

---



---

COMMENT

---



---



---



---

CONTEXT, CONTENT, INTENT: SOCIAL MEDIA'S ROLE IN  
TRUE THREAT PROSECUTIONS

---



---

MEGAN R. MURPHY<sup>†</sup>

INTRODUCTION .....	734
I. BACKGROUND ON THE TRUE THREAT DOCTRINE .....	736
A. <i>Watts Through Black—Uncertain Protection</i> .....	736
B. <i>Elonis—Failure to Clarify</i> .....	737
C. <i>After Elonis—Opportunities to Elaborate</i> .....	739
II. SUBJECTIVE INTENT AS A NECESSARY COMPONENT OF FIRST AMENDMENT PROTECTION .....	741
III. HOW CONTEXT CAN CLARIFY INTENT IN ONLINE SPEECH .....	746
A. <i>Accounting for Speaker Identity</i> .....	748
B. <i>Accounting for Audience</i> .....	748
C. <i>Accounting for Platform-Based and Technological Constraints</i> .....	751
IV. THE NEED FOR SPECIFICITY IN PROSECUTING TRUE THREATS .....	756
A. <i>Elonis on Remand</i> .....	756
B. <i>Other Courts' Application of Subjective-Intent Standards</i> .....	759
V. ADMITTING MORE CONTEXTUAL EVIDENCE TO FACILITATE MORE ACCURATE DETERMINATIONS OF INTENT .....	763
CONCLUSION .....	766

---

<sup>†</sup> Executive Editor, Volume 168, *University of Pennsylvania Law Review*. J.D. Candidate, 2020, University of Pennsylvania Law School; B.A., 2015, Hamilton College. I am grateful to Professor Seth Kreimer for his guidance and feedback on early drafts. I also thank my classmates in Professor Kreimer's Spring 2019 seminar on the First Amendment in the Twenty-First Century for their thoughtful questions and the editors of the *University of Pennsylvania Law Review* for patient and thorough editing.

## INTRODUCTION

The doctrine that carves out “true threats” from First Amendment protection has been unclear, in its scope and operation, since the exception was first recognized more than half a century ago. This category of unprotected speech was recognized by the Supreme Court in 1961, in a decision that identified “true threats” as distinct from other, protected, potentially threatening speech, but did not articulate a standard which lower courts could apply to distinguish the two.<sup>1</sup> In the fifty years since, the Court has addressed the constitutional bounds of the true threat doctrine only once, clarifying that true threats require some showing of intent.<sup>2</sup> But even that more precise articulation of the standard<sup>3</sup> left unresolved whether the relevant intent standard is what the speaker of a threat actually, subjectively intended, or what a reasonable listener would have understood the speaker to intend. Rather than revisit the doctrine and resolve the intent-requirement question definitively, the Supreme Court has repeatedly declined opportunities to clarify, instead deciding threat cases on statutory, rather than constitutional, grounds,<sup>4</sup> and, more recently, declining to hear new threat cases as they arise.<sup>5</sup>

Even as the development of the true threat doctrine has stagnated, the nature of communication has changed dramatically with the advent of the Internet and the increasing prevalence of social media. These new platforms only heighten the stakes of inconsistencies in the true threat doctrine and highlight the need for a clear standard that tells courts whose intent governs whether a statement is a threat.<sup>6</sup> While in-person communication has never been immune to misunderstandings and erroneous interpretations, speech that is mediated by information networks and binary code amplifies the potential for a receiving user to interpret a statement as conveying something different than what the speaker intended to convey.<sup>7</sup> Where the potential for

<sup>1</sup> See *Watts v. United States*, 394 U.S. 705, 708 (1969); see also *infra* notes 12–16 and accompanying text (discussing *Watts* and the beginnings of the true threat exception in greater detail).

<sup>2</sup> *Virginia v. Black*, 538 U.S. 343, 359 (2003) (citing *Watts*, 394 U.S. at 708; *R.A.V. v. City of St. Paul*, 505 U.S. 377, 388 (1992)).

<sup>3</sup> See *id.* (“True threats’ encompass those 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.”); see also *infra* notes 17–23 and accompanying text.

<sup>4</sup> See *Elonis v. United States*, 135 S. Ct. 2001, 2012–13 (2015) (resolving the appeal on statutory, rather than First Amendment, grounds).

<sup>5</sup> See *infra* text accompanying notes 38–43 (discussing the Court’s denials of petitions for certiorari filed by defendants prosecuted under threat statutes after *Elonis*).

<sup>6</sup> See, e.g., Thomas “Tal” DeBauche, Note, *Bursting Bottles: Doubting the Objective-Only Approach to 18 U.S.C. § 875(c) in Light of United States v. Jeffries and the Norms of Online Social Networking*, 51 HOUS. L. REV. 981, 1016 (2014) (noting that application of the federal threat statute to online speech “expose[d] fractures” in the doctrine).

<sup>7</sup> See, e.g., Lyrissa Barnett Lidsky & Linda Riedemann Norbut, #I 🍻 U: *Considering the Context of Online Threats*, 106 CALIF. L. REV. 1885, 1908 (2018) (“[T]he gun emoji . . . looks like a space

misinterpretation is great, the principles underlying the First Amendment should require that a speaker's actual intent to threaten be proved as an element of conviction, to sufficiently protect individuals whose speech might otherwise be construed as threatening due to interpretive errors, idiosyncratic usage, identity-based stereotyping, or other misrepresentation.<sup>8</sup> But the same online speech that reveals these weaknesses in the true threat doctrine can also be relevant in resolving them, by providing a record of how online speech was created, interpreted, and modified by a speaker and her audience.<sup>9</sup> This evidence, where it is available, should be put to use by prosecutors and defense attorneys to enable factfinders to more accurately assess a defendant's intent.

This Comment proceeds as follows: Part I outlines the development of the true threat doctrine up to the current juncture. Part II briefly explains how a carefully applied subjective-intent requirement should operate, in conjunction with a requirement that the punishable speech be objectively threatening, to protect the First Amendment rights of speakers by maintaining a protective zone for individuals to engage in expressive—even hyperbolic—speech without fear of prosecution for speech that inadvertently misses the mark. Part III canvasses social science literature to suggest ways in which courts can more meaningfully—and more consistently—engage with the context surrounding utterances posted on social media, which can provide more evidentiarily grounded insight into a posting user's intent if he is prosecuted for an alleged threat. Part IV assesses the extent to which courts hearing Internet-mediated threat cases are already using contextual evidence to support findings of subjective intent, and identifies shortfalls in the reasoning of these decisions where more contextual evidence would facilitate more consistent, accurate, and speech-protective outcomes. Part V considers one existing proposal for procedural reform in light of current practices.

---

pistol on some platforms and like a revolver on others.”); *see also Full Emoji List, v12.0*, UNICODE, <http://www.unicode.org/emoji/charts/full-emoji-list.html> [<https://perma.cc/G4RH-59KV>] (last visited Dec. 18, 2019) (showing this and other variations in how the same emoji display across different platforms). How emojis display on different platforms can significantly affect how they are perceived by readers. *See* Jonathan Geneus, Note, *Emoji: The Caricatured Lawsuit*, 16 COLO. TECH. L.J. 431, 451 (2018) (“Objectively, an officer pointed his gun at a cartoon. Subjectively, however, a white officer pointed a gun at a black male cartoon.”).

<sup>8</sup> *See generally, e.g.*, Adam Dunbar & Charis E. Kubrin, *Imagining Violent Criminals: An Experimental Investigation of Music Stereotypes and Character Judgments*, 14 J. EXPERIMENTAL CRIMINOLOGY 507 (2018) (finding that study participants who read “violent” song lyrics were more likely to find them reflective of “bad character” if told that they were lyrics to a rap song, rather than a song of another genre, or that the writer was black, rather than of another race).

<sup>9</sup> *See, e.g.*, The Sedona Conference, *The Sedona Conference Primer on Social Media, Second Edition*, 20 SEDONA CONF. J. 1, 44-52 & nn.94, 100 (2019) (discussing various methods of accessing and preserving both “static images of social media data” and the underlying “metadata, logging data, or other information,” and noting that an individual post on Twitter or Facebook contains “over 20 specific metadata items”).

## I. BACKGROUND ON THE TRUE THREAT DOCTRINE

“True threats’ encompass those 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.”<sup>10</sup> Applying the label of “true threat” to an utterance places it outside the protection of the First Amendment.<sup>11</sup> Cognizant that applying this exception broadly would run counter to the First Amendment purpose of limiting government interference with expression, the Supreme Court has always recognized the need to limit the scope of the true threat exception.<sup>12</sup> But the Court has so far resisted clarifying the means by which legislatures and courts should balance a government’s interest in preventing individuals from living in fear after credible threats of violence with the imperative that distasteful or unpopular speech that is not imminently harmful is still subject to the protections of the First Amendment.

### A. *Watts Through Black—Uncertain Protection*

In *Watts v. United States*, the Supreme Court upheld the constitutionality of a statute prohibiting “knowingly and willfully . . . [making] any threat to take the life of or to inflict bodily harm upon the President of the United States.”<sup>13</sup> It nonetheless reversed the conviction of the defendant-petitioner, who had been overheard in a small group discussion at an antiwar rally on the National Mall, saying, “If they ever make me carry a rifle the first man I want to get in my sights is L.B.J.”<sup>14</sup> The Court did not resolve the question raised by the divided court of appeals below, as to whether a successful prosecution required a showing that the defendant intended to carry out his threat.<sup>15</sup> Instead, the Court held that the threat at issue could not be a “true ‘threat’” because, in the context of political debate in which it occurred, “regarding the expressly conditional nature of the statement and the reaction of the listeners,” it could not have been *interpreted* as any more than a crude form of political speech.<sup>16</sup>

---

<sup>10</sup> *Virginia v. Black*, 538 U.S. 343, 359 (2003) (citing *Watts v. United States*, 394 U.S. 705, 708 (1969)); *R.A.V. v. City of St. Paul*, 505 U.S. 377, 388 (1992)).

<sup>11</sup> *United States v. Alvarez*, 567 U.S. 709, 717-18 (2012).

<sup>12</sup> *Watts*, 394 U.S. at 708 (“[W]e must interpret the language Congress chose ‘against the background of a profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open [sic], and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials.’” (quoting *N. Y. Times Co. v. Sullivan*, 376 U.S. 254, 270 (1964))).

<sup>13</sup> *Id.* at 705 (quoting 18 U.S.C. § 871(a)).

<sup>14</sup> *Id.* at 706.

<sup>15</sup> *Id.* at 707-08.

<sup>16</sup> *Id.* at 708.

Two decades later, in *R.A.V. v. City of St. Paul*, the Court clarified in dicta the reasoning behind the outcome in *Watts*.<sup>17</sup> That opinion explains that the harm caused by threats is the fear of violence they instill in individuals, the disruption such fear can cause, and the possibility that the threatened violence will actually occur.<sup>18</sup> The Court relied on this explanation in *Virginia v. Black*, the first case after *Watts* to squarely address the constitutionality of a statute prohibiting threats.<sup>19</sup> There, the Court held unconstitutional a statute that prohibited burning a cross with the intent to intimidate a person or group, where burning a cross could constitute prima facie evidence of that intent to intimidate.<sup>20</sup> In light of *R.A.V.*'s explanation of the harm posed by threats, intimidation constituted a "type of true threat, where a speaker directs a threat to a person or group of persons with the intent of placing the victim in fear," and that *some* cross burnings could—and historically, had—fit within this definition.<sup>21</sup> But the second piece of the statute, which provided that the fact of burning a cross would constitute prima facie evidence of the defendant's intent to intimidate, "strip[ped] away the very reason why a State may ban cross burning with the intent to intimidate."<sup>22</sup> A jury would not need to find that a defendant in fact *had* a subjective intent to intimidate in order to convict her, making it possible that the defendant could be convicted for an act of "lawful political speech at the core of what the First Amendment is designed to protect."<sup>23</sup>

#### B. *Elonis—Failure to Clarify*

After *Black*, which struck down a statute that criminalized speech without requiring that the speaker have a subjective intent to threaten, the question nevertheless remained open whether the First Amendment *requires* such a showing of subjective intent in a prosecution for a true threat.<sup>24</sup> The Court's

---

<sup>17</sup> 505 U.S. 377 (1992).

<sup>18</sup> See *id.* at 388 (listing these as the reasons "why threats of violence are outside the First Amendment"). The decision ultimately turned on a different question of First Amendment doctrine, the "fighting words" exception, and held unconstitutional a statute that prohibited only fighting words aimed at members of certain protected classes as an impermissible content-based restriction. *Id.* at 381.

<sup>19</sup> 538 U.S. 343 (2003).

<sup>20</sup> *Id.* at 347-48.

<sup>21</sup> *Id.* at 360.

<sup>22</sup> *Id.* at 364-65.

<sup>23</sup> *Id.* at 365; see also *id.* at 365-66 ("As the history of cross burning indicates, a burning cross is not always intended to intimidate. . . . The prima facie provision makes no effort to distinguish . . . between a cross burning done with the purpose of creating anger or resentment and a cross burning done with the purpose of threatening or intimidating a victim.")

<sup>24</sup> See Petition for Writ of Certiorari at 16-20, *Elonis v. United States* (No. 13-983), 135 S. Ct. 2001 (2015) (noting a circuit split and "widespread confusion" as to whether *Black* meant that the First Amendment required a showing of subjective intent in addition to an objective, "reasonable speaker" showing).

decision twelve years later in *Elonis v. United States*, its next and most recent on the subject of true threats, declined to answer the constitutional question, instead relying again on principles of statutory interpretation to read a subjective-intent requirement into the general federal threat statute, 18 U.S.C. § 875(c).<sup>25</sup>

Elonis was convicted under § 875(c) for posting “graphically violent” rap lyrics on a Facebook page, first under his real name and then using a pen name.<sup>26</sup> While some of the lyrics Elonis posted referred to clearly fictitious people and situations, others referenced his coworkers and “included crude, degrading, and violent material about his soon-to-be ex-wife,” and caused the people referenced to fear for their safety.<sup>27</sup> The district court rejected Elonis’s request for a jury instruction that a conviction required proof of Elonis’s “inten[t] to communicate a true threat,” a question raised by Elonis’s testimony that his lyrics were fictitious—like those of the rapper Eminem, who has never been prosecuted for recording songs with lyrics that included specific “fantasies of killing his ex-wife.”<sup>28</sup> The jury was instead instructed to convict Elonis if a reasonable person would foresee that the lyrics would be interpreted as a threat, and on that basis they found him guilty of violating § 875(c).<sup>29</sup>

In vacating Elonis’s conviction and remanding the case for further proceedings under a subjective-intent standard, the Supreme Court rejected the notion that convictions under this particular statute could be supported by only a showing that the defendant “himself knew the contents and context of [the Facebook posts that formed the basis for his prosecution], and a reasonable person would have recognized that the posts would be read as genuine threats.”<sup>30</sup> Such a conviction would impermissibly rely on a negligence standard, which is out of place in the context of criminal prosecution.<sup>31</sup>

In separate dissents, Justices Alito and Thomas criticized the majority for leaving open the question whether a showing of recklessness could also

---

<sup>25</sup> 135 S. Ct. 2001, 2012, 2013 (2015); *see also* 18 U.S.C. § 875(c) (2018) (prohibiting “transmit[ting] in interstate or foreign commerce any communication containing any threat to kidnap any person or any threat to injure the person of another”).

<sup>26</sup> *Elonis*, 135 S. Ct. at 2004-05.

<sup>27</sup> *Id.* at 2005-06.

<sup>28</sup> *Id.* at 2007-08. One of Elonis’s posts referencing his ex-wife—and the restraining order she had obtained against him—included the following lyrics: “Fold up your [protection-from-abuse order] and put it in your pocket / Is it thick enough to stop a bullet? / . . . And if worse comes to worse / I’ve got enough explosives / to take care of the State Police and the Sheriff’s Department.” *Id.* at 2006. In a song titled “Kim,” the name of his ex-wife, Eminem raps: “Ha-ha, gotcha! Go ahead, yell! / Here, I’ll scream with you, ‘Ah! Somebody help!’ / Don’t you get it, bitch? No one can hear you! / Now shut the fuck up, and get what’s comin’ to you! / You were supposed to love me! / Now bleed, bitch, bleed!” *Eminem — Kim Lyrics*, GENIUS LYRICS, <https://genius.com/Eminem-kim-lyrics> [<https://perma.cc/J4SH-8XMK>].

<sup>29</sup> *Elonis*, 135 S. Ct. at 2007.

<sup>30</sup> *Id.* at 2011.

<sup>31</sup> *Id.*

support a conviction under § 875(c).<sup>32</sup> Justice Thomas objected to the majority opinion's effect of displacing the precedent of nine circuits, which he characterized as requiring proof of "general intent" for conviction under § 875(c), and providing no specific guidance as to what standard replaced it.<sup>33</sup> Justice Thomas further contended that neither *Watts* nor *Black* had come to any conclusion on a constitutional requirement to consider the subjective intent behind a threat—and that in any event, an objective test, which asked how a reasonable speaker in the defendant's position would have expected his utterance to be received, would confer sufficient protection because it would "for[e] jurors to examine the circumstances in which a statement is made."<sup>34</sup> Justice Alito argued that the Court should have clarified that a mens rea of recklessness was sufficient, both to support a criminal conviction and to interpret § 875(c) in a manner that did not violate the First Amendment.<sup>35</sup> Because the harm with which threat statutes are concerned does not change based on the speaker's mental state, in Justice Alito's reading, the applicability of the true-threat exception to First Amendment protection could not turn on this distinction.<sup>36</sup> Instead of requiring a mens rea of knowledge or purpose, Justice Alito proposed that evaluating speech in the proper context would be sufficient to protect speakers who engage in artistic expression not meant to put anyone in fear of harm.<sup>37</sup>

### C. After *Elonis*—Opportunities to Elaborate

The Court has not yet faced the two questions left open by its decision in *Elonis*: whether showing recklessness as to a statement's possible interpretation is a sufficiently high bar to support a conviction under § 875(c), and whether the Constitution requires a showing of subjective intent at all for threat convictions. Justice Sotomayor's "reluctant[] concur[rence]" in the 2017 denial of certiorari in *Perez v. Florida* recognized that some clarification is needed on

---

<sup>32</sup> See *id.* at 2013 ("Both Justice Alito and Justice Thomas complain about our not deciding whether recklessness suffices for liability under Section 875(c)." (capitalization altered)).

<sup>33</sup> *Id.* at 2018 (explaining that the previous standard had "require[d] no more than that a defendant knew he transmitted a communication, knew the words used in that communication, and understood the ordinary meaning of those words in the relevant context"); see also *id.* at 2022 ("The majority . . . casts my application of general intent as a negligence standard disfavored in the criminal law.").

<sup>34</sup> *Id.* at 2027 (quoting *United States v. Jeffries*, 692 F.3d 473, 479-80 (6th Cir. 2012)).

<sup>35</sup> *Id.* at 2014-16.

<sup>36</sup> *Id.* at 2015-16.

<sup>37</sup> *Id.* at 2016-17; see also *id.* at 2016 ("[C]ontext matters. 'Taken in context,' lyrics in songs that are performed for an audience or sold in recorded form are unlikely to be interpreted as a real threat to a real person. Statements on social media that are pointedly directed at their victims, by contrast, are much more likely to be taken seriously." (quoting *Watts v. United States*, 394 U.S. 705, 708 (1969)) (citation omitted)).

both questions.<sup>38</sup> In April 2019, the Court denied a petition for certiorari in *Knox v. Pennsylvania*.<sup>39</sup> That petition had preserved and squarely presented the question “whether, to establish that a statement is a true threat unprotected by the First Amendment, the government must show that a ‘reasonable person’ would regard the statement as a sincere threat of violence, or whether it is enough to show only the speaker’s subjective intent to threaten.”<sup>40</sup> The petition sought review of a conviction under Pennsylvania’s terroristic threat statute, interpreted by the Pennsylvania Supreme Court to require a showing of *only* subjective intent.<sup>41</sup> In arguing that clarification of the First Amendment standard for true threats is urgently needed, the petition outlined the disparate standards currently applied across the federal circuits and state courts.<sup>42</sup> The majority of states and some federal courts of appeals apply an objective-only standard while a minority of states and other circuits apply a subjective-only standard, with the result that in some cases, the standard to be applied depends on “which prosecutor decides to bring charges.”<sup>43</sup>

Three Justices have now articulated an interest in clarifying the implications of the *Elonis* decision: Justices Thomas and Alito in dissents to that decision and Justice Sotomayor in her concurrence in the denial of certiorari in *Perez*. But the Court still summarily denied the petition for certiorari in *Knox*. In doing so, the Court again declined to actually clarify the constitutional requirement that applies to statutes criminalizing threats. As the Court remains silent on the questions raised by true threat prosecutions, further questions accumulate behind that one: Not only are lower courts left without an authoritative constitutional standard, but they are left to apply these divergent standards to an increasingly broad range of social media contexts which raise new and complicated questions about how speech should be regulated. As a first step toward recalibrating the scope of the true threat exception, the Court should take the earliest opportunity to recognize a uniform constitutional standard that requires proving both a subjective intent to cause fear of harm *and* the objective, reasonable foreseeability of such harm, in order to support a conviction for making a threat.

---

<sup>38</sup> 137 S. Ct. 853, 854-55 (2017) (Sotomayor, J., concurring in denial of certiorari) (“In an appropriate case, the Court should affirm that [a jury cannot convict solely on the basis of the words a defendant used]. The Court should also decide precisely what level of intent suffices under the First Amendment.”).

<sup>39</sup> 139 S. Ct. 1547 (2019) (summary order).

<sup>40</sup> Petition for Writ of Certiorari at i, *Knox*, 139 S. Ct. 1547 (No. 18-949), 2019 WL 277506 [hereinafter Cert Petition, *Knox*].

<sup>41</sup> *Id.* at 6-7.

<sup>42</sup> *Id.* at 9-14.

<sup>43</sup> *Id.*



## II. SUBJECTIVE INTENT AS A NECESSARY COMPONENT OF FIRST AMENDMENT PROTECTION

Clarification of the intent standard will require addressing two distinct sources of indeterminacy. First, imposition of any intent requirement prompts a question of what proof will suffice. At its core, intent is an internal phenomenon which produces no physical evidence and is knowable only by the person who experiences it<sup>44</sup>—if at all.<sup>45</sup> And it will rarely be in a defendant's interest to testify that she possessed the intent required to convict her of a charged offense.<sup>46</sup> As a result, proof of intent much more frequently turns on circumstantial, objective evidence meant to persuade a jury that the defendant's intent to bring about an outcome is the most reasonable explanation for the way she acted under the circumstances.<sup>47</sup> Disputes over whether the circumstantial evidence offered is sufficient to show the requisite intent can arise in any area of law where intent is relevant; these disputes will frequently involve narrow line-drawing based on the facts at hand.<sup>48</sup>

And such determinations are complicated further in prosecutions for crimes effected through speech. Where a prosecutor establishing intent on a murder charge, for instance, may point to laws of the physical universe in

---

<sup>44</sup> *E.g.*, *Long v. State*, 867 N.E.2d 606, 614 (Ind. Ct. App. 2007); RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 1 cmt. c, illus. 3 (AM. LAW INST. 2005); David Crump, *What Does Intent Mean?*, 38 HOFSTRA L. REV. 1059, 1071-72 & nn.82-84 (2010) (citing *Commonwealth v. Pond*, 846 A.2d 699, 707 (Pa. Super. Ct. 2004)).

<sup>45</sup> For a brief overview of the “self-serving” cognitive bias and other ways in which individuals’ ability to accurately assess their own motivations are clouded, see Thomas Shelley Duval & Paul J. Silvia, *Self-Awareness, Probability of Improvement, and the Self-Serving Bias*, 82 J. PERSONALITY & SOC. PSYCHOL. 49, 49-50 (2002).

<sup>46</sup> *Cf.* Crump, *supra* note 44, at 1072 (“[I]ntent is easily denied or rebutted, even when it exists, and sometimes the denial is accompanied by convincing belief on the part of the actor.”).

<sup>47</sup> *See id.* at 1072 (“[T]he law evaluates intent by what the actor does, which means that the law evaluates intent by circumstantial evidence.”). The chasm between what subjective intent *is* and how it is proved in a court of law highlights the need for a precise definition of the standard by which intent is found to exist, as this standard will also determine what kind of rebuttal arguments are judged sufficient to overcome the circumstantial evidence presented. *See id.* at 1072, 1074-78 (describing how different forms of rebuttal of intent operate to undercut the persuasive value of circumstantial evidence).

<sup>48</sup> *See, e.g.*, *United States v. Souder*, 436 F. App'x 280, 292 (4th Cir. 2011) (declining to overturn the grant of a motion for a new trial on fraud charges where that decision was based on “the district court[s] cho[ice] to credit the Defendants’ innocent explanations for their actions over the sinister interpretation posited by the government”); *People v. Johnson*, 27 N.Y.2d 119, 123 (1970) (concluding, where a defendant’s testimony at trial denying an intent to commit burglary when he entered the premises contradicted his statement to police and extrinsic evidence from the time of his arrest, that the resolution of the issue was “open to the People to test credibility by the inconsistencies in his statement”); *see also generally* Baruch Weiss, *What Were They Thinking?: The Mental States of the Aider and Abettor and the Causer Under Federal Law*, 70 FORDHAM L. REV. 1341 (2002) (weighing arguments for and against applying the same intent standard to prosecutions for aiding and abetting criminal activity as for principal offenses).

establishing that a defendant must have known a certain outcome was probable—and therefore was more likely to have been done intentionally than unintentionally<sup>49</sup>—no such universal background rules exist for the causes and effects of language use.<sup>50</sup> Instead, the bounds of conversation are governed by the context in which they occur, including the participants' relationships to one another, their subjective understandings of the situation, and any previous interactions that have occurred between them.<sup>51</sup> At the same time, though, participants rarely make explicit within a conversation the rules by which the conversation will operate.<sup>52</sup> Where the 'rules' of a conversation are never fixed or stated out loud but are nonetheless crucial to understanding what has been said, a careful examination of context is crucial if one party is at risk of criminal conviction for the content she contributed to the conversation.<sup>53</sup>

It is not a novel observation that speech and language are central to many human functions,<sup>54</sup> but it is one that has yet to be fully felt in the operation

---

<sup>49</sup> *But see* Crump, *supra* note 44, at 1066-67 (characterizing the presumption that “a man is presumed to intend the natural and probable consequences of his acts[.]” as imposing a “preponderance” or “awareness of likelihood” standard on criminal liability (internal quotation marks omitted)).

<sup>50</sup> Trial lawyers use analogies to physical actions to explain intent to juries. *See id.* at 1073 & n.95 (describing the author's experience of explaining evidence of intent to a jury through the example of a person who got up and walked out of the courtroom as intending to leave). But because speech is so contextually determined, it is difficult to imagine a comparably simple analogy that would suffice to explain when a threat could be presumably intended.

<sup>51</sup> *See, e.g.*, H. P. Grice, *Logic and Conversation*, in 3 PETER COLE & JERRY L. MORGAN, SYNTAX AND SEMANTICS: SPEECH ACTS 41, 45-46 (1975) (proposing “maxims” of conversation that interlocutors must presume of one another, to explain how communication is effective in the absence of identifiable rules); Lidsky & Norbut, *supra* note 7, at 1909-10 (describing how “discourse conventions” that govern online speech may vary across platforms and between social groups on the same platform, causing online speech to be misconstrued).

<sup>52</sup> *See* PETER L. BERGER & THOMAS LUCKMANN, THE SOCIAL CONSTRUCTION OF REALITY: A TREATISE IN THE SOCIOLOGY OF KNOWLEDGE 172 (1966) (“Most conversation . . . takes place against the background of a world that is silently taken for granted. Thus, an exchange such as, ‘Well, it’s time for me to get to the station’, and ‘Fine, darling, have a good day at the office’, implies an entire world *within which* these apparently simple propositions make sense.”); Erving Goffman, *Replies and Responses*, 5 LANGUAGE IN SOC’Y 257, 259-60, 268 (1976) (describing the embedded meanings and assumptions that allow “Milk and sugar?” to be a complete, intelligible response to the question “Have you got coffee to go?”).

<sup>53</sup> *Cf.* Caleb Mason, *Framing Context, Anonymous Internet Speech, and Intent: New Uncertainty About the Constitutional Test for True Threats*, 41 SW. L. REV. 43, 101-10 (2011) (recognizing that an objective-intent-only standard, without a requirement to show specific speaker intent, would render threats a “perlocutionary” crime).

<sup>54</sup> *See, e.g.*, Amanda Shanor, *The New Lochner*, 2016 WIS. L. REV. 133, 176-77 (noting the “simple but often overlooked fact” that “nearly all human action . . . operates through communication” and the associated tendency of modern regulation to operate through speech rather than conduct); Bonnie Urciuoli & Chaise LaDousa, *Language Management/Labor*, 2013 ANN. REV. ANTHROPOLOGY 175, 176 (describing the rise of language skills in the contemporary United States economy as both necessary skills and as a commodity in themselves).

of First Amendment protections<sup>55</sup>—especially as increasingly more speech occurs in the public or semipublic forums of the Internet.<sup>56</sup> The increase in the volume of public expression and the changing media through which it occurs give rise to warring impulses as to how that speech should be evaluated for First Amendment purposes. On one hand, the ease with which an individual can dash off a tweet, Facebook post, or email to her federal representative supports a view that online speech taken as a whole is less formal, or even less important, than the written forms that preceded it, which required more physical effort to produce and cost more to disseminate.<sup>57</sup> Crucially, even the register of the speech that prevails on these platforms can contribute to the notion that online speech is less serious, meaningful, or valuable than speech disseminated through other media.<sup>58</sup>

On the other hand, this apparently casual and largely frivolous speech tends to remain publicly available, “recorded and searchable,” making it far more accessible to audiences intended and unintended alike, indefinitely.<sup>59</sup> And when an instance of online expression like an insult or a threat is discovered by an individual targeted by it (or a concerned third party), there can be little hope of convincing that person that the speech raises less cause

<sup>55</sup> See Shanor, *supra* note 54, at 176 (observing that because speech and expressive conduct permeate so many areas of human action, a strong version of “the First Amendment possesses near total deregulatory potential”).

<sup>56</sup> See, e.g., Edison Research, *Infinite Dial 2019* (Mar. 6, 2019), <https://www.slideshare.net/webby2001/infinite-dial-2019> [<https://perma.cc/56H5-3GP8>] (navigate to slide 5 of 62) (estimating that in early 2019 there were 223 million social media users in the United States); see also Jacob Rowbottom, *To Rant, Vent and Converse: Protecting Low Level Digital Speech*, 71 CAMBRIDGE L.J. 355, 359 (2012) (“With the growth of digital communications, there is more content being published than before . . . This content includes not only professionally produced content, but all the amateur content, conversations and comments that are made by users.”).

<sup>57</sup> See, e.g., *SI03, Inc. v. Bodybuilding.com, LLC*, No. 07-6311, 2008 WL 11348459, at \*5 (D. Idaho Dec. 23, 2008) (advancing the “general understanding that Internet blogs, message boards, and chat rooms are, by their nature, typically casual expressions of opinion”); John Cluverius, *How the Flattened Costs of Grassroots Lobbying Affect Legislator Responsiveness*, 70 POL. RES. Q. 279, 280-81, 286-88 (2017) (describing legislators’ efforts to adapt to the higher volume of constituent contacts made possible by email and other mass communication tools by searching for new heuristics to indicate that a particular issue is of high importance).

<sup>58</sup> See, e.g., *Rocker Mgmt. LLC v. John Does*, No. 03-33, 2003 WL 22149380, at \*2 (N.D. Cal. 2003) (relying on the general appearance and style of messages on a message board, including “grammar and spelling errors” and an absence of standard capitalization, to decide that no reasonable observer would have taken a purportedly defamatory statement as fact).

<sup>59</sup> Rowbottom, *supra* note 56, at 355-56. Even an individual who realizes after the fact that a social media post was in poor taste may not be able to completely erase the record of the expression from the Internet. See *id.* at 356 (describing the prosecution of a man who removed a “joke” page within hours of creating it); see also SNAP INC. LAW ENFORCEMENT GUIDE 4, 9-10 (last updated Sept. 21, 2018), <https://storage.googleapis.com/snap-inc/privacy/lawenforcement.pdf> [<https://perma.cc/54WF-SX8K>] [hereinafter SNAP LAW ENFORCEMENT GUIDE] (describing the methods by which Snapchat messages, which are designed to disappear after twenty-four hours, can in certain circumstances be recovered for longer periods of time).

for concern because it occurred online or did not employ standard grammar.<sup>60</sup> This inconsistent view of online speech—as generally less valuable or meaningful than other forms of speech, but with the potential to do acute, potentially legally actionable harm to individuals or discrete groups of hearers—creates a trap for the unwary social media user.

The Supreme Court has long recognized that “First Amendment freedoms need breathing space to survive,” and that this requirement puts the onus on governments to draft laws criminalizing speech with “narrow specificity.”<sup>61</sup> Clearly that specificity should be required of mens rea as well as actus reus requirements<sup>62</sup>: a statute that clearly delineates the type of statement to be punished but that does not make clear whether a merely reckless, rather than knowing or purposeful utterance, will be punished is likely to have a chilling effect on the overall amount of speech produced, compared to a statute that clarifies that reckless statements are not covered. And if the speech chilled by that vague statute would be legal and subject to constitutional protection, then a criminal statute that is vague as to intent has the effect of restricting speech in violation of the First Amendment.

Some convictions under threat statutes that do not require a showing of subjective intent to threaten would also have been upheld if a subjective-intent showing were required.<sup>63</sup> But in the jurisdictions that have continued to require only an objective, reasonable-observer standard for threat statutes not altered by *Elonis*,<sup>64</sup> some defendants have been convicted for threats made without a subjective intent to threaten. These defendants have been sanctioned for speech that, under a subjective definition of “true threat,” would arguably be within the heart of the First Amendment’s protective sphere.<sup>65</sup>

---

<sup>60</sup> For examples of ungrammatical, “casual” online speech that nevertheless prompted the targeted entities to commence litigation, see generally *Sl03, Inc.*, 2008 WL 11348459, and *Rocker Mgmt. LLC*, 2003 WL 22149380. See also *Elonis v. United States*, 135 S. Ct. 2001, 2005-07 (2015) (describing grammatically incorrect online speech that caused the speaker’s ex-wife to fear for her safety).

<sup>61</sup> *NAACP v. Button*, 371 U.S. 415, 433 (1963) (citing *Cantwell v. Connecticut*, 310 U.S. 296, 311 (1940)).

<sup>62</sup> See 21 AM. JUR. 2D *Criminal Law* § 112 (2019) (“Mens rea is generally an essential element of any criminal offense.”).

<sup>63</sup> For example, it would be difficult to dispute that a defendant who sent seven specific threats of violence to an individual, her family, and an embassy over the course of two months, “accompanied by threatening behavior,” did not also evince the subjective intent of putting the targeted individual and her family in fear of violence. See *United States v. Jordan*, 639 F. App’x 768, 770-71 (2d Cir. 2016).

<sup>64</sup> See Cert Petition, *Knox*, *supra* note 40, at 11-13 & nn. 4-5 (collecting cases from jurisdictions that apply an objective standard alone).

<sup>65</sup> See, e.g., *United States v. Dierks*, No. 17-2065, 2017 WL 4873067, at \*1, \*3-4 (N.D. Iowa Oct. 27, 2017) (allowing a prosecution for true threat to proceed under the Eighth Circuit’s objective-hearer standard for speech directed at an elected official, over objections that the threats were conditional, physically impossible to carry out, and actually intended to be “impolite criticism” like the protected speech at issue in *Watts*); *Lidsky & Norbut*, *supra* note 7, at 1886-87 (describing the attempted prosecution of a young man in Texas for what appeared, in context, to be a “hyperbolic response to provocation,” a “rant,” or even a poorly calibrated joke); see also Brief of Amici Curiae

The fact that speech made carelessly and not clearly outside the protection of the First Amendment can carry a sentence of, for example, six years in prison<sup>66</sup> suggests that the objective standard does not offer sufficient protection. Particularly where the kind of speech likely to be prosecuted under threat statutes may be made off-the-cuff and in the heat of an argument, and where the effects of a speech act are not as neatly foreseeable as the consequences of physical actions, the absence of formal protection for inadvertently threatening speech has the potential to chill a significant amount of speech. At least some of that silenced speech would be politically productive or facilitative of an individual's development of personal conscience.<sup>67</sup> Allowing convictions for true threats to stand without a showing of subjective intent leaves unpopular or borderline speech that could plausibly be viewed as having a threatening effect underprotected, relative to speech that verges on the outer bounds of the First Amendment's protection for other reasons.<sup>68</sup>

An objective component, while not sufficient on its own to protect the First Amendment interests at the edge of the true threat doctrine, is still necessary in any true threat standard. The Pennsylvania law in the *Knox* case, which allowed a conviction on a showing that the speaker acted with intent to terrorize or intimidate, but without a demonstration that any observer actually experienced fear or terror as a result, runs counter to the primary justification for punishing true threats: that they do, in fact, cause identifiable harm to some

---

Michael Render ("Killer Mike") et al. in Support of Petitioner at 16-17, *Knox v. Pennsylvania*, 139 S. Ct. 1547 (2019) (No. 18-949), 2019 WL 1115837 [hereinafter Brief of Michael Render ("Killer Mike") et al.] (recognizing prior uses of the phrase "fuck the police" in rap songs as a way to express criticism of police).

<sup>66</sup> Associated Press, *Iowa Man Gets 6 Years for Threatening U.S. Sen. Joni Ernst*, DES MOINES REG. (June 14, 2018, 6:26 AM), <https://www.desmoinesregister.com/story/news/crime-and-courts/2018/06/13/joseph-hilton-dierks-iowa-joni-ernst-cedar-rapids-capitol-police-federal-harassment-twitter/700225002> [<https://perma.cc/QY8V-WLPS>].

<sup>67</sup> Cf. *Red Lion Broad. Co. v. FCC*, 395 U.S. 367, 390 (1969) ("It is the purpose of the First Amendment to preserve an uninhibited marketplace of ideas in which truth will ultimately prevail . . ."); *Sherbert v. Verner*, 374 U.S. 398, 412 (1963) ("The harm is the interference with the individual's scruples or conscience—an important area of privacy which the First Amendment fences off from government.").

<sup>68</sup> See, e.g., *NAACP v. Claiborne Hardware Co.*, 458 U.S. 886, 927-28 (1982) (holding that although "references to the possibility that necks would be broken . . . conveyed a stern[] message," "[a]n advocate must be free to stimulate his audience with spontaneous and emotional appeals"); *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 271-72 (1964) ("[E]rroneous statement[s] [are] inevitable in free debate, and . . . must be protected . . ."); *Smith v. California*, 361 U.S. 147, 153 (1959) ("[I]f the bookseller is criminally liable [under the obscenity statute] without knowledge of the [book's] contents, . . . he will tend to restrict the books he sells to those he has inspected; and thus the State will have imposed a restriction upon the distribution of constitutionally protected as well as obscene literature."); see also Lyrissa Barnett Lidsky, *Nobody's Fools: The Rational Audience as First Amendment Ideal*, 2010 U. ILL. L. REV. 799, 841-42 (characterizing *Greenbelt Cooperative Publishing Ass'n v. Bresler*, 398 U.S. 6 (1970), as endorsing the idea that "speakers should not be held liable for 'misreadings' of their speech by idiosyncratic or unsophisticated audience members" (emphasis added)).

who hear them.<sup>69</sup> Additionally, the vast majority of evidence of even subjective intent will be objective, circumstantial evidence which already lends itself to an objective analysis of the content of the speech.<sup>70</sup> This finding is not only constitutionally required,<sup>71</sup> but also should impose only a minimal additional burden in a prosecution for any speech act that could be reasonably regarded as a threat—especially as compared to the magnitude of the defendant’s interest in avoiding conviction for constitutionally protected activity.

The objective-intent standard ensures that speech must actually create some risk of harm before it is punishable. The subjective-intent standard protects speakers who use language idiosyncratically—or in a way unfamiliar to others in a community—by requiring that a speaker know that his message may be seen as a threat.<sup>72</sup> In the contemporary linguistic environment of the United States, where many linguistic communities occupy the same physical and discursive spaces and there is no basis for a claim that one use of a phrase is objectively correct,<sup>73</sup> the First Amendment should be understood to require *both* standards be met before a person can be criminally punished on the basis of her speech.

### III. HOW CONTEXT CAN CLARIFY INTENT IN ONLINE SPEECH

The requirement of subjective and objective mens rea standards will not resolve the other question left open by *Elonis*: what evidence will suffice to prove

---

<sup>69</sup> Commonwealth v. Knox, 190 A.3d 1146, 1158-61 (Pa. 2018) (interpreting *Black* to reject an objective standard and relying instead on deducing the defendant’s subjective intent from the content and context of his speech); Cert Petition, *Knox*, *supra* note 40, at 18 (“[V]irtually every exception to the First Amendment’s protections includes a baseline requirement that the speech in question be objectively harmful.”); see also R.A.V. v. City of St. Paul, 505 U.S. 377, 388 (1992) (listing “protecting individuals from the fear of violence, from the disruption that fear engenders, and from the possibility that the threatened violence will occur,” as the harms against which statutes banning true threats protect).

<sup>70</sup> See *supra* text accompanying notes 44–47; but see *infra* note 143 (proposing that evidence of an audience’s reaction can be probative of either subjective intent or a reasonable observer’s reaction, depending on when the defendant became aware of the reaction).

<sup>71</sup> See Cert Petition, *Knox*, *supra* note 40, at 18 (“[T]he objective standard is the constitutionally mandated minimum protection for all speech onto which th[e] Court has layered *additional* protections.”).

<sup>72</sup> Cf. Kunz v. New York, 340 U.S. 290, 313 (1951) (Jackson, J., dissenting) (“The law of New York does not segregate, according to their diverse nationalities, races, religions, or political associations, the vast hordes of people living in its narrow confines. Every individual in this frightening aggregation is legally free to live, to labor, to travel, when and where he chooses.”); see also *id.* at 301 (“A hostile reception of his subject certainly does not alone destroy one’s right to speak. . . . [I]n a free society all sects and factions, as the price of their own freedom to preach their views, must suffer that freedom in others.”).

<sup>73</sup> See Richard A. Epstein, *Linguistic Relativism and the Decline of the Rule of Law*, 39 HARV. J.L. & PUB. POL’Y 583, 589 (2016) (describing twentieth-century language philosophers’ rejection of “the notion that language has consistent and coherent usage”); Arturo Madrid, *Official English: A False Policy Issue*, 508 ANNALS AM. ACAD. POL. & SOC. SCI. 62, 63 (1990) (“Contrary to popular belief, American society never enjoyed a golden age in which we all spoke English; we never were all one linguistically.”).

a subjective intent to put some audience member in fear of serious harm. Since 2016, convictions under 18 U.S.C. § 875(c) have required a showing of some specific intent as to the threat alleged.<sup>74</sup> But the Supreme Court has yet to hear an appeal of a conviction under this new standard; it has thus provided no guidance as to what evidence supports a conviction under a subjective-intent standard—beyond the bedrock principle that in evaluating threats, “context matters.”<sup>75</sup> The context of how a defendant communicated a threatening message has the potential to illuminate, among other things, why the defendant made the utterance at issue, why he chose the words (or pictures or characters) he chose, whom he intended to reach with his message, and how he expected the message to be understood by his intended audience. With the advent of social media, speakers have an enormous, novel opportunity to craft messages such that the choice of platform, audience, and form may actually reflect a speaker’s conscious thought about how a message would be delivered<sup>76</sup>—but at the same time, of course, platforms like Facebook, Twitter, and message boards are also used impulsively by speakers posting without such conscious objectives.<sup>77</sup> Where evidence that can provide insight into speakers’ intents exists, and where some showing of intent to threaten is a constitutionally required element of a threat prosecution, not using available contextual evidence to better understand speakers’ intents constitutes an inexplicable oversight by the criminal justice system. This Part draws from existing research in the fields of sociology and linguistics to propose ways in which context—especially online—can provide evidence of what a speaker intended to communicate in making a public statement.

---

<sup>74</sup> See *supra* text accompanying notes 25–31 (describing the holding in *Elonis*).

<sup>75</sup> *Elonis v. United States*, 135 S. Ct. 2001, 2016 (2015) (Alito, J., concurring in part and dissenting in part); see also Lidsky & Norbut, *supra* note 7, at 1888 (“[E]stablishing the full context of [the defendant’s online comments made on Facebook] should be an essential part of determining whether he made terroristic threats or merely talked trash with a fellow video gamer.”).

<sup>76</sup> See, e.g., *The Secret Formula to Creating a Viral Instagram Post*, SOCIALINSIGHT.IO, <https://socialinsight.io/the-secret-formula-to-creating-a-viral-instagram-post/> [<https://perma.cc/PG94-RGGQ>] (last visited Nov. 15, 2019) (advising that “go[ing] viral” requires “know[ing] what photos to post, when to upload them, and how to connect with other users on the network to gain more followers and likes”); see also Kevin Roose & Keith Collins, *Who’s Winning the Social Media Midterms*, N.Y. TIMES (Oct. 18, 2018), <https://www.nytimes.com/interactive/2018/10/18/us/politics/social-election.html> [<https://perma.cc/S38D-CSDL>] (describing the success of various political campaigns at increasing their candidates’ social media engagement).

<sup>77</sup> See, e.g., Stephanie McNeal, *People Are Freaking out over This Girl’s Sad Picture of Her Grandpa*, BUZZFEED NEWS (Mar. 20, 2016, 6:23 PM), <https://www.buzzfeednews.com/article/stephaniemcneal/paw-paw-why> [<https://perma.cc/QA7N-WC7M>] (interviewing a woman who posted a captioned photo of her grandfather “not expecting it to even get a retweet,” and describing the viral reception the photo received); see also mstem, *Thinking About the People Behind the Viral Videos*, MIT CTR. FOR CIVIC MEDIA (May 7, 2012), <https://civic.mit.edu/2012/05/07/thinking-about-the-people-behind-the-viral-videos/> [<https://perma.cc/86UF-DSZ6>] (collecting stories of people whose Internet speech attracted unexpected, outsized attention and contemplating the consequences of “go[ing] viral”).

### A. *Accounting for Speaker Identity*

It is widely acknowledged that gender-based, class-based, ethnic, geographic, social, and other groups to which a person belongs can affect and be affected by how she uses language.<sup>78</sup> There is also reason to believe that individuals maintain internally consistent writing styles throughout their lifetimes, remaining somewhat constant across subject matters as well.<sup>79</sup> The implications of these findings on threat prosecutions are admittedly weak but nevertheless worth noting. Preliminary research suggests that variations in a person's writing style can reveal changes in personality or mental health;<sup>80</sup> while this research is not remotely advanced enough to read the objective evidence of a threatening communication to reveal subjective intent to threaten, it nonetheless bolsters the argument that a careful, close reading of a written statement can provide useful information about a speaker's intent. Additionally, although it is unlikely that a person could go long in the world without being alerted that her idiosyncratic personal writing style made her innocent statements appear threatening, such a person (if she exists) should have the opportunity to show a lack of notice to disprove that she knew or intended that her statement would be interpreted as a threat.

### B. *Accounting for Audience*

Although individual writing styles are relatively internally consistent, people do modulate the way they present themselves in online interactions, just as they might in person,<sup>81</sup> based on whom they understand to be their

---

<sup>78</sup> See, e.g., Umashanthi Pavalanathan & Jacob Eisenstein, *Audience-Modulated Variation in Online Social Media*, 90 AM. SPEECH 187, 187 (2015) ("Linguistic differences are robust enough to support unnervingly accurate predictions of [identity] characteristics based on writing style—with algorithmic predictions in some cases outperforming those of human judgments." (citation omitted)); Gretchen McCulloch, *Move Over Shakespeare, Teen Girls Are the Real Language Disruptors*, QUARTZ (Aug. 7, 2015), <https://qz.com/474671/move-over-shakespeare-teen-girls-are-the-real-language-disruptors> [<https://perma.cc/BM6N-LYTM>] ("William Labov[] observed that women lead 90% of linguistic change—in a paper he wrote 25 years ago. Researchers continue to confirm his findings. It takes about a generation for the language patterns started among young women to jump over to men."); see also generally Matthew L. Newman et al., *Gender Differences in Language Use: An Analysis of 14,000 Text Samples*, 45 DISCOURSE PROCESSES 211 (2008).

<sup>79</sup> See James W. Pennebaker & Laura A. King, *Linguistic Styles: Language Use as an Individual Difference*, 77 J. PERSONALITY & SOC. PSYCHOL. 1296, 1308 (1999) (using Linguistic Inquiry and Word Count ("LIWC") analysis to examine writing done by individuals over the course of years and concluding that "the ways people express themselves in words are remarkably reliable across time and situations").

<sup>80</sup> See *id.* (finding that where study participants wrote about their own emotions, the variation in their writing styles "predict[ed] such things as illness as well or better than the [traditional] five-factor [personality] dimensions"); see also, e.g., *Watts v. United States*, 394 U.S. 705, 708 (1969) (evaluating the purported threat in light of its "expressly conditional nature").

<sup>81</sup> See, e.g., Benjamin Bailey, *Switching*, 9 J. LINGUISTIC ANTHROPOLOGY 241, 242 (1999) (describing the phenomenon of intentional "code-switching" as a linguistic adjustment to the



audiences.<sup>82</sup> Such adjustments may result from a desire to emphasize certain aspects of one's personality, or "self";<sup>83</sup> to establish rapport with an audience, whether that audience is a known individual or a group of known individuals<sup>84</sup> or a larger audience of online "friends" or "followers";<sup>85</sup> or to demonstrate a sense of belonging within a certain discourse community.<sup>86</sup> These variations can affect the way in which a message is conveyed: a student who posts a message on her Facebook page, which she uses to interact with friends her own age, that she is "dying," even "literally" so,<sup>87</sup> may be best understood as engaging in hyperbole to convey some strong emotional state, rather than conveying that she is actually dying or contemplating death.<sup>88</sup> But that same message by the same speaker might be understood differently in a different context: Posted on a message board for people undergoing cancer treatment or spoken over the phone to an emergency dispatcher, for example, "I'm dying" might be better understood as conveying its more literal meaning.

---

identities of participants in a conversation or in recognition of a setting); Bonnie Urciuoli, Book Review, 33 LANGUAGE SOC'Y 153, 153, 156 (2004) (reviewing BENJAMIN H. BAILEY, LANGUAGE, RACE, AND NEGOTIATION OF IDENTITY (2002)) (describing how Dominican-American students' distinct uses of language with different interlocutors contributed to their own identity formation and how others perceived them).

<sup>82</sup> See, e.g., Pavalanathan & Eisenstein, *supra* note 78, at 189 ("[I]ndividuals modulate their linguistic performance as they use social media affordances to control the intended audience of their messages.").

<sup>83</sup> See Liam Bullingham & Ana C. Vasconcelos, *The Presentation of Self in the Online World: Goffman and the Study of Online Identities*, 39 J. INFO. SCI. 101, 107-08 (2013) (describing how one woman's two blogs varied in style, according to their purposes of highlighting the professional and personal "side[s]" of her personality respectively).

<sup>84</sup> See Cristian Danescu-Niculescu-Mizil et al., *Mark My Words! Linguistic Style Accommodation in Social Media*, 20 INT'L WORLD WIDE WEB CONF. PROC. 745, 746, 750 (2011) (concluding that in one-on-one conversations on Twitter, users' linguistic styles tended to converge and become more similar to one another's); see also generally David Jacobson, *Interpreting Instant Messaging: Context and Meaning in Computer-Mediated Communication*, 63 J. ANTHROPOLOGICAL RES. 359 (2007) (describing the way that shared context can alter how an individual expresses himself to an interlocutor).

<sup>85</sup> See Pavalanathan & Eisenstein, *supra* note 78, at 201-05 (finding that Twitter users were less likely to use geographically specific slang when directing tweets to a wider audience than when sending messages to a particular user, especially a user from the same geographic area).

<sup>86</sup> See Lidsky & Norbut, *supra* note 7, at 1909-10 (discussing the ways in which speech conventions vary by age, gender, and profession on the same social media platform).

<sup>87</sup> *Id.* at 1912.

<sup>88</sup> See *id.* at 1912-13 (acknowledging the "generational convention" of teenage hyperbole regarding death and speculating that it has been caused in part by the performative, audience-driven nature of social media); Jacobson, *supra* note 84, at 365 (observing different responses to a subject's use of the phrase "please kill me" as an exaggeration in a time of stress," which varied in part based on how well each interlocutor knew the speaker). Speech about bodily harm or death may be uniquely susceptible to misinterpretation. Understood literally, such speech reasonably provokes a sense of alarm in hearers; at the same time, some observers suggest that younger Americans have developed a particularly macabre sense of humor and are more likely to joke about death than older generations, creating situations rife with potential for misunderstanding. See, e.g., Elizabeth Bruenig, *Why Is Millennial Humor So Weird?*, WASH. POST (Aug. 11, 2017), [https://www.washingtonpost.com/outlook/why-is-millennial-humor-so-weird/2017/08/11/64af9cae-7dd5-11e7-83c7-5bd546f0d7e\\_story.html?noredirect=on&utm\\_term=.9e0f54d8b516](https://www.washingtonpost.com/outlook/why-is-millennial-humor-so-weird/2017/08/11/64af9cae-7dd5-11e7-83c7-5bd546f0d7e_story.html?noredirect=on&utm_term=.9e0f54d8b516) [perma.cc/FHJ9-SZXJ].

Audience-based misinterpretation can occur on platforms where a dyadic, or one-on-one, conversation takes place in the view of other observers: for example, if Facebook user *A* comments publicly on the public post of another user, *B* (as opposed to using the direct, private messaging function), and *B* responds with a sub-comment to *A*'s comment. While this exchange takes place between *A* and *B*, and may be clearly understood by them to operate under the discursive norms they have already established between the two of them, the fact remains that the conversation is observable by other Facebook users who may not know, for instance, that *A* means "I'm stressed" when he says "Please kill me."<sup>89</sup> A similar problem of interpretation resulted in a prosecution for making a terroristic threat when an observer of a Facebook conversation between two teenagers interpreted one teen's statements as a threat, rather than hyperbolic trash talk related to a video game the two played together.<sup>90</sup>

Jordan Strauss proposes two more criteria, indicative of whether a communication shows a subjective intent to invoke fear of violence, which consider a speaker's relationship to his audience: whether the communication identifies a specific target and, if so, whether it was reasonably foreseeable that the communication would reach the target.<sup>91</sup> Implicit in the overlap between these criteria are several gradations of likelihood that the communication demonstrates subjective intent. A speaker may name a target in a communication directed, privately or publicly, at that target, or may communicate the purported threat publicly, not *to* the target but in a manner that the target might come across it; finally, he might make a threat at a particular, individual target but in a manner that makes it unlikely for the target to ever encounter the threat (for example, a private, direct Facebook message to a third party).<sup>92</sup> Each of these general scenarios encompasses

---

<sup>89</sup> See Jacobson, *supra* note 84, at 365.

<sup>90</sup> Lidsky & Norbut, *supra* note 7, at 1886-88. In that case, after reading a violent statement that Justin Carter, a teenage boy in Texas, had posted on his public Facebook page, a Canadian woman tracked down Carter's address in order to report the post to the police in his locality. *Id.* at 1886. Carter's post, which included the phrase "I think I[m going to] SHOOT UP A KINDERGARTEN," had been directed at a friend and was apparently part of an ongoing "war of words" with a friend, who had "trashed" him first." *Id.* at 1886-88. The friend to whom the post was directed did not contact the police. *Id.* at 1888. Carter spent four months in jail before being released on bail, and the charges stemming from the arrest were pending for five years before being dismissed in exchange for a guilty plea. *Id.* at 1886-87.

<sup>91</sup> Jordan Strauss, *Context Is Everything: Towards a More Flexible Rule for Evaluating True Threats Under the First Amendment*, 32 SW. U. L. REV. 231, 264-65, 267-68 (2003).

<sup>92</sup> Counter to Strauss's argument, the analysis should not end if the speaker targets a group. In such a case, analysis should proceed as to the nature of the group, including whether the group targeted is a discrete and identifiable group and whether the group targeted is geographically or otherwise accessible to the speaker such that any members hearing the threat might feel targeted. In some instances the fact that a speaker stops short of identifying a single, discernible individual target might indicate that the speaker lacks a plan for carrying out any purported threat. But in other cases, speech targeting a group whose members are relatively small in number, visibly or

potentially infinite variations, where under the specific facts at issue the method of communication results in a greater or lesser likelihood that the speaker intended to put a target in fear of violence. But these factors are nonetheless possible to discern in the choices that speakers make in their social media use, for which lawyers should account in making the case that a communication does or does not show a subjective intent to make a true threat.

Further evidentiary potential for social media records comes from the fact that on most platforms, audience responses are recorded and can become interpretive tools for an original utterance. In other words, a speaker's intent in a given social media post can become clear through the responses the communication attracts and the way the speaker responds to those responses. If a user on Twitter, for example, publishes a tweet, then receives a response that appears to misunderstand her intent, she ordinarily has an opportunity to clarify the intent of her first tweet in a second one, a reply to the response.<sup>93</sup> How much evidentiary weight should be attached to the lack of such a clarifying reply—that is, whether the respondent's interpretation should be presumed to be in line with the original speaker's intent, absent a statement to the contrary—will likely be a case-specific question that accounts for the user's general habits, including how closely monitored the account is, whether the user regularly replied to interlocutors on the account, and how visible the interpretive response was to the original speaker. But replies to a post are easily discoverable and have the potential to explain how a post was understood and whether the speaker disagreed with that reception,<sup>94</sup> and this evidence should therefore be accounted for in threat prosecutions when available.

### C. Accounting for Platform-Based and Technological Constraints

The structural constraints and default settings that determine the form of expression that can take place on a given social media platform may also provide evidence of what a user intended her communications on that

---

otherwise identifiable, and located near the speaker may be sufficiently definite to support a finding that the threat was actually intended to put group members in fear.

<sup>93</sup> See, e.g., Shannon Romano, PhD (@sromano23), TWITTER (Nov. 10, 2018, 5:26 AM), <https://twitter.com/sromano23/status/1061248740596559872> [<https://perma.cc/H4B3-QYHY>] (“Maybe it was poor word choice. Sorry. Didn’t mean to imply your parent didn’t do a wonderful job raising you.”); see also T. S. Eliot, *The Love Song of J. Alfred Prufrock*, in COLLECTED POEMS 1909–1962, at 7 (1963) (“That is not it at all / That is not what I meant, at all.” (internal quotation marks omitted)).

<sup>94</sup> See Cinzia Padovani, *The Media of the Ultra-Right: Discourse and Audience Activism Online*, 15 J. LANGUAGE & POL. 399, 412 (2016) (using the comments posted under a press release by a far-right political organization to “decod[e]” the press release by examining, in part, “[w]hat element of the original message resonated most among the audience”); see also *United States v. Dutcher*, 851 F.3d 757, 762 (7th Cir. 2017) (using comments from concerned friends to reach the conclusion that a Facebook post including a threat against President Obama had been taken seriously by observers who knew the defendant).

platform to express. For example, Professors Lidsky and Norbut note that a platform like Twitter, with a relatively low character limit for a single message, may frustrate a user's ability to engage in nuanced figurative language like "sarcasm, hyperbole, or jests"—making misinterpretation of such utterances more likely.<sup>95</sup> But in recent years, Twitter has become more advanced in its design and allows users to "thread" sequential messages, so that a user who sends multiple 280-character tweets in a row can ensure that they will appear in the correct order and appear, at least somewhat, like a paragraph instead of a series of disjointed thoughts out of context.<sup>96</sup> Component tweets of a Twitter "thread" may still prove more easily taken out of context than any given sentence in a longer Facebook or blog post, though, given that the entire thread may not always load all at once.<sup>97</sup> Members of the legal community interested in assessing the contents of a tweet should be aware that the full context of a tweet may not be available at first glance, and that this may have an effect on the way a tweet was perceived by relevant parties.

Another platform-specific feature is the retweet function, also unique to Twitter. This function, when employed by user *A*, posts a message originally tweeted by user *B* on user *A*'s account—essentially linking to user *B*'s message but in a manner that displays the full content of the link on user *A*'s page.<sup>98</sup> The link's continued viability depends on user *B*'s continued decision not to

<sup>95</sup> Lidsky & Norbut, *supra* note 7, at 1910.

<sup>96</sup> See Aliza Rosen & Ikuhiro Ihara, *Giving You More Characters to Express Yourself*, TWITTER: BLOG (Sept. 26, 2017), [https://blog.twitter.com/en\\_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html](https://blog.twitter.com/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html) [<https://perma.cc/D3WH-2Z92>] (announcing that the maximum length of a Twitter post would double, from 140 characters to 280 characters); *Help Center: How to Create a Thread on Twitter*, TWITTER, <https://help.twitter.com/en/using-twitter/create-a-thread> [<https://perma.cc/MVC8-BBGD>] (last visited Dec. 19, 2019) ("Sometimes we need more than one Tweet to express ourselves. A thread on Twitter is a series of connected Tweets from one person. With a thread you can provide additional context, an update, or an extended point by connecting multiple Tweets together."); see also Aja Romano, *Twitter's New Thread Feature Takes Us One Step Closer to Longform Tweeting*, VOX (Dec. 15, 2017, 12:10 PM), <https://www.vox.com/culture/2017/12/15/16771922/twitter-new-threading-feature> [<https://perma.cc/RN3L-GXVD>] ("The new display should make it much easier to actually read and write threads."). Some early research suggests that the increased length of Twitter posts changed the content shared on the site, improving the "quality of discourse" by allowing room for the development of more nuanced ideas within a single tweet. Yphtach Lelkes, Kokil Jaidka & Alvin Zhou, *Twitter Got Somewhat More Civil When Tweets Doubled in Length. Here's How We Know.*, WASH. POST: MONKEY CAGE (Sept. 17, 2019, 7:00 AM), <https://www.washingtonpost.com/politics/2019/09/17/twitter-got-somewhat-more-civil-when-tweets-doubled-length-heres-how-we-know> [<https://perma.cc/MP9J-2ZLX>].

<sup>97</sup> See Romano, *supra* note 93 ("[W]ith particularly long threads . . . , when you click on the first tweet, Twitter will only display the first 200. When you click on the most recent tweet, you can only see a few tweets back. Lost in the middle are about 75 tweets that are essentially totally inaccessible." (emphasis omitted)).

<sup>98</sup> See *Help Center: Retweet FAQs*, TWITTER, <https://help.twitter.com/en/using-twitter/retweet-faqs> [<https://perma.cc/9MBC-4RKQ>] (last visited Dec. 18, 2019) ("Retweets look like normal Tweets with the author's name and username next to it, but are distinguished by the Retweet icon . . . and the name of the person who Retweeted the Tweet." (emphasis omitted)).

delete the tweet, though user *A* may decide to undo the retweet, removing the link from user *A*'s profile.<sup>99</sup> As Twitter gained popularity, a dispute arose, especially among people employed by traditional media outlets, regarding what a retweet “meant.” It is now common to see a phrase along the lines of “Retweets do not signal endorsement” on journalists’ profiles, indicating that anyone viewing their tweets should not presume that they agree with, endorse, or vouch for the truth of any tweets from others to which they have linked on their own profiles.<sup>100</sup> In at least one legal context, a Twitter user has been held personally responsible for the content of another user’s tweet, which she retweeted.<sup>101</sup> And somewhat surprisingly, a prosecution for a true threat under 18 U.S.C. § 875(c) was allowed to proceed on evidence that the defendant had retweeted threatening messages.<sup>102</sup> While this prosecution was arguably influenced by other factors and this conclusion may not be generalizable,<sup>103</sup> the conclusion reached by the magistrate judge in that case should nevertheless give observers pause. The prospect of prosecution for a

---

<sup>99</sup> *Id.*

<sup>100</sup> See Charlie Warzel, *Meet the Man Behind Twitter’s Most Infamous Phrase*, BUZZFEED (Apr. 15, 2014, 2:51 PM), <https://www.buzzfeednews.com/article/charliewarzel/meet-the-man-behind-twiters-most-infamous-phrase> [<https://perma.cc/2G76-MWDW>] (interviewing an early adopter of the disclaimer, who explained that he “wanted to be clear that a retweet did not necessarily indicate agreement. Nor did it mean [he] was confirming what another news organization was reporting.” (internal quotation marks omitted)).

<sup>101</sup> Callum Borchers, *Retweets ≠ Endorsements? Oh, Yes, They Do, Say the Hatch Act Police*, WASH. POST: THE FIX (Oct. 4, 2017), [https://www.washingtonpost.com/news/the-fix/wp/2017/10/04/retweets-endorsements-hatch-act/?utm\\_term=.8f6309cb6ffb](https://www.washingtonpost.com/news/the-fix/wp/2017/10/04/retweets-endorsements-hatch-act/?utm_term=.8f6309cb6ffb) [<https://perma.cc/NZM7-WTA2>] (describing the decision of the United States Office of Special Counsel to issue a warning to United Nations Ambassador Nikki Haley, whose retweet of a tweet supporting a candidate for political office was interpreted as an endorsement of the candidate in violation of the Hatch Act).

<sup>102</sup> See *United States v. Yassin*, No. 16-03024-01, 2017 WL 1324141, at \*4-5 (W.D. Mo. Feb. 23, 2017) (report and recommendation of magistrate judge), adopted by No. 16-3024-01 (Apr. 6, 2017) (rejecting the defendant’s argument that because the retweeted threats did not consist of “her words, . . . as a matter of law, it [could not] be alleged that she made threats”). The discussion of subjective intent in this opinion was unnecessary, as the Eighth Circuit applies a reasonable-listener test to true threat prosecutions, but nonetheless took a strong position that a user who retweets a threatening communication “intends to convey the message that she agrees with the ‘tweet.’” *Id.* at \*5 (quoting Bethany C. Stein, Comment, *A Bland Interpretation: Why a Facebook “Like” Should Be Protected First Amendment Speech*, 44 SETON HALL L. REV. 1255, 1277 (2014)) (internal quotation marks omitted).

<sup>103</sup> See, e.g., Katie Zavadski, *The American Anti-Vaccine Mom Turned ISIS Superstar*, DAILY BEAST (July 12, 2017, 6:54 PM), <https://www.thedailybeast.com/the-american-anti-vaccine-mom-turned-isis-superstar> [<https://perma.cc/4W9G-4CVL>] (describing Yassin’s alleged ties to the Islamic State and noting that in addition to retweeting other ISIS-affiliated users’ threats, she was alleged to have “produce[d] . . . original content” including sharing photos and addresses of U.S. military personnel); see also generally U.N. Security Council, *Report of the Secretary-General on the Threat Posed by ISIL (Da’esh) to International Peace and Security and the Range of United Nations Efforts in Support of Member States in Countering the Threat* ¶¶ 4-34, U.N. Doc. S/2016/92 (Jan. 29, 2016) (describing the international community’s assessment of the threats posed by the Islamic State in 2016 and highlighting the role of online social media in the organization’s efforts to recruit international support).

retweet—when it is arguably unsettled what the “meaning” of a retweet is,<sup>104</sup> and thus a defendant could have intended the action of retweeting to mean something entirely different than a full-throated endorsement of the original tweet’s message—would seem to leave too much discretion in the hands of individual members of the judicial branch. Such discretion, exercised without more precise guidance from experts and without systemic guidance from the Supreme Court, is likely to replicate the systemic biases of other parts of society, which can be mediated through and even exacerbated by language differences.<sup>105</sup> This result would undermine the promise of the First Amendment, as well as the Constitution’s equal protection and due process guarantees, and should be guarded against in the development of the true threat doctrine and its application to speech on the Internet.

Other relevant features of specific platforms include the disappearing-photo feature on platforms like Snapchat.<sup>106</sup> As Lidsky and Romano note, the fact that a photo message is visible for only a limited period of time could make any threat sent over Snapchat either “more or less ominous.”<sup>107</sup> They also note the evidentiary problems inherent in proving a threat sent over Snapchat: because the message disappears once viewed, investigators looking into a perceived threat have to rely on the memory of the participant, which may be unreliable in the immediate aftermath due to the viewer’s emotional response—perhaps a response that is outsized in comparison to the emotional reaction the speaker intended or anticipated producing.<sup>108</sup> If a Snapchat message is recovered<sup>109</sup>—a process which takes a considerable amount of time<sup>110</sup>—the very fact that the sender used Snapchat, or a similar time-limited messaging service, could support a claim that the sender intended to threaten the recipient. The

---

<sup>104</sup> See John Dickerson, *What RTs Mean*, JOHN DICKERSON, <https://johndickerson.com/blog/what-rt-mean> [<https://perma.cc/gY7W-NFAV>] (last visited Jan. 9, 2019) (providing one Twitter user’s explanation of several such possible “meanings” he intends to convey by retweeting another user’s tweet).

<sup>105</sup> See, e.g., Brief of Michael Render (“Killer Mike”) et al., *supra* note 65, at 19-23 (describing social science research demonstrating the effect of an association with rap music on potential jurors’ perception of a defendant); John Eligon, *Speaking Black Dialect in Courtrooms Can Have Striking Consequences*, N.Y. TIMES (Jan. 25, 2019), <https://www.nytimes.com/2019/01/25/us/black-dialect-courtrooms.html> [<https://perma.cc/6LA4-DE6C>] (describing a significantly higher rate of transcription errors in courtroom testimony given with African-American English grammar and vocabulary compared to testimony given in standard American English).

<sup>106</sup> Lidsky & Norbut, *supra* note 7, at 1910.

<sup>107</sup> *Id.*

<sup>108</sup> *Id.*; see also *supra* notes 61-68 and accompanying text (explaining the need for a subjective-intent requirement as a safe harbor for expression that unintentionally crosses a line into threatening).

<sup>109</sup> Although a photo or message sent through Snapchat disappears from the recipient’s inbox once it has been viewed, in some cases content is retained on Snapchat’s servers for up to thirty days and can be made available to law enforcement agencies. SNAP LAW ENFORCEMENT GUIDE, *supra* note 59, at 4, 9-10.

<sup>110</sup> Lidsky & Norbut, *supra* note 7, at 1911.

sender's choice of medium could be cited to argue that the sender knew that the message's content was illicit or otherwise desired to evade detection.

More generally, factfinders should take into account the range of options available to a speaker on a given platform. On a platform that does not allow for traditional forms of textual emphasis like bolding or underlining, a user's choice to use all-capitalized text for emphasis may look less threatening than it would in another context.<sup>111</sup> Where a Snapchat user had the option to use a "neutral yellow" version of the emoji that shows a man running, but instead chose a version of the emoji with dark brown skin, it may be reasonable for a jury to draw inferences from that choice as to the user's intent.<sup>112</sup> Evaluation of these choices requires further research, and may raise even more questions: Such choices by users may be weighted differently depending on other evidence, such as whether the user was aware of the full complement of choices the platform offered.<sup>113</sup> Additionally, departure from a default formatting choice may be appropriately weighted as more demonstrative of a user's intent than using the default formatting option.<sup>114</sup> But where this

---

<sup>111</sup> Compare Paul Lukas, *All-Caps Typography Is Doomed*, NEW REPUBLIC (June 23, 2013), <https://newrepublic.com/article/113578/using-all-caps-worst-form-emphasis> [<https://perma.cc/EQ9J-B3T3>] ("The standard criticism is that using all-caps is akin to shouting."), with u/Voyager5589, *How Do We Still Not Have Italics in iMessage?*, REDDIT, [https://www.reddit.com/r/apple/comments/8d71xp/how\\_do\\_we\\_still\\_not\\_have\\_italics\\_in\\_imessage](https://www.reddit.com/r/apple/comments/8d71xp/how_do_we_still_not_have_italics_in_imessage) [<https://perma.cc/4A2V-4A2V>] (last visited Nov. 15, 2019) (lamenting that in the absence of other forms of emphasis, users of the iMessage messaging platform are "left with just using all caps"); see also Gretchen McCulloch, *The Meaning of All Caps—in Texting and in Life*, WIRED (July 23, 2019, 9:00 AM), <https://www.wired.com/story/all-caps-because-internet-gretchen-mcculloch> [<https://perma.cc/28N3-5ZGB>] ("Typewriters and early computer terminals . . . wouldn't let you type italics and underlines or change font sizes (for that matter, many social media sites still don't). This created a vacuum into which the preexisting but relatively uncommon shouty caps expanded.").

<sup>112</sup> Geneus, *supra* note 7, at 451.

<sup>113</sup> It is not implausible that a software update could add new default settings to a communication platform without making users aware. See, e.g., John Patrick Pullen, *The Ultimate Guide to Apple's New Messages App*, TIME (Oct. 18, 2016), <http://time.com/4534887/apple-messages-imessage-tips-tricks/> [<https://perma.cc/9RVL-748T>] (noting that iPhone users might have noticed "a surprise or two" in the text messages they sent after updating to a new version of the iPhone operating system, including messages automatically accompanied by "a burst of confetti," and explaining this and other new features to users who were not otherwise made aware of them). In fact, in researching this topic in the spring of 2019, this author learned for the first time of several extratextual features her iPhone has been capable of executing since 2016.

<sup>114</sup> Take the example of using an emoji of a running human, and the choice to use a version of that emoji with dark brown skin rather than either the "neutral yellow" emoji or any other of "various shades of white and brown." See Geneus, *supra* note 7, at 431-32, 451 (internal quotation marks omitted). Choosing an emoji with any of the more realistic skin tones requires an additional step in the typing process, where the cartoonish yellow emojis are made the default choice. Matt Klein, *How to Change Emoji Skin Tones on iPhone and OS X*, HOW-TO GEEK (June 30, 2016, 12:34 PM), <https://www.howtogeek.com/260800/how-to-change-emoji-skin-tones-on-iphone-and-os-x> [<https://perma.cc/JE9D-4LKA>]; see also Zara Rahman, *The Problem with Emoji Skin Tones That No One Talks About*, DAILY DOT (Dec. 18, 2018, 5:50 AM), <https://www.dailydot.com/irl/skin-tone-emoji>

information may add further context and further elucidate the reasons behind a user's expressive social media content, it should also be considered as relevant to determining her subjective intent in disseminating that content.

#### IV. THE NEED FOR SPECIFICITY IN PROSECUTING TRUE THREATS

Lower courts' attempts to apply subjective-intent requirements to Internet statements, whether related to charges brought under § 875(c) after *Elonis* or under other statutes, so far reveal a lack of preparation for the charge of fully weighing evidence about a defendant's subjective intent.

##### A. *Elonis on Remand*

On remand from the Supreme Court, following the Court's announcement that 18 U.S.C. § 875(c) encompassed a subjective-intent requirement, the Third Circuit in *United States v. Elonis* reviewed the record again to consider whether its prior application of the statute was nevertheless harmless.<sup>115</sup> The court assessed the evidence adduced at Elonis's trial, in light of the jury instruction prescribed by the Supreme Court: that the jury could convict if it found that Elonis "transmit[ed] a communication for the purpose of issuing a threat, or with knowledge that the communication w[ould] be viewed as a threat."<sup>116</sup> Any error would be harmless if a jury properly instructed under the new standard could nevertheless have found Elonis guilty, because the facts shown at trial established Elonis's subjective intent such that no rational jury could find that Elonis had committed the relevant acts without also finding the required mens rea.<sup>117</sup>

The Third Circuit decided that under either a knowledge or recklessness standard, a jury faced with the evidence presented at Elonis's trial would have convicted him on each of four counts of violating 18 U.S.C. § 875(c).<sup>118</sup> The first of these counts was based on threatening statements directed at Elonis's ex-wife, which he made on Facebook over the course of slightly more than a month.<sup>119</sup> In concluding that no rational juror would have believed Elonis had he testified to a lack of knowledge or intent that his posts would make his ex-wife feel threatened, the court relied on the "graphic nature" of the messages,

---

[<https://perma.cc/62ZT-2LS4>] (discussing some of the reasons a user might choose an emoji of a particular skin color and how such choices can affect the message received by the sender's audience).

<sup>115</sup> 841 F.3d 589, 597-601 (3d. Cir. 2016).

<sup>116</sup> *Id.* at 596 (quoting *United States v. Elonis*, 135 S. Ct. 2001, 2012 (2015)).

<sup>117</sup> *Id.* at 598. The Third Circuit cited Justice Alito's concurring opinion in *Elonis* as the impetus for its consideration of harmless error on remand. *Id.* at 596 (citing *Elonis*, 135 S. Ct. at 2018 (Alito, J., concurring in part and dissenting in part)).

<sup>118</sup> *Id.* at 598.

<sup>119</sup> *Id.*



reproducing the text of the posts in full.<sup>120</sup> But the court also referred to developments in the real world between the two posts, specifically that Elonis's ex-wife sought a restraining order after his first allegedly threatening approach, and that he continued to post similar messages after a restraining order was issued against him—including one referencing the restraining order.<sup>121</sup> The court found that Elonis's decision to continue with similar posts following the restraining order made it “less credible still” that he had disseminated the post without either the purpose or knowledge that his ex-wife would interpret the posts as threats—and that “[n]o rational juror could conclude otherwise.”<sup>122</sup>

Similarly, in affirming Elonis's conviction on another count, for threatening to injure employees of the Pennsylvania State Police and a county sheriff's department, the court quoted language from Elonis's posts that gave rise to the indictment.<sup>123</sup> It held that Elonis's experience with the threats he had made about his ex-wife—and other, previous threats made on Facebook toward former co-workers that had resulted in his termination—made it more likely that a jury would find that Elonis knew that a reasonable person could interpret these posts, too, as threats.<sup>124</sup> In so concluding, the court rejected Elonis's arguments that the form of his posts—rap lyrics—and the forum through which he posted them—Facebook, which he argued was “a medium that magnifies the potential for disconnect between the speaker's intent and the audience's understanding”—would have allowed a jury to find that Elonis did not know that his posts would be taken as threats.<sup>125</sup> Because Elonis had already seen the effect that earlier lyrical Facebook posts had on his ex-wife, and had witnessed the consequences of those posts in the form of a restraining order issued against him, the court concluded that Elonis was “clearly aware” of how his audience would interpret this subsequent post.<sup>126</sup>

---

<sup>120</sup> *Id.* at 593-94, 599.

<sup>121</sup> *Id.* at 598.

<sup>122</sup> *Id.* at 599.

<sup>123</sup> *Id.* at 599-600.

<sup>124</sup> *Id.* at 600.

<sup>125</sup> *Id.*

<sup>126</sup> *Id.* The court also found that the language Elonis used in this later threat against law enforcement officials was “[i]f anything, . . . more explicit” than those that had resulted in his termination and the restraining order. *Id.* at 599. The court did not explain this conclusion in further detail, but it would seem to contradict the paradigmatic reasoning in *Watts* that a conditional statement is less objectively threatening than a statement not couched in conditional terms. *See id.* (“Despite that, he posted yet another violent message stating his intention to detonate explosives near State Police officers and the Sheriff's Department if ‘worse comes to worse.’”); *cf. Watts v. United States*, 394 U.S. 705, 708 (1969) (“We agree with petitioner that his only offense here was ‘a kind of very crude offensive method of stating a political opposition to the President.’ Taken in context, and regarding the expressly conditional nature of the statement and the reaction of the listeners, we do not see how it could be interpreted otherwise.”).

As to the third count of violating § 875(c), which charged Elonis with communicating a threat to injure a kindergarten class of elementary school children, the court relied solely on the language Elonis used and the context of the statement.<sup>127</sup> The court found Elonis's language "graphic and specific" and "narrow[ed] . . . to kindergarten classes"; in particular, the court was troubled by the post's conclusion, a "haunting question that suggests he will carry out his threat imminently."<sup>128</sup> The court then assessed that post in the context of "the understandable sensitivity regarding school shootings in this country," of which it assumed without evidence Elonis had some awareness.<sup>129</sup> On the fourth count of violating § 875(c), the court returned to relying on Elonis's prior experiences with his Facebook posts as "overwhelming[]" evidence that Elonis knew how his post that purportedly threatened an FBI agent would be interpreted.<sup>130</sup>

The *Elonis* remand decision, therefore, shows a court that engages inconsistently with the types of social media evidence outlined in Part III. On the majority of the counts it reviewed, the court looked to the real-world responses to Elonis's posts, ranging from the grant of a protective order against him to a visit from FBI agents, to demonstrate that it would have been impossible for Elonis *not* to know that his similar, subsequent Facebook posts could be viewed as threats.<sup>131</sup> This conclusion relies on similar reasoning to my suggestion that subsequent comments provide an interpretive frame for a post.<sup>132</sup> It arguably makes an even stronger showing to the same end. Where a person making a Facebook post might not see subsequent comments on that post,<sup>133</sup> Elonis was visited by officers in person and therefore could not claim that he was not on notice of the reactions to his lyrical threats.<sup>134</sup> This overwhelming evidence based on real-world interactions outweighed the persuasive weight of Elonis's platform-based argument about the uncertain nature of Facebook posts and his context-based argument about the uncertain status of threatening statements made in rap music.<sup>135</sup> Because he had previously made similar statements on the same platform using a similar

---

<sup>127</sup> *Elonis*, 841 F.3d at 600.

<sup>128</sup> *Id.* The court reproduced the post at issue in its entirety before this analysis, and it follows here: "That's it, I've had about enough / I'm checking out and making a name for myself / Enough elementary schools in a ten mile radius to initiate the most heinous school shooting ever imagined / And hell hath no fury like a crazy man in a kindergarten class / The only question is . . . which one?" *Id.*

<sup>129</sup> *Id.*

<sup>130</sup> *Id.* at 600-01. Though the court again reproduced the relevant text in full, its analysis did not elaborate at all on the import of the language or format of the post.

<sup>131</sup> See *supra* notes 119-126 and accompanying text (outlining the court's findings on these grounds).

<sup>132</sup> See *supra* notes 93-94 and accompanying text (discussing the way responses clarify the meaning of an original post).

<sup>133</sup> See *supra* text accompanying notes 93-94.

<sup>134</sup> *Elonis*, 841 F.3d at 600.

<sup>135</sup> *Id.*

lyrical form, and had received feedback that members of his social circle found posts phrased in that manner threatening, a reasonable jury could have found that he acted with knowledge that similar reactions would result from his similar, subsequent statements.

The decision on remand falls short of persuasive, however, in its reexamination of the count charging Elonis with threatening to injure a class of kindergartners. The decision mentions the “graphic nature” of Elonis’s language and the “haunting question” with which the post ends, but goes no further in articulating what evidentiary value this language could have to a jury determining whether Elonis knew or intended the post to be a threat.<sup>136</sup> It is not enough under a subjective standard to simply *identify* the nature of the language in a social media threat. At most, a judge’s pronouncement that a post contains “graphic” language speaks to the perception of a reasonable observer—effectively duplicating the objective-intent standard, reading the specific, necessary protections of a subjective-intent requirement out of the true-threat analysis.<sup>137</sup> As mentioned above, there is no evidence that definitively links any specific word or phrasing choice to a specific mental state; however, courts should require *some* reasoning in accepting a prosecutor’s contention that a defendant’s use of a word or phrase is probative as to that individual defendant’s state of mind.

#### B. Other Courts’ Application of Subjective-Intent Standards

In the limited number of other cases where courts have considered alleged threats made via social media under a subjective-intent standard, they have approached social media evidence much like the Third Circuit in its decision in *Elonis* on remand. Courts familiar with principles of notice and fair warning are almost universally comfortable with assessing when a defendant’s prior conduct, particularly his prior online speech and the reactions of others to that speech, should have warned him that similar speech acts would be received by audiences as threatening. In *United States v. Dutcher*, a defendant who announced on his Facebook page his intent to travel to a nearby town where the President was scheduled to speak and assassinate the President was convicted under a subjective-intent standard.<sup>138</sup> His conviction was upheld in part because, rather than recanting the threatening statement when Facebook friends commented, expressing worry and urging him to “[s]tay calm,” the defendant continued to post Facebook updates in a similar vein.<sup>139</sup> Although

---

<sup>136</sup> *Supra* text accompanying notes 128–130.

<sup>137</sup> *See supra* notes 81–88, 93–94 & 105 (describing how different social groups use language differently, including using the same phrase or action to mean different things).

<sup>138</sup> 851 F.3d 757, 760–61 (7th Cir. 2017).

<sup>139</sup> *Id.* at 762.

the commenters' expressions of alarm in *Dutcher* were not so severe a reaction as the law-enforcement responses in *Elonis*, where the FBI monitored the *Elonis*'s Facebook account and visited him at his house,<sup>140</sup> the comments on the defendant's posts in *Dutcher* were nevertheless sufficiently clear in demonstrating alarm that the defendant's subsequent posts could be considered knowing.<sup>141</sup>

Even in conducting this relatively familiar analysis, though, some courts are unclear as to the function of evidence of an audience's reaction,<sup>142</sup> which can be admitted in different contexts to demonstrate both objective and subjective components of a true threat.<sup>143</sup> In integrating the more novel

---

<sup>140</sup> *Elonis v. United States*, 135 S. Ct. 2001, 2006 (2015).

<sup>141</sup> *Dutcher*, 851 F.3d at 762; see also *State v. Taupier*, 193 A.3d 1, 25 (Conn. 2018) (“[I]f the defendant had been unaware when he sent the first e-mail that it would be interpreted as a serious threat, he would have reacted quite differently to [his acquaintance’s] characterization of the e-mail as ‘disturbing’ and his admonition to the defendant to refrain from making such statements.”). But see *People v. Khan*, 127 N.E.3d 592 (Ill. Ct. App. 2018). In the *Khan* case, the prosecution relied on the fact that the defendant, like *Elonis*, had previously been visited by police investigating threatening statements he had made online. *Id.* at 595. Both posts reported as threatening were made on Facebook, and the fact that police had interviewed the defendant after a 2010 post could reasonably have called into question the defendant’s credibility in insisting, in 2013, that “[e]veryone knows if you post something on Facebook it’s a joke.” *Id.* at 595-96. But in upholding the conviction, the Illinois Appellate Court did not refer to this evidence of past conduct in its analysis. See *id.* at 603. Instead it relied on the content of the defendant’s message, “which told people that he came to school every day with a gun and was going to use it on somebody,” the fact that fear was a foreseeable consequence of posting such a message from an anonymous Facebook account, and holding that juries were entitled to infer that the defendant “intended the natural and probable consequences of his act.” *Id.*

<sup>142</sup> Compare *Commonwealth v. Knox*, 190 A.3d 1146, 1159-60 (Pa. 2018) (relying heavily on the reactions of the police officers named in an allegedly threatening rap song once they had discovered the song), with *United States v. LaFontaine*, No. 15-77, 2015 WL 5999834, at \*3 (N.D. Iowa Oct. 14, 2015) (excluding on relevancy grounds evidence showing how recipients had responded to the defendant’s phone calls, on the basis that “[a] recipient’s reaction to an alleged threat is not relevant to the charge of communicating threats” after *Elonis*).

<sup>143</sup> As *Elonis* and *Dutcher* demonstrate, evidence of an audience’s reaction to *prior* purported threats can be highly relevant as to a defendant’s subjective intent in making *subsequent* statements understood as threats. See *supra* notes 121-126, 131-135, 138-141, and accompanying text. But an audience reaction that *follows* a threat can only be relevant to whether a reasonable observer would find the defendant’s statement threatening, as persuasive evidence that at least some people did (or did not) find the statement to articulate a threat. To find the latter type of reaction evidence relevant to a defendant’s subjective intent would assume that a defendant can anticipate with a high degree of accuracy her audience’s eventual responses. But to exclude the former type of reaction evidence (or its absence, as the case may be) could prevent a defendant from demonstrating that similar past statements had given her no reason to know that certain language or expressions could be interpreted as threatening. Cf. *United States v. Bagdasarian*, 652 F.3d 1113, 1122 (9th Cir. 2011) (“Nobody who read the message board postings, however, knew that he had a .50 caliber gun or that he would send the later emails. Neither of these facts could therefore, under an *objective* test, ‘have a bearing on whether [Bagdasarian’s] statements might reasonably be interpreted as a threat’ by a reasonable person in the position of those who saw his postings . . . .” (quoting *United States v. Parr*, 545 F.3d 491, 502 (7th Cir. 2008))). The lack of clarity in the standard regarding audience reaction evidence can inhibit defendants from making their strongest cases, either way.

features of Internet communication in their analysis of subjective intent, courts are even more inconsistent. The fact that a speaker may know concretely which individuals make up his online audience, based on the privacy settings applied to his social media accounts, can be a potent tool for discovering whether a defendant intended his online message to reach an individual named in the message—or indeed, whether he intended his message to reach anyone at all.<sup>144</sup> In *United States v. Wheeler*, the Tenth Circuit overturned a conviction because the defendant had testified credibly that he thought he had removed all of his Facebook “friends,” rendering his profile completely private, before posting messages “urging his ‘religious followers’ to ‘kill cops.’”<sup>145</sup> *Wheeler*, then, stands clearly for the proposition that under a subjective standard, a threat made by a defendant who believes he has no audience at all cannot be a punishable true threat.

But where a defendant has posted a message in a forum where she knows that *someone* might see it, courts’ approaches vary widely. In *United States v. Stock*, the Third Circuit assumed without explanation that an anonymous message posted on the website “Craig’s List,” identifying a target by only his initials and his profession, could constitute a true threat.<sup>146</sup> This is not a self-evident conclusion: “Craig’s List” (or “Craigslist”) is not a social media platform like Facebook or Twitter, in which users create stable profiles and receive notifications when other users interact with those profiles or associated content.<sup>147</sup> Instead, Craigslist functions more like an online classified ads section, where users post standalone advertisements seeking or offering furniture, housing, or companionship,<sup>148</sup> or simply expressing themselves;<sup>149</sup> any communication takes place off the site through users’ existing communication channels.<sup>150</sup> Additionally, “[p]eople usually only go to Craigslist when they want something specific,”<sup>151</sup> as opposed to a social networking site which a user might check daily. The court in *Stock* considered none of these distinctions, apparently finding irrelevant to the intent element

---

<sup>144</sup> See *supra* text accompanying notes 89–90 (highlighting the potential for misunderstanding when unintended audiences see a post without understanding its conversational context).

<sup>145</sup> 776 F.3d 736, 738–39, 741 (10th Cir. 2015).

<sup>146</sup> 728 F.3d 287, 299 (3d Cir. 2013).

<sup>147</sup> See Josh Constine, *Facebook Launches Marketplace, a Friendlier Craigslist*, TECHCRUNCH (Oct. 3, 2016, 8:01 AM), <https://techcrunch.com/2016/10/03/facebook-marketplace-2> [<https://perma.cc/2HYX-WVHS>] (contrasting the then-new Facebook Marketplace, an online buy-sell forum associated with users’ Facebook accounts, with the anonymous, “dead-simple” Craigslist platform).

<sup>148</sup> See generally *Philadelphia*, CRAIGSLIST, <https://philadelphia.craigslist.org> [<https://perma.cc/PN44-24H6>] (last visited Dec. 17, 2019).

<sup>149</sup> See, e.g., *Rants and Raves*, CRAIGSLIST, <https://philadelphia.craigslist.org/d/rants-raves/search/rnr> [<https://perma.cc/6NKT-HDXG>] (last visited Dec. 17, 2019).

<sup>150</sup> *How to Reply to Craigslist Postings*, CRAIGSLIST, [https://www.craigslist.org/about/help/replying\\_to\\_posts](https://www.craigslist.org/about/help/replying_to_posts) [<https://perma.cc/GL4F-SCN2>] (last visited Dec. 17, 2019).

<sup>151</sup> Constine, *supra* note 147.

whether the defendant anticipated that his target would ever visit Craigslist and discover the message.<sup>152</sup> Taking the opposite approach, the Sixth Circuit in *United States v. Jeffries* looked to the defendant's distribution of his music video with messages like "[g]ive this to the judge for court," taking this statement of intent that the video reach the judge as part of a "single communication" with the video itself.<sup>153</sup>

In drawing conclusions from the specific language a defendant used, several courts take the approach taken by the Third Circuit in the remanded *Elonis* decision in concluding, essentially, that the language was received as threatening and it was therefore intended as a threat.<sup>154</sup> In *Knox*, the Pennsylvania Supreme Court drew a comparison to *Watts* in noting that the language of the statements at issue was "mostly unconditional," but did not go further to explain how the nonconditional statements actually demonstrated an intent that they be interpreted as threats.<sup>155</sup> In *D.C. v. R.R.*, a civil suit involving high school students, a California court directly quoted the language of the posts at issue and concluded that the "message [was] unequivocal" in its "serious expression of intent to inflict bodily harm," "conveyed no less than three times by the phrases 'rip out your fucking heart,' 'want[] to kill you,' and 'pound your head in with an ice pick.'"<sup>156</sup> Over the defendant's contention that he had been "in a playful mood" when he posted the messages and intended them as a joke, the court concluded that on the face of the posts there was no "jocular intent"; instead, "[a]n intent to harm [was] evident."<sup>157</sup> This tautological approach to the subjective-intent standard

---

<sup>152</sup> See generally *United States v. Stock*, 728 F.3d 287 (3d Cir. 2013).

<sup>153</sup> 692 F.3d 473, 482 (6th Cir. 2012); see also *Commonwealth v. Knox*, 190 A.3d 1146, 1160 (Pa. 2018) (concluding that although the defendant had not sent the allegedly threatening song directly to the police officers named in it, this did not preclude the defendant from having intended the officers to hear it); cf. *United States v. Patillo*, 431 F.2d 293, 297 (4th Cir. 1970) (requiring a heightened showing of imminent harm where a threat to the President is "uttered without communication to the President intended," because the rationale behind punishing such threats is not implicated when the President neither learns of the threat nor is put in imminent danger because of it).

<sup>154</sup> See *supra* text accompanying notes 138–139.

<sup>155</sup> *Knox*, 190 A.3d at 1159–60.

<sup>156</sup> 182 Cal. App. 4th 1190, 1219 (Cal. Ct. App. 2010).

<sup>157</sup> *Id.* at 1206, 1220–21. The opinion in *D.C. v. R.R.* also assumed without any apposite supporting facts that the defendant had "composed" the statements at issue "over a period of at least several minutes," concluding on this basis that "the content of the message and its transmission show[ed] deliberation on the part of the author." *Id.* at 1219. When the Ninth Circuit came to a similar conclusion regarding the weight of the speech in *Bagdasarian*, it did so on the basis of specific findings that demonstrated that the defendant had continued to monitor reactions to his speech—and that he subsequently took up the same conversation weeks later. *United States v. Bagdasarian*, 652 F.3d 1113, 1130 (9th Cir. 2011). Without such specific findings, the court in *D.C.* relied only on its own assumptions about Internet use, without evidence and contrary to the views of other courts in similar contexts, which come to the conclusion that the process behind posting online speech is less deliberative and more casual than speech in other mediums. *D.C.*, 182 Cal. App. 4th at 1219. *But see, e.g.,* *SI03, Inc. v. Bodybuilding.com, LLC*, No. 07-6311, 2008 WL 11348459, at \*5 (D. Idaho Dec.

is not universal: other courts conduct more thorough analyses of defendants' specific uses of language in context, providing reasoning for the conclusions about subjective intent they then draw.<sup>158</sup> But the fact that this intensive context-based intent analysis is applied selectively—or at the very least, inconsistently—poses a problem for the vindication of First Amendment rights of defendants whose speech may be afforded greater First Amendment protection based on either their own identities or the identities of the judges who hear their cases.

#### V. ADMITTING MORE CONTEXTUAL EVIDENCE TO FACILITATE MORE ACCURATE DETERMINATIONS OF INTENT

In order to improve their analyses of defendants' subjective intents, courts should be encouraged to interpret relevance rules liberally and admit more social media evidence. In the examples of reasoning from evidence discussed above, the social media evidence courts failed to consider was generally of the sort that, if admitted, would have been helpful to the defendant in providing more context. Especially in the early days of prosecutions under subjective-intent standards, and in light of the evidence in this Comment that courts have not yet fully appreciated the probative effects of contextual evidence, courts should err on the side of speech protection in admitting contextual evidence. The associated risks of confusion and delay are not insubstantial, but even a temporary period of over-admitting contextual evidence would provide appellate courts with the opportunity to observe outcomes and provide guidance for lower courts applying the new standard.

The argument for adopting a liberal posture toward the admission of contextual evidence is further supported by the principle that intent is a question of fact, best resolved by juries.<sup>159</sup> Even where judges are the primary

---

23, 2008) (“Internet blogs, message boards, and chat rooms are, by their nature, typically casual expressions of opinion.”).

<sup>158</sup> See, e.g., *State v. Carroll*, 196 A.3d 106, 121 (N.J. Super. 2018) (concluding in the absence of any specific threatening language that the defendant had merely “expressed disdain” for the party named in her posts and instructing the prosecution that a conviction would require showing that specific “cohorts or allies of [the] defendant would understand her expression of hope [that someone “blow’ the[] glasses . . . off his face”] as a request or command and would act on it; and that [the] defendant intended that reaction”); see also *Bagdasarian*, 652 F.3d at 1121 (concluding that the context of the defendant’s statements about the President, on a financial news message board, “blunt[ed] any perception that statements made there were serious expressions of intended violence”).

<sup>159</sup> See *United States v. Yassin*, No. 16-03024-01, 2017 WL 1324141, at \*4 (W.D. Mo. Feb. 23, 2017) (report and recommendation of magistrate judge), adopted by No. 16-3024-01 (Apr. 6, 2017) (“[T]he question of whether the allegations . . . constitute a true threat is a question reserved for the jury and thus would seem premature at [the motion to dismiss] stage . . . .”); accord *Universal Calvary Church v. City of New York*, No. 96-4606-99, 2000 WL 1538019, at \*22 (S.D.N.Y. Oct. 17, 2000) (“The question of intent is a question of fact for the jury.” (citing *Casimir v. Hoffman*, 213 N.Y.S.2d 499, 500 (N.Y. App. Div. 1961))).

arbiters on the question of intent, more evidence can only contribute to better-informed decisions. Whether a judge or jury will make the ultimate findings of fact, though, judges must also be careful to emphasize the analytical distinction between the evidence that supports the subjective-intent finding and evidence that demonstrates how an observer would have reacted.<sup>160</sup> A clear delineation between these separate findings is necessary: if factfinders are not clearly instructed on these standards, the reasonable-observer evidence may elide into evidence of what the individual speaker actually intended, reinstating the underprotective, tautological reasoning identified as problematic in Part IV.<sup>161</sup>

As an alternative, Professors Lidsky and Norbut have proposed the creation of an affirmative defense for context, which in their explanation would allow defendants, who are in the best position to know what context will be helpful in persuading juries that they lacked intent, to best rebut a prosecution's showing of intent to threaten.<sup>162</sup> Along with this defense, they would add a preliminary hearing before the trial judge where a defendant could present his contextual evidence and might get an indictment dismissed quickly if the context reveals that he did not intend to make a threat.<sup>163</sup> But together, these proposals eliminate the protective elements of criminal procedure without offering defendants much in return. First, intent is an element of any statute criminalizing threats, so while defendants may be in the best position to *produce* this evidence, the burden to persuade on the issue of intent properly belongs with the prosecution. And second, there is no reason to expect that a pretrial hearing on subjective intent will be resolved in the defendant's favor any more often than the post-trial and remanded decisions evaluated in Part IV. Instead of providing defendants with an early opportunity to vindicate their rights, then, this procedure could become just one more procedural hurdle to clear before a defendant can make her case to a jury, the members of which are somewhat more likely to be of her peer group than the average judge.<sup>164</sup> Instead, making clear that contextual

---

<sup>160</sup> See, e.g., *supra* note 143 (explaining that evidence of audience reaction can be relevant to both standards, depending on the sequence of events at issue).

<sup>161</sup> Recent Case, *First Amendment—True Threat Doctrine—Pennsylvania Supreme Court Finds Rap Song a True Threat*—Commonwealth v. Knox, 190 A.3d 1146 (Pa. 2018), 132 HARV. L. REV. 1558, 1564 (2019) (“In using the listeners’ reactions to Knox’s speech to discern Knox’s subjective intent, the court rendered its insistence on finding specific [i.e., subjective] intent meaningless. . . . The emphasis in *Knox* on the listeners’ fearful reactions . . . vitiated the subjective prong of the inquiry.” (internal footnote omitted)). But see *supra* note 143 and accompanying text (positing that a properly framed investigation of listener reactions can serve as evidence of a speaker’s intent in subsequent statements).

<sup>162</sup> Lidsky & Norbut, *supra* note 7, at 1925-26.

<sup>163</sup> *Id.*

<sup>164</sup> Compare Shannon Greenwood, Andrew Perrin & Maeve Duggan, *Social Media Update 2016*, PEW RESEARCH CTR. (Nov. 11, 2016), <https://www.pewinternet.org/2016/11/11/social->



evidence is relevant to the issue of intent, but leaving the burden of proving intent with the prosecution and the ultimate decision to a jury whenever practicable, will provide more consistent and more conceptually coherent protection for defendants who made threatening statements without the requisite intent under a properly recalibrated true threats doctrine.

Neither the minimal adjustments proposed here nor the more significant shift outlined above, to a recognition that the First Amendment requires subjective intent to be a necessary element of a threat conviction, is likely to be outcome-determinative in a significant number of cases. In *Elonis*, for example, even under the stringent examination of contextual evidence and evaluation of subjective intent I propose, there appears to be sufficient evidence to convict the defendant of at least two counts of violating 18 U.S.C. § 875(c). But the best argument for announcing a clear, uniform standard in this context is not that it offers any individual defendant a greater chance of acquittal. Instead, there is additional systemic value in requiring prosecutors, juries, and judges to rely on evidentiarily based, logically relevant evidence, and to give accurate, non-pretexual reasons, when they argue for and impose punishments for speech. Instead of appealing to jurors' own biases as to which, if any, words are beyond the pale in polite society, prosecutors and defenders should be given incentives to draw on social media archives, where available, and craft arguments about subjective intent that rely on how a defendant has spoken in the past, how his interlocutors have responded, and what kinds of communicative media have been accessible to him—in order to show what he actually knew and meant to say. Requiring such analytical clarity in prosecutions for speech would be a step in the direction of ensuring that speech is not punished for its bad effects or based on potentially biased conjecture about a speaker's mental state, but only where a speaker demonstrably intended to cause harm and actually did so.

---

media-update-2016/ [https://perma.cc/Q6K5-CNNC], *with Age and Experience of Judges*, FED. JUD. CTR., <https://www.fjc.gov/history/exhibits/graphs-and-maps/age-and-experience-judges> [https://perma.cc/8KLK-WBHX] (last visited Dec. 18, 2019). Of course, age is not the only factor that determines a person's knowledge and patterns of usage of the Internet and social media. See, e.g., Nicole Zillien & Eszter Hargittai, *Digital Distinction: Status-Specific Types of Internet Usage*, 90 SOC. SCI. Q. 274, 287 (2009) (finding that "social status is significantly related to various types of capital-enhancing uses of the Internet, suggesting that those already in more privileged positions are reaping the benefits of their time spent online more than users from lower socioeconomic backgrounds"). But on most relevant axes, a multi-member jury will be more representative of the general population than an individual judge.

## CONCLUSION

The true threat doctrine has developed for over a decade without an adequate review of its operation within the context of the First Amendment. In that time, states and federal jurisdictions have adopted a variety of approaches to filling in the gap left by the Supreme Court's decisions as to what intent the Constitution requires. Judges at all levels have also been left without authoritative guidance as to the added complications of applying the less-than-clear old decisions to the new contexts of social media. The result, as outlined above, is a system of overlapping, inconsistent rules, under which one individual may be liable for an act that she may not have considered to have any expressive value at all, while another individual who expressed a similar sentiment might be found innocent—based on geographic location, the prosecuting authority, and even more specific factors like an individual judge's attitude toward and experience with social media.

To begin to untangle the complicated patchwork of true threat standards, the Supreme Court should act, at the next opportunity, to reestablish some uniformity in the true threat doctrine by announcing that the true threat standard requires measured, informed consideration of the context in which a speech act occurs, to assess whether a speaker actually intended to threaten harm. In the meantime, though, judges and juries who are already required to evaluate threatening statements under a subjective-intent standard may take it upon themselves to use insights from social science and media research to more accurately assess intent. In requiring that factfinders rely on sufficient, logically relevant evidence when they determine that a defendant subjectively intended to make a threat, this limitation on the true threat exception increases protections against arbitrary or unjustified punishment for speech, increasing the likelihood that criminal consequences for speech are imposed only where some actually culpable conduct has been shown.