How often is Employee Anger an Insider Risk I? Detecting and Measuring Negative Sentiment versus Insider Risk in Digital Communications

Eric Shaw, Ph.D.*
Suite 514
5225 Connecticut Ave. NW
Washington, DC 20015
eshaw@msn.com
Maria Payri, M.A.
Michael Cohn, M.A.
Ilene R. Shaw, R.N.

*Corresponding Author

Keywords: Insider Risk, Digital Communication, Disgruntlement, Detecting, Negative Sentiment, Threat, Employee Anger, Workplace Violence, Content Analysis, Automated Psychological assessment

ABSTRACT

This research introduced two new scales for the identification and measurement of negative sentiment and insider risk in communications in order to examine the unexplored relationship between these two constructs. The inter-rater reliability and criterion validity of the Scale of Negativity in Texts (SNIT) and the Scale of Insider Risk in Digital Communications (SIRDC) were established with a random sample of email from the Enron archive and criterion measures from established insiders, disgruntled employees, suicidal, depressed, angry, anxious, and other sampled groups. In addition, the sensitivity of the scales to changes over time as the risk of digital attack increased and transitioned to a physical attack was also examined in an actual case study. Inter-rater reliability for the SNIT was extremely high across groups while the SIRDC produced lower, but acceptable levels of agreement. Both measures also significantly distinguished the criterion groups from the overall Enron sample. The scales were then used to measure the frequency of negative sentiment and insider risk indicators in the random Enron sample and the relationship between the two constructs. While low levels of negative sentiment were found in 20% of the sample, moderate and high levels of negative sentiment were extremely rare, occurring in less than 1% of communications. Less than 4% of the sampled emails displayed indicators of insider risk on the SIRDC. Emails containing high levels of insider risk comprised less than one percent or the sample. Of the emails containing negative sentiment in the sample, only 16.3%, also displayed

indicators of insider risk. The odds of a communication containing insider risk increased with the level of negative sentiment and only low levels of insider risk were found at low levels of negative sentiment. All of the emails found to contain insider risk indicators on the SIRDC also displayed some level of negative sentiment. The implications of these findings for insider risk detection were then examined.

1. INTRODUCTION

News reports continually feature insider episodes including acts of workplace violence, espionage, sabotage, theft of intellectual property, harassment, violations of financial rules and leaks of sensitive information by employees or other insiders. Recent research on the relatively low frequency of organizational reporting of many classes of non-violent insider offenses to law enforcement indicates that these public cases are only the tip of the insider iceberg (Computer Security Institute [2011] 15th Annual Computer Security Institute 2010/2011 Survey, CSI www.GoCSI.com). Whether public or private, digital communications frequently form the core of investigative leads and evidence in insider cases. All too often, as in the recent cases of the Boston Marathon bombers (The New York Times, 2013), and other dangerous insiders (e.g., Pittsburgh gym killer George Sodini [The Telegraph, 2009], accused anthrax killer Bruce Ivins [Expert Behavioral Analysis Panel, 2010], etc.), investigators find that the writing was on the virtual wall in terms of the warning signs of insider risk discovered. While there are many available approaches for detecting and analyzing technical anomalies indicative of insider risks related to theft of intellectual property, espionage, leaks, and fraud (unusual copying, use of prohibited memory devices, etc.), there are few tools for detecting content indicative of insider risk in digital communications. Improved detection and assessment of risk indicators in content might facilitate more effective investigation, prevention, and management of insider and other problems by helping investigators prioritize leads based on technical detection signals and assigning priority to individuals who are also disgruntled. In addition, human scales that measure risk can serve as a valuable benchmark for computerized risk assessment detection approaches.

Such improvements in online risk detection tools seem particularly important as more of our lives have moved from face-to-face and telephonic communication to online exchanges. In addition, research on the lack of coworker and supervisor reporting of observed risk indicators indicates that security awareness programs seeking to increase such reports are facing significant cultural obstacles (Wood and Marshall-Mies, 2003).

1.1 Negative Sentiment and Insider Risk

Employee frustration and anger have long been associated with aggression and violence in the workplace (Glomb and Liao, 2003; Hershcovis et al., 2007; Hershcovis and Barling, 2010), as well as turnover, absenteeism, accidents on the job, alcohol consumption, and other high-risk health behaviors (O'Neil et. al. 2009). Holton (2009) also found an association between anger and fraud, and Band and colleagues (Band et al., 2006) found similar links to sabotage and espionage. Occupational health researchers who study a range of counter-productive work behaviors (CWBs), from taking long lunches to workplace violence, have consistently found a strong link between negative emotions and CWBs (Brief and Weiss, 2002; Dalal, 2005; Sakurai and Jex, 2012; Schat and Kelloway, 2005).

However, not all forms of anger or negative sentiment pose a risk of such counter-productive behavior. For example, Averill (1983) estimated that only around 10% of incidents involving anger result in aggressive or violent outcomes. Occupational health researchers surveying large groups of employees for the presence and frequency of CWBs routinely note such contributing factors as individual baseline anger (trait negative affectivity), job type, and job autonomy, all of which can contribute to negative sentiment without increasing the risk of CWBs. In addition, occupational health researchers have found that other factors, such as social and supervisor support, can mediate the link between negative emotions and CWBs (Sakurai and Jex, 2012). These researchers have also documented a range of possible employee reactions to negative emotions that may substitute for insider acts, including withdrawal, avoidance, regulation of negative emotion through diet, smoking, exercise, or leaving the workplace. Clearly, an individual's disposition, job stress, freedom to react to stress, and previous experience can set the stage for his reactivity to perceived aversive stimuli and modify the likelihood of CWBs. There are also probably few organizations where some form of negative sentiment regarding working conditions or related issues is not part of the background "noise" in employee communications.

Thus, the use of anger or negative sentiment alone, or the routine use of low levels of negative sentiment as an indicator of insider risk or other CWBs may result in false positive reports, distracting attention from more serious cases. Based on their workplace data, Calhoun and Weston (2008) have argued that authors expressing disgruntlement and even threats (so-called "Howlers") may be at much *less* risk for actual attacks than those who more carefully plan without any form of expressed warning (so called "Hunters").

There is also a strong ethical and scientific tradition in clinical psychology test development of ensuring that clinical measures are highly specific in their ability to identify criterion groups of concern, versus false positives. Thus, measures of depression, anti-social behavior, or attention deficit disorders should identify at

least 80% of persons with these syndromes in clinical samples and differentiate them from persons who do not have these disorders. Similarly, effective detection tools for insider risk should, theoretically, be able to differentiate persons with negative sentiment who do and do not pose a threat of insider activity. While we have not yet achieved this level of diagnostic specificity or selectivity, the deployment of both a rating scale for negative sentiment and a separate scale for insider risk will allow improved investigation of the relationship between these two phenomena and help us better understand areas of overlap and independence between the constructs. This area is so unexplored that we do not know the answer to such basic questions as: (1) What is the frequency of communication with negative sentiment or insider risk in a random sample of organizational emails? (2) What percentage of digital communications with significant negative sentiment also contains insider risk indicators? (3) What percentage of messages associated with demonstrated insider risk also contain negative sentiment? (4) Are particular types of negative sentiment more likely to be associated with different types of insider risk such as violence, espionage, leaks, fraud, or theft of intellectual property? Without separate rating scales for these two phenomena these issues have been difficult to address empirically.

Currently, there are no published, validated, specialized rating scales for the detection and measurement of negative sentiment in communications for use by analysts of insider risk and investigators of insider violations. While on the surface, the detection of negative sentiment may seem a rather straight-forward pursuit, there are so many ways in which individuals can express negative feelings and judgments without the use of overtly negative terms that the detection and rating of negative expressions can actually be quite complicated. A validated measurement system for use by analysts and investigators might eliminate threats to reliability derived from subjective "expert" or other human judgments, buttress the credibility of these judgments with empirical support, improve analyst detection rates, sensitize analysts to changes over time that might signal increased risk, and help investigators narrow a field of suspects according to objectively measured levels of disgruntlement contributing to subject motivation. Computerized, content-based risk detection methods may always be challenged to detect the many nuanced aspects of human negative expression such as sarcasm ("that worked out well"), irony ("you got as good as you gave") and even non-negative forms of negative expression such as protest ("I've always done my best for the Company"). Human-based coding schemes that capture these more subtle forms of negative emotion will provide a critical benchmark for these particularly challenging tasks. Even if the proposed scales prove more useful in research than in applied evaluations and investigations, familiarity with these measures might still improve analyst and investigator sensitivity, as well as coworker reporting.

To address some of these gaps in the literature, this paper describes two observational scales designed to detect and measure levels of negative sentiment and insider risk, respectively, within digital or other content. The derivation of these scales, their inter-rater reliability, and their performance with criterion groups are presented. Scale results with a random sample of employee communications from the Enron archive combined with a random sample of established insider communications are also examined to shed light on the unexplored relationship between negative sentiment and insider risk in digital communications.

The balance of this paper describes the Scale of Negativity in Text (SNIT) and the Scale of Insider Risk in Digital Communication (SIRDC), as well as the research design and results obtained from tests of their inter-rater reliability and performance with criterion groups. The SNIT and SIRDC results are then explored to address the relationship between expressed negative sentiment and insider risk.

2. DETECTING AND MEASURING NEGATIVE SENTIMENT AND INSIDER RISK

2.1 Description of the SNIT and SIRDC

2.1.1 The Scale for Negativity in Text (SNIT)

The Scale for Negativity in Text was designed to help researchers and investigators detect and score the frequency of negative feelings and attitudes in communication. While the SNIT codes straight-forward judgments and feelings with negative connotations, it was also designed to detect and score more subtle and complex forms of expression communicating negativity. For example, the SNIT identifies and codes the frequency of negative judgments and feelings, as well as terms that add emphasis or power to these sentiments. So, in the example "I deeply resent your intrusion," "resent" would be coded as a negative feeling, and "deeply" would be scored as an adverbial intensifier—a term that increases the power of the feeling of resentment (Weintraub, 1981, 1989). Other non-verbal intensifiers may include exclamation marks, underlining, emoticons, or other "non-verbal" forms of emphasis. In addition, "your intrusion" would be coded as a direct accusation, criticism, or attack against a specific individual or group. Other examples of direct negative sentiment scored by the SNIT include statements of opposition or negation ("I won't do that") (Weintraub, 1981, 1989), direct and indirect threats, use of curses, foul language or other slurs, dehumanizing sexual material, sarcasm, rhetorical questions or negative irony, negative religious or ethnic attacks, and provocations or taunts.

The SNIT also identifies and scores more subtle and complex expressions of negative sentiment, including appeals, pleas, requests, or demands that communicate author discomfort without expressing overt negativity. For example, the phrase "please listen to me" does not contain any overtly negative content but may represent a statement of author discomfort. In addition, the SNIT identifies and scores neutral or even positive statements that imply negativity, protest, criticism, or opposition without direct expressions of negativity. For example, the statement "she left me" does not include any overtly negative material, but the context can indicate author disappointment with the event. The phrase "I have always done my best for the Company" also does not contain any overtly negative content. However, given the appropriate context, it could be coded by the SNIT as a non-negative statement of protest. Table 1 in the Appendix displays these 16 SNIT categories and examples of coded terms for each variable.

As an example from insider communication, the excerpt below was taken from Army Private Bradley Manning's correspondence with a hacker contact during the period in which he was accused of leaking classified material to Wikileaks (Hansen, 2011).

> *i cant believe what im confessing to you :'(ive been so isolated so long… i just wanted to be nice, and live a normal life… but events kept forcing me to figure out ways to survive… smart enough to know whats going on, but helpless to do anything…*

Table 1 identifies examples of terms from this passage that would be coded on the SNIT by category.

2.1.2 The Scale of Insider Risk in Digital Communication (SIRDC)

The SIRDC contains seven components related to the detection and scoring of insider risks, including such insider acts as violence, sabotage, espionage, IP theft, and damaging leaks. These seven components include:

- Process variables that indicate the extent to which subject behavior that could be directly associated with, or contribute to, the accomplishment of insider actions is present and/or increasing (preparations, rehearsals, etc.);
- Psychological State variables that indicate the extent that subject attitudes, beliefs, and feelings are consistent with individuals who have committed insider acts;
- Personal Predisposition variables that indicate the extent to which the subject's observed history, experiences, personal characteristics, and contacts mirror those of previous insider subjects;
- Personal Stressors;
- Professional Stressors;

- Concerning Behaviors, such as violations of workplace or other rules, traditions, laws, policies, or procedures that indicate the extent to which the subject has had difficulty controlling his behavior consistent with expectations, in a manner similar to other insiders; and
- Mitigating factors indicating that the subject's level of insider risk may be modified by personal or other characteristics that reduce the level of risk.

Table 1 Examples of Terms and SNIT Code by Category

SNIT Coding for Bradley Manning Passage by Content and SNIT Category	**SNIT Category Scored**
Can't	Negation or Opposition Statement
Confessing	Non-Negative negative
So	Adverbial intensifier
isolated	Negative feeling
So	Adverbial intensifier
long	Non-negative Negative
:	Non-verbal emphasis
…	Non-verbal emphasis
just	Adverbial intensifier
wanted to be nice	Non-negative negative
live a normal life	Non-negative negative
…	Non-Verbal emphasis
But	Negation/Opposition
forcing me	Negative Evaluator
figure out ways to survive	Non-negative negative
…	Non-Verbal emphasis
But	Negation/Opposition
helpless	Negative feeling
anything	Adverbial intensifier

2.1.2.1 Process Variables: In the area of violence prediction, the Association of Threat Assessment Professionals (ATAP, 2006) lists process variables as threat indicators that suggest that a subject is actually approaching an act of violence. Variables such as evidence of escalation, attack rehearsal, actual attack preparations, and weapon acquisition are included in this category (ATAP, 2006, p. 7). The SIRDC includes many of these variables as predictive of increased violence risk but also utilizes these process measures more broadly to denote other types of insider attack preparations. For example, a subject in an angry or violent state of mind is more likely to act on his rage, but the type of action he or she chooses may be better predicted from his or her personality, experience,

and access to tools that can facilitate his destructive goals. Subject fantasies about insider acts, rehearsal, and feelings that other options are not available also carry over from violence to broader insider act predictors on the SIRDC. In addition, some extreme states of mind associated with violence also are included as broader insider risk predictors, such as suicidal thoughts, a loss of cognitive control or inhibitions due to substance abuse or other factors, and depersonalization of a potential target, making it easier to attack.

2.1.2.2 At-Risk Psychological States: The next section of the SIRDC includes at-risk psychological states that may not be as extreme as those cited above, but that refer to explicit impairments of rational thought that have been associated with insider actions. These states of mind include content indicators of:

- Suspiciousness,
- Accusations,
- Victimization,
- Excessive Blame,
- Obsession with a potential target or situation;
- Rationalization of destructive actions, and
- A range of injustice attributions regarding an organization and/or its leadership.

For example, an individual who is suspicious that he is going to be terminated due to some perceived bias against him may manifest suspicion, feelings of victimization, blame his supervisor or other coworkers, describe perceived injustices he attributes to organizational leaders, and rationalize his actions prior to leaving the organization and stealing its intellectual property.

The passage below was taken from an insider who leaked proprietary information from an organization and eventually sought extortion money to cease his activities. It would be coded on the SIRDC for examples of injustice attributions, suspiciousness, victimization and excessive blame.

> *It was not enough for government officials to destroy my financial resources; but they also had to destroy my reputation and violate my civil rights, including my first amendment rights, by threatening those who associated with me. For all I have been accused of, I should have been locked away for life. Where was the Justice Department then? Because I have endured multiple punishments without ever having any opportunities to know what the charges were or even have one proceeding to present evidence on my behalf, its payback time.*

2.1.2.3 Personal Predispositions: Personal predispositions are derived from previous research on the critical pathway travelled by insiders as they move through a series of steps over time within their organization taking them closer

to committing sabotage, espionage, IP theft, violence or fraud. This critical pathway has been described by Shaw and Fischer (2005); Band and colleagues (Band et al., 2006); Shaw, Fischer, and Rose (2009); and Shaw and Stock (2011). Personal predispositions describe characteristics and experiences of individuals prior to joining their organizations that, in the presence of other precipitants, appear to make them more vulnerable to participating in insider violations. These personal predispositions include:

- Serious mental health disorders or medical conditions impacting perception, judgment, impulsiveness and decision-making, such as alcoholism, attention deficit disorders, anxiety disorders, depression, etc.
- Personality disorders, social skills problems and/or significant biases in personal and professional decision-making that consistently negatively impact personal and professional relationships producing observable, maladaptive personal and professional behaviors.
- A history of rule violations ranging from such serious conduct as prosecuted legal violations and convictions (DWIs) to less serious offenses such as chronic tardiness, violations of dress code or hygiene regulations, ignoring organizational policies and practices, or other violations of organizational protocol.
- Social network risks, such as a family history of criminal activity or membership in an adversary group, or any pre-employment (at location from which espionage is committed) contact (face-to-face, telephone or digital) with members of an adversarial at-risk group.

2.1.2.4 Personal Stressors: Within the critical pathway framework, personal and professional stressors are seen as activating personal predispositions and contributing to an increased likelihood of Concerning Behaviors and insider risk. Personal stressors are defined as changes in personal or social responsibilities or conditions requiring significant energy for adaptation, which do not involve direct workplace or financial issues—including death of a family member, marriage, divorce, births, moves, etc. These events or experiences are derived from the Holmes-Rahe Scale of stressful life events (Holmes and Rahe, 1967).

2.1.2.5 Professional Stressors: Professional stressors include changes in professional, school, and/or work conditions or responsibilities that require significant energy for adaptation, exclusive of financial and personal implications—changes in affiliation, graduation, attending a new institution or obtaining a new job, demotion, termination, promotion, transfer, retirement, consulting work, taking side jobs, etc. The loss or gain of income related to new school, job, promotion, or termination is scored separately in financial stressors.

In the Manning excerpt above, he references plans for his discharge, being demoted from intelligence to supply duties, and his reaction to news coverage of his alleged leaks. Both these negative and "positive" events would be coded as professional stressors within this section.

2.1.2.6 Concerning Behaviors: Concerning Behaviors are violations of policies, standard procedures, professional conduct, accepted practice, rules, regulations, or law through action or inaction (failure to report), which have been observed by managers, supervisors, coworkers, or reported to these individuals by others. For example, failure to submit a time sheet in a timely manner, unreported travel, misuse of expense accounts, violations of hygiene or dress, refusal to follow supervisor instructions, going around a supervisor to a superior, coworker or supervisor conflicts, in some settings, filing complaints or protests against other employees or supervisors, etc. This category of variables is also derived from work on the Critical Pathway to insider risk described above by Shaw and Fischer (2005); Band and colleagues (Band et al., 2006); Shaw, Fischer, and Rose (2009); and Shaw and Stock (2011).

References to the occurrence of any of the following behaviors or failures to act are recorded as concerning behaviors including violation of accepted policies and practices governing:

- Interpersonal conduct;
- Use of information technology or other technical systems;
- The protection of sensitive or classified information;
- Physical security;
- Financial conduct;
- Personnel security;
- Travel rules; and
- Social network contacts and affiliations.

References to mental health or addiction problems that could impact judgment or behavior are also included in this section. In addition, other categories and individual actions may be included tailored to the definition of "concerning" that applies to the organization involved. As noted above, there is considerable overlap between personal predispositions and concerning behaviors, with the only difference in some of these categories consisting of the timing of the behavior. Behaviors or experiences noted prior to the subject's joining the organization are categorized as personal predispositions (characteristics he brought to the organization), while similar issues noted while on the job are considered concerning behaviors.

Digital references to any of these issues are coded within the Concerning Behavior category, including signs of conflict with others across digital media. For example, portions of the excerpt below, taken from an insider who sabotaged

servers at a financial institution, would be coded for such concerning behaviors as interpersonal conflict with a supervisor and failing to follow instructions regarding the operation of an information technology resource. Additional codes would include professional stressors for the references to being fired, relieved of duty, or quitting.

> *Until you fire me or I quit, I have to take orders from you. I'll sit with K after I've written some procedures on what he can do. Just like he cannot have LAN supervisor password until he is a trained LAN expert, I won't give him Sybase ROOT access until he has been trained to be of some minimal use. If you order me to give him root access, then you have to permanently relieve me of any duties on that machine. I can't be a garbage cleaner if someone screws* up.

Examples of coded material using both the SNIT and the SIRDC are included in the Appendix. In the next section the research design used to assess the scales' inter-rater reliability and performance is described.

3. RESEARCH DESIGN

The first objective of this research was to assess the inter-rater reliability and criterion group validity of the SNIT and SIRDC scales. For this purpose, two email samples were selected at random from the full, publically available Enron email corpus (http://www.edrm.net/projects/dataset) using Net's Random class protocol (http://msdn.microsoft.com/en-us/library/system.random.aspx). Three coders—a psychiatric nurse and two individuals with master's degrees in Political Psychology, but no advanced threat assessment or clinical psychological training—performed the coding.

SNIT and SIRDC inter-rater reliability was first tested with 75 randomly selected emails from the Enron email archive. The SNIT's initial inter-rater reliability as determined by the Intraclass Correlation Coefficient (ICC) (Koch & Norman, 1982) which was used due to the presence of continuous versus nominal or ordinal data, was .975 ($p \leq 0.05$). SIRDC's initial ICC was .862 ($p \leq 0.05$). A preliminary analysis of the SNIT scale's subcategories showed that four measured overlapping manifestations of negative sentiment and could be merged and redefined to improve inter-rater reliability.

We anticipated that a relatively low proportion of Enron emails would contain negative sentiment and insider risk. Therefore, we sought additional subjects from criterion groups with known negative sentiment, and in some cases, insider actions, to test scale inter-rater reliability as well as scale performance with these more expressive and complicated subjects. For this purpose, we collected communication samples from the following 13 sources:

- Ten communications from subjects who subsequently committed insider

attacks were randomly selected from the first author's investigative case archive. These were individuals who expressed themselves online prior to or during their insider actions;

- Thirteen disgruntled employees from the Enron archive who complained, but did not subsequently take insider actions according to Google searches of their names. These subjects were selected by a visual scan of subject headers followed by inspection of the email from the original Enron archive;
- Seventeen time series emails from an online stalker who subsequently attacked his target physically, taken from the first author's case archive. These emails were selected to assess scale sensitivity to subject changes over time and especially any indication of heightened risk;
- Five publically available emails from Bruce Ivins, who was implicated but never tried in the 2001 Anthrax attacks, taken from a public FBI report on the investigation (U.S. Department of Justice [2010] Amerithrax Investigative Summary, http://www.justice.gov/amerithrax/docs/amx-investigative-summary.pdf);
- Ten communications selected by a clinician for representation of depressive communication from participants in public but anonymous online chat rooms for anxiety and depression (http://www.depressionhaven.org/phpBB2/viewtopic.php?t=8397 (http://www.anxietyzone.com/index.php/board,4.0.html?PHPSESSID=e9a40ba11ef20150e3649cc1aca3456f);
- Ten communications selected by a clinician for representation of anger from participants in a public but anonymous online chat room for anger (http://www.anxietyzone.com/index.php/topic,31323.0.html);
- Ten communications selected by a clinician for representation of financial stress from participants in public but anonymous online chat rooms for financial stress (http://www.indeed.com/forum/cmp/Target/target-is-bogus/t8403; http://www.anxietyzone.com/index.php/topic,51034.0.html; http://www.experienceproject.com/stories/Hate-Living-Paycheck-To-Paycheck/2161275);
- Ten communications selected by a clinician for representation of suicide risk from participants in public but anonymous online chat room for suicide risk (http://www.yourlifeyourvoice.org/AskIt/Pages/default.aspx);
- Ten communications selected by a clinician for representation of substance abuse from participants in a public but anonymous online chat room for substance abuse problems (http://www.anxietyzone.com/index.php/board,13.0.html?PHPSESSID=e9a40ba11ef20150e3649cc1aca3456f);

- Ten communications selected by a clinician as representative of disgruntled employee communication from participants in public but anonymous online chat rooms for disgruntled employees complaining about work stress (http://www.anxietyzone.com/index.php/topic,51034.0.html and http://www.experienceproject.com/stories/Hate-Living-Paycheck-To-Paycheck/2161275);
- The content from an OpEd piece in the Wall Street Journal by Greg Smith complaining about the ethical climate at Goldman Sachs as he resigned (Smith, 2012),
- Email communications from alleged US Army leaker Bradley Manning with his hacker contact taken from public coverage of the incident (Hansen, 2011); and
- An email from U.S. Navy sailor Paul Hall, who changed his name to Hassan Abu-Jihaad, communicating from his destroyer to an established website sympathetic to Al Qaeda. Hall was convicted of providing material support to terrorists in March 2008 (United States District Court, 2009).

Although other researchers have used similar data to study disgruntled populations (Holton, 2009), we were concerned that use of a convenience sample of publically posted comments from individuals speaking with potential anonymity on self-help chat boards might not generalize to content found in organizational email. However, we were anxious to test scale performance with self-identified criterion groups with problems known to arise in the workplace. Subsequent criterion group materials derived from workplace communications might improve the generalizability of this sample. To ameliorate these concerns we included actual content from disgruntled Enron employees. Since discussing this dilemma with personnel responsible for monitoring email content at several government and commercial organizations, we were assured that employees are surprisingly frank in their discussions of personal emotions and life crises and that it would not be unusual to discover employee discussions of suicide, financial distress, and other serious concerns while searching for references to policy or security violations.

To assess the face and criterion validity of the scales we planned to compare the mean results from these established groups with 1000 randomly selected emails from the Enron archive with the hypothesis that the criterion groups would score significantly higher on both measures. This data set was reduced to 994 emails when duplicates were discovered. We did not make predictions regarding the relative ranking of groups on the SNIT or the SIRDC scales due to the limited nature of the samples and the unknown expressive characteristics of insiders versus the other criterion groups. However, we did expect some portion of the samples from actual insiders (including Private Manning, Bruce Ivins, Abu

Jihaad and the Online Stalker) to score higher on the SIRDC. To address the scales' sensitivity to changes over time we charted the scores of the Online Stalker in 17 emails to his intended victim leading up to, and after, his physical attack.

To address the earlier questions regarding the frequency of negative sentiment and insider risk indicators in organizational email we calculated the simple percentage of subjects expressing a range of these variables. To address the question of the proportion of subjects with different ranges of negative sentiment that also manifested insider risk indicators, we performed a simple review of the percentage of High, Medium and Low SNIT subjects that also manifested scores on the SIRDC. To determine the proportion of subjects with different levels of insider risk that did or did not also display negative sentiment, we reviewed the distribution of SNIT scores among subjects with High and Low SIRDC scores.

4. RESULTS

Inter-rater reliability for the SNIT and SIRDC across the randomly selected Enron subjects was .919 and .915, respectively. As noted above, this result may have been inflated due to the low frequency of subjects manifesting negative sentiment and insider risk and the subsequent frequency of zero scores. However, we were encouraged that coders could agree on whether these often subtle variables were present or absent in this sample. Inter-rater reliability across the more complicated criterion groups averaged .969 ($p \leq 0.05$) for the SNIT scale and .731 ($p \leq 0.05$) for the SIRDC scale.

While inter-rater reliability for the SNIT proved excellent measured against an ICC criteria of .70, the performance of the longer and more complicated SIRDC was less consistent among coders. Inter-rater reliability scores for four of the 13 groups performed just below the .70 criteria, indicating the need for further work on the SIRDC scales to improve coder reliability. This may also be a product of the lower frequency of SIRDC scores in this sample, limiting coder experience with the scale. However, the global ICC score of .731 indicated that coder reliability overall was acceptable.

4.1 What is the Frequency of Emails Containing Negative Sentiment in a Randomly Selected Sample of Corporate Communication?

Table 2 below displays the frequency of emails with High, Medium, Low and No negative sentiment as measured by the SNIT within the randomly selected Enron email sample.

Table 2 Distribution of negative sentiment in ENRON Sample of High, Medium, Low, and No SNIT score groups

SNIT Group	# Emails Per Group and Percentage of Total	Mean SNIT Score
High (SNIT Score ≥31)	6 (0.6%)	44.39
Medium (SNIT Score ≥16 and ≤30)	9 (0.9%)	19.93
Low (SNIT Score ≥1 and ≤15)	207 (20.82%)	2.56
No Negative Sentiment	772 (77.67%)	0

Illustrative examples of email excerpts from each category containing negative sentiment are displayed in the Appendix. It was notable that many of the insider risk issues captured concerned potential fraud, interpersonal conflict, litigation, as well as organizational and interpersonal disgruntlement. As can be seen in Table 2, while low levels of negative sentiment were common, moderate and high levels were extremely rare in this sample, occurring in less than 1% of communications. However, if this finding were extrapolated to the large number of emails contained in organizational systems, this rate of discovery would be equivalent to finding 6,000 emails high in negative sentiment within a million email cache.

4.2 What is the Frequency of Emails Containing Insider Risk Indicators in a Randomly Selected Sample of Corporate Email?

Table 3 below displays the distribution of insider risk indicators recorded on the SIRDC within the randomly selected Enron email.

Table 3 Distribution of SIRDC Scores in Enron Sample

Group	Number and Percentage of Email	Mean Group SIRDC Score
High SIRDC (≥9)	2 (.2%)	12.3
Low SIRDC (<9)	34 (3.4%)	1.4
No SIRDC score	958 (96.4%)	0

Illustrative examples of email excerpts representative of both the High and Low SIRDC groups are contained in the Appendix. While almost 22% of sampled emails displayed negative sentiment, less than 4% displayed indicators of insider risk. Emails containing high levels of insider risk according to the SIRDC

comprised less than one percent or the cache. Although just over three percent of emails contained low levels of SIRDC scored risk, these scores were extremely low, as displayed in Table 3.

4.3 What Percentage of Emails Containing Negative Sentiment Also Contain Some Level of Insider Risk Indicator?

Table 4 below displays the distribution of insider risk at different levels of negative sentiment. Of the 994 emails in the Enron sample, 222 or 22% contained some level of negative sentiment according to the SNIT. Of these 222, 36 or 16.3% also displayed indicators of insider risk. This finding indicates that only a very low percent of emails with negative sentiment also contain indicators of insider risk. However, the data in Table 5 also indicates that the odds that an email will contain insider risk increase as the level of negative sentiment rises. As Table 4 displays, two-thirds of the emails with high levels of negative sentiment contained insider risk indicators, while only 13.5% of emails with low negative sentiment contained insider risk indicators. However, the distribution of insider risk across emails with negative sentiment does not appear to be straightforward in this sample. As Table 6 shows, the two emails highest in insider risk appeared in the Medium SNIT group.

Table 4 Distribution of Insider Risk across High, Medium, and Low Negative Sentiment Groups in the Enron Sample

SNIT Group	Number of Emails with SIRDC Scores and Percent of Total Emails in Group with SIRDC Score
SNIT High (31 or greater) N=6	4 (all low) 66.6%
SNIT Medium (16-30) N=9	4 (2 high, 2 low) 44.4%
SNIT Low (1-15) N=207	28 (all low)13.5%
All SNIT Emails N=222	36 (16.2% of all emails in sample)

4.4 What Percentage of Emails Containing Insider Risk Content Also Contains Some Level of Negative Sentiment?

Table 5 below examines the 36 emails from the Enron sample that contained either high or low levels of insider risk indicators on the SIRDC to determine how many also contained negative sentiment as measured by the SNIT. As the table indicates, all of the emails with insider risk also contained negative sentiment. As the mean SNIT scores show, the level of negative sentiment was higher for the High insider risk group than for the Low risk group.

Combined with the data displayed in Table 4, this pattern of results indicates that overall, emails with negative sentiment are far less likely to contain insider risk indicators, while all emails with insider risk indicators contain negative sentiment. In addition, these results indicate that the odds of an email with negative sentiment containing insider risk indicators increases with the level of negative sentiment and that the level of negative sentiment found in emails with insider risk increases as the level of insider risk increases. An important implication of this finding is that any approach utilizing negative sentiment alone to locate the communications of individuals at-risk for insider actions will be handicapped by a very high, built-in, false positive rate. Another finding with practical implications is that emails low in negative sentiment contained exclusively low levels of insider risk. This finding may help analysts prioritize their search for at-risk individuals by avoiding this group.

Table 5 Distribution of SNIT Scores in High and Low SIRDC Groups

SIRDC Group	**Number of Emails with SNIT Scores and (Percent of Emails in Group with SNIT Score)**	**Mean SNIT Scores**
SIRDC High ≥9 N= 2	2 (100%)	26
SIRDC Low (1-9) N=34	34 (100%)	11.21

4.5 Do the SNIT and the SIRDC Successfully Differentiate Groups with Known Levels of Negative Sentiment and Insider Risk from a Randomly Generated Sample of Organizational Email?

As Tables 6 and 7 indicate, the SNIT and SIRDC scores for the criterion groups were significantly different than those for the overall Enron sample.

4.6 Are the SNIT and SIRDC Sensitive to Changes over Time in an Actual Insider as the Risk of Action Moving from Online Harassment to Physical Assault Increased?

Figure 1 depicts changes over time in the SNIT and SIRDC scores of emails from an online stalker to his victim. These scores were normalized for number of words to control for email length. This harassment turned from hostile and threatening communications to an actual assault on the victim's car just after Email 12 but prior to Email 13 (on Valentine's Day). As can be seen in Figure 1, the subject's SNIT and SIRDC scores peaked just prior to the attack and then declined immediately afterwards. It was also informative to observe the independence of the SNIT and SIRDC at the early stages of the case when the communications from the stalker were emotionally distraught but not threatening. As his frustration grew, so did the threatening nature of his email and thus, his SIRDC score. Table 5 in the Appendix provides illustrative excerpts

and SNIT and SIRDC values for the three emails demonstrating the SNIT's and SIRDC's relative independence (emails 3 and 17) and overlap (email 12).

Table 6 Mean SNIT Scores by Criterion Groups and Enron Sample

SNIT Scores for Criterion Groups and Enron Sample	**Mean SNIT Score**	**Standard Deviation**	**Significance Value p≤**
Depressed Chat Participants	98.23	56.36	.001
Disgruntled with Job Chat Participants	83.53	51.28	.001
Disgruntled Enron Employees Email	79.64	44.25	.000
Angry Chat Participants	69.27	43.01	.001
Substance Abuse Chat Participants	64.30	38.45	.001
Suicidal Chat Participants	60.33	26.09	.000
Ten Actual Insider's correspondence	47.68	33.37	.002
Financial Distress Chat Participants	38.77	38.91	.002
Five Emails from Bruce Ivins	35.87	11.23	.002
Mean for 17 emails from Online Stalker	28.16	21.22	.000
Overall Mean for Criterion Groups	60.58	36.42	.000
Greg Smith Op Ed Attacking Goldman Sachs	163.0	Na	Na
Bradley Manning Disgruntled Chat with Hacker	84.0	Na	Na
Abu Jihaad	68.67	36.02	Na
Enron 994	.98	.15	Na

Table 7 Mean SIRDC Scores by Criterion Groups and Enron Sample

SIRDC Scores for Criterion Groups and Enron Sample	**Mean SIRDC Score**	**Standard Deviation**	**Significance Value p≤**
Depressed Chat Participants	25.70	14.14	.000
Disgruntled with Job Chat Participants	18.33	13.49	.002
Disgruntled Enron Employees Email	26.79	20.72	.001
Angry Chat Participants	13.13	8.07	.001
Substance Abuse Chat Participants	15.67	9.62	.001
Suicidal Chat Participants	12.67	4.72	.000
Ten Actual Insider's correspondence	17.93	15.65	.006
Financial Distress Chat Participants	7.40	4.69	.001
Five Emails from Bruce Ivins	17.87	6.78	.004
17 emails from Online Stalker	14.04	10.74	.000
Overall Means for Criterion Groups	16.99	7.33	.000
Greg Smith Op Ed Attacking Goldman Sachs	46	Na	
Bradley Manning Disgruntled Chat with Hacker	34.0	Na	
Abu Jihaad	34.67	30.66	
Enron 994	.07	.55	

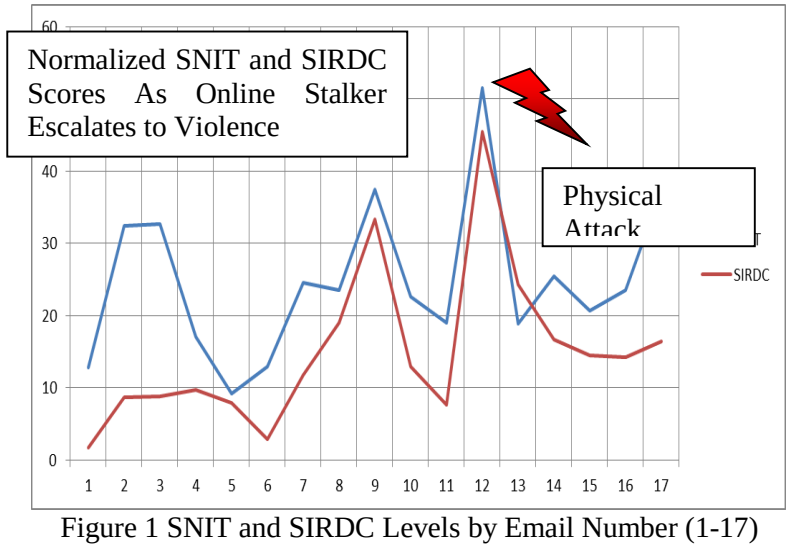

Figure 1 SNIT and SIRDC Levels by Email Number (1-17)

5. CONCLUSIONS

This preliminary research on the SNIT and SIRDC indicated that both scales displayed acceptable levels of inter-rater reliability for a randomly selected sample of 994 emails from the Enron archive and for a subset of at-risk criterion subjects presenting more robust and complex forms of negative sentiment and insider risk. The SNIT and SIRDC successfully differentiated communications from criterion groups, including subjects self-identified as angry, depressed, anxious, suicidal, addicted to substances, disgruntled with their employment, struggling with financial stress, and disgruntled Enron employees, from a randomly selected group of Enron employee communications. In addition, the SNIT and SIRDC also differentiated the 994 Enron controls from the communications of ten known insiders, as well as Private Bradley Manning, Dr. Bruce Ivins, former Goldman Sachs Director Greg Smith and convicted terrorist accomplice, former U.S. Navy sailor Paul Hall (aka Hassan Abu-Jihaad). The SNIT and SIRDC also proved sensitive to changes over time in a case of 17 communications from a jilted online stalker to his former lover, and both SNIT and SIRDC scores peaked just prior to this subject escalating from harassing communications to a physical attack on the victim's property.

This research also examined the distribution of negative sentiment and insider risk as measured by the SNIT and SIRDC in a randomly selected sample of 994 Enron communications and the heretofore unexplored relationship between negative sentiment and insider risk. Based on this sample from an organization which was suffering significant stress over the time frame examined, negative sentiment proved relatively rare, appearing in only 22% of examined emails. Moreover, all but 1.5% of these emails containing negative sentiment scored in the low range of the SNIT. Less than one percent of these emails contained negative sentiment in the high range.

Communications containing insider risk measured by the SIRDC in this sample were even more rare, with only 3.5% of emails registering any SIRDC score. Of those 36 emails discovered, only two or 0.2% of the sample, contained high levels of SIRDC indicators. The relatively rare occurrence of signs of negative sentiment and insider risk indicates the importance of using samples from known criterion groups when testing the sensitivity of human or automated systems designed to detect negative sentiment or insider risk. Samples from naturally occurring email caches are unlikely to contain realistic representations of desired target groups and therefore will not test any system's ability to detect these communications. Test samples with only low levels of SNIT or SIRDC communications also are unlikely to be representative of real insider communications.

Another important theoretical and practical question examined in this research concerned the relationship between negative sentiment and insider risk, specifically, how frequently emails with negative sentiment also contain insider

risk indicators. This question has significant implications for analysts and investigators concerned with insider risk identification and assessment. In this Enron sample, only 16.3% of emails with any level of negative sentiment also contained any level of insider risk indicator. While the odds of discovering insider risk in emails containing negative sentiment increased with the level of negative sentiment, investigators using any level of negative sentiment alone to discover communications with insider risk would appear to handicap themselves with a significant burden of false positives. However, all of the emails scoring in any range of the SIRDC for insider risk contained some level of negative sentiment. Another practical finding cited above is that all of the emails with low levels of negative sentiment that had SIRDC scores were in the low range. Pending some evidence that low insider risk scores escalate over time, investigators may want to prioritize their search resources against higher levels of negative sentiment.

This finding indicates that negative sentiment is an integral part of insider risk, as would be expected given the high rates of disgruntlement and negative psychological states associated with those responsible for insider events. However, more complex and sensitive paradigms than negative sentiment alone will be required to detect insider risk without the problem of significant numbers of false positives. This challenge will be discussed further in following articles, where the concept of perceived Victimization as an insider risk factor is examined.

5.1 Progress toward an At-Risk Insider Target Group

Further analysis of this data and future research will explore the characteristics of negative sentiment associated with insider risk compared to negative sentiment in general, as well as the other psycholinguistic markers of insider risk that can be identified in digital communications. In the meantime, results of this research indicate that communications with moderate-to- high levels of negative sentiment are more likely to contain insider risk indicators than communications with low levels of negative sentiment. It appears from a preliminary review of our data that these high SNIT and SIRDC scores are capturing many of the psychological conditions and manifestations of disgruntlement that have been found to contribute to insider risk (Shaw and Stock, 2011). These at-risk cases were easily differentiated from the more general Enron population sample and may constitute a readily identifiable at-risk group of extreme interest to investigators. These high-to-moderate negative sentiment communications are also likely to contain a lower percentage of false positive leads that will unnecessarily burden analysts and investigators.

ACKNLOWLEDGEMENTS

****This work was supported by a grant from the U.S. Department of Defense, however, the findings, conclusions, and opinions expressed are solely those of the authors. We also want to acknowledge the support of readers Drs. Marcus Rogers and Stephen Band.****

REFERENCES

Association of Threat Assessment Professionals. (2006). Risk assessment guideline elements for violence-Considerations for assessing the risk of future violent behavior. Retrieved from htttp://atapworldwide.org/associations/8976/files/documents/RAGE-V.pdf.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, *38*: 1145-1160.

Band, S., Cappelli, D., Fischer, L, Moore, A, and Trezciek, R. (2006). Comparing insider IT sabotage and espionage: A model based approach. Carnegie Mellon, NY: *Technical Report, Software Engineering Institute*.

Brief, A. P., and Weiss, H. M. (2003). Organizational behavior: Affect in the workplace. *Annual Review of Psychology*, *53*: 279-307.

Calhoun, F. and Weston, S. (2008). *Threat assessment and management strategies: Identifying the howlers and hunters*. Weston CRC Press.

Computer Security Institute 2011 15th Annual Computer Security Institute 2010/2011 Survey, CSI. Retrieved from www.GoCSI.com.

Dalal, R. S. (2005). Meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, *90*: 1241-1255.

Glomb, T. M., and Liao, H. (2003). Interpersonal aggression in work groups: Social influence, reciprocal, and individual effects. *Academy of Management Journal*, *46*: 486-496.

Hansen, E. (2011). Manning-Lamo Chat Logs Revealed. Retrieved from http://www.wired.com/threatlevel/2011/07/manning-lamo-logs on July 13, 2011.

Hershcovis, M. S., and Barling, J. (2010). Towards a multi-foci approach to workplace aggression: A meta-analytic review of outcomes from different perpetrators. *Journal of Organizational Behavior*, *31*: 24–44. doi: 10.1002/job.621.

Hershcovis, M. S., Turner, N., Barling, J., Inness, M., LeBlanc, M. M., Arnold, K. A., et al. (2007). Predicting workplace aggression: A meta-analysis. *Journal of Applied Psychology*, *92*: 228-238.

Holton, C. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decisions Support Systems*, *4*: 853-864.

Homles, T., Rahe, R., and Homes-Rahe. (1967). Social readjustment rating scale. *Journal of Psychosomatic Research*, *2*.

Koch, Gary. (1982). Intraclass correlation coefficient, in *Encyclopedia of Statistical Sciences*, Samuel Kotz and Norman L. Johnson, 4. New York, NY: John Wiley & Sons. pp. 213–217.

MacLennan, R. N. (2003). Interrater reliability of police training simulations. *Canadian Journal of Police and Security Services*, *1*: 202-209.

O'Neil, O. A., Vandenberg, R. J., DeJoy, D. M., and Wilson, M. G. (2009). Exploring relationships among Anger, perceived organizational support, and workplace outcomes. *Journal of Occupational Health Psychology*, *3*: 318-333.

Sakurai, K. and Jex, S. M. (2012). Coworker incivility and incivility targets work effort and counterproductive work behaviors: The moderating role of supervisor social support. *Journal of Occupational Health Psychology*, *17*: 150-161.

Schat, A. C., and Kelloway, E. K. (2005). Workplace aggression, in *Handbook of Work Stress*, Barling, J., Kelloway, K., and Frone, M, editors. Sage Publications, p. 189-218.

Shaw, E., and Fischer, L. F. (2005). Ten tales of betrayal: The threat to corporate infrastructures by information technology insiders analysis and observations. *Defense Personnel Security Research Center, PERSEREC. Technical Report 05-13*. September 2005.

Shaw, E., Fischer, L. F., and Rose, A. E. (2009). Insider risk evaluation and audit. *Technical report 9-02, Defense Personnel Security Research Center*.

Shaw, E., and Stock, H. (2011). Behavioral risk indicators of malicious insider theft of intellectual property: Misreading the writing on the wall. *Symantec Corporation, White Paper*, December 7, 2011. http://investor.symantec.com/phoenix.zhtml?c=89422&p=irol-newsArticle

Smith, G. (2012). Opinion/Editorial: Why I Am Leaving Goldman Sachs. *The New York Times*, March 14, 2012, page A27.

U.S. Department of Justice (2010). Amerithrax Investigative Summary. Retrieved from http://www.justice.gov/amerithrax/docs/amx-investigative-summary.pdf.

United States District Court. (2009). District of Connecticut v. Hassan Abu-Jihaad. *NO. 3:07CR57 .[MRK]: MEMORANDUM OF DECISION* Dated at New

Haven, Connecticut: March 4, 2009. Retrieved from http://jurist.law.pitt.edu/pdf/abujihaadhassan.pdf, pg. 12-20.

Weintraub, W. (1989). *Verbal Behavior in Everyday Life*. New York, NY: Springer.

Weintraub, W. (1981). *Verbal Behavior: Adaptation and Psychopathology*. New York, NY: Springer.

Wood S., and Marshall-Mies, J. C. (2003). Improving supervisor and coworker reporting of information of security concern. *Defense Personnel Security Research Center*. *PERS-TR-02*-3. Monterey, CA.

APPENDIX

Table 1

Scale of Negativity In Texts (SNIT) Item Descriptions	
1. Negative Evaluators	Judgments, beliefs, attributes with negative connotations
2. Negative Feelings	Emotions, feelings with negative connotations
3. Adverbial Intensifiers	Add emphasis or power to expressedSentiment—"very," "so," "too."
4. Non-verbal emphasis	Add power or emphasis to negative content with symbols like !, all caps, underline, numbers, etc.
5. Negation or Opposition Statements	Create negativity thru the addition of terms like no, not, never, n't, impossible, or phrases like "over my dead body."
6. Sarcasm or Negative Irony	Nice try. Thanks for sharing. Couldn't happen to a nicer guy.
7. Rhetorical Question	How's that working for you? Did you think before you spoke?
8. A direct accusation, criticism or attack toward a specific individual or group.	"You don't understand how much misery you have caused," "you have ruined our culture," "I don't know how you can sleep at night keeping your millions and leaving thousands without jobs."
9. Threats—Direct, specific and indirect general threats that may involve violence and non-violent coercive threats such as lawsuits, leaks, illicit communications blackmail or other "white collar" non-violent acts.	Direct threat— "I'll get even with you tomorrow," "next time I see you I'm going to rearrange your face." Indirect threat—lack specifics about the target, timing or means and are more general in nature. For example, "they'll get what they deserve," "someone will take care of them." Coercive threats—"if you do not comply we will go to the press or file suit in court."
10. Use of curses, foul language, racial, political, religious, sexual or other slurs	Shit, asshole, whore, raghead, kike, Etc.
11. Dangerous Religious, Political, Racial or other Beliefs	All non-believers are doomed, Mudpeople need to be extinguished
12. Sexual Material	Score 1 point for inappropriate sexual content and additional points for dehumanizing, objectifying content

13. Provocations, taunts, challenges, dares, confrontational command to an individual or group.	"bring it," "I dare you," "You and what army," "Watch me."
14. A Direct personal appeal, plea, or address with negative connotation indicating the author is uncomfortable, upset or anxious	"listen to me," "Level with me," "trust me," "believe me," "I implore you to consider."
15. Direct demand for recompense, reparation, justice, or conciliatory actions denoting the author's upset and assertive or aggressive state	"after 20 years you owe me," "who's going to pay for this," "this is your responsibility to fix," "there should be an accounting."
16. Non-negative statements that imply criticism, negativity, protest, opposition or express these indirectly	"good try," "I was going to ask her out but then she left," "another night the same." "but I've always loved you," "I've worked hard all my life," "I've done my best for the Company."
Total	

Table 2

Overt signs of **violent, angry or vengeful state of mind—**references to anger or frustration linked to violence or vengeance through some type of hostile act which may involve violence or other insider actions. "You won't get away with this," "You will pay," "You've hurt so many of us—there will be an accounting." Must have direct or strongly implied connection to an insider act ranging from violence to leaks, etc.
Signs of **escalation** of negative feeling, anger, frustration, desperation across communications. Examples might include a communication regarding anger and depression which subsequently moves to a threat of harm. There may be rare occasions when escalation is apparent within communications. However, this should be sufficiently dramatic to assure that this is not just natural venting emerging over time within the communication and should include escalation from some type of feeling or evaluation to some type of threatened action or fantasy rather than a slightly more extreme feeling or criticism.
Signs of **addictive behavior** (alcohol, illegal drugs, prescribed medications, sexual activity, gambling, media or game addiction, pornography) that may impact judgment, motivation, vulnerability to compromise or impulse control.
Signs of **fantasy** about negative insider-related acts—"I wish I could put you in my place," "I wonder what it would be like to shut you down," "I wonder how long you would last if the public knew how you really operate."
Signs of **suicidal or self-destructive** thoughts or feelings—references to suicidal mood, plans (resignation regarding doom or inevitability of suicide). Pay attention to references to aggressive actions that may result in suicide through others such as suicide by cop as in the cases of Major Hassan or Sodini (the Pa Gym Killer). Do not code depressive feelings here if they do not include direct references to suicidal behavior, plans or fantasies. Code depression in Mental Health issues.
Signs of **Planning** of negative insider or related acts—discussion of materials, equipment, steps needed, results, etc.
Rehearsal of negative insider or related acts or negative act practiced, approached, or attempted without execution. For example, a shooter who brings his guns to the organization planning to attack or a leaker who copies material but then does not attack or send the information. An insider who rehearsed removing materials from work by carrying out similar data or equipment.
Signs of **deterioration** in cognitive state, concentration, attention, self-control or other mental functions.
Depersonalization of potential victims or targets—language suggesting objectification, dehumanization of persons or groups making it easier to attack or betray them.
Signs of **diminishing inhibitions—**references to negative behavior indicating an increase in lack of judgment, control, vulnerability to impulsive actions versus a decline in cognitive functions such as concentration and attention.

Signs of perceived **inability to pursue other options**—references to path being blocked, feeling no choice but insider acts.
Suspicion—the author expresses suspicion regarding others' negative attitudes or behavior toward him or those with whom he identifies without any specific accusations, statements of victimization or blame. "I don't trust Mike," "Watch your back around that guy," "If it walks like a duck, talks like a duck…"
Accusations—the author identifies a specific individual or group as responsible for his own or other identified persons' mistreatment. "Jay made very sexist statements (about me or our female employees)." "Mr. Smith you have no idea how much misery you have caused employees." Differs from Injustice Attribution which negatively characterizes the organization or leadership for its behavior, policies or practices with impact across the organization and its clients/customers, etc.
The author states that they feel specifically, directly and personally **victimized**, taken advantage of, by persons or groups independent of whether these individuals or groups are named. The passage may be scored for both accusations and victimization if the author is specific in describing a direct action against him and an individual or group to blame. "Bill's unfair review has destroyed my chances of promotion this year." Differs from Injustice Attribution which negatively characterizes the organization or leadership for its behavior, policies or practices with impact across the organization and its clients/customers, etc. rather than a specific accusation of victimization.
Rationalizing or Projective Blame—blame exaggerated or rationalized without probable foundation to make the author feel better. Beyond a specific complaint the author describes an individual or group as responsible for a global set of problems he has encountered over time, reflecting an effort to externalize responsibility. "My supervisor ruined my life and destroyed my marriage." "You laying me off lead to the death of my son." "Your negative review two years ago ruined my entire career."
General Injustice Attributions—Rather than a specific complaint about an individual or group's actions, injustice attributions impacting specific individuals or groups reflect systemic problems with organizational procedures, values, enforcement or leadership through specific actions or inactions. They involve specific reports of events, procedures, actions or inactions that are reflective or organizational or leadership problems rather than individual or small group behaviors. They differ from Professional stressors or general unfairness at work. Includes specific complaints about unjust decisions resulting from organizational leadership, policies or practices impacting employees beyond the author (but may include him). Examples can include a failure to respond to complaints or take action regarding a complaint. In other examples, the Subject may believe: • wrongful behavior and unfair advantages or connections are systematically rewarded (managers selling stock ahead of bad news); • lack of work is rewarded while hard work is not. • There is unequal versus equal treatment of employees. • There is equal or nondiscriminatory treatment when it should be individualized and different.

• Good behavior or innocence is punished, • Punishment is displaced onto persons who either do not deserve it or do not deserve the severity of the punishment. • The punishment is disproportionate to the act or intent. • Wrongful behavior goes unpunished and management makes arbitrary rules.
Signs subject is **obsessed with situation** as manifested in repeated references or complaints about the issue within or across communications, statements that the subjects is constantly monitoring the situation, or primed to react to related events.
Signs of rationale for insider action or activity—"Everyone else is doing it," "I was instrumental in creating this information," "It can't be traced back to me," "It makes my life easier not to have to recreate this at my next job," "They have it coming." "The organization is powerless to protect its interests and assets and therefore deserves to be taken advantage of."
Signs of **mental health problems**—include overt references to serious levels of depression, anxiety or other mental health disorders or references to treatment, referral for evaluation or other indicators of the existence of a mental health problem. Do not infer the existence of a diagnosable mental health problem from implied or direct references to less serious feelings or states. References to suicidal ideation, plans, or attempts may also be coded.
Signs of **personality issues** associated with negative Actions (lack of conscience, narcissism, psychopathy, social isolation or avoidance, entitlement, impulsiveness, difficulties getting along with others, etc.)
References to **previous violations** of policies, practices, laws, accepted procedures.
References to **social network risks** in the form of contact, communication with or relationship with persons and/or groups associated with adverse or competitive intentions or actions against organization or personnel. References to family members, social contacts or others involved in adversarial, illegal or anti-social acts.
Section 4. References to personal stressors directly impacting the author—do not infer that something was stressful. Code specific references to events, circumstances or perceived situations generally identified as stressful (death of spouse, new job, move, divorce, break-up, etc.) or identified by the author as perceived so.
Section 5. References to professional stressors directly impacting the author— do not infer that something was stressful. Code specific references to events, circumstances or perceived situations identified as professionally stressful by the author or generally accepted as stressful –failure to get a raise, promotion, a transfer, demotion, layoffs at work, cuts in hours, benefits, etc.

Section 6. Concerning Behaviors
References to **Concerning Behaviors—**recent violations of policy, practice, law, Ethics, standards of interpersonal behavior, information security, finances, personnel security, etc. while in current or past position. Different from previous violations which occurred before recent employment. DWI as young adult might be Previous Violation under personal predispositions while a recent DWI would be scored as a Concerning Behavior.
References to unusual **travel that could involve contacts with adversaries**
Section 7. Inhibiting or Mitigating Factors (score negative points for each item, subtracted from Insider risk score)
References to **inhibiting religious or ethical beliefs** or optimistic attitudes that could inhibit insider actions. "I am young enough to start over." "I'd love to get even but that would mean being as nasty as he was."
References to **social support**, dependents who could be impacted or act as inhibitors of negative actions
References to **successful treatment**, counseling or other Inhibiting services or assets (legal, financial, social)
References to concerns about possible insider action on **career, reputation, effects on others**
References to **use of sanctioned channels** for complaint, protest such as a letter to the CEO, writing a complaint within channels, filing a lawsuit, complaining to HR, etc.
Qualitative Adjustment for Insider Risk in Author Not Captured Above—if you feel that some aspect of the ratings do not capture the true level of insider risk expressed in this author, add 1-10 points, with an explanation. For example, is the author's language intensely threatening within a very short passage. Or, does the author go on producing a lengthy list of complaints that are not fully captured. Alternatively, does the author's score on one versus other categories raise significant concerns, such as the presence of a significant mental health disorder (Paranoid Schizophrenia with command hallucinations, Anti-social personality but without known concerning behaviors or past offenses mentioned).

For complete copies of the SNIT and SIRDC as well as instructions, contact the corresponding author.

Table 3 Examples of High, Medium and Low Enron SNIT email excerpts by score

Goup	Email SNIT Score	Email Excerpt
High (>30)	42.3	*"I'm so sad!!! I'm so depressed about the whole thing…He feels like the bad guy. And I feel like a bad father!!!"*
	36.0	*"Girl, remind me, just in case I have a memory lapse that I will never, ever go back over to that church again…they're fools… I was so embarrassed that my chuckle came out!"*
	32.3	*"Damnit Jeff. I don't have time!!!!!!!!!!!!!!!! The Panic is my waste of time trying to get things organized. That's the panic…"*
Medium (16-30)	30.0	*"Diomedes is going to jump all over me for even sending the ESA memo to Metts. I told him that this message affects more than us and he thinks he has free reign within this region… We don't want to come off looking dishonest to our own people. And believe me, the rumors are out there… We would be talking about half the international people that there is no future for them. Whew!"*
	23.0	*"No one seems sure if Bill is the acting originator for this contract or not… Also I would appreciate receiving clarification of why we don't define the purchase amount as the total and not just the balance remaining. It could create discrepancies in the document… this seems inconsistent to me… How is this possible?"*
	16.0	*"Brad, I really don't know where to begin other than just to say it. Enron is having some extremely difficult times now… this is not a good time for you or anyone else to try and seek employment here. I am sorry if I'm letting you down. I was only wanting to help. I will always keep you in mind, and hopefully when things turn around here, I will be able to address you coming here."*
Low (1-15)	11.7	*"I wanted to call you at home, but I am never really sure when you are sleeping and I don't want to call and wake you up.*
	1.7	*"As I see it, we really have not choice but to join this system if we are going to participate in the JDG. Shortly they will stop circulating work product, etc. via email and will rely upon this website instead. The costs are somewhat unclear but it is intended to be an economy of scale cost sharing concept."*

Table 4 Examples of High and Low SIRDC emails

Group	**SIRDC Score**	**Email Excerpt**
High (≥9)	12.3	*"Enron and you made millions out of the pocketbooks of California's consumers and from the efforts of your employees… while you netted well over a $100 million, many of Enron's employees were financially devastated when the company declared bankruptcy and then retirement plans were wiped out… As a result, there are thousands of consumers who are unable to pay their basic energy bills and the largest utility in the state is bankrupt. The NY Times reported that you sold $100 million worth of Enron stock while aggressively urging the company's employees to keep buying it. Please donate this money to the funds set up to help repair the lives of those Americans hurt by Enron's under handed dealings."*
Low (<9)	6.7	*"Diomedes is going to jump all over me for even sending the ESA memo to Metts. I told him that this message affects more than us and he thinks he has free reign within this region… We don't want to come off looking dishonest to our own people. And believe me, the rumors are out there… We would be talking about half the international people that there is no future for them. Whew!"*
	4.7	"*Yes, unfortunately, we aren't going to get ours until the 5th. No explanation was given. We may get the numbers a day or so ahead of time but I haven't gotten them yet."*
	2.7	*"I question whether we should ask Capt. Sawant to put anything in writing concerning how to beef up his second report until we talk further… I am concerned that anything he puts in writing may be discoverable in a U.K. arbitration proceeding...."*

Table 5 Stalker email excerpts by SNIT and SIRDC scores

Email #	**Email Excerpt**	**SNIT Scores**	**SIRDC Score**
Email 3	Who would ever be attracted to you. And you think you are going to get a banker. All you are going to get is an asshole who will treat you like the ugly slut that you are. And the funny thing is that you probably think you're cute. Well honey, you're far far from it.	32.4	8.7
Email 12 (Just prior to attack)	No, we did not forget about you. You are next bitch!	51.5	44.5
Email 17	From one bitch to another, you can wear all the black pants and black outfits that you want, it still doesn't hide your fat ass… nothing can hide the fact that you have absolutely NO tits… And please, don't ever decide to have children.	37.3	16.5