THE JOURNAL OF
DIGITAL FORENSICS,
SECURITY AND LAW

Journal of Digital Forensics,
Security and Law

Volume 11 | Number 3 Article 7

9-30-2016

Toward Online Linguistic Surveillance of Threatening Messages

Brian H. Spitzberg
San Diego State University, California, USA.

Jean Mark Gawron
San Diego State University, California, USA.

Recommended Citation

Spitzberg, Brian H. and Gawron, Jean Mark (2016) "Toward Online Linguistic Surveillance of Threatening Messages," *Journal of Digital Forensics, Security and Law*: Vol. 11 : No. 3 , Article 7.

DOI: https://doi.org/10.15394/jdfsl.2016.1418
Available at: http://commons.erau.edu/jdfsl/vol11/iss3/7

TOWARD ONLINE LINGUISTIC SURVEILLANCE OF THREATENING MESSAGES

Brian H. Spitzberg
San Diego State University
California, USA

Jean Mark Gawron
San Diego State University
California, USA

ABSTRACT

Threats are communicative acts, but it is not always obvious what they communicate or when they communicate imminent credible and serious risk. This paper proposes a research- and theory-based set of over 20 potential linguistic risk indicators that may discriminate credible from non-credible threats within online threat message corpora. Two prongs are proposed: (1) Using expert and layperson ratings to validate subjective scales in relation to annotated known risk messages, and (2) Using the resulting annotated corpora for automated machine learning with computational linguistic analyses to classify non-threats, false threats, and credible threats. Rating scales are proposed, existing threat corpora are identified, and some prospective computational linguistic procedures are identified. Implications for ongoing threat surveillance and its applications are explored.

Keywords: risk assessment, computational linguistics, cyber-harassment, threats

1. THREATENING COMMUNICATION

Since 911, the threat of terrorism has become ubiquitously ingrained in the minds of governments and the publics they represent and seek to protect. Between a third and half of the U.S. population is worried about themselves or a family member being victimized by terrorism (http://www.gallup.com/poll/4909/terrorism-united-states.aspx). At the same time, societies have become increasingly sensitized to interpersonal acts of terrorism, ranging from intimate partner violence, sexual coercion and rape, hate speech, slurs, micro-aggressions, psychological abuse, bullying, stalking, cyber-harassment, trolling, bombings and mass shootings. As diverse as this landscape of interpersonal and institutional terrorism is, at least one form of speech act is common, although neither unique nor necessary, to all of these forms of aggression: the act of threatening communication.

In this analysis, the focus is on making threats, as opposed to posing a threat or creating a sense of threat. For example, many individuals and groups may pose a threat to cyber-security, but may not communicate a threat indicating intent to enact harm

(Fachkha et al., 2012). Other research examines how to create a sense of threat or fear in others (Peters, Ruiter, & Kok, 2013; Peters, Ruiter & Kok, 2014). Further, although a variety of forms and contexts of threats are reviewed, the focus of this analysis will be on *interpersonal* threats—threats expressed by one person to another person. Finally, the interest of this analysis is more on identifying linguistic threats themselves, rather than identifying the threatener (see: Abbasi & Chen, 2008; Hadjidj et al., 2009), even though each of the two approaches may have much to offer the other. The purpose of this analysis is to examine the nature of expressed threats, identify some of their linguistic features potentially amenable to machine learning, and to provide a rating scale that could assist in validating training sets of threats for such machine learning and classification. To the extent that a reasonably accurate threat surveillance system could be developed, it could significantly enhance the identification, assessment, and potential interventions associated with existing case-based threat management situations.

There are laws proscribing communicated threats as a particular form of unprotected speech (e.g., 18 U.S.C. § 875(c)). The Supreme Court, however, when assessing the legality and seriousness of threats, tends toward an abundance of caution in regulating such speech. In recent cases on threats in electronic media (e.g., Watts v. United States, and Elonis vs. United States), the court has ruled that there is an objective intent standard burden of proof—the prosecution has an expectation to demonstrate a *mens rea* requirement that the communicator intended the message as a threat, and that it would be understood by the target as a threat (Maras, 2015). Yet, the Court has not specified the standard for determining what communication features constitute a "true threat." In so doing, the Court left the seriousness of threats to be determined by their context and the subjective intent of the speaker.

Threat is often associated with fear, dread, terror, anxiety, and apprehension (Shen & Dillard, 2014), and the sense of threat is no doubt an evolved sensitivity with adaptive value and neural substrates (Pichon, de Gelder & Grèzes, 2009). When uttered or expressed, however, threats are a *prima facie* indicator of risk, although not everything perceived as a threat or as threatening is predicated on an intentional threat message (Rick, Mania, Gaertner, McDonald, & Lamoreaux, 2010; Smith & Morra, 1994; Spitzberg, in press; Surface, 2011). Threats reveal a complex relationship to actual harm. From a forensic perspective, threats are generally understood as a risk indicator of potentially violent, criminal or terrorist behavior (Meloy, Hoffmann, Guldimann, & James, 2012). As Meloy (2000) pessimistically summarizes: "Most individuals do not act on their threats. Threats may increase, decrease, or have no relationship to subsequent violence" (p. 166). The degree to which threats are systematically predictive of violence varies from context to context. For example, threats are more associated with violence in workplace, school, and intimate relationships than in public figure contexts (Jenkins, 2009). Even among attacks on college campuses, threats were apparent in only 13% of attacks (U.S. Secret Service, 2010).

Other research, however, demonstrates some value of threats as predictors of subsequent violence. Threats have been identified as risk indicators of potentially violent, criminal or terrorist behavior (Meloy, et al., 2012), stalking (Churcher & Nesca, 2013; Spitzberg & Cupach, 2014), and femicide (Campbell et al., 2003; Glass, Laughon, Rutto, Bevacqua & Campbell, 2008). Studies have been conducted on threats against stalking victims (Spitzberg & Cupach, 2014), intimate partners (Brewster, 2000; Campbell, et al., 2003; Glass, Laughon, Rutto, Bevacqua & Campbell, 2008; Palarea, Zona, Lane, & Langhinrichsen-Rohling, 1999), clinicians (Brown, Dubin, Lion, & Garry, 1996), nurses (Maier, 1996), social workers (Newhill, 2002), psychotherapists (Bernstein, 1981), clinical staff (Doren, Miller, & Maier, 1993; Hillbrand, 2001; Sandberg, McNiel, & Binder, 1998), lawyers (Brown & MacAlister, 2006), organizations (Bulling & Scalora, 2008; Moore, Mundie, Collins, 2013; Mundie, Moore, & McIntire, 2012; Seger, 1993), students (Nekvasil & Cornell, 2012), schools (Bondü & Scheithauer, 2014; Borum, Cornell, Modzeleski & Jimerson, 2010; Cornell, 2010; Drysdale & Modzeleski, 2010; Lindberg, Oksanen, Sailas, & Kaltiala-Heino, 2012; Meloy, Hoffmann, Roshdi, & Guldimann, 2014; Trump, 2015; Sokolow, Lewis, Schuster, Swinton, & Van Brunt, 2015), celebrities (Dietz, Matthews, Van Duyne & Martell, 1991b; Twemlow, Fonagy, Sacco, & Vernberg, 2008; U.S. Secret Service, 2002), judges and politicians (Dietz, et al., 1991a; Every-Palmer, Barry-Walsh & Pathé, 2015; Fein & Vossekuil, 1999; Schoeneman-Morris et al., 2007), royalty (James et al., 2008; James et al., 2009a; James et al., 2009b; James et al., 2010; van der Meer, Bootsma & Meloy, 2012), and a variety of public figures (Baumgartner, Scalora & Plank, 2001; Kropp, Hart & Lyon, 2008; Meloy, 2011; Meloy et al., 2004; Schoeneman et al., 2011; Sinclair, 2009).

Certain types of threats, such as death threats (Barnes, Gordon, & Hudson, 2001; MacDonald, 1968; Morewitz, 2010) and bomb threats (Mazur, 1983; Häkkänen, 2006; Zaitsu, 2010) have particular cultural, professional and scholarly currency. Although school attackers rarely directly threaten the school, the vast majority of cases reveal information that disquiets others, usually peers, although a very small minority of such bystanders ever recognize the seriousness of such signals (Polluck et al., 2008; McCann, 2001, 2002). To the extent that such "disquieting" or signaling information is available in online contexts such as social media, surveillance of such media may provide invaluable information as a tool for recognizing and potentially avoiding such threats from becoming violent (Scalora, 2014; Vudhiwat, 2002). Such possibilities of advance threat assessment has generated an extensive scholarly and practioner interest in threat surveillance and prediction (e.g., Borum, Fein, Vossekuil, & Berglund, 1999; Davis, 2001; Davis, Stewart & Siota, 2001; Dunn, 2008; Fein, Vossekuil, Pollack & Borum, 2000, 2002; Glasgow & Schouten, 2014; Jackson, 2012; Meloy, Hoffmann, Roshdi, Glaz-Ocik, & Guldimann, 2014; Meloy, Hoffmann, Guldimann, & James, 2012; Simons & Cook, 2014; Storey, Givas, Reeves, & Hart, 2011; White & Cawood, 1998).

2. THREATS AND CHARACTERISTICS OF THREATS

Threats are a trope recognized since ancient times, which reflects a basic speech act (*perclusio*), although "threats are not necessarily, or even typically, verbal" (Salguerio, 2010, p. 215). For the purposes of this project, and in accord with the kinds of available data for analysis, only discursive and transcribable texts will be considered. Although many threats are nonverbal in nature (e.g., burning a cross on someone's lawn; sending an ominous gift or image, such as a picture of an ex-girlfriend with rifle crosshairs drawn on her face; menacing approach behavior, Crowner, Peric, Stepcic & Lee, 2005), the contingency and preferred outcome features of threats seem likely to be expressed linguistically. Furthermore, with the ubiquity and anonymizing capabilities of new media technologies, the communication of threats has become more efficient, and the potential for social status implications vastly expanded due to the potential mass-communication features of such media (e.g., threats of revealing sexting images to broader audiences; Hadnagy & Fincher, 2015).

2.1 The Structure and Themes of Threats

From a communication or pragmatics perspective, threats, even when nonverbally enacted, are distinct and varied speech acts (see: Beller, Bender, & Song, 2009; Milburn & Watman, 1981; Murdock, Bradac, & Bowers, 1984; O'Hair, Bernard, & Roper, 2011). "From a speaker-oriented as well as functional perspective, a verbal threat constitutes a linguistic strategy that is used to manipulate or even coerce the addressee into (not) doing something that is an undesirable outcome for him/her" (Limberg, 2009, p. 1378).

Threats have been studied from a linguistic and speech act philosophy perspective (e.g., Beller, Bender, & Song, 2009; Beller, Bender, & Kuhnmünch, 2005; Fraser, 1975; Kissine, 2008; López-Rousseau, Diesendruck & Benozio, 2011). In Searle's (1975) categorization of speech acts, threats and promises are considered commissives, which are illocutionary acts intended to commit the issuer to a particular course of action. Threats may seek a purely instrumental goal (e.g., compliance with a particular request or demand), or they may simply seek to terrorize and evoke fear in the service of a personal gratification and arousal motive.

Typically, scholars concur that threats involve (a) relevance and implications for the recipient(s), (b) which are negatively valenced by the recipient(s), and (c) evaluated by the recipient in regard to the preparatory or credibility conditions (i.e., that the issuer knows the recipient understands and negatively valences the implied effects of the threat, that the issuer intends and is able to enact the effect through some course of action, and that the issuer will prevent the effect upon recipient compliance; Gill & Ben-Shahar, 2005; Martínez-Cabeza, 2009). That is, threats are typically directive rather than commissive acts—acts that seek to influence a recipient rather than necessarily commit the issuer to a particular course of action (Salgueiro, 2010). Yet, threat assessment experts commonly envision threats and violence as motivated primarily by either instrumental or expressive motives (e.g., Hamel, Desmarais, & Nicholls, 2007; McEllistrem, 2004; Meloy, 2002; Tweed & Dutton, 1998), suggesting that some threats serve little tangible instrumental function. The issuer also may intend indirect rather than direct control over the threat outcomes. For example, a political candidate suggesting "second amendment" options for dealing with an opponent is re-directing the source of the implied threat. Furthermore, despite the explicit connection between the speech act, and the actual actions implied, most Western jurisprudence recognizes a fundamental distinction between act and speech, making threats a problematic legal category of crime (Bar-Gill & Ben-Shahar, 2005; Feinstein, 1996; Martínez-Cabeza, 2009), especially in societies with freedom of speech rights.

Given these characteristics, although laypersons may see many diverse events or situations as "threatening," to make issue or enact a threat implies a speech act that can be characterized by several explicit or implicit features:

1. Intentionality: The issuer intends to achieve one or more conscious and identifiable outcomes, thereby making threats a subset of persuasive speech acts intended as forms of influence;

2. Negative valence: The act implies some harm(s) or undesirable consequence(s) to the target(s);
3. Implicit or explicit issuer control: The issuer is actually in control of, and/or attempts to communicate self-efficacy and control over, the means of the occurrence of the harm(s);
4. Issuer's Preferred Outcome: The issuer suggests or specifies a demand or course of action on the part of the target that may avert the harm;
5. Contingency: The issuer suggests or specifies that the probability or severity of the harm is probabilistically related to the target's behavior. That is, the target may avert the harm by complying with or fulfilling the issuer's preferred outcome;
6. Credibility and willingness: The issuer's efficacy (i.e., capability of enacting or enabling the harm) and the likelihood or probability of instantiating the harm are either implied or specified as part of the message;
7. Subjunctive Mood: Threats tend to be directed toward future possibilities, even though they often refer to past perceived wrongs or transgressions, and threats may presage future contingencies through present action (e.g., vandalism in the present may be a message of what may happen in the future if demands are not met).

From this pragmatic approach, threats are typically conceptualized as a form of conditional speech intended to influence or gain compliance from a target recipient or agent, even when the proximal motive may be expressive in nature. Such inquiries have often focused on differentiating threats from predictions (Kissine, 2008), promises, advice, warnings (e.g., López-Rousseau et al., 2011; Wood & Quinn, 2003), and anger (Frick, 1986; Sinaceur, van Kleef, Neale, Adam, & Haag, 2011; Sinaceur & Neal, 2005). For example, warnings say that there is a risk of a bad event occurring that is not under the control of the speaker (as in a friend or family member telling their daughter "you are headed for trouble" or "anyone who dresses like that is asking for it"). In contrast, a threat is a statement of a punishment under the control of the threatener that is implicitly or explicitly contingent upon the noncompliance of the target with the threatener's demands ("if you don't do what I ask, I will make you regret it").

Another potential asymmetry is between promises and threats. Promises tend to obligate behavior upon compliance based on positively-valenced outcomes, whereas threats relinquish the issuer from obligation upon compliance based on negatively-valenced outcomes, even though in essence, "a threat is always accompanied by a promise and vice versa, thereby making obligation as consubstantial to threats as to promises" (Salgueiro, 2010, p. 224; see also Castelfranchi & Guerini, 2007). Promises also tend to imply an acquiescence of the receiver, who can "deactivate" the promise, whereas threats are more unilaterally contracted in effect or implication (Salgueiro, 2010). Another common but not necessary asymmetry is that it is common in actual speech for speakers to employ the name of the speech act in their speech (e.g., "I promise you that...," "I'm warning you...," "My advice is to...," etc.), whereas issuers rarely use the word "threat" in their spoken or written threats, although targets may tend to apply the label to the act or use it as a credibility marker (e.g., "This is no idle threat I'm making"). Furthermore, recipients may often label the speech act in context (e.g., "Are you threatening me?").

There may be typological differences across certain contexts of threats. For example, research is progressing in identifying the linguistic profile of potential threat-relevant crimes, including cyber-bullying (Dinakar, Recichart Lieberman, 2011; Hatakeyama, Masui, Ptaszynski & Yamamoto, 2016; Komuda, Ptaszynski, Rzepka & Araki, 2016; Latham, Crockett & Bandar, 2010; Lieberman, Dinakar & Jones, 2011; Nandhini & Sheeba, 2015; Nitta et al., 2013; Ptazynski et al., 2010;Ptaszynski, Masui, Kimura, Rzepka & Araki, 2015a; Pstaszynski, Masui, Kimura, Rzepka, & Araki, 2015b; Raisi & Huang, 2016; Van Royen, Poels, Daelemans & Vandebosch, 2015; Xu, Jun, Zhu, & Bellmore, 2012), rape (Woodhams & Grant, 2006), suicidality (e.g., Colombo, Burnap, Hodorog & Scourfield, 2016; Desmet & Hoste, 2012, 2013; Egnoto & Griffin, 2016; Handelman & Lester, 2007) in Twitter domains (O'Dea, Larsen, Batterham, Calear, & Christensen, 2016; Sueki, 2015), threats against public figures (e.g., Hoffmann, 2009; Meloy, Mohandie, & Green, 2008; Meloy, Sheridan, & Hoffman, 2008), school shooter threats (Meloy, Hoffmann, Roshdi, & Guldimann, 2014; Bondü & Scheithauer, 2014; Lindberg, Oksanen, Sailas, & Kaltiala-Heino, 2012; Meloy, Hoffman, Roshdi, & Guldimann, 2014; Sulkowski, 2010; Tiongco, 2015; Van Brunt, 2015), or terrorist threats (Cohen et al., 2014; Weinstein et al., 2009). Threats in such contexts may be substantially different from more relational or interpersonal threats.

2.2 The Language of Threats

Several typologies of threats and threateners have been proposed. Reminiscent of the instrumental/expressive dichotomy, an early empirically-based typology of threateners derived from a study of over 3,000 threats against federal officials differentiated "hunters" and "howlers" (Calhoun, 1998, p. xix): hunters "act in furtherance of committing intended violence" (Calhoun & Weston, 2009, p. 22) whereas howlers "communicate inappropriately, ominously, even threateningly, or ... communicate emotionally but ... never act violently" (Calhoun & Weston, 2009, p. 28; Calhoun & Weston, 2008), a typology later refined into "screamers," "shielders," "shockers," "schemer," and "signalers" (Warren, Mullen & McEwan, 2014). Another approach distinguished "real threats" from "bluffs," "latent threats" and "nonthreats" (Chung & Pennebaker, 2011). O'Toole (2004) distinguished direct threats, indirect threats, veiled threats, and conditional threats. Turner and Gelles (2003) identify threat communication characteristics of (a) organized versus disorganized, (b) fixation, (c) focus on self as wronged and on source of responsibility, and (d) action imperative, and time imperative.

The traditional psychological approach to threatening behavior seeks to understand the threatener (e.g., Schoeneman et al., 2011; Scalora et al., 2002; Warren, Mullen & Ogloff, 2011; Warren, Ogloff & Mullen, 2013; Warren, Mullen, Thomas, Ogloff & Burgess, 2008), and often imputes motives, psychological states, stages of preparation or actions toward violence to the speaker based on the nature of the threats.

More recently, research has begun to investigate the linguistic, pragmatic and contextual features of threatening communications, and the links between such features and threat outcomes. Geurts, Granhag, Ask and Vrij (2016) found, contrary to expectations, that bluffing threateners used more "how" or implementation language in their threat messages than actualizer threateners. The FBI Behavioral Analysis Unit, among other factors, seeks to identify mode of delivery, evidence of staging (i.e., purposeful manipulation of the message, such as the use of the pronoun "we"), motive, level of veracity (i.e., true intent), resolution to violence (i.e., justification, acceptance of consequences, ability to carry out the threat), and imminence (Simons & Tunkel, 2014). Schoeneman-Morris et al. (2007) compared email to letter threats to members of Congress and found that emails were more likely to emphasize governmental issues, use obscenity, and reveal disorganization in language, and less likely to evidence psychological disorders or problematic approach behavior. Schoeneman et al. (2011) also investigated communication features that characterized threateners who engaged in problematic approach behavior toward political officials. They found that approacher communications revealed longer handwritten correspondence, references to specific events, demands, noting personal stressors, violation of their rights, and expressing intentions to approach. In contrast, threatening language itself was unrelated to actual approach.

Threats no doubt present substantial challenges to standardized search and identification criteria. Threats, like most language, are highly contextual. Consider, for example, the following two exchanges between hypothetical persons A and B:

A: I'm having a party at my place on Friday. Do you know where I live?

B: *I know where you live. I'll see you soon.*

A: You are frightening me. Leave me alone. If I see you again I'll call the police, I swear!

B: *I know where you live. I'll see you soon.*

The content of B's speaking turn is identical in both interchanges, but clearly takes on a more threatening implication in the second exchange. Yet, by a priori notions of threat, there is little in the explicit or surface content of B's statement that seems particularly sinister. Whether or not threat content can be identified independent of such contextualizing information is an empirical question.

Assuming that threats can be reliably identified, the other major challenge is to distinguish threats in regard to their credibility. Spitzberg and Cupach's (2014) summary of 16 studies of stalker threats identified a false positive rate of 60% and a false negative rate of 18%, similar to estimates by Meloy (1999, 2002) and Resnik (2007). In a study of open source lone actor terrorists, Meloy and Gill (2016) found that only 22% engaged in pre-event warning behaviors that were considered directly communicated threats. Thus, many threats appear to have relatively little relation to the violence they portend. The credibility, or seriousness, of threats may be highly contextual. The prevailing wisdom is that judgments of threat message credibility is highly contextual and case-specific, requiring intensive evaluation of all case materials. There may still be significant practical value to more general forms of threat message identification in large text or 'big data' environments.

Computational linguistics is a rapidly advancing field that investigates ways of parsing elements of language, usually written text, to identify underlying dimensions and elements (e.g., Joacchims, 1998; Salton & Buckley, 1988). Progress is being accomplished in discourse analysis in the discrimination of arguments (e.g., Bex, Atkinson & Bench-Capon, 2014; Faulkner, 2015), narratives (Kypridemou & Michael, 2014), beliefs, motives, justifications (Prentice, Rayson, & Taylor, 2012), emotions (Oster, 2010; Westbury, Keith, Briemeister, Hofmann, & Jacovs, 2015), conflict (e.g., Kaya, Ozkaptan, Salah & Gurgen, 2015), sarcasm (e.g., Kovaz, Kreuz, & Riordan, 2013), impoliteness (Marco, 2008), group formation and membership (Tsou et al., 2014), and intention (e.g., Feng, 2015).

Only a few computational linguistics studies have been applied to threatening communications (Carter, 2010, Gales, 2010, 2011, 2015; Glukhov & Martynova, 2015; Smith, 2006, 2008; Tiongco, 2015; Watt, Kelly & Llamas, 2013), although several scholars have commented on the potential value of such analyses on threat messages (e.g., Cohen, Johansson, Kaati, & Mork, 2014; Leonard, 2005/2006; Sanfilippo, 2010). Taylor et al. (2013) investigated the emails of "insider threats" in a game simulation, and found that language became more self-focused, more negative in affective tone, and demonstrated more cognitive processing load compared to normal coworker participants. Glukhov and Martynova (2015) selected a corpus of 525 threats spoken in interpersonal contexts in fictional texts. They content-analyzed these threats for several features, including the nature of the fear appeal implied by the threat. They concluded that although threats to health or physical security were more represented in the corpus, threats to social identity were more efficient in achieving concessions for the fictional characters.

Carter (2010) extracted corpora of terrorist and non-terrorist threats from public websites. The terrorist corpus consisted of 4,059 words, and the non-terrorist corpus consisted of 2,172 words. These two corpora were each subdivided into those sentences containing clear threatening utterances. Simple word count metrics were assessed on pronoun usage and sentence structure (negative command, command, command-then statements, if-then statements, questions, and declarative statements). The results are entirely descriptive, but showed that the second-person nominative pronoun "you" (and lemmatized to include "you'll" and "you're") were most common. Grammatically, the subjective "I" and the objective "you" were the most common uses of pronouns. Declarative statements were most typical of specific threat grammar (e.g., "Now you're dead!" and "For this and other injustices, you will pay the ultimate price!").

Glasgow and Schouten (2014) examined a corpus of 60 documents sent to judges that raised safety concerns. Although only 3 of the documents "made clear threats of violence" (p. 41), 5 had vague threats of violence, and 16 threatened legal action, and another 8 threatened reputational attacks. Glasgow and Schouten applied a content and word software (LIWC; Chung & Pennebaker, 2011; Pennebaker, Francis, & Booth, 2001; http://www.liwc.net/) that seeks evidence of emotional states of writers, and a topic model that statistically aggregates topical themes (e.g., see Weinstein, Frazier, & Bongar, 2009). The authors found little ability to differentiate serious from non-serious threats, although the corpus was recognized as under-powered. Sanfilippo, McGrath and Bell (2014) report a computer modeling approach using frame analysis (Goffman, 1974), in which content themes and features are processed from terrorist messages, including: (a) moral disengagement, (b) message delivery, (c) seek resonance, (d) violence and contention, (e) call to arms, (f) social isolation, and (g) violation of sacred values (see also, Sanfilippo, 2010; Sanfilippo, McGrath & Whitney, 2011).

An ambitious project by Gales (2010a, 2010b, 2011, 2015) obtained a corpus consisting of 470 threat letters from the Academy Group, a consulting behavioral analysis organization employing former FBI Special Agents. The project sought to analyze threats through the lens of speaker stance and appraisal. *Stance* represents "the ways in which speakers and writers linguistically demonstrate their commitment to or attitudes about a person or proposition" (Gales, 2011, p. 27). *Appraisal* involves linguistic markers of speaker *attitude* ("how feelings are mapped within texts," p. 30), *engagement* ("how writers

... dialogically position themselves with respect to their audience or to propositions referenced within the text," p. 30), and *graduation* ("to demonstrate greater or lesser degrees of positive or negative feelings," p. 30). From this perspective, she theorized that "stances relating to the emotions of the writer are outlined through the systems of attitude, while stances relating to the writer's level of commitment or investment are highlighted through the system of engagement" (pp. 30-31). Her case studies indicated, contrary to common predictions, that threatener language demonstrated ambivalent attitudes (i.e., disfavor of both the target's and self's actions) and ambivalent graduation (i.e., through heteroglossic utterances such as "may"). In a separate analysis of 397 threats (128,774 total words) from the same source, stance was used to differentiate threats in stalking cases, harassment cases, and defamation cases. Stalking threats were particularly characterized by prediction modals of *will, would, shall, be going to*, a strong co-occurrence of these predictions modals and pronouns (e.g., *I/we*, $r = .88$), trigrams (i.e., *I will be* and *I will have* indicating volition and possessiveness), verb-controlled *that*-complement clauses indicating certainty (e.g., *you know that*) and intention (e.g., *want, need, like*). Suggestive of the role of the credibility pragmatic of threats, Gales (2015) found that "verbs of certainty, which are linked to the epistemic function of language, are considerably more frequent in all categories of threats, in general" (p. 189).

Smith (2008) examined a corpus of 96 FBI threatening communication cases, classified as (1) no action by the threatener, (2) stalking or approaching, or (3) harmful action. She found several language content variables related significantly to action taken, including threatening to reveal detrimental information, threatening to stalk, using persuasion, repeatedly mentioned love or marriage or romance, used polite threatening tone, and words associated with prejudices regarding religion. Threat document features also predicted action taken, including typed or handwritten notes (vs. computer printed) and inappropriate capitalization, and using a true return address. She has more recently begun to incorporate various linguistic metrics into a software package for assessing seriousness of threats that demonstrates good discriminatory power with this same threat corpus (Smith, Woyach & O'Toole, 2014). This computational linguistic system is most immediately exemplary to the current project. It employs an algorithm of seven weighted factors (www.threattriage.com), some of which can be extracted automatically from the language of a threat text: prior contacts, paranoid expressions, polite tone, mentions of love—marriage—or romance, specifying the target, specifying the harm for the victim, and conceptually complex language). The language complexity variable is considered an indicator of planning capacity, which is interpreted as a proxy for intent. These seven factors demonstrated significant discrimination of threat-to-problematic action or seriousness in a data set of 89 FBI threat cases. The threat triage system continues to add closed cases to refine the algorithm and accuracy of the system.

Also, exemplary of this project's objectives, research by Tiongco (2015) sought to develop and validate a more holistic rating scale. The Communicated Threat Analysis Scale (CTAS) was intended as a holistic rating scale to assess the seriousness of a threat. CTAS seeks to assess five characteristics associated with threats: organization versus disorganization, fixation, time imperative, action imperative, and focus. Two exemplary closed-case threats were used as stimuli, one credible and one not credible. The CTAS was also compared to a known threat assessment instrument with similar guided holistic subjective format (WAVR-21; Meloy, White & Hart, 2013). The 18-item Likert-type scale demonstrated marginal to unacceptable reliability of subscales, although the scale and its subscales could be argued to be indexes rather than scales, thereby not requiring internal consistency (Streiner, 2003). Construct validity coefficients between the CTAS and the WAVR were generally nonsignificant or modest in effect size, indicating little evidence of validity for the CTAS. There were also few differences manifested between the credible and the noncredible threat, or between the expert and lay raters.

Van Brunt (2015) also proposed a holistic rating scale of written messages. It is comprised of five factors, each with multiple sub-items: fixation and focus (specification of a target), hierarchical thematic content (narrative construction of the writer as a superior status protagonist), action and time imperative (indication of progression toward action through chronemic and spatial cues), pre-attack planning (subtle or explicit cues related to plan details related to threatened action), and injustice collecting (indications of a scorecard of having been wronged). This system is an entirely qualitative rating system, although some of its sub-items could be generated as template search ontologies or linguistic algorithms in big data contexts, such as target name repetition, graphic language, weapons mentions, and violence (e.g., Purohit et al., 2016). Such rating scales may be particularly relevant to validating training sets of threats for machine learning and classification, as well as heuristics for case assessment.

There are probably other relevant features not yet identified (Leonard, 2005/2006). For example, certain metrics would be calibration-based, such as sudden pattern changes or "bursts" of preoccupation with a particular topic, entity or person (Meloy & O'Toole, 2011). Some forensic approaches capitalize on establishing baseline distributions of a given communicator, and scan for significant pattern deviations or discrepancies (e.g., Abbasi & Chen, 2008; Hadjidj, et al., 2009). Pennebaker and Chung (2005) demonstrate that there may be distinct patterns of affective tone before, during, and following a crisis (e.g., a terrorist attack). Furthermore, several of these features cannot be captured in single messages, but can only be validly understood in a broader context of a 'campaign' or 'relationship' in which a given message establishes its credibility in the context of a broader set of message exchanges.

Threats are clearly complex communicative phenomena. In everyday speech, as a commissive, threats are most characterized by their false positives—a failure to commit an act that is promised (Spitzberg & Cupach, 2014). Such failure pragmatically places them more in the role of directive—influence attempts (i.e., directives). As such, a failure to commit an act is often taken as an ironic sign of the effectiveness of the speech act—the target's compliance foregoes the need to enact the harm implied by the speech act. Even though threats tend to demonstrate very high rates of false positives, they may yet reveal significant diagnostic and perhaps even predictive information about prospective acts of aggression. As Smith et al. (2014, p. 322) conclude: "A growing body of literature shows that a significant minority of threateners do approach or become violent subsequent to threatening...Research also indicates that the way people use language can have value for discerning their intent and future actions."

3. A RATING APPROACH: PRELIMINARY SCALES

A preliminary sketch of potential variables that might differentiate the credibility or seriousness of threats follows. These are based on familiarity with communication research, stalking research, and experience with the Association of Threat Assessment Professionals. These particular items, and others yet to be formulated, can be translated into rating scales (see Appendix 1), and treated as an index of threat credibility and seriousness. The result would be a THReat Evaluation & Assessment of Discourse (THREAD) index:

1. Feasibility: is the threat fulfillment possible (e.g., threatening to bring on the plague vs. spreading rumors)?
2. Capability/expertise: is there textual evidence the threatener is able to carry out the threat (Gales, 2011)?
3. Extremity/intensity: what is the severity of the consequences or scope of those threatened?
4. Evidence of prior perpetration efficacy and consistency—is there textual evidence that the threatener has issued, and followed through with, prior relevant threats?
5. Self-expressed agency/efficacy: does the threatener express confidence and a sense of self-efficacy in carrying out the threat (Gales, 2010a, 2010b; Schoeneman-Morris et al., 2007)?
6. Conditional probability in the verb phrases and contingency phrases: does text shift in verb tenses and conditionality (Gales, 2010a, 2010b)?
7. Immediacy/imminence: what is the time horizon of the language and implied harm?
8. Knowledge of target: how much information and/or insight into the target/victim is manifest in the threat (Smith, 2008)?
9. Complexity of plan(s): how complicated is the expressed threat, and how much cognitive processing is displayed in the speech construction (Smith, 2008; Taylor et al., 2013)?
10. Verbal/nonverbal features—to what extent does the text incorporate nonverbal elements?
11. Self-focus: are there shifts from other- or collective-based references to self-focused reference (Meloy, 2011; Taylor et al., 2013)?
12. Us-Them/You-I dichotomies: is the theme of pitting self-versus-other prominent (e.g., blame, attribution; see Carter, 2010; Gawron et al., 2012)?
13. Reference to relevant others as targets: are others, such as mutual children, pets, family, etc. included in the threat?
14. Linguistic divergence: to what degree does the person's speech style diverge from, rather than accommodate to, ingroup norms and/or interlocutor's 'turns at talk' or intermediary communications (Taylor et al., 2013)?
15. Sentiment deterioration or escalation: is there an increase in, or degree of contamination of speech with negative affect, particularly anger-based terminology (Meloy, 2011; Taylor et al., 2013)?
16. Goal-linking: is there evidence in the language of higher-order goal linking of the target with victim life objectives and/or values, and/or implicit proprietariness or entitlements (Meloy, 2011; Schoeneman-Morris et al., 2007; Spitzberg & Cupach, 2014)?
17. Identification/fixation: to what extent do words or phrases indicate fixation,

preoccupation, and personal identity fusion with a topic, entity, or person (Meloy & O'Toole, 2011; Spitzberg & Cupach, 2014)?

18. Philosophical embeddedness: are the threats embedded in a broader ideological manifesto (Schoeneman-Morris et al., 2007)?
19. Last resort terminology: to what extent do words or phrases indicate that all options have been exhausted, that death would be preferable to the status quo, etc. (Meloy & O'Toole, 2011)?
20. Coherence/organization: is there textual evidence that the threatener has engaged in planning, preparation, has an overall vision of implementing the threat?
21. Delusional content: is there content suggesting psychoses or lack of mental competence, especially references indicating persecutory beliefs, paranoid ideation, and Axis I and II disorders (Taylor et al., 2013)?
22. Finality fantasies: are there "end-game," suicide fantasies or images, suggested (Meloy, 2011)?

A preliminary operationalization draft of these dimensions is displayed in Appendix A, currently formatted as a set of semantic differential scales.

4. A COMPUTATIONAL LINGUISTICS APPROACH

Chung and Pennebaker (2011) provide a useful survey of computational approaches to the analysis of threat message texts. They identify several broad classes of approach; (a) word pattern analysis, approaches such as LSA (Landauer & Dumais 1997) and topic analysis (Steyvers & Griffiths, 2007) which analyze the co-occurrence patterns of words in text classes of interest; and (b) word count strategies, which identify psychologically salient classes to which words belong, and compile word counts for these classes. The classes may be both semantically defined (for example, social words or family words), and functionally defined (for example, first person pronouns). The semantic classes pertain to what is sometimes called content analysis and the functional classes to what is sometimes called style analysis. Both types of analysis have been effective in predicting a wide variety of textual properties. One of the most influential exemplars of this style of approach is Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2007). Another is Bucci's Discourse Attribute Analysis Program (Bucci & Maskit, 2005). The equally influential approach of Biber (1988) is somewhat more abstract; using factor analysis to combine a large variety of textual features, Biber succeeds in finding "linguistic fingerprints" for broad text genres like newspaper stories and romance novels.

Hancock et al. (2010) outline another approach they refer to as Social Language Processing (SLP). SLP shares features with the word-counting approach and may be thought of as building on it, while adding aspects of the machine learning paradigm. SLP is a method of classifying texts according to some social construct, for example, classifying threat messages to predict whether they will lead to a physical approach of the victim by the threatener, or to violence against the victim. SLP consists of three stages: (1) linguistic feature identification, (2) linguistic feature extraction, and (3) statistical classifier development. The first stage requires the identification of grammatical or psychological features of language that might be associated with the construct in question. In the second, feature extraction stage, the discovered features are extracted from a set of texts whose properties with respect to the social construct in question are known. This set is known as the training set. In the third stage, the learning stage, texts in the training set are classified according to the social construct, by optimizing weights for the features. This stage combines two processes, weight assignment and feature selection. In feature selection, features may be eliminated to eliminate noise or merged to account for feature interactions.

The difficult part of applying this paradigm is stage one, finding useful and extractable features. A resource like the LIWC dictionary is the endpoint of a process like a stage one process, but the features in LIWC are only a starting point. Each application has its own set of useful features, and some demonstrably useful features may involve linguistically complex actions such as describing financial problems, or announcing a significant anniversary, which are SLP problems in their own right. An example of an approach to stage one is the work of Miah et al. (2014), which uses a sentence similarity measure to cluster words associated with particular stages in child exploitation chats. Once words with strong associations with a particular stage are found, a LIWC dictionary is built, but with new features specific to child exploitation chats.

The threat message literature has identified a number of text features, of various levels of complexity, which might plausibly play a role in a threat assessment classifier, either to predict approach or violence.

Gales (2010a) analyzes threat messages, trying to identify those that are most likely to produce fear or anxiety in their recipients. A corpus-based approach is used to focus on what are known as appraisal features, linguistic features that express or reveal the author's evaluative stance toward the subject. The features examined have considerable computational potential, because they can be extracted with relative ease. They include specific trigrams such as "I will have" or "I will be", verbs with that-complement clauses, prediction modals such as "will", and adverbials of stance expressing certainty, likelihood, attitude, and style (for example, "frankly", "kind of"), and verbs of intention. All of Gale’s features are what are referred to here as content features. Not all predictive text features bear on the content of the text.

Of the various text variables Smith (2006) studies, the following showed some positive correlation with subsequent violent action: threateners (1) giving their real return address, either partial or complete (2) using a typewriter, (3) using inappropriate capitalization, and (4) handwriting the threat. Note that two of the three are non-content textual features. Smith also used software that conducted content analysis to identify psychological states: Gottshalk’s (2000; Gottschalk & Bechtel, 2000) PCADS and Herman's (2003) Profiler Plus.

Schoeneman-Morris, Scalora, Chang, Zimmerman and Garner (2007) discussed several text variables of considerable utility in predicting approach by the threatener using a corpus of threats on members of Congress. They identified the following content features in order of predictive power: discussion of personal themes, making a request for help, mention of entitlements owed the subject, mentions of matters of finance, discussion of injustice, discussion of government policy or human rights, identifying oneself, mention of stressors, appeals to patriotism, expression of an intent to approach, mention of upcoming anniversary, and discussion of contact plans. Schoeneman et al. also identified some non-content text features with predictive power, including all caps in messages and general disorganization of the text.

Meloy (2011) identified a number of features found consistently to predict approach. Although focusing on non-text features, Meloy does identify several features and communicative properties that might possibly be detected automatically, including request for help, entitled reciprocity (the claim that something is owed the subject), and grandiosity (imagined importance, or the wish to achieve importance). The first two coincide with features discussed by Schoeneman et al. (2007). Grandiosity and narcissism open a new text domain that may be important.

Recognizing abstract features of text like grandiosity or narcissism may fall between personality classification and recognizing psychological state. The literature on psychological content analysis has addressed both classes of problems. In 2005, a pioneering work by Argamon et al. (2005/2006) classified neuroticism and extraversion using linguistic features such as function words, deictics, appraisal expressions, and modal verbs. One year later, Oberlander and Nowson (2006) classified extraversion, stability, agreeableness, and conscientiousness of blog authors using n-gram features. Mairesse et al. (2007) reported a long list of correlations between the Big Five personality traits (Norman 1963) and LIWC Features (Pennebaker, Francis, & Booth 2001). Celli and Rossi (2012) used a very simple list of features to try to sort Twitter users into three classes (secure, neurotic, and balanced) using profile and timeline information. They successfully applied several the text features from Mairesse et al.'s (2007) data to their classification task (Table 4). These features may well apply to other psychological classification tasks, including recognizing grandiosity (e.g., use of exclamation/question marks, negative/positive emoticons, and number of long words).

Summarizing, the most promising approach to the computational problem of threat assessment is some variant of the SLP approach. Pursuing this paradigm seriously requires significant work on identifying a useful feature set. The work on textual threat assessment features suggests a number of easily extractable text features may be useful, but it also suggests that more abstract features may help, and abstract features like grandiosity pose classification problems of their own.

Such approaches are distinct from forensic efforts to identify threateners (e.g., Abbasi & Chen, 2008; Jadjidj et al., 2009). The contrast, however, is informative of potential connections between the approaches. The term stylometric analysis (SA) is generally used for text classification focusing on identifying some property of the author of a text, such as level of linguistic competence, gender, psychological profile, or just the author's identity. SA has played a role in Psychology, Language Pedagogy, Forensic Analysis, and Literary Studies. It has used a variety of text features (e.g., lexical, ngram, syntactic, and orthographic). Stylometric features may be extracted and clustered for a collection of texts to create "writeprints" for anonymous authors (e.g., Iqbal et el., 2010) or for problems of author identification or authentication. These approaches may be fruitfully combined with machine learning methods (Koppel, Schler & Argamon, 2009), such as support vector machines (SVMs; Diederich 2003; De Vel 2001; Li et al. 2006), neural networks (Merriam 1995; Tweedie, Singh & Holmes, 1996; Zheng, Li, Huang & Chen, 2006), and decision trees (Apte et al 1998; Abbasi & Chen 2005).

All these machine learning methods have also been successful in a distinct class of text analysis problems focusing on properties of the texts rather than properties of the authors; the most relevant problems are sentiment analysis and affect identification (Poria, Cambria & Gelbukh, 2015, Severyn & Moschitti, 2015, Teng et al. 2015). In this broader context, the success of neural networks, especially Convolutional Neural Networks (CNNs), is important. CNNs map word-level representations of sentences or documents into fairly low-dimensional representations of the entire sentence or document. They thus take into account, or try to take into account, the composition of word meanings into more complex messages. CNNs have been shown to be of significant help in sentiment analysis, although the shortcomings of a word-oriented approach have long been apparent in sentiment analysis, because diverse features of context may affect the final effect, such as when sarcasm is used.

The particular problem of threat analysis can be viewed as combining the two approaches of author-oriented analysis and text-oriented analysis. The psychological profile of the author is a significant factor, as is the content of the message. To this may be added a third component, identification of a particular kind of relationship, the predator-prey relationship, between the author and addressee. In two out of three of these components, it is entirely possible that key information is not encoded in the message, and that extra-textual features such as that provided by an author profile may prove essential. The multi-modal nature of the evidence is one respect in which the problem of threat assessment differs from many other text classification problems. Another is that a multiple component system trained to address the three components of the problem separately may have the best success because the architectures best suited to each problem are different. For example, the identification of personality types or author types seems to benefit from class-specific feature sets (Abbasi and Chen 2008, Poria et al. 2015). Finally, the best approach may be a "rating-based" approach that seeks to assign a numerical threat level (1-5). This is not simply a 5-class classification problem, since the training algorithm should exploit the fact that a 4 is closer to a 5 than to a 1. Thus the "metric labeling" technique of Pang and Lee (2005), which they apply to SVMs, may be of help. The process of factoring the problem into simpler parts, each of which may be its own more tractable machine learning problem, is productive. There are well known ensemble-learning techniques for co-training such separate learners. Abbasi and Chen (2008) and Poria et al. (2015) provide good examples.

One final point worth noting: An important component of the progress made in text classification over the last few years has been the increasing use of dimensionality reduction. Dimensionality reduction has its mathematical roots in Principal Components Analysis (PCA) and the closely related Singular Value Decomposition. PCA has been applied to authorship identification by using feature covariances over sliding text windows to compute author-specific patterns. Recent work using neural net trained word vectors (Mikolov et al. 2013) has introduced another "deep learning"-based form of dimensionality reduction, and though the amount of data required to train such word embeddings takes us well beyond the size of any plausible forensically tagged dataset, various practical methods of adapting such vectors to specific tasks have been proposed. For example, Tang, Wei, Qin, Liu and Zhou (2014) proposed a method of training the vectors with sentiment tags, to learn sentiment-specific word vectors. Similarly, the work of Poria et al. and Severyn and Moschitti, cited above begins with the word2vec vectors trained by Mikolov et al., and uses CNNs to train a sentence level sentiment analyzer, in effect training up a set of contextually sensitive word features relevant to sentiment classification. This provides some hope that deep learning may provide ways of detecting features of texts that have significant subtlety, including the many gradations of predator language, if we can supply the proper training sets.

The basic research agenda involves the following procedures. First, corpora of threatening messages and texts will be needed. Several such corpora exist in threat management institutions, both public and private. The more vetted as to outcome, the more useful they will be. Second, such threat corpora will be rated by experts and laypersons, using the rating scales in Appendix 1, or some version of them. Third, a corpus of mundane written text will be identified and collected for group discrimination purposes. Fourth, a variety of computational linguistic analyses will be used to (a) identify the most prominent features of the threat corpora that (b) distinguish it from the mundane textual discourse, (c) examine the extent to which such features also predict the expert and layperson ratings, and (d) identify the degree to which expert ratings are more predictable than layperson ratings. Numerous language corpora exist that might serve as the control archive (e.g., Brezina & Gabllosova, 2015; Drude, Broeder & Trilsbeek, 2014; Garfinkel, Farrell, Roussev, & Dinolt, 2009). The larger the corpora, the more stable the results are likely to be, and greater the opportunity to examine unique discriminating features of different types of threats (e.g., public figure vs. institutional vs. intimate partner, bombing vs. school shootings, etc.). Furthermore, to the extent that exemplar gold standard threat messages can be identified in reasonable numbers, they can be used to train machine learning protocols, which can then be used to refine the threat discrimination process on an ongoing basis.

5. CONCLUSION

There are two potentially practical immediate possible outcomes of successfully pursuing this project, assuming that linguistic indicators provide any statistically significant and substantial precision in identifying serious or credible threats: (1) a holistic rating scale could be valuable to various agencies, institutions, and law enforcement in providing a relatively efficient holistic and consistent approach to evaluating specific communication events and threat messages; (2) the development of an open-ended but annotated corpus of threats would become useful in subsequent studies. Indefinite but plausible outcomes would include potential findings that more credible or dangerous threats may be distinguishable by particular features that are easily identified. To the extent that such approaches can be automated, they can be built into social media surveillance dashboards (http://vision.sdsu.edu/hdma/), and corpora of threat messages can be exponentially increased, substantially facilitating assessment validity efforts (e.g., Fitzgerald, 2007; see also https://sites.google.com/site/tammygales/forensic-linguistic-data#threats and https://vault.fbi.gov/threats-against-members-of-congress). Phenomena ranging from school bullying to school shootings, mass shootings, and terrorist events may become more predictable, and thus more preventable.

REFERENCES

Abbasi, A. & Chen, H. (2005). Identification and comparison of extremist-group Web forum messages using authorship analysis. *IEEE Intelligent Systems, 20*, 5, 67–75.

Abbasi, A. & Chen, H. (2008). Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS) 26*, 2, 7.

Abbasi, A. & Chen, H. (2008). CyberGate: A design framework and system for text analysis of computer-mediated communication. *MIS Quarterly 32*(4), 811-837.

Apte, C., Damerau, F., Weiss, S. M. (1998). Text mining with decision trees and decision rules. Proceedings of the Conference on Automated Learning and Discovery: Learning from Text and the Web. Workshop 6: Learning from Text and the Web. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=5E0DDF906ABE950272ED4A129D7E84B3?doi=10.1.1.39.6018&rep=rep1&type=pdf

Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005/2006). Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America* (pp. 1–16). St. Louis, MO: Interface. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.60.6697

Barnes, M. T., Gordon, W. C., & Hudson, S. M. (2001). The crime of threatening to kill. *Journal of Interpersonal Violence, 16*, 312-319.

Baumgartner, J. V., Scalora, M. J., & Plank, G. L. (2001). Case characteristics of threats toward state government targets investigated by a Midwestern State. *Journal of Threat Assessment, 1*, 41-60.

Beller, S., Bender, A., & Jie, S. (2009). Conditional promises and threats in Germany, China, and Tonga: Cognition and emotion. *Journal of Cognition & Culture*, *9*(1/2), 115-139. doi: 10.1163/156853709X414674

Beller, S., Bender, A., & Kuhnmünch, G. (2005). Understanding conditional promises and threats. *Thinking & Reasoning, 11*(3), 209-238. doi: 10.1080/13546780442000141

Bernstein, H. A. (1981). Survey of threats and assaults directed toward psychotherapists. *American Journal of Psychotherapy*, *35*(4), 542-549.

Biber, D. (1991). *Variation across speech and writing*. New York, NY: Cambridge University Press.

Bondü, R., & Scheithauer, H. (2014). Leaking and death-threats by students: A study in German schools. *School Psychology International*, *35*(6), 592-608. doi: 10.1177/0143034314552346

Borum, R., Cornell, D. G., Modzeleski, W., & Jimerson, S. R. (2010). What can be done about school shootings? A review of the evidence. *Educational Researcher*, *39*(1), 27-37. doi: 10.3102/0013189X09357620

Borum, R., Fein, R., Vossekuil, B., & Berglund, J. (1999). Threat assessment: Defining an approach for evaluating risk of targeted violence. *Behavioral Sciences and the Law, 17*, 323-337.

Brewster, M. P. (2000). Stalking by former intimates: Verbal threats and other predictors of physical violence. *Violence and Victims, 15*, 41-54.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the "new general service list". *Applied Linguistics*, *36*(1), 1-22.

Brown, G. P., Dubin, W. R., Lion, J. R., & Garry, L. J. (1996). Threats against clinicians: A preliminary descriptive classification. *Bulletin of the American Academy of Psychiatry & the Law*, *24*(3), 367-376.

Brown, K. N., & MacAlister, D. (2006). Violence and threats against lawyers practicing in Vancouver, Canada. *Canadian Journal of Criminology and Criminal Justice*, *48*(4), 543-571. doi:10.3138/cjccj.48.4.543

Bucci, W., & Maskit, B. (2005). Building a weighted dictionary for referential activity. In Y. Qu, J. Shannon, & J. Wiebe (Eds.), *Computing attitude and affect in text* (pp. 49–60). Dordrecht, The Netherlands: Springer.

Bulling, D., Scalora, M., Borum, R., Panuzio, J., & Donica, A. (2008). *Behavioral science guidelines for assessing insider threats.* Lincoln, NE: The University of Nebraska Public Policy Center. Paper 37. http://digitalcommons.unl.edu/publicpolicypublications/37

Calhoun, F. S. (1998). *Hunters and howlers: Threats and violence against federal judicial officials in the United States, 1789-1993* (USMS No. 80). Washington, DC: U.S. Department of Justice, United States Marshals Service.

Calhoun, F. S., & Weston, S. W. (2008). On public figure howlers. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 105-122). New York, NY: Oxford University Press.

Calhoun, F. S., & Weston, S. W. (2016). *Threat assessment and management strategies: Identifying the howlers and hunters* (2nd ed.). Boca Raton, FL: CRC Press/Taylor & Francis.

Campbell, J., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M., & ... Laughon, K. (2003). Risk factors for femicide in abusive relationships: results from a multisite case control study. *American Journal of Public Health*, *93*(7), 1089-1097. doi:10.2105/AJPH.93.7.1089

Carter, N. R. (2010). *We shall be watching you, you're going to die*, and other threats: A corpus-based speech act approach. *UTA Working Papers in Linguistics, 3.* https://uta-ir.tdl.org/uta-ir/bitstream/handle/10106/5192/Threat-corpus-48-61.pdf?sequence=1&isAllowed=y

Castelfranchi, C., & Guerini, M. (2007). Is it a promise or a threat? *Pragmatics & Cognition*, *15*(2), 277- 311.

Celli, F., & Rossi, L. (2012). Long chains or stable communities: The role of emotional stability in Twitter conversations. In *Proceedings of the workshop on semantic analysis in social media* (pp. 10–17). Association for Computational Linguistics.

Chung, C. K., & Pennebaker, J. W. (2011). Using computerized text analysis to assess threatening communications and behavior. In C. Chauvin (Ed.), *Threatening communications and behavior: Perspectives on the pursuit of public figures* (pp. 3-32). Washington, DC: National Academies Press.

Churcher, F. P., & Nesca, M. (2013). Risk factors for violence in stalking perpetration: A meta-analysis. *FWU Journal of Social Sciences*, *7*(2), 100-112.

Cohen, K., Johansson, F., Kaati, L., & Mork, J. C. (2014). Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence, 26*(1), 246-256. doi:10.1080/09546553.2014.849948

Colombo, G. B., Burnap, P., Hodorog, A., & Scourfield, J. (2016). Analysing the connectivity and communication of suicidal users on twitter. *Computer Communications*, *73,* 291-300. http://dx.doi/org/10.1016/j.comcom.2015.07.018doi:10.1016/j.comcom.2015.07.018

Crowner, M. L., Peric, G., Stepcic, F., & Lee, S. (2005). Assailant and victim behaviors immediately preceding inpatient assault. *Psychiatric Quarterly, 76,* 243-256. Doi: 10.1007/s11126-005-2977-2

Darrow, C. D. (2014). Targeted threats: An examination of thematic content and approach behavior displayed by mentally ill and non-mentally ill contactors. *Dissertation Abstracts International*, *74*. http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1060&context=psychdiss

Davis, J. A. (2001). The assessment of potential threat: A second look. *Journal of Police and Criminal Psychology, 16*(1), 1-10.

Davis, J. A., Siota, R., & Stewart, L. (1999). Future prediction of dangerous and violent behavior: Psychological indicators and considerations for conducting and assessment of potential threat. *Canadian Journal of Clinical Medicine, 6*(3), 44-57.

Davis, J. A., Stewart, L. M., & Siota, R. (2001). Future prediction of dangerousness and violent behavior: Psychological indicators and considerations for conducting an assessment of potential threat. In J. A. Davis (Ed.), *Stalking crimes and victim protection: Prevention, intervention, threat assessment, and case management* (pp. 261-282). Boca Raton, FL: CRC Press.

Desmet, B., & Hoste, V. (2012). Combining lexico-semantic features for emotion classification in suicide notes. *Biomedical Informatics Insights*, *5*(Suppl. 1), 125-128. doi:10.4137/BII.S8960

Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. *Expert Systems with Applications*, *40*(16), 6351-6358. doi:10.1016/j.eswa.2013.05.050

Dietz, P. E., Matthews, D. B., Martell, D. A., Stewart, T. M., Hrouda, D. R., & Warren, J. (1991a). Threatening and otherwise inappropriate letters to members of the United States Congress. *Journal of Forensic Sciences, 36*, 1445-1468.

Dietz, P. E., Matthews, D. B., Van Duyne, C., Martell, D. A., Parry, C. D. H., Stewart, T., Warren, J., & Crowder, J. D. (1991b). Threatening and otherwise inappropriate letters to Hollywood celebrities. *Journal of Forensic Sciences, 36*, 185-209.

Dinakar, K., Reichart, R., & R. Lieberman, R. (2011). Modeling the detection of textual cyberbullying. *Proceedings of the International Conference on Weblog and Social Media - Social Mobile Web Workshop.* Barcelona, Spain.

Doren, D. M., Miller, R., & Maier, G. J. (1993). Predicting threatening psychopathic patient behavior in an inpatient milieu. *International Journal of Offender Therapy and Comparative Criminology*, *37*(3), 221-229. doi: 10.1177/0306624X9303700303

Drude, S., Broeder, D., & Trilsbeek, P. (2014). The Language Archive and its solutions for sustainable endangered languages corpora. *Book 2.0*, *4*(1/2), 5-20.

Drysdale, D. A., & Modzeleski, W. (2010). *Campus attacks: Targeted violence affecting institutions of higher education.* Washington DC: U.S. Secret Service, U.S. Department of Education, and Federal Bureau of Investigation.

Dunn, J. (2008). Operations of the LAPD Threat Management Unit. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 325-342). New York, NY: Oxford University Press.

Egnoto, M. J., & Griffin, D. J. (2016). Analyzing language in suicide notes and legacy tokens: Investigating clues to harm of self and harm to others in writing. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, doi:10.1027/0227-5910/a000363

Every-Palmer, S., Barry-Walsh, J., & Pathé, M. (2015). Harassment, stalking, threats and attacks targeting New Zealand politicians: A mental health issue. *Australian & New Zealand Journal of Psychiatry*, *49*(7), 634-641. doi: 10.1177/0004867415583700

Fein, R. A., Vossekuil, B., & Holden, G. A. (1995, September). *Threat assessment: An approach to prevent targeted violence.* National Institute of Justice Research in Action (NCJ 155000). Washington, DC: U.S. Department of Justice.

Fein, R. A., Vossekuil, B., Pollack, W. S., & Borum, R. (2002). *Threat assessment in schools: A guide to managing threatening situations and to creating safe school climates.* Washington DC: U.S. Secret Service and U.S. Department of Education.

Fein, R., & Vossekuil, B. (2000). *Protective intelligence and threat assessment investigations: A guide to managing threatening situations and to creating safe school climates.* Washington, DC: U.S. Secret Service and U.S. Department of Education.

Fitzgerald, J. R. (2007). The FBI's Communicated Threat Assessment Database. *FBI Law Enforcement Bulletin*, *76*(2), 6-9.

Fraser, B. (1975). Warning and threatening. *Centrum, 3*, 169-190.

Gales, T. (2010b). Ideologies of violence: A corpus and discourse analytic approach to stance in threatening communications. *International Journal of Speech, Language & the Law*, *17*(2), 299-302. doi:10.1558/ijsll.v17i2.299

Gales, T. (2011). Identifying interpersonal stance in threatening discourse: An appraisal analysis. *Discourse Studies*, *13*(1), 27-46. doi: 10.1177/1461445610387735

Gales, T. (2015). The stance of stalking: a corpus-based analysis of grammatical markers of stance in threatening communications. *Corpora*, *10*(2), 171-200. doi:10.3366/cor.2015.0073

Gales, T. A. (2010a). *Ideologies of violence: A corpus and discourse analytic approach to stance in threatening communications.* Unpublished Ph.D. dissertation, Department of Linguistics, University of California, Davis, CA.

Garfinkel, S., Farrell, P., Roussev, V., & Dinolt, G. (2009). Bringing science to digital forensics with standardized forensic corpora. *Digital Investigation*, *6*S2-S11. doi:10.1016/j.diin.2009.06.016

Geurts, R., Granhag, P. A., Ask, K., & Vrij, A. (2016). Taking threats to the lab: Introducing an experimental paradigm for studying verbal threats. *Journal of Threat Assessment and Management, 3*, 53-64. http://dx.doi.org/10.1037/tam0000060

Gill, O., & Ben-Shahar, O. (2005). Credible coercion. *Texas Law Review*, *83*(3), 717-780.

Glasgow, K., & Schouten, R. (2014). Assessing violence risk in threatening communications. *Workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 38-45). Baltimore, MD: Association for Computational Linguistics. http://acl2014.org/acl2014/W14-32/pdf/W14-3205.pdf

Glass, N., Laughon, K., Rutto, C., Bevacqua, J., & Campbell, J. C. (2008). Young adult intimate partner femicide: An exploratory study. *Homicide Studies*, *12*(2), 177-187.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Cambridge, MA: Harvard University Press.

Gottschalk, L. A. (2000). The application of computerized content analysis of natural language in psychotherapy research now and in the future. *American Journal of Psychotherapy, 54*(3), 305-311.

Gottschalk, L. A., & Bechtel, R. J. (2000). Pcad 2000: *Psychiatric content analysis and diagnosis*. Technical report, GB Software LLC, Corona Del Mar, CA.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211-244. doi:10.1037/0033-295X.114.2.211

Hadjidj, R., Debbabi, M., Lounis, H., Iqbal, F., Szporer, A., & Benredjem, D. (2009). Towards an integrated e-mail forensic analysis framework. *Digital Investigations 5*, 124-137.doi.10.1016/j.diin.2009.01.004

Hadnagy, C., & Fincher, M. (2015). *Phishing dark waters: The offensive and defensive sides of malicious E-mails*. Indianapolis, IN: John Wiley & Sons.

Häkkänen, H. (2006). Finnish bomb threats: Offence and offender characteristics. *International Journal of Police Science & Management*, *8*(1), 1-8.

Hamel, J., Desmarais, S. L., & Nicholls, T. L. (2007). Perceptions of motives in intimate partner violence: Expressive versus coercive violence. *Violence and Victims, 22*, 563-576.

Hancock, J. T., Beaver, D. I., Chung, C. K., Frazee, J., & Pennebaker, J. W., Graesser, A., & Cai, Z. (2010). Social language processing: A framework for analyzing the communication of terrorists and authoritarian regimes. *Behavioral Sciences of Terrorism and Political Aggression, 2*(2), 108–132.

Handelman, L. D., & Lester, D. (2007). The content of suicide notes from attempters and completers. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, *28*(2), 102-104. doi:10.1027/0227-5910.28.2.102

Hatakeyama, S., Masui, F., Ptaszynski, M., & Yamamoto, K. (2016). Statistical analysis of automatic seed word acquisition to improve harmful expression extraction in cyberbullying detection. *International Journal of Engineering and Technology Innovation, 6*(2), 165-172.

Hermann, M. G. (2003). Assessing leadership style: Trait analysis. In J. M. Post (Ed.), *The psychological assessment* *of political leaders with profiles of Saddam Hussein and Bill Clinton* (pp. 178–212). Ann Arbor, MI: University of Michigan Press.

Hillbrand, M. (2001). Threatening and non-threatening verbal aggression as predictors of physical aggression in violent psychiatric patients. *Journal of Threat Assessment, 1*, 63-74.

Hoffmann, J. (2009). Public figures and stalking in the European context. *European Journal on Criminal Policy & Research*, *15*(3), 293-305. doi: 10.1007/s10610-009-9104-0

Hoffmann, J., & Sheridan, L. (2008). Stalking, threatening, and attacking corporate figures. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 123-142). New York, NY: Oxford University Press.

Iqbal, F., Binsalleeh, H., Fung, B. C.M., & Debbabi, M. (2010). Mining writeprints from anonymous e-mails for forensic investigation. *Digital Investigation 7*(1), 56-64. doi:10.1016/j.diin.2010.03.003

Jackson, G. M. (2012). *Predicting malicious behavior: Tools and techniques for ensuring global security.* Indianapolis, IN: John Wiley & Sons.

James, D. V., Kerrigan, T. R., Forfar, R., Farnham, F. R., & Preston, L. F. (2010). The fixated threat assessment centre: Preventing harm and facilitating care. *Journal of Forensic Psychiatry and Psychology, 21,* 521-536.

James, D. V., McEwan, T. E., MacKenzie, R. D., Meloy, J. R., Mullen, P. E., Pathé, M. T., & ... Darnley, B. J. (2010). Persistence in stalking: A comparison of associations in general forensic and public figure samples. *Journal of Forensic Psychiatry & Psychology*, *21*(2), 283-305. doi: 10.1080/14789940903388994

James, D. V., Mullen, P. E., Pathé, M. T., Meloy, J. R., Farnham, F. R., Preston, L., & Darnley, B. (2008). Attacks on the British Royal Family: The role of psychotic illness. *Journal of the American Academy of Psychiatry & the Law*, *36*(1), 59-67.

James, D. V., Mullen, P. E., Pathé, M. T., Meloy, J. R., Preston, L. F., Darnley, B., & Farnham, F. R. (2009). Stalkers and harassers of royalty: The role of mental illness and motivation. *Psychological Medicine*, *39*(9), 1479-1490. doi:10.1017/S0033291709005443

Jenkins, D. M. (2009). When should threats be seen as indicative of future violence? Threats, intended violence, and the intimacy effect. In F. S. Calhoun & S. W. Weston (Eds.), *Threat assessment and management strategies: Identifying howlers and hunters* (pp. 151-199). Boca Raton, FL: CRC/Taylor & Francis.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Proceedings of the European Conference on Machine Learning (ECML).* New York, NY: Springer.

Kissine, M. (2008). From predictions to promises: How to derive deontic commitment. *Pragmatics & Cognition, 16*, 471-491. doi: 10.1075/p&c.16.3.03kis

Komuda, R., Ptaszynski, M., Rzepka, R., & Araki, K. (2016, July). Recognizing and converting cockney rhyming slang for cyberbullying and crime detection. *IJCAI 2016 International Workshop on Language Sense on Computer*. New York, NY. http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/data/Komuda-cameraready.pdf

Kontostathis, A., Edwards, L., & Leatherman, A. (2010). Text mining and cybercrime. In M. W. Berry & J. Kogan (Eds.), *Text mining: Applications and theory.* Chichester, UK: John Wiley & Sons, Ltd.

Koppel, M., Schler, J. & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology 60*(1), 9-26. doi: 10.1002/asi.20961

Kovacevic, A., & Nikolic, D. (2015). In M. M. Cruz-Cunha & I. M. Portela (Eds.), *Handbook of research on digital crime, cyberspace security, and information assurance* (pp. 277-290). Hershey, PA: Information Science Reference/IGI Global.

Kropp, P. R., Hart, S. D., & Lyon, D. R. (2008). Risk assessment of public figure stalkers. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 343-362). New York, NY: Oxford University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211-240. doi:10.1037/0033-295X.104.2.211

Larionovs, A., Teilans, A., & Grabusts, P. (2015). CORAS for threat and risk modeling in social networks. *Procedia Computer Science, 43*, 26-32. doi: 10.1016/j.procs.2014.12.005

Latham, A., Crockett, K., & Bandar, Z. (2010, January). A conversational expert system supporting bullying and harassment policies. *Proceedings of the Second International Conference on Agents and Artificial Intelligence* (pp. 163–168). Frente Lisboa, Portugal: INSTICC (Institute for Systems and Technologies of Information, Control and Communication.

Leonard, R. A. (2005/06). Forensic linguistics: Applying the scientific principles of language analysis to issues of the law. *International Journal of the Humanities, 3*, 1447-9559.

Lieberman, H., Dinakar, K., & Jones, B. (2011). Let's gang up on cyberbullying. *Computer, 44*, 93–96.

Limberg, H. (2009). Impoliteness and threat responses. *Journal of Pragmatics*, *41*(7), 1376-1394. doi:10.1016/j.pragma.2009.02.003

Lindberg, N., Oksanen, A., Sailas, E., & Kaltiala-Heino, R. (2012). Adolescents expressing school massacre threats online: Something to be extremely worried about? *Child and Adolescent Psychiatry and Mental Health*, *6*doi:10.1186/1753-2000-6-39

López-Rousseau, A., Diesendruck, G., & Benozio, A. (2011). My kingdom for a horse: On incredible promises and unpersuasive warnings. *Pragmatics & Cognition*, *19*(3), 399-421.

MacDonald, J. M. (1968). *Homicidal threats.* Springfield, IL: Charles C. Thomas.

Maier, G. (1996). Managing threatening behavior. The role of talk down and talk up. *Journal of Psychosocial Nursing and Mental Health Services*, *34*(6), 25-30.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30,* 457–500.

Maras, M-H. (2015). Unprotected speech communicated via social media: What amounts to a true threat? *Journal of Internet Law*, *19*, 3-9.

Marco, M. A. (2008). Influence of situational factors on the codification and interpretation of impoliteness. *Pragmatics*, *18*(4), 757-773.

Martínez-Cabeza, M. A. (2009). Dangerous words: Threats, perlocutions and strategic actions. In B. Lewandowska-Tomaszczyk, & P. Stalmascczyk (Eds.), *Cognitive approaches to language and linguistic data* (pp. 269-283). Frankfurt, GDR: Peter Lang.

Mazur, A. (1983). Bomb threats against American nuclear-energy facilities. *Journal of Political & Military Sociology*, *11*(1), 109-121.

McCann, J. T. (2001). The relationship between threats and violence in juvenile stalking. *Journal of Threat Assessment, 1*, 81-90.

McCann, J. T. (2002). *Threats in schools: A practical guide for managing violence*. New York, NY: Haworth.

McEllistrem, J. E. (2004). Affective and predatory violence: A bimodal classification system of human aggression and violence. *Aggression and Violent Behavior, 10*, 1-30.

Meloy, J. R. (1999). Stalking: An old behavior, a new crime. *Psychiatric Clinics of North America, 22*, 85-99.

Meloy, J. R. (2000). *Violence risk and threat assessment*. San Diego, CA: Specialized Training Services.

Meloy, J. R. (2002). Pathologies of attachment, violence, and criminality. In A. M. Goldstein & I. B. Weiner (Eds.), *Handbook of psychology* (Vol. 11: Forensic psychology, pp. 509-526). Hoboken, NJ: John Wiley & Sons.

Meloy, J. R. (2011). Approaching and attacking public figures: A contemporary analysis of communications and behavior. In C. Chauvin (Ed.), *Threatening communications and behavior: Perspectives on the pursuit of public figures* (pp. 75-106). Washington, DC: National Academies Press.

Meloy, J. R., & Gill, P. (2016). The lone-actor terrorist and the TRAP-18. *Journal of Threat Assessment and Management, 3*, 37-52. http://dx.doi.org/10.1037/tam0000061

Meloy, J. R., & O'Toole, M. E. (2011). The concept of leakage in threat assessment. *Behavioral Sciences and the Law, 29*(4), 29(4), 513-527. doi: 10.1002/bsi.986

Meloy, J. R., & O'Toole, M. E. (2011). The concept of leakage in threat assessment. *Behavioral Sciences & The Law*, *29*(4), 513-527. doi:10.1002/bsl.986

Meloy, J. R., Hoffmann, J. Roshdi, K., Glaz-Ocik, J., & Guldimann, A. (2014). Warning behaviors and their configurations across various domains of targeted violence. In J. R. Meloy & J. Hoffmann (Eds.), *International handbook of threat assessment* (pp. 39-53). New York, NY: Oxford University Press.

Meloy, J. R., Hoffmann, J., Guldimann, A., & James, D. (2012). The role of warning behaviors in threat assessment: An exploration and suggested typology. *Behavioral Sciences & the Law, 30*, 256-279. doi:10.1002/bsl.999

Meloy, J. R., Hoffmann, J., Roshdi, K., & Guldimann, A. (2014). Some warning behaviors discriminate between school shooters and other students of concern. *Journal of Threat Assessment and Management*, *1*(3), 203-211. doi: 10.1037/tam0000020

Meloy, J. R., James, D. V., Farnham, F. R., Mullen, P. E., Pathe, M., Darnley, B., & Preston, L. (2004). A research review of public figure threats, approaches, attacks, and assassinations in the United States. *Journal of Forensic Sciences (Wiley-Blackwell)*, *49*(5), 1086-1093.

Meloy, J. R., Mohandie, K., & Green, M. (2008). A forensic investigation of those who stalk celebrities. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 37-54). New York, NY: Oxford University Press.

Meloy, J. R., Sheridan, L, & Hoffman, J. (2008). Public figure stalking, threats, and attacks: The state of the science. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 3-34). New York, NY: Oxford University Press.

Meloy, J. R., White, S. G., & Hart, S. (2013). Workplace Assessment of Targeted Violence Risk: The Development and reliability of the WAVR-21. *Journal of Forensic Sciences (Wiley-Blackwell)*, *58*(5), 1353-1358. doi:10.1111/1556-4029.12196

Merriam, T. V. N. & Matthews, R. A. J. (1994). Neural computation in stylometry II: An application to the works of Shakespeare and Marlowe. *Literary and Linguistic Computing, 9*, 1–6.

Miah, M., Rahman, W., Yearwood, J., & Kulkarni, S. (2015). Constructing an inter-post similarity measure to differentiate the psychological stages in offensive chats. *Journal of the Association for Information Science and Technology, 66*(5), 1065–1081.

Mikolov, T., Yih, W-t., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. Proceedings of HLT-NAACL-2013 (pp. 746-751). Association for Computational Linguistics. http://www.aclweb.org/anthology/N13-1090

Milburn, T. W., & Watman, K. H. (1981). *On the nature of threat: A social psychological analysis.* New York, NY: Praeger.

Moore, A. P., Mundie, D. A., & Collins, M. L. (2013, July). A system dynamics model for investigating early detection of insider threat risk. *Conference Proceedings of the 31st International Conference of the System Dynamics Society.* Cambridge, MA. ISBN 978-1-935056-12-06

Morewitz, S. J. (2010). *Death threats and violence: New research and clinical perspectives.* New York, NY: Springer.

Mundie, D. A., Moore, A. P., & McIntire, D. (2012). Building a multidimensional pattern language for insider threats. *Proceedings of the Conference on Pattern Languages of Programs.* Tucson, AZ.

Murdock, J. I., Bradac, J. J., & Bowers, J. W. (1984). Effects of power on the perception of explicit and implicit threats, promises, and thromises: A rule-governed perspective. *Western Journal of Speech Communication, 48*, 344-361.

Nandhini, B. S., & Sheeba, J. I. (2015). Online social network bullying detection using intelligence techniques. *Procedia Computer Science, 45*, 485-492. doi: 10.1016/j.procs.2015.03.085

Nekvasil, E. K., & Cornell, D. G. (2012). Student reports of peer threats of violence: Prevalence and outcomes. *Journal of School Violence*, *11*(4), 357-375. doi:10.1080/15388220.2012.706764

Newhill, C. E. (2002). Client threats toward social workers: Nature, motives, and response. *Journal of Threat Assessment, 2*, 1-19.

Nitta, T., Masui, F., Ptaszynski, M., Kimura, Y., Rzepka, R., & Araki, K. (2013, October). Detecting cyberbullying entries on informal school websites based on category relevance maximization. *Proceedings of the 6th International Joint Conference on Natural Language Processing* (IJCNLP 2013, pp. 579-586). Nagoya, Japan. http://aclweb.org/anthology/I/I13/I13-1066.pdf

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, *66*(6), 574-583. doi:10.1037/h0040291

O'Dea, B., Larsen, M., Batterham, P., Calear, A., & Christensen, H. (2016). Talking suicide on Twitter: Linguistic style and language processes of suicide-related posts. *European Psychiatry*, *33*S329. doi:10.1016/j.eurpsy.2016.01.727

O'Hair, H. D., Bernard, D. R., & Roper, R. R. (2011). Communication-based research related to threats and ensuing behavior. In C. Chauvin (Ed.), *Threatening communications and behavior: Perspectives on the pursuit of public figures* (pp. 33-73). Washington, DC: National Academies Press.

Oberlander, J., & Nowson, S. (2006). Whose thumb is it anyway?: Classifying author personality from weblog text. In *Proceedings of the COLING/ACL on Main conference poster sessions* (pp. 627–634). Association for Computational Linguistics.

Oster, U. (2010). Using corpus methodology for semantic and pragmatic analyses: What can corpora tell us about the linguistic expression of emotions? *Cognitive Linguistics*, *21*(4), 727-763. doi:10.1515/COGL.2010.023

O'Toole, M. E., & National Center for the Analysis of Violent Crime (U.S.). (2000). *The school shooter: A threat assessment perspective*. Quantico, VA: FBI Academy.

Palarea, R. E., Zona, M. A., Lane, J. C., & Langhinrichsen-Rohling, J. (1999). The dangerous nature of intimate relationship stalking: Threats, violence, and associated risk factors. *Behavioral Sciences and the Law, 17*, 269-283.

Pang, B. & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 115-124). Association for Computational Linguistics. doi: 10.3115/1219840.1219855

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, J. W. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahwah, NJ: Lawrence Erlbaum Associates.

Peters, G. Y., Ruiter, R. C., & Kok, G. (2013). Threatening communication: A critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health Psychology Review*, *7*(Suppl 1), S8-S31. doi:10.1080/17437199.2012.703527

Peters, G. Y., Ruiter, R. C., & Kok, G. (2014). Threatening communication: A qualitative study of fear appeal effectiveness beliefs among intervention developers, policymakers, politicians, scientists, and advertising professionals. *International Journal of Psychology*, *49*(2), 71-79. doi:10.1002/ijop.12000

Pichon, S., de Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage,* *47*(4), 1873-1883. doi:10.1016/j.neuroimage.2009.03.084

Polluck, W. S., Modzeleski, W., & Rooney, G. (2008). *Prior knowledge of potential school-based violence: Information students learn may prevent a targeted attack.* Washington DC: U.S. Secret Service and U.S. Department of Education.

Poria, S., Cambria, E., & Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, (pp. 2539–2544).

Prentice, S., Rayson, P., & Taylor, P. J. (2012). The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics*, *17*(2), 259-286. doi:10.1075/ijcl.17.2.05pre

Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., & Araki, K. (2010, April). Machine learning and affect analysis against cyber-bullying. *Proceedings of the 36th Annual Convention of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour* (pp. 7-16). Leicester, UK. http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/data/AISB2010_Cyberbullying_paper.pdf

Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., Araki, K., & Momouchi, Y. (2010). In the service of online order: Tackling cyber-bullying with machine learning and affect analysis. *International Journal of Computational Linguistics Research, 1*(3), 135-154.

Ptaszynski, M., Masui, F., Kimura, Y., Rzepka, R., & Araki, K. (2015, November). Extracting patterns of harmful expressions for cyberbullying detection. P*roceedings of 7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'15), The First Workshop on Processing Emotions, Decisions and Opinions* (EDO 2015, pp. 370-375). Poznan, Poland. http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/data/EDO-3.pdf

Ptaszynski, M., Masui, F., Kimura, Y., Rzepka, R., & Araki, K. (2015, July). *Brute force works best against bullying.* IJCAI 2015 Workshop on Intelligent Personalization (IP 2015), Buenos Aires. http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/data/Brute_Force_Works_Best.pdf

Purohit, H. H., Banerjee, T., Hampton, A., Shalin, V. L., Bhandutia, N., & Sheth, A. (2016). Gender-based violence in 140 characters or fewer: A #BigData case study of Twitter. *First Monday, 21*(1), 1.

Raisi, E., & Huang, B. (2016). *Cyberbullying identification using participant-vocabulary consistency.* 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications. New York, NY.

Resnick, P. J. (2007). Stalking risk assessment. In D. A. Pinals (Ed.), *Stalking: Psychiatric perspectives and practical approaches* (Group for the Advancement of Psychiatry and the Law, pp. 61-84). New York, NY: Oxford University Press.

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, *10*, 336-353. doi: 10.1207/s15327957pspr1004_4

Riek, B. M., Mania, E. W., Gaertner, S. L., McDonald, S. A., & Lamoreaux, M. J. (2010). Does a common ingroup identity reduce intergroup threat? *Group Processes & Intergroup Relations*, *13*(4), 403-423. doi: 10.1177/1368430209346701

Salgueiro, A. B. (2010). Promises, threats, and the foundations of speech act theory. *Pragmatics, 20*, 213-228.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24,* 513–523.

Sandberg, D. A., McNiel, D. E., & Binder, R. L. (1998). Characteristics of psychiatric inpatients who stalk, threaten, or harass hospital staff after discharge. *American Journal of Psychiatry, 155*, 1102-1105.

Sandberg, D. A., McNiel, D. E., & Binder, R. L. (2002). Stalking, threatening, and harassing behavior by psychiatric patients toward clinicians. *Journal of the American Academy of Psychiatry and the Law, 30*, 221-229.

Sanfilippo, A. (2010, December). *Content analysis for proactive protective intelligence* (PNNL-20062). Springfield, VA: Pacific Northwest National Laboratory/Battell.

Sanfilippo, A., McGrath, L., & Bell, E. (2014). Computer modeling of violent intent: A content analysis approach. In J. R. Meloy & J. Hoffmann (Eds.), *International handbook of threat assessment* (pp. 224-235). New York, NY: Oxford University Press.

Sanfilippo, A., McGrath, L., & Whitney, P. (2011). Violent frames in action. *Dynamics of Asymmetric Conflict, 4*, 103-112. http://dx.doi.org/10.1080/17467586.2011.627933

Scalora, M. J. (2014). Electronic threats and harassment. In J. R. Meloy & J. Hoffmann (Eds.), *International handbook of threat assessment* (pp. 214-224). New York, NY: Oxford University Press.

Scalora, M. J., Zimmerman, W. J., & Wells, D. G. (2008). Use of threat assessment for the protection of the United States Congress. In J. R. Meloy, L. Sheridan, & J. Hoffman (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 425-434). New York, NY: Oxford University Press.

Schoeneman, K. A., Scalora, M. J., Darrow, C. D., McLawsen, J. E., Chang, G. H., & Zimmerman, W. J. (2011). Written content indicators of problematic approach behavior toward political officials. *Behavioral Sciences & the Law*, *29*(2), 284-301. doi:10.1002/bsl.977

Schoeneman-Morris, K. A., Scalora, M. J., Chang, G. H., Zimmerman, W. J., & Garner, Y. (2007). A comparison of email versus letter threat contacts toward members of the United States Congress. *Journal of Forensic Sciences*, *52*(5), 1142-1147. doi:10.1111/j.1556-4029.2007.00538.x

Seger, K. A. (1993). Violence in the workplace: An assessment of the problem based on responses from 32 large corporations. *Security Journal, 4* (3), 139—149.

Severyn, A., & Moschitti, A. (2015). UNITN: Training deep convolutional neural network for Twitter sentiment classification. *Proceedings of the 9th International Workshop on Semantic Evaluation* (SemEval 2015, pp. 464-469), Association for Computational Linguistics. https://pdfs.semanticscholar.org/496f/395d4d4038e85ba666691382f717b83c564b.pdf

Simons, A., & Cook, A. N. (2014). The assessment of anonymous threatening communications. In J. R. Meloy & J. Hoffmann (Eds.), *International handbook of threat assessment* (pp. 195-213). New York, NY: Oxford University Press.

Sinaceur, M., & Neale, M. (2005). Not all threats are created equal: How implicitness and timing affect the effectiveness of threats in negotiations. *Group Decision & Negotiation*, *14*(1), 63-85. doi: 10.1007/s10726-005-3876-5

Sinaceur, M., Van Kleef, G. A., Neale, M. A., Adam, H., & Haag, C. (2011). Hot or cold: Is communicating anger or threats more effective in negotiation? *Journal of Applied Psychology*, *96*(5), 1018-1032. doi: 10.1037/a0023896

Sinclair, H. C. (2009). Stalking, threatening, and attacking public figures: A review. *Journal of Police and Criminal Psychology*, *24*(2), 139-140. doi: 10.1007/s11896-009-9047-x

Smith, M. D., & Morra, N. N. (1994). Obscene and threatening telephone calls to women: Data from a Canadian national survey. *Gender & Society, 8*, 584-596.

Smith, S. S. (2006). *From violent words to violent deeds: Assessing risk from FBI threatening communications.* Unpublished Ph.D. dissertation, Georgetown University, Washington, DC.

Smith, S. S. (2008). From violent words to violent deeds: Assessing risk from FBI threatening communication cases. In J. Meloy, L. Sheridan, J. Hoffmann (Eds.), *Stalking, threatening, and attacking public figures: A psychological and behavioral analysis* (pp. 435-455). New York, NY, US: Oxford University Press.

Smith, S. S., & Shuy, R. W. (2002). Forensic psycholinguistics. *FBI Law Enforcement Bulletin*, *71*(4), 16.

Sokolow, B. A., Lewis, W. S., Schuster, S. K., Swinton, D. C., & Van Brunt, B. J. (2014). *Threat assessment in the campus setting* (The NaBITA 2014 whitepaper). Berwyn, PA: NaBITA. https://www.gtc.edu/sites/default/files/files/documents/2014-NaBITA-Whitepaper-Text-with-Graphics.pdf

Spitzberg, B. H. (in press). Acknowledgement of unwanted pursuit, threats, assault and stalking in a college population. *Psychology of Violence*.

Spitzberg, B. H., & Cupach, W. R. (2014). *The dark side of relationship pursuit: From attraction to obsession and stalking* (2nd ed.). New York, NY: Routledge.

Storey, J. E., Gibas, A. L., Reeves, K. A., & Hart, S. D. (2011). Evaluation of a violence risk (threat) assessment training program for police and other criminal justice professionals. *Criminal Justice and Behavior, 38*, 554-564.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, *80*(3), 217-222.

Sueki, H. (2015). The association of suicide-related Twitter use with suicidal behaviour:

a cross-sectional study of young internet users in Japan. *Journal of Affective Disorders*, *170*155-160. doi:10.1016/j.jad.2014.08.047

Sulkowski, M. L. (2011). An investigation of students' willingness to report threats of violence in campus communities. *Psychology of Violence*, *1*(1), 53-65. doi:10.1037/a0021592

Surface, J. L. (2011). Not all threats are equal. *Clearing House*, *84*(4), 150-154.

Tang, D., Wei, F., Qin, B., Liu, T., & Zhou, M. (2014). Coooolll: A deep learning system for Twitter sentiment classification. *Proceedings of the 8th International Workshop on Semantic Evaluation* (SemEval 2014, pp. 208-212). http://www.aclweb.org/anthology/S14-2033

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* *29*(1) 24–54. doi: 10.1177/0261927X09351676 http://jls.sagepub.com

Taylor, P. J., Dando, C. J., Ormerod, T. C., Ball, L. J., Jenkins, M. C., Sandham, A., & Menacere, T. (2013). Detecting insider threats through language change. *Law and Human Behavior, 37*(4), 267-275. doi: 10.1037/lhb0000032Bar-

Tiongco, J. A. (2015). *An approach to measure communicated threats: Developing a rating scale using a threat analysis model.* Unpublished dissertation, California School of Forensic Studies, Alliant International University, San Diego, CA.

Trump, K. (2015, February 9). Study finds rapid escalation of violent school threats. http://www.scchoolsecurity.org/2015/02/study-finds-rapid-escalation-violent-school-threats/

Turner, J. T. & Gelles, M. G. (2003). *Threat assessment: A risk management approach.* New York, NY: Haworth.

Turner, J. T. (2003). *Threat assessment: A risk management approach*. Binghampton, NY: Haworth.

Tweed, R. G., & Dutton, D. G. (1998). A comparison of impulsive and instrumental subgroups of batterers. *Violence and Victims, 13*, 217-230.

Tweedie, F. J., Singh, S., and Holmes, D. I. (1996). Neural network applications in stylometry: The Federalist papers. *Computers and the Humanities, 30*, 1, 1–10. http://www.jstor.org/stable/30204514

Twemlow, S. W., Fonagy, P., Sacco, F. C., & Vernberg, E. (2008). Assessing adolescents who threaten homicide in schools. *Clinical Social Work Journal*, *36*(2), 131-142. doi: 10.1007/s10615-007-0101-9

U.S. Secret Service, & U.S. Department of Education (2002). *Threat assessment in schools: A guide to managing threatening situations and to creating safe school climates.* Washington, DC: Authors.

U.S. Secret Service, U.S. Department of Education, & Federal Bureau of Investigation. (2010, April). *Campus attacks: Targeted violence affecting institutions of higher education.* Washington DC: Authors.

Van Brunt, B. (2015). Violence Risk Assessment of the Written Word (VRAW^2). *Journal of Campus Behavioral Intervention, 3,* 12-25. https://schoolshooters.info/sites/default/files/vraww.pdf

van der Meer, B. B., Bootsma, L., & Meloy, R. (2012). Disturbing communications and problematic approaches to the Dutch Royal Family. *Journal of Forensic Psychiatry & Psychology*, *23*(5/6), 571-589. doi:10.1080/14789949.2012.727453

Van Royen, K., Poels, K., Daelemans, W., & Vandebosch, H. (2014). Automatic monitoring of cyberbullying on social networking sites: From technological feasibility to desirability. *Telematics and Informatics.*

Van Royen, K., Poels, K., Daelemans, W., & Vandebosch, H. (2015). Automatic monitoring of cyberbullying on social networking sites: From technological feasibility to desirability. *Telematics & Informatics,* *32*(1), 89-97. doi:10.1016/j.tele.2014.04.002

Vudhiwat, C. (2002, September). Developing threats: Cyberstalking and the criminal justice system. *Crime & Justice International*, 9-10, 28-29.

Warren, L. J., MacKenzie, R., Mullen, P. E., & Ogloff, J. R. P. (2005). The problem behavior model: The development of a stalkers clinic and a threateners clinic. *Behavioral Sciences & the Law, 23*, 387-397.

Warren, L. J., Mullen, P. E., & McEwan, T. E. (2014). Explicit threats of violence. In J. R. Meloy & J. Hoffmann (Eds.), *International handbook of threat assessment* (pp. 18-38). New York, NY: Oxford University Press.

Warren, L. J., Mullen, P. E., & Ogloff, J. P. (2011). A clinical study of those who utter threats to kill. *Behavioral Sciences & the Law*, *29*(2), 141-154. doi:10.1002/bsl.974

Warren, L. J., Mullen, P. E., Thomas, S. M., Ogloff, J. P., & Burgess, P. M. (2008). Threats to kill: A follow-up study. *Psychological Medicine*, *38*(4), 599-605. doi: 10.1017/S003329170700181X

Warren, L. J., Ogloff, J. P., & Mullen, P. E. (2013). The psychological basis of threatening behaviour. *Psychiatry, Psychology and Law*, *20*(3), 329-343. doi:10.1080/13218719.2012.674716

Watt, D., Kelly, S., & Llamas, C. (2013). Inference of threat from neutrally-worded utterances in familiar and unfamiliar languages. *York Papers in Linguistics (Series 2, Issue 13)*, 99-120. http://www.york.ac.uk/language/ypl/ypl2issue13/YPL2_2013_Issue_13_Complete.pdf

Weinstein, H., Frazier, D., & Bongar, B. (2009). Why are they attacking us? Decoding the messages of Al-Qaeda terrorists targeting the United States and Europe. *Revue Internationale de Psychologie Sociale, 22*(3), 65-85.

Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2015). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. *Quarterly Journal of Experimental Psychology*, *68*(8), 1599-1622. doi:10.1080/17470218.2014.970204

White, S. G., & Cawood, J. S. (1998). Threat management of stalking cases. In J. R. Meloy (Ed.), *The psychology of stalking* (pp. 295-315). San Diego, CA: Academic Press.

Wood, W., & Quinn, J. M. (2003). Forewarned and forearmed? Two meta-analysis syntheses of forewarnings of influence appeals. *Psychological Bulletin*, *129*(1), 119-138. doi: 10.1037/0033-2909.129.1.119

Woodhams, J., & Grant, T. (2006). Developing a categorization system for rapists' speech. *Psychology, Crime & Law*, *12*(3), 245-260. doi:10.1080/10683160500151134

Xu, J-M., Jun, K-S., Zhu, X., & Bellmore, A. (2012). Learning from bullying traces in social media. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 656-666). Stroudsburg PA: Association for Computational Linguistics.

Xu, J-M., Jun, K-S., Zhu, X., & Bellmore, A. (2012). *Learning from bullying traces in social media*. Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 656-666). Stroudsburg PA: Association for Computational Linguistics.

Zaitsu, W. (2010). Bomb threats and offender characteristics in Japan. *Journal of Investigative Psychology & Offender Profiling*, *7*(1), 75-89. doi: 10.1002/jip.106

Zheng, R., Li, J., Huang, Z., & Chen, H. (2006). A framework for authorship analysis of online messages: Writing-style features and techniques. *Journal of the American Society for Information Science and Technology, 57*(3), 378–393.

APPENDIX A:

Preliminary THReat Evaluation & Assessment of Discourse (THREAD) Index

THREATS

A. **Feasibility**: How *capable* is the threatener to carry out the threat fulfillment possible (e.g., threatening to bring on the plague is not very feasible, whereas spreading disparaging rumors is relatively feasible)?

1. **INFEASIBLE:_1_:_2_:_3_:_4_:_5_:_6_:_7_:FEASIBLE**

B. **Capability/expertise**: Is there evidence the threatener is able to carry out the threat?

2. **INCAPABLE:_1_:_2_:_3_:_4_:_5_:_6_:_7_:CAPABLE**

C. **Extremity/intensity**: How severe are the potential consequences or scope of harm to those threatened (e.g., a practical joke intended to embarrass is relatively minor, whereas threats to kill you and your family are relatively serious)?

3. **NEGLIGIBLE/MINOR:_1_:_2_:_3_:_4_:_5_:_6_:_7_:EXTREMELY SERIOUS**
4. **SLIGHT:_1_:_2_:_3_:_4_:_5_:_6_:_7_:INTENSE**

D. **Self-efficacy**: Does the threatener express confidence and a sense of self-efficacy in carrying out the threat?

5. **INSECURE:_1_:_2_:_3_:_4_:_5_:_6_:_7_:SELF-CONFIDENT**
6. **UNCERTAIN:_1_:_2_:_3_:_4_:_5_:_6_:_7_:SELF-ASSURED**

E. **Prior efficacy**: Is there evidence that the threatener has issued, and followed through with, prior relevant threats?

7. **INEXPERIENCED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:EXPERIENCED**

F. **Linguistic Conditionality**: Is the threat phrased provisionally with highly conditional probability in the verb phrases and contingency phrases (e.g., this “may” or “might” happen) or with highly certain and probable types of phrases (e.g., this “will” or “absolutely is going to” happen)?

8. **IMPROBABLE PHRASING:_1_:_2_:_3_:_4_:_5_:_6_:_7_:PROBABLE PHRASING**
9. **UNCERTAIN PHRASING:_1_:_2_:_3_:_4_:_5_:_6_:_7_:CERTAIN PHRASING**

G. **Immediacy/imminence**: What is the time horizon of the language and implied harm (threatening to make you regret something in your future seems off in the distance, whereas threatening to show up tonight is relatively immediate)?

10. **DISTANT:_1_:_2_:_3_:_4_:_5_:_6_:_7_:IMMEDIATE**
11. **NON-URGENT/NON-IMMINENT:_1_:_2_:_3_:_4_:_5_:_6_:_7_:URGENT/IMMINENT**

H. **Knowledge of target**: How much information and/or insight into the target/victim is manifest in the threat?

12. **UNACQUAINTED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:ACQUAINTED**
13. **IGNORANT:_1_:_2_:_3_:_4_:_5_:_6_:_7_:KNOWLEDGEABLE**

I. **Inclusion of others**: Are others, such as relevant or mutual children, pets, family, etc., included in the threat?

14. **EXCLUSIVE TO TARGET: 1_:_2_:_3_:_4_:_5_:_6_:_7_:INCLUSIVE OF OTHERS**

J. **Referential foci**: Is the threat focused from a self-focus or perspective? Is there a vivid and/or repeated fixation on self vs. other, or one group against another?

15. **FOCUSED ON OTHER(S):_1_:_2_:_3_:_4_:_5_:_6_:_7_:SELF-FOCUSED**
16. **COLLECTIVELY FOCUSED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:FOCUSED ON US/THEM OR YOU-I**

K. **Linguistic deviation:** To what extent does the language diverge or differ from the language of the person or group being threatened?

17. **ACCOMMODATIVE LANGUAGE:_1_:_2_:_3_:_4_:_5_:_6_:_7_:DIVERGENT LANGUAGE**

L. **Plan Complexity**: How complicated is the expressed threat (are there many steps, rigid sequences of steps, or multiple endeavors required to carry out the threat, or is the threat relatively simple and straightforward.

18. **COMPLEX:_1_:_2_:_3_:_4_:_5_:_6_:_7_:SIMPLE**
19. **CIRCUITOUS:_1_:_2_:_3_:_4_:_5_:_6_:_7_:STRAIGHTFORWARD**

M. **Message Mode**: Is the threat purely verbal, or are there also nonverbal (e.g., objects, visual elements such as drawings or photographs, etc.) components of the threat?

20. **EXCLUSIVELY VERBAL:_1_:_2_:_3_:_4_:_5_:_6_:_7_:NONVERBAL AND/OR VERBAL**

N. **Goal-linking**: Is there evidence in the language of higher-order goal linking of, or (inter)dependency on the target with threatener's life objectives and/or values (e.g., "I can't be happy without you," "There is no one in the world for me but you," etc.), or are the threats unlinked to the target person (e.g., "Bad things are going to happen")?
21. **UNLINKED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:LINKED**
22. **INDEPENDENT:_1_:_2_:_3_:_4_:_5_:_6_:_7_:(INTER)DEPENDENT**

O. Identification fixation: To what extent do words or phrases indicate fixation, preoccupation, and personal identity fusion with a topic, entity, or person?
23. **DIFFUSED IDENTITY:_1_:_2_:_3_:_4_:_5_:_6_:_7_:PREOCCUPIED IDENTITY**

P. **Coherence/organization**: is there evidence that the threatener has engaged in planning, preparation, and/or has an overall organizing vision of implementing the threat, or is the threat disorganized, chaotic, and ill thought out?
24. **INCOHERENT:_1_:_2_:_3_:_4_:_5_:_6_:_7_:COHERENT**
25. **DISORGANIZED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:CHAOTIC**

Q. **Sentiment deterioration**: Is there an increase in, or degree of emphasis on speech with increasingly negative, anger-based terminology?
26. **AFFECT NEUTRAL OR BALANCED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:INCREASINGLY ANGRY**

R. **Delusional content**: Does the content suggest psychoses or lack of mental competence (are there indications of unrealistic visions, conspiracy theories, illusions, fantasies, or other psychotic content)?
27. **DELUSIONAL:_1_:_2_:_3_:_4_:_5_:_6_:_7_:ACTUALITY**
28. **FANTASTICAL:_1_:_2_:_3_:_4_:_5_:_6_:_7_:GROUNDED**

S. **Embeddedness**: Are the threats embedded in a broader manifesto, or isolated fragmented thoughts or outbursts?
29. **FRAGMENTED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:PHILOSPHICALLY EMBEDDED**
30. **ISOLATED:_1_:_2_:_3_:_4_:_5_:_6_:_7_:IDEOGICALLY EMBEDDED**

T. **Finality fantasies**: Are there "end-game," suicide fantasies or images, suggested (e.g., "If I can't have you, no one can," "I'll take you and me down together," "It will all end soon"), or is the language more optimistic (e.g., "Life would be so wonderful with you in it," "I believe we would make the most amazing couple," etc.)?
31. **HOPEFUL:_1_:_2_:_3_:_4_:_5_:_6_:_7_:HOPELESS**
32. **ENCOURAGING:_1_:_2_:_3_:_4_:_5_:_6_:_7_:FATALISTIC**

DEPENDENT VARIABLES: Respond to the next 5 items on a 7-point scale from:

STRONGLY DISAGREE (0) to STRONGLY AGREE (7)

Holistic Credibility Rating:

33. The speaker presents a credible threat.
34. The speaker intends to carry out their threat.
35. The speaker is likely to carry out their threat.
36. The speaker seems determined to do something harmful to someone or something.

Holistic Danger Rating:

37. The speaker seems dangerous.
38. I would be afraid (i.e., experience fear) if I received this message.
39. The speaker appears to be preparing to do something violent.
40. I view this as a serious and/or imminent threat.

Holistic Threat Ranking:

0 = Not a serious threat
1 = A minor threat
2 = A serious but not imminent threat
3 = An imminent and severe threat