stalking and Intimate Partner Femicide*, 3 HOMICIDE STUD. 300, 311–16 (1999).

37. Douek & Lakier, *supra* note 34.

38. 143 S. Ct. 2106, 2120–21 (2023) (Sotomayor, J., concurring in part and concurring in the judgment) (highlighting that, with stalking, “[r]epeatedly forcing intrusive communications directly into the personal life of ‘an unwilling recipient’ . . . enjoys less protection”) (citing *Rowan v. Post Office Dept.*, 397 U.S. 728, 738 (1970)).

39. *See* Transcript of Oral Argument at 21, *Counterman v. Colorado*, No. 22-138 (U.S. argued Apr. 19, 2023).

40. *Id.* at 77, 15, 84. By contrast, Justice Sotomayor seemed likely to require subjective intent. *See* *Perez v. Florida*, 580 U.S. 1187, 1189 (2017) (Sotomayor, J., concurring in denial of petition for writ of certiorari) (“Together, *Watts* and *Black* make clear that to sustain a threat conviction without encroaching upon the First Amendment, States must prove more than the mere utterance of threatening words—some level of *intent* is required.”).

41. Transcript of Oral Argument, *supra* note 39, at 77.

sentiments from their opinions in *Elonis*, in which they advocated for a mens rea standard below specific intent.⁴²

Several Justices also shunned Colorado's contextual objective test. Justices Gorsuch, Thomas, and Barrett all questioned the reliability of what a "reasonable person" might find threatening.⁴³ Though the Justices referred to a divide between subjective and objective tests, it is worth noting that lower courts are further divided in their approach to the objective standard, with some using a reasonable listener standard and others using a reasonable speaker standard.⁴⁴

Amici overwhelmingly focused on whether a subjective intent standard was superior to an objective standard. For instance, the American Civil Liberties Union argued that, without assessing a speaker's subjective intent, protections for art and satire that incorporate threatening language would be on the chopping block.⁴⁵ They further argued that the problem would compound for online speech, because the initial speaker has little control over the audience and reach of their speech.⁴⁶ By contrast, a cohort of First Amendment scholars argued that the First Amendment does not protect objectively terrorizing speech; they stressed that other categorical exceptions like obscenity and fighting words do not require subjective intent, and that threats fail to contribute to the marketplace of ideas.⁴⁷ They also emphasized effects in the social media context, where "online abuse can have a totalizing and devastating impact upon victims, causing chilling of their own speech, sharing, and engagement online."⁴⁸

B. The Counterman Opinions

Justice Kagan authored the sparse 7–2 opinion for the Court. Justice Sotomayor wrote a concurring opinion in which Justice Gorsuch

42. See *Elonis*, 575 U.S. at 746 (Alito, J., concurring) ("I would hold that a defendant may be convicted... if he or she consciously disregards the risk that the communication transmitted will be interpreted as a true threat"); *id.* at 753 (Thomas, J., dissenting) (advocating for "[o]ur default rule in favor of general intent").

43. Justice Barrett suggested "nowadays people would be more sensitive" if a law professor were to read aloud lynching threats while teaching about the realities of the Jim Crow South, to the point that a "reasonable" law student could "fear for their safety because they don't understand it." *Id.* at 80, 82. This hypothetical illustrated the Justices' worry that an "eggshell audience" could enable a true threats conviction, regardless of the law professor's intent. *Id.* at 81.

44. See Kenneth L. Karst, *Threats and Meanings: How the Facts Govern First Amendment Doctrine*, 58 STAN. L. REV. 1337, 1348 (2006).

45. Brief for American Civil Liberties Union et al. as Amici Curiae Supporting Petitioner at 27, *Counterman v. Colorado*, 143 S. Ct. 2106 (2023) (No. 22-138), 2023 WL 2431342.

46. *Id.* at 28.

47. Brief of First Amendment Scholars as Amici Curiae in Support of Respondent at 6–7, 18, *Counterman v. Colorado*, 143 S. Ct. 2106 (2023) (No. 22-138), 2023 WL 2815329.

48. *Id.* (quoting Jonathon W. Penney, *Understanding Chilling Effects*, 106 MINN. L. REV. 1451, 1478 (2022)) (internal quotation marks omitted).

joined in part, and Justices Thomas and Barrett each authored dissenting opinions. The majority first held that some level of subjective understanding beyond an objective standard is required to allow sufficient “breathing space” and prevent a chilling effect on speech.⁴⁹ In discussing the insufficiency of an objective standard, the Court drew support from the recklessness standard in *New York Times Co. v. Sullivan*,⁵⁰ the Court’s landmark defamation case that established public figures could only recover when the defendant spoke with actual malice or recklessness.⁵¹

The Court then held that the correct minimum standard for true threats is recklessness. In particular, the Court adopted the recklessness formulation in Justice Alito’s *Elonis* concurrence: a true threat is established when “a speaker is aware ‘that others could regard his statements as’ threatening violence and ‘delivers them anyway.’”⁵² The majority opinion compared true threats with two other categories of unprotected speech: defamation and incitement.⁵³ The Court reiterated that recklessness has struck the right balance in defamation cases, and contrasted this with the necessity of a higher mens rea for incitement, since “incitement to disorder is commonly a hair’s-breadth away from political ‘advocacy.’”⁵⁴ Interestingly, the majority never mentioned the Internet or online communications.

Justice Sotomayor filed a concurring opinion, writing that the Court should not have reached the question of whether recklessness is a sufficient mens rea for true threats, but only needed to establish that some level of subjective mens rea is required.⁵⁵ She would have disentangled the true threats issue altogether, and resolved the case on stalking grounds.⁵⁶ Though she did not want the Court to establish a recklessness standard, Justice Sotomayor pragmatically advised lower courts on how they should administer the new standard. She counseled against using Justice Alito’s formulation for recklessness from *Elonis*, in favor of *Sullivan*’s formulation that requires “a high degree of awareness that a statement was probably threatening.”⁵⁷ She worried about the practical ramifications of the discrepancy between a recklessness standard for true threats and an intent standard for incitement. Namely, she stressed that incitement to imminent lawless

49. *Counterman v. Colorado*, 143 S. Ct. 2106, 2118 (2023).

50. 376 U.S. 254 (1964).

51. *See id.* at 280.

52. *Counterman*, 143 S. Ct. at 2118 (citing *Elonis v. United States*, 575 U.S. 723, 746 (2015) (Alito, J., concurring in part and dissenting in part)).

53. *Id.* at 2114; *see CHEMERINSKY, supra* note 11.

54. *Counterman*, 143 S. Ct. at 2118 (referring to a desire to avoid repeating past mistakes that jeopardized the protection of “mere advocacy of force or lawbreaking from legal sanction”).

55. *Id.* at 2120 (Sotomayor, J., concurring in part and concurring in the judgment).

56. *Id.* at 2121 (Sotomayor, J., concurring in part and concurring in the judgment).

57. *Id.* at 2131 (Sotomayor, J., concurring in part and concurring in the judgment).

activity will inherently threaten somebody, and so prosecutors could have an easier time convicting on true threats charges and will charge incitement cases as true threats.⁵⁸ While the majority never mentions the Internet, Justice Sotomayor noted that “[d]ifferent corners of the [I]nternet have considerably different norms around appropriate speech.”⁵⁹ The next Part will shed further light on how courts might approach these different Internet norms when adopting subjective mens rea standards to evaluate online speech.

Justice Barrett’s dissenting opinion advocated for an objective standard. In oral arguments, she had probed this stance from the opposing standpoint, suggesting that “nowadays people would be more sensitive” and such “eggshell audiences” would make it easier to obtain a true threats conviction.⁶⁰ Her dissent argued that nearly every category of unprotected speech should be restricted using an objective standard.⁶¹ Justice Thomas, despite his engagement with true threats in *Virginia v. Black*, authored a dissent that focused solely on his distaste for *New York Times Co. v. Sullivan*.⁶²

Overall, only Justice Sotomayor’s opinion considered that the conduct at issue in *Counterman* took place exclusively online — more specifically, on social media. As discussed in Part IV, the online environment entails precise and legally under-explored consequences for human interactions and behaviors.

IV. ONLINE THREATS: NEW NORMS FOR A NEW CONTEXT?

In the new recklessness regime, the contextual analysis of *Watts* and *Black* will remain indispensable in true threats cases.⁶³ As Justice Barrett noted in her dissent, “[w]hen context is ignored, true threats cannot be reliably distinguished from protected speech. The reverse also holds: When context is properly considered, constitutional concerns abate.”⁶⁴ This Part describes the characteristics of online speech on social media, focusing specifically on hateful and threatening

58. See *id.* at 2129 (Sotomayor, J., concurring in part and concurring in the judgment) (“prosecutors could now simply charge such offenses as true threats. This is particularly worrisome because the standard for recklessness decreases the lower the “social utility” of the conduct”).

59. *Id.* at 2122 (Sotomayor, J., concurring in part and concurring in the judgment).

60. See Transcript of Oral Argument, *supra* note 39, at 82, 80.

61. *Counterman*, 143 S. Ct. at 2134–35 (Barrett, J., dissenting) (considering “fighting words,” “false, deceptive, or misleading” commercial speech, and “obscenity”).

62. See *id.* at 2132–33 (Thomas, J., dissenting) (citing *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 280 (1964)).

63. See *Watts v. United States*, 394 U.S. 705, 708 (1969); *Virginia v. Black*, 538 U.S. 343, 367 (2003). In *Counterman*, the Court did not perform a contextual analysis itself, remanding the case to Colorado.

64. 143 S. Ct. at 2138 (Barrett, J., dissenting).

expressions. It also introduces existing research about the link between online and offline hate.

A. The Offline Impacts of Online Hate

Online speech, particularly on social media, is characterized not only by the potential scale at which one's content can spread and the ease of accessing anyone (including via commenting on their content or sending them a message), but also by anonymity: diminished identifiability and absence of physical presence.⁶⁵ The combination of these characteristics with social identity dynamics results in online environments where individuals engage in self-stereotyping as group members, increasing conformity to group norms.⁶⁶ Similarly, users view others through the lens of their (voluntary or involuntary) group memberships, all while facing less accountability for their online actions thanks to their relative anonymity.⁶⁷ A final factor is the instantaneity of such speech.⁶⁸ In a nutshell, it is easier to say something thoughtlessly online, and the potential damage can be far wider than the same statement would have in an offline context.

There is burgeoning research linking online hateful content with offline harm. Yet these findings are absent from the true threats discussion. For example, Professors Müller and Schwarz identify a causal link between exposure to anti-refugee sentiment on social media and offline anti-refugee crimes in Germany between January 2015 and February 2017 on a weekly basis.⁶⁹

65. See TEO KEIPI, MATTI NÄSI, ATTE OKSANEN & PEKKA RÄSÄNEN, ONLINE HATE AND HARMFUL CONTENT: CROSS-NATIONAL PERSPECTIVES 2 (2016).

66. *Id.* at 32. Describing the development of the social identity model of deindividuation effects ("SIDE") by Lea and Spears as "an effort . . . to bring both positive and negative effects of anonymity into a common framework. This was carried out through experiments involving anonymity and computer-mediated communication in a laboratory setting. Participants' identity group membership was reinforced and interaction was studied in conditions of anonymity and identifiability. Results showed that anonymity combined with group membership increased conformity to group norms, thus establishing group self-awareness." *Id.*

67. *Id.* at 33 ("The SIDE model linked to the aspects of social identity described earlier helps to explain why the comment section below an otherwise benign video or news article can turn into a warzone of politics, race or anything else. Here, social identity groups clash on the basis of stereotyping, seeing the "other" as an oversimplified representative of an opposing group."); *see also id.* at 33 fig.2.1.

68. See Alexander Brown, *What is so Special About Online (As Compared to Offline) Hate Speech?*, 18 ETHNICITIES 297, 304–06 (2018).

69. See Karsten Müller & Carlo Schwarz, *Fanning the Flames of Hate: Social Media and Hate Crime*, 19 J. EUROPEAN ECON. ASS'N 4, 2131–2167 (2021). This study was completed prior to the implementation of "NetzDG," a German law that came into effect in October 2017. *See* Netzwerkdurchsetzungsgesetz [Network Enforcement Act], Sep. 1, 2017, BUNDESGESETZBLATT, Teil I [BGBL. I] (Ger.) It requires large platforms to, *inter alia*, remove obviously illegal content within 24 hours, or face a fine. *Id.* §2.

Müller and Schwarz introduce a framework to assess the effect that exposure to online hate on social media had on local hate crime rates, independently of preexisting offline hateful sentiment. To test causality, Müller and Schwarz leverage local Internet outages and national Facebook outages during the period. The idea is: if online anti-refugee sentiment triggers localized offline events against these populations according to the local prevalence of followers of far-right political parties, a local Internet outage — difficulties or impossibility of accessing the Internet — would significantly diminish instances of offline hate.⁷⁰ Indeed, it is unlikely that anti-refugee sentiment itself would be affected by transitory Internet or Facebook issues. Furthermore, the framework addresses concerns about potential reverse causality: If offline hate crimes exclusively drove online hate expression, Internet outages should show no impact on offline outcomes.

Online anti-refugee sentiment at the national level, for a given week, is measured by the number of posts containing the term “Flüchtlinge” (refugee) on the Facebook group page for the Alternative für Deutschland (“AfD”) party.⁷¹ The AfD is the most popular far-right political party in Germany.⁷² The local impact of this online activity is measured using the ratio of AfD followers publicly indicating a given municipality on their Facebook profile, and the population of this municipality.⁷³ Anti-refugee crime data comes from two sources: the Amadeu Antonio Foundation, and Pro Asyl, a pro asylum NGO.⁷⁴ Finally, instances of local Internet outages were collected from user reports as listed on Heise Online, a leading information technology news outlets in German-speaking countries.⁷⁵ From this list, only the

70. See Müller & Schwarz, *supra* note 69, at 2133, 2147–2155. “Significantly” in a quantitative setting usually means that there is less than a small probability (often 1% or below) that the effect would have been observed by chance. See NAT’L INST. STDS. & TECH., *What Are Statistical Tests?*, in ENGINEERING STATISTICS HANDBOOK § 7.1.3. (2012), <https://www.itl.nist.gov/div898/handbook/prc/section1/prc13.htm> [https://perma.cc/336X-F2L4].

71. The authors of the study confirmed by a manual inspection of such content that it overwhelmingly expressed anti-refugee sentiment. See *id.* at 2137–2139; *id.* app. at 3 fig.A.1.

72. *Id.* at 2137.

73. *Id.* at 2139–2140.

74. *Id.* at 2136–2137.

75. See *Jährliches t3n-Ranking der Tech-Medien in Deutschland / Chip bleibt Nr. Eins, PC Welt erstmals in den Top Fünf* [Annual t3n Ranking of the Tech Media in Germany / Chip Remains No. One, PC World in the Top Five for the First Time], FINANZNACHRICHTEN (Dec. 13, 2016), <https://www.finanznachrichten.de/nachrichten-2016-12/39440234-jaehrliches-t3n-ranking-der-tech-medien-in-deutschland-chip-bleibt-nr-eins-pc-welt-erstmals-in-den-top-fuenf-007.htm> [https://perma.cc/QZG3-RGJ7]. For examples of Internet outages and their coverage in German newspapers or specialized websites, see Müller & Schwarz, *supra* note 69, app. at 13–15.

313 outages that lasted at least 24 hours and affected a significant part of the population are used.⁷⁶

Müller and Schwarz find that “internet outages appear to mitigate the entire effect of social media,” in the sense that, “for a given level of anti-refugee sentiment [at the national level], there are fewer attacks in municipalities with high [AfD] Facebook usage during an internet outage than in municipalities with low [AfD] Facebook usage *without* an outage.”⁷⁷

These results seem to be specific to neither the online platform nor the geography where the study was conducted.⁷⁸ The study points to a significant link between online exposure to hateful sentiment and the offline actualization of such sentiments. It is likely that multiple mechanisms are at play, such as local coordination, as Müller and Schwarz found,⁷⁹ or norm shifting around what people perceive as socially acceptable behaviors. Indeed, another study by Karel, Linke, Holland, and Hendrickson points to the role of social media in shaping users’ views of behavioral norms.⁸⁰ Their work studies the link between local hard-right civil unrest and prior video activity on Parler.⁸¹ The authors measure the extent to which the content produced by the platform’s “elites” reflects prior user content for a given location and

76. Müller & Schwarz, *supra* note 69, at 2140–41. The number of user reports in a given week and municipality is scaled by the local population. Only outages with a scaled number of reports in the top quartile of reported outages are kept. The duration of outages is the minimum between that reported by users and 3 weeks.

77. *Id.* at 2152. This effect is not (or not *only*) caused by general internet access: the interaction between internet usage indicators and outages is either not significant or show an inverse relationship. *Id.* at 2152–53. The effect is also significant when replacing local outages with national Facebook outages. *Id.* at 2153–55.

78. The two following works investigate the impact of Twitter usage in the US, and VKontakte penetration in Russia, respectively, finding consistent results: Karsten Müller & Carlo Schwarz, *From Hashtag to Hate Crime: Twitter and Antiminority Sentiment*, 15 AM. EC. J.: APPLIED ECON. 270 (2023); Leonardo Bursztyn et al., *Social Media and Xenophobia: Evidence from Russia* (Nat'l Bureau of Econ. Rsch., Working Paper No. w26567, 2019).

79. See Müller & Schwarz, *supra* note 69, at 2163. Hand-coding the number of perpetrators for roughly one quarter of anti-refugee crimes, Müller and Schwarz find that the relationship between hate crimes and the interaction of the local ratio of AfD followers and national refugee posts is significant for crimes with 4 or more perpetrators, and not significant for crimes with 3 or fewer perpetrators. This is consistent with the coordination hypothesis.

80. See Daniel Karel, Andrew Linke, Edward Holland & Edward Hendrickson, “Born for a Storm”: Hard-Right Social Media and Civil Unrest, 88 AM. SOC. REV. 322, 333–35, 342 (2023).

81. Karel et. al. define “hard-right” as “not . . . an extreme of the political spectrum (e.g., “far-right” or “right-wing”); most content on HRSM [hard-right social media] is not extreme relative to the modern conservative movement. Instead, [the authors] mean that the majority of content is, first, socially and politically conservative and, second, more recalcitrant than other conservative perspectives. It is mainly the content of the contemporary conservatives who resist compromise with the political center or left. Additional features of the content arise from the relational element. Namely, hard-right content often glorifies being “banned” or “deplatformed” from mainstream social media—creating a social capital characteristic of HRSM.” *Id.* at 325 (internal citations omitted).

month.⁸² The study concludes that a significant link exists between this variable and subsequent local hard-right civil unrest in the United States in 2020.⁸³

B. Tailoring Contextual Analyses to the Online Setting

As illustrated above, the online context is unique; online threatening speech can impact many more people than the targeted victim and on a boundless geographic scale. Regardless of the mens rea standard they apply, courts have consistently undertaken a contextual analysis in true threats cases, as introduced in *Watts*.⁸⁴ In his *Elonis* dissent, Justice Alito noted that context cues must change for the Internet, since “lyrics in songs that are performed for an audience or sold in recorded form are unlikely to be interpreted as a real threat,” while “[s]tatements on social media that are pointedly directed at their victims, by contrast, are much more likely to be taken seriously.”⁸⁵ So how does one perform a contextual analysis in the online context? What are comparable online context clues to the conditional phrasing,⁸⁶ the particular communal setting, and the universally understood response of laughter — all supporting evidence in *Watts*?

82. “Elites” is close to the “colloquial idea of social media ‘influencers.’” *Id.* at 326 n.4. Platforms with strong hard-right activity often promote their elites: “For example, some display a special symbol alongside elites’ accounts, making visible their status, such as Gab’s ‘PRO account’ designation and Parler’s ‘gold badge’ icon, which was given to users like U.S. Representative Marjorie Taylor Greene; former government official, scholar, and pundit Hugh Hewitt; and Enrique Tarrio, the leader of the Proud Boys, a far-right group. Platforms can also prompt new users to follow elites at the sign-up stage (as Parler does) and algorithmically amplify their content on users’ feeds.” *Id.* at 326.

83. *Id.*; *see id.* at 340 tbl.4; *see also id.* § B.3 at 19–21 (Supp. 2023), https://journals.sagepub.com/doi/suppl/10.1177/00031224231156190/suppl_file/sj-pdf-1-asr-10.1177_00031224231156190.pdf.

84. *See, e.g.*, *Counterman v. Colorado*, 143 S. Ct. 2106, 2114 (2023) (“The ‘true’ in that term [true threats] distinguishes what is at issue from jests . . . when taken in context”) (citing *Watts v. United States*, 394 U.S. 705, 708 (1969)); *Elonis v. United States*, 575 U.S. 723, 747 (2015) (Alito, J., concurring in part and dissenting in part) (citing *Watts*, 394 U.S. at 708) (“But context matters.”); *United States v. Bly*, 510 F.3d 453, 459 (4th Cir. 2007) (citing *Watts*, 394 U.S. at 707–08) (“In so ruling, the Court looked to and relied upon several contextual factors Unlike in *Watts*, the Letter was not addressed to a public audience and, . . . , it was delivered privately to specific individuals.”); *United States v. Parr*, 545 F.3d 491, 497 (7th Cir. 2008) (citing *Watts*, 394 U.S. at 708) (“Parr stated repeatedly and consistently that he was going to bomb the federal building in Milwaukee, and nothing in the context required the jury to find that he was joking or using hyperbole.”); *Planned Parenthood of the Columbia/Willamette, Inc. v. Am. Coalition of Life Activists*, 290 F.3d 1058, 1085 (9th Cir. 2002) (“This is not political hyperbole. Nor is it merely ‘vituperative, abusive, and inexact.’ [citing *Watts*, 394 U.S. at 708]. In the context of the poster pattern, the posters were precise in their meaning to those in the relevant community of reproductive health service providers. They were a true threat.”).

85. *Elonis*, 575 U.S. at 747 (Alito, J., concurring in part and dissenting in part).

86. In *Watts*, the degree to which the threat underlying the statement was dependent on the likelihood of other events weighed against criminalizing the speech in question. 394 U.S. at 706 (“if they ever make me carry a rifle...”).

In its analysis in *People v. Counterman*, the Colorado Court of Appeals considered “multiple factors, including whether the communication was direct, public, or private; its platform, method, and characteristics of conveyance; and its impact on the intended or foreseeable recipient.”⁸⁷ While this approach accounted for the nature of the Internet as a platform for speech, judicial tests could benefit from even finer tailoring to online norms. A better approach might factor in whether the communication was sent as a direct message (“DM”), posted within a closed group or posted publicly, and whether the implicated individuals knew each other.⁸⁸ Indeed, one article has proposed splitting the contextual analysis emerging from *Watts* in two.⁸⁹ This approach would first consider the “literary context,” followed by the “social context.” For online speech, the “literary context” would amount to “looking beyond the few words that are allegedly threatening . . . to everything else the defendant said.”⁹⁰ The “social context” would correspond to the analysis of the communication medium and the involved parties’ relationship.⁹¹

A considerable challenge for contextual analysis in the online setting is the ease with which one can deprive content of its original context. Professor Lawrence Lessig, writing about offline meaning-making, defines “meaning” as “the product of a text in a particular context,” such that “we can change meaning by changing either text or context.”⁹² Changing the context is comparatively easy to do in the online setting as content is re-shared or even appropriated wholesale. This, in turn, will influence how the content is perceived. Individuals will interpret it in isolation according to their own background: “[F]or the purposes of threatening speech, a hyper-individualized reading of a message lacks the benefit of a true communal response.”⁹³

87. *People v. Counterman*, 497 P.3d 1039, 1046–49 (Colo. App. 2021) (citing *People in Int. of R.D.*, 464 P.3d 717, 721 (Colo. 2020)).

88. See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act). The Digital Services Act defines “dissemination to the public” as: “making information available, at the request of the recipient of the service who provided the information, to a potentially unlimited number of third parties.” Further, “where access to information requires registration or admittance to a group of recipients of the service, that information should be considered to be disseminated to the public only where recipients of the service seeking to access the information are automatically registered or admitted without a human decision or selection of whom to grant access.” Art. 3. Recital 14. It explicitly rules out “emails or private messaging services” and includes “public groups or open channels.”

89. Renee Griffin, *Searching for Truth in the First Amendment’s True Threat Doctrine*, 120 MICH. L. REV. 721, 750 (2022).

90. *Id.*

91. *Cf. id.*

92. Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943, 961 (1995).

93. Fuller, *supra* note 27, at 53.

A second challenge for contextual analysis in the online context is the necessity to understand the semiotics of particular platforms and even the vernacular of particular groups. For example, the Pepe the Frog meme might mean innocuous malaise when shared on MySpace, support for Nazism on the 4chan “/pol/” imageboard in 2020,⁹⁴ feeling good about a workout on 4chan’s “/fit/” message board in 2013,⁹⁵ and something else when a presidential candidate posts it on X (formerly Twitter).⁹⁶ Another example is identifying that, when a user posts on Threads — Meta’s new alternative to X — content is much more likely to reach strangers, since there is no dedicated feed for “following” accounts.⁹⁷ Only a deeper analysis will identify with confidence online settings similar to the “small discussion groups” context that was a crucial factor in *Watts*.⁹⁸ It will be essential for judicial tests to reference such factors to incentivize both parties to call for (and heed) the input of experts of online social dynamics and specific online communities.⁹⁹

V. A RECKLESSNESS STANDARD IN PRACTICE

Two main points support a general usage of the recklessness standard. First, objective tests allow for bias or prejudice to be injected into a reasonableness standard. What one deems “reasonable” is influenced by one’s experiences, which are themselves filtered through an individual’s characteristics, such as race and socioeconomic background. Many people in the United States — including in the aggregate, but also as jurors or judges — might decide that a “reasonable person” would find cross burning to not be a true threat,¹⁰⁰ but that rap lyrics are.¹⁰¹ As Justice Holmes noted, jurors “will

94. */pol/*, WIKIPEDIA, <https://en.wikipedia.org/wiki/pol/> (last visited July 11, 2023) [<https://perma.cc/ELP4-HSHY>].

95. See generally *4chan*, WIKIPEDIA, <https://en.wikipedia.org/wiki/4chan> (last visited Aug. 6, 2023) [<https://perma.cc/LZR4-4WWV>].

96. See FEELS GOOD MAN (Giant Pictures 2020) (documenting efforts of the creator of “Pepe the Frog” to salvage his character from being an icon of hatred and bigotry).

97. See Shannon Connellan, *How to Change Your Threads Feed to Accounts You Actually Follow*, MASHABLE (July 6, 2023), <https://mashable.com/article/thread-people-you-follow-feed> [<https://perma.cc/BAD5-SVUS>].

98. 394 U.S. 705, 706 (1969).

99. Scholars have spoken to the flow of hateful content between different online contexts. See, e.g., Savvas Zannettou, Joel Finkelstein, Barry Bradlyn & Jeremy Blackburn, *A Quantitative Approach to Understanding Online Antisemitism*, 14 PROC. INT’L AAAI CONF. ON WEB & SOC. MEDIA 786 (2020) (analyzing the use of the antisemitic “Happy Merchant” meme across different platforms for contextual meaning).

100. For instance, a majority on the Virginia Supreme Court, in its ruling that the U.S. Supreme Court eventually overturned in *Virginia v. Black*, worried penalties for “the act of burning a cross alone, with no evidence of intent to intimidate” would lead to “[s]elf-censorship.” See *Black v. Commonwealth*, 553 S.E.2d 738, 778 (Va. 2001).

101. See Clay Calvert, Emma Morehart & Sarah Papadelias, *Rap Music and the True Threats Quagmire: When Does One Man’s Lyric Become Another’s Crime?*, 38 COLUM. J.L.

introduce into their verdict a certain amount — a very large amount . . . of popular prejudice.”¹⁰² This dynamic complicates the fair trial of “true threat” crimes, which are often intertwined with hate speech and likely to be inflammatory. A recklessness test sidesteps the bias creep that accompanies objective tests.

Second, in the online setting, a recklessness standard could better account for the existence of a causal link (discussed in Part IV) between exposure to online hate speech and offline harm. Making threats online to a wide audience increases the likelihood of offline harm.¹⁰³ Speaking online thus carries more attendant risks than speaking offline. Since a recklessness standard measures conscious disregard of a substantial risk, and the special norms of online speech amplify the risks of harm on a greater scale, online threats may be more likely to be found reckless.

Conversely, online anonymity, informality, and pervasive reinforcement of users’ identity group memberships might increase the difficulty of proving that a speaker consciously disregarded risks of threatening speech and therefore acted recklessly.¹⁰⁴ In its recklessness discussion, the *Counterman* majority cited *Voisine v. United States*.¹⁰⁵ In *Voisine*, the Court held that “consciously disregarding” a risk is not an accident, but instead must be a “deliberate decision.”¹⁰⁶ An online culture of impulsive commenting might not encourage deliberate decision-making.¹⁰⁷ This could make it harder to prove a speaker’s conscious disregard for the risk created by *online* true threats.

The thoughtlessness of online speech contrasts with the increased risks at scale described above. These two opposite effects could balance each other out when conducting a recklessness appraisal; however, courts will be able to dial the magnitude of these effects up or down, based on how they measure recklessness in online spheres. The majority opinion in *Counterman* defined the test for conscious

& ARTS 1, 24 (2014) (urging courts to “(1) take into account the actual knowledge and background with rap music when the target or victim is a rap-literate target and (2) attribute some minimal understanding of rap’s conventions — the understanding that a hypothetical reasonable person would have — to the rap-ignorant target,” when discussing *United States v. Jeffries*, 692 F.3d 473 (6th Cir. 2012); *see also* Transcript of Oral Argument, *supra* note 39, at 82 (“to get back to rap music, a concert makes it unreasonable to view yourself as being threatened given what is going on”).

102. Oliver Wendell Holmes, *Law in Science and Science in Law*, 12 HARV. L. REV. 417, 459–60 (1899).

103. *See* discussion *supra* notes 69–82 (analyzing the work of Müller & Schwarz).

104. *See, e.g.*, KEIPI, NÄSI, OKSANEN & RÄSÄNEN, *supra* note 65, at 32–34 (Chapter 2).

105. 579 U.S. 686 (2016).

106. *Id.* at 694.

107. Psychologists have explored the “online disinhibition effect” for many years, where users “self-disclose or act out more frequently or intensely than they would in person.” *See, e.g.*, John Suler, *The Online Disinhibition Effect*, 7 CYBERPSYCHOLOGY & BEHAV. 321, 321 (2004) (examining “dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and minimization of authority”).

disregard as: “[A] speaker is aware ‘that others could regard his statements as’ threatening violence and ‘delivers them anyway.’”¹⁰⁸ But Justice Sotomayor’s concurrence introduced the concept of gradations of stringency within recklessness. She encouraged courts to implement a *Sullivan* recklessness test for online speech that requires a “high degree of awareness” by the threatener.¹⁰⁹ The very fact that some Internet communities revel in off-the-cuff threats could militate against a speaker possessing a “high degree of awareness” that his comments are threatening.¹¹⁰ If courts were to pursue the stricter approach advanced by Sotomayor, they might foster an Internet ecosystem in which norms of threatening speech self-reinforce the permissibility of such speech. That approach would essentially delegate to (private) platforms’ terms of service the protection of the marketplace of ideas from the socially corrosive effects of true threats.¹¹¹

Lastly, it is worth considering the downstream effects of the mental state requirement against the backdrop of the American criminal justice system. A recklessness standard could better preserve individuals’ rights compared to an intent requirement. Civil liberties advocates tend to align on an intent requirement, which makes speech-based prosecutions more difficult,¹¹² while groups more concerned about the impacts of hateful content advocate for an objective test to capture more hateful and threatening speech.¹¹³ If an objective test were implemented, it would be easier to prosecute true threats. Given that the current state of the criminal system disproportionately targets Black and Brown individuals for prosecution, a recklessness standard might thread the needle, capturing more threatening content than an intent standard, while also limiting the ease of speech-related convictions that could result under an objective standard. This consideration is not hypothetical — federal prosecutors filed true threats charges in ten cases that targeted protesters supporting racial justice in 2020.¹¹⁴

108. *Counterman v. Colorado*, 143 S. Ct. 2106, 2118 (2023) (citing *Elonis v. United States*, 575 U.S. 723, 746 (2015) (Alito, J., concurring in part and dissenting in part)).

109. *See id.* at 2130–31 (Sotomayor, J., concurring in part and dissenting in part).

110. *See id.* at 2122 (Sotomayor, J., concurring in part and concurring in the judgment) (“Different corners of the internet have considerably different norms around appropriate speech. Online communication can also lack many normal contextual clues, such as who is speaking, tone of voice, and expression.”).

111. For an analysis on the marketplace of ideas theory of free speech, its origins in Justice Holmes’s discussion of truth emerging through competition, and its critics, see ERIC BARENDE, *FREEDOM OF SPEECH* 11 (2007).

112. *See, e.g.* Brief for American Civil Liberties Union et al. as Amici Curiae Supporting Petitioner, *Counterman v. Colorado*, 143 S. Ct. 2106 (2023) (No. 22-138), 2023 WL 2431342; Brief of Amici Curiae Electronic Frontier Foundation & Student Press Law Center in Support of Petitioner, *Counterman*, 143 S. Ct. 2106 (No. 22-138), 2023 WL 2381310.

113. *See, e.g.* Brief of First Amendment Scholars, *supra* note 47.

114. MUDASSAR TOPPA & PRINCESS MASILUNGAN, *STRUGGLE FOR POWER: THE ONGOING PERSECUTION OF BLACK MOVEMENT BY THE U.S. GOVERNMENT* 28 (Ramzi Kassem & Naz Ahmad eds., 2021), <https://m4bl.org/wp-content/uploads/2021/08/Struggle->

As courts adjust to the new *Counterman* recklessness standard for true threats, specific analyses of the unique online context, as discussed above, will prove indispensable in attaining the adequate “breathing space” that animates the decision.

VI. CONCLUSION

Given the impact of online hate,¹¹⁵ there is a need to unite all stakeholders’ strengths to build and maintain an online environment where speech is free from the chilling effects of abuse, threats, and incitement to hatred. The *Counterman* decision shows that there is new judicial interest in moderating certain malicious speech. The true threats doctrine is a useful lens through which to probe the boundaries of civic discourse — especially regarding the line between when speech reinforces or is destructive to democratic deliberation.

Future scholarship may probe solutions to reduce the harmful impacts of online hate building off the analysis presented in this Note, including along the following dimensions:

1. Research and Educate: fund more research on the links between online hate and offline harm and inform young people of the consequences of hate speech on victims, for example, at the psychological level;¹¹⁶
2. Moderate: recognize and act on the fact that product and policy decisions on online platforms impact the probability that users will share hateful content;¹¹⁷ and
3. Regulate: pursue different levels of regulation, be it *ex ante* or *ex post*, at the user or at the platform level, or under the form of codes of conduct or law.¹¹⁸

For-Power-The-Ongoing-Persecution-of-Black-Movement-by-the-U.S.-Government.pdf [https://perma.cc/8QVV-SDFD]. It seems likely that local prosecutors in different parts of the country might operate in a similarly discriminatory fashion, with less scrutiny than that which exists at the federal level.

115. See *supra* Part IV.

116. See, e.g., KEIPI, NÄSI, OKSANEN & RÄSÄNEN, *supra* note 65, at 78–83 (Chapter 5).

117. There are multiple tools that are available in addition to the removal of hateful content to encourage users towards civil discourse, e.g. displaying community rules at the top of discussions. See, e.g., Nathan Matias, *Preventing Harassment and Increasing Group Participation Through Social Norms in 2,190 Online Science Discussions*, 116 PROC. NAT'L ACAD. SCIS. 9785, 9785–9789 (2019). On the subject of product features, it is interesting to note that in the *Counterman* case, “to increase her fan base, C.W. relied on a Facebook feature to automatically accept all friend requests.” Brief for Respondent at 4, *Counterman v. Colorado*, 143 S. Ct. 2106 (2023) (No. 22-138). Furthermore, “[s]everal times, C.W. blocked Counterman from messaging her. Each time, Counterman created a new profile (always as either Bill or Billy Counterman).” Brief for Petitioner at 6, *Counterman*, 143 S. Ct. 2106 (No. 22-138) (internal citations removed). A (simple) mitigation of such a risk would be to modify the “accept all friend requests” feature or add the option to demand manual confirmation for users who are similar to previously blocked users.

118. See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single market for Digital Services and amending Directive 2000/31/EC

Any proposed solutions should aim to create and maintain a healthy online marketplace of ideas where all citizens' speech is welcome.

(Digital Services Act). Since November 2022, the Digital Services Act provides a legal framework at the EU level for the regulation of providers of intermediary services in relation to societal risks, including their tackling of illegal hate speech. Codes of conduct are integrated in this framework as one of the tools for the application of the Regulation (arts. 45–47).