PNNL-20062

Content Analysis for Proactive Protective Intelligence

Antonio Sanfilippo

December 2010

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

This document was printed on recycled paper.
(9/2003)

PNNL-20062

Content Analysis for Proactive Protective Intelligence

A Sanfilippo

December 2010

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

1.0 Executive Summary

The ability to assess the severity of abnormal communications to protectees has never been more topical. Since Mr. Obama took office, the rate of threats against the president has risen to an average of over 30 a day, a 400 per cent increase as compared to the previous administration [1]. Allocation of significant law enforcement resources to the pursuit of every threat is unrealistic and would not result in an efficient use of resources. This paper outlines a plan for developing and validating a PPI approach that helps prioritize threats and prevent targeted violence to USSS protectees through the analysis of abnormal communications to USSS protectees.

Only a handful of studies on abnormal communications and approaches to prominent people have been carried out to-date and no compatible terminologies have emerged [2]. Yet, these studies have demonstrated that the occurrence of problematic approaches can be meaningfully related to quantifiable behavioral factors observed in abnormal communications and therefore offer a serviceable basis on which to develop a PPI approach.

Our prospected PPI approach leverages existing work on abnormal communications and approaches to prominent people [2] and is based on the *Frames in Action* platform [3,4,5].[1] The *Frames in Action* platform integrates information extraction, content analysis and computer modeling methods with theories of radical rhetoric to make qualitative and quantitative assessments of violent intent from the study of verbal communication behavior. It can be effectively used to model a subject's intent to harm a protectee through the analysis of the subject's abnormal communications to the protectee, as outlined below.

A. Create a dataset of abnormal communications to protectees, with an indication of which communications escalated in a problematic approach.

B. Use insights from existing studies on abnormal communications and approaches [2] to tune the information extraction and content analysis capabilities available in the *Frames in Action* platform to the protective intelligent domain.

C. Process the dataset developed in step A with the calibrated information extraction and content analysis capabilities from step B to correlate categories of verbal behavior with the occurrence, or lack thereof, of a problematic approach.

D. Use the categorical correlations obtained in step C to build and test computer models capable of assessing the likelihood of abnormal communications to a protectee to escalate into problematic approaches.

Further details on the *Frames in Action* platform and how it can be applied to the protective intelligence domain using insights from existing work on abnormal communications and approaches to prominent people are provided below.

[1] The *Frames in Action* platform has been developed by PNNL scientists within the *Motivation and Intent* and the *Comparative Case Studies of Radical Rhetoric* projects sponsored by the Human Factors/Behavioral Sciences Division, at the Science & Technology Directorate within the Department of Homeland Security.

2.0 Frames in Action: Modeling Radical Rhetoric to Identify Violent Intent

The *Frames in Action* approach relies on the co-expression of rhetoric and action features in discourse to infer computational models that identify messages from communication sources that have violent intent and estimate their proximity to an attack. It combines insights from Frame Analysis [6,7,8,9] and theories that explain the emergence of violence with reference to factors such as moral disengagement [10], the violation of sacred values [11] and social isolation [12] in order to developed a predictive computational approach to radical rhetoric.

The objective of Frame Analysis is to understand the communicative and mental processes that explain how groups, individuals and the media try to influence their target audiences, and how target audiences respond. The *Frames in Action* approach provides a computational implementation of Frame Analysis [3, 4] that is based on the integration of text mining techniques from computational linguistics [13] and content analysis methods from sociology, psychology and political science [14]. The same approach is used to integrate the import of factors contributing to the detection of violent intent such as moral disengagement, the perceived violation of sacred values and social isolation using additional theoretical insights from anthropology and psychology. The approach comprises the following steps.

1. Define a *Frames in Action* annotation scheme that enables the identification of communication events according to insights from Frame Analysis and other theories relevant to the detection of violent intent. Each code in the scheme links up either to another code or a list of words, e.g. FRAME→CRITICIZE→{*accuse, blame, …*}
2. Validate the *Frames in Action* annotation scheme using inter-annotator agreement.
3. Develop a text-mining application that applies *Frames in Action* annotations automatically to naturally occurring text.
4. Evaluate the result of automatic *Frames in Action* annotation using quantitative measures.
5. Use statistical and machine learning techniques to infer models of violent intent from *Frames in Action* annotations.
6. Use the inferred models to identify messages from terrorist communication sources and estimate proximity of the date of issue of the messages to an attack.

Steps 1-4 are described in [8, 9]. Steps 5 and 6 are described in [10], where we discuss a specific application of *Frames in Action* to a dataset containing texts issued by four groups: Central al Qa'ida (AQC), al Qa'ida in the Arabian Peninsula (AQAP), Hizb ut-Tahrir (HuT), and the Movement for Islamic Reform in Arabia (MIRA) [15]. AQC and AQAP are known for their extremist violence. Hizb ut-Tahrir and MIRA share the same ideology and goals with AQC and AQAP but have not engaged in terrorism. Each document in the collection contains metadata that include annotations about

- the communication source of the document: AQC, AQAP, HuT, or MIRA
- the proximity of the date of issue of the document to a terrorist attack: 30/60/90/120 days
- the identity of the attacker: AQC, AQAP, AQC or AQC-affiliated group (AFF), AQC or AQC-affiliated group or AQC-inspired group (INS).

We first divide the dataset in two subsets: document from terrorist sources (AQC, AQAP) and documents from non-terrorist sources (MIRA, HuT). Using the automated *Frames in Action* text-mining application mentioned in step 3 above, we map each data subset into *Frames in Action* annotations and calculate the statistical relevance of each *Frames in Action* annotation to each of the two data subsets, as shown in Figure 1 for the subset of terrorist texts (only annotations with p-values ≤ 0.05 are shown). *Frames in*

Action annotations that are instrumental in identifying terrorist texts fall into four broad semantic categories:

- *moral disengagement*
- *violation of sacred values*
- *social isolation*
- *violence and contention*.

Moral disengagement [10] occurs when people choose and urge others to engage in inhumane conduct (e.g. genocide) to achieve a goal believed to be morally right (e.g. the Nazi idea of racial-biological purity). One embodiment of moral disengagement is the removal of ethical restrictions against violence through acts of *dehumanization*. By negating the human identity of the victim, the assailant asserts his moral superiority and reduces identification with the targets of harmful acts. Dehumanization of the victim thus frees the assailant of moral sanctions against harming the victim. In our reference dataset, moral disengagement is expressed by communicative acts such as attributing blame to the victim, and conveying hatred, disgust and fear of the victim.

The violation of sacred values [11] occurs when ideals of love, honor, justice and religion come under secular assault and people struggle to protect their private selves and public identities from moral contamination. In our reference dataset, the violation of sacred values is linguistically marked by the co-occurrence of military and religious terms, which marks the propensity for armed struggle under the apprehension that religious freedom is under secular attack.

Within a terrorist context, social isolation results from the requirement that a recruit cut off ties to family, friends, and anyone else outside the organization. It is one of the main strategies in ensuring the recruit's commitment to the organization and his/her readiness to action. It is widely regarded as a key factor promoting terrorism [12]. In our reference dataset, social isolation is marked by the occurrence of events conveying abandonment, confinement, withdrawal and isolation.

Reference to violence and contention is another important marker of radicalization and violent intent and it is generally linguistically expressed by acts such a fighting, attacking and killing.

Once established that *Frames in Action* annotations are useful in identifying texts originating from terrorist sources, we use machine learning techniques to infer a decision-tree model from the data [15] that identifies terrorist texts in terms of *Frames in Action* annotation features. Training on 90% of the data to learn the decision-tree model and testing the decision-tree model on remaining 10% using 10-fold cross-validation, we obtained the results shown in Figure 2. These results indicate that a reliable decision-tree model can be inferred from data using the *Frames in Action* approach to identify texts issued by terrorist communication sources.

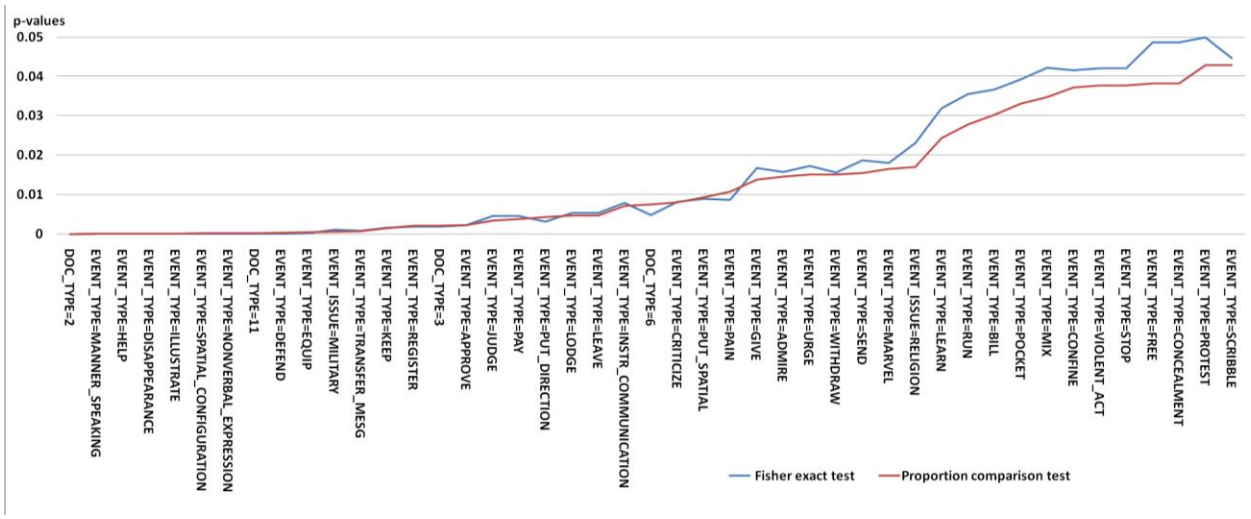

Figure 1: *Frames in Action* annotations highly correlated with terrorist texts.

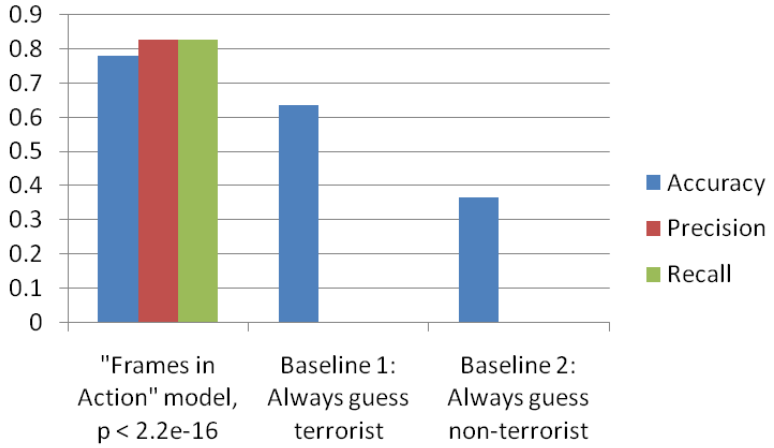

Figure 2: Evaluation results of *Frames in Action* decision-tree model of terrorist discourse. Baseline 1 assumes that documents always originate from a terrorist source; baseline 2 assumes the contrary. Accuracy, precision and recall are defined in terms of true positives (TPs), true negatives (TNs), false positives or "false alarms" (FPs), and false negatives or "misses" (FNs):

$$accuracy = \frac{TPs+TNs}{TPs+Tns+FPs+FNs}, precision = \frac{TPs}{TPs+FPs},\ recall = \frac{TPs}{TPs+FNs}.$$

We use the same approach to develop models which estimates the proximity of an AQC, AQAP, INS or AFF attack from the date of issue of a terrorist document. We divide documents from terrorist sources into 16 subsets according whether they are issued within 30/60/90/120 days from a terrorist attack and whether the attack is perpetrated by an AQC/AQAP/INS/AFF source. We use the 16 document subsets to develop and evaluate 16 models of proximity of an attack to the data of issue of a document. Training on 90% of the data and testing the resulting decision-tree models on the remaining 10% using 10-fold cross-validation, we obtained the results shown in Figure 2. The results of this evaluation, as shown in Figure 3, are more speculative in nature as compared to the results shown in Figure 2. This is primarily due to data sparseness in the training materials, as highlighted by the fact that only four of the sixteen models can be regarded as having statistical significance ($p < 0.05$). Still, it is worth noticing that our models outperform both baselines in three of the four statistically relevant models. This shows that the approach provides useful insights even in the much harder task of anticipating the timeframe of an attack, and it suggests that significant gains can be obtained with a larger training data set.

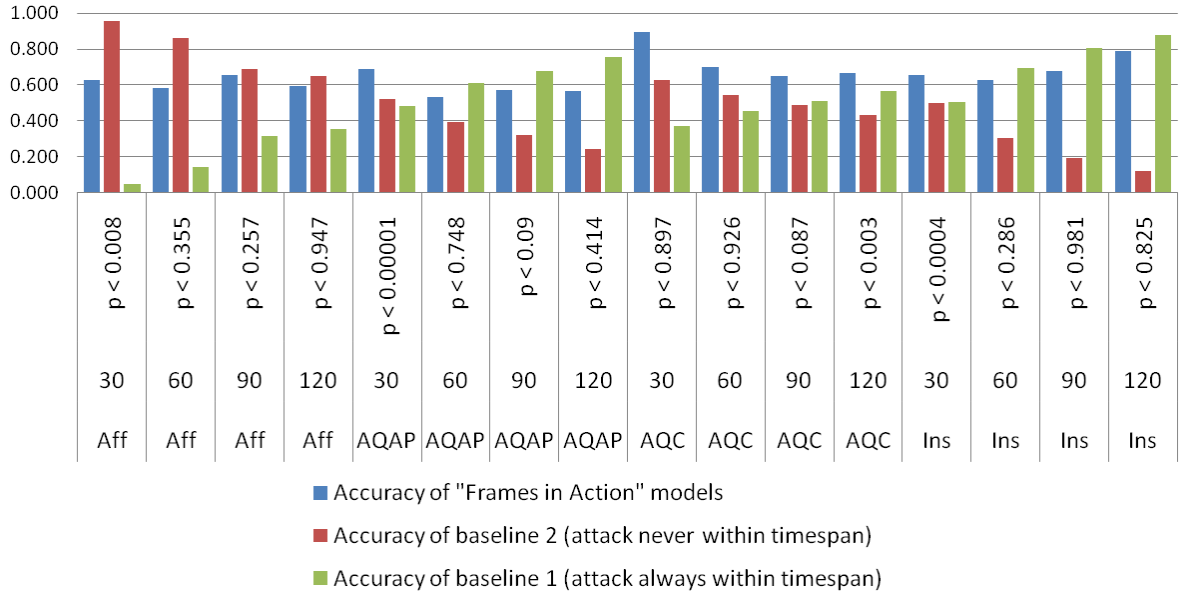

Figure 3: Evaluation results of decision-tree models of proximity of an attack to the date of issue of a terrorist document**.** Baseline 1 assumes that documents are always issued within a given timespan from a terrorist attack; baseline 2 assumes the contrary. Accuracy is defined in terms of true positives, true negatives, false positives, and false negatives (see Figure 2).

We are currently investigating ways to derive equations that express the contribution of *Frames in Action* features to the identification of a terrorist communication source and the proximity of an attack to the date of issue of a terrorist communication from the decision-tree models discussed above. These equations will allow us to develop dynamic models that generate time series simulations of violent intent and proximity of an attack, using modeling techniques such as system dynamics [17].

3.0 Existing Work on Abnormal Communications and Approaches

As reported in [2], six studies have been carried out to date on abnormal communication and approach to prominent people: one on the British royal family [18], two on Hollywood celebrities [19, 20], and three on the U.S. Congress [21,22,23]. The studies relate the occurrence of problematic approaches to factors such as mental illness, threatening or antagonistic communication, requests for help, multiple means of communication, and multiple contacts and targets. Details of these studies and their results are given in Table 1 and Table 2. The six studies do not use compatible terminologies and rely on manual annotation. Consequently, the approaches developed in these studies do not provide a unified framework and are resource-intensive. Yet, despite these limitations, they demonstrate that the occurrence of problematic approaches can be meaningfully related to quantifiable behavioral factors observed in abnormal communications and therefore offer a useful point of departure for the development of a PPI approach.

Table 1: Data on six comparative studies of abnormal communicator subjects who approached and did not approach public figures (adapted from [2], p. 2).

Study	Targets	Sampled Universe	Sampling Method	Number of Approachers	Number of Nonapproachers
James et al. (8)	British Royal Family	8001 police files	Random stratified	53	53
Dietz et al. (1)	Hollywood celebrities	1559 approach 1272 nonapproach	Random stratified	107	107
Dietz et al. (3)	U.S. Congress	U.S. Capitol police files Number not given	Random stratified	43	43
Scalora et al. (4)	U.S. Congress	4387 Capitol police files	Chronological 1993–1999	986	3401
Scalora et al. (5)	U.S. Congress	1500 U.S. Capitol police files 1998–1999	Random	104	212
Meloy et al. (2)	Hollywood celebrities	271 files from Enter. Corp.	Nonrandom and random	61 (nonrandom)	61 (random)

Table 2: Comparison of headings associated with escalation (approach vs. non-approach) across six public figure studies (adapted from [2], p. 3).

Study	James et al. (7)			Dietz et al. (1)			Dietz et al. (3)			Scalora et al. (4)			Scalora et al. (5)			Meloy et al. (2)		
Target	British Royal Family			Hollywood Celebrities			U.S. Congress			U.S. Congress			U.S. Congress			Hollywood Celebrities		
	Nonapproach (%)	Approach (%)	χ^2, p	Nonapproach (%)	Approach (%)	χ^2, p	Nonapproach (%)	Approach (%)	χ^2, p	Nonapproach (%)	Approach (%)	χ^2, p	Nonapproach (%)	Approach (%)	χ^2, p	Nonapproach (%)	Approach (%)	χ^2, p
Serious mental illness	69.8	92.5	11.135, 0.004	72	72	NS	81	98	4.47, <0.04	27.7	35.1	20.24, <0.0001	39.6	59.6	11.22, <0.001	21	36	3.366, <0.067
Any threatening communication	15.1	1.9	9.00, 0.009	22	23	NS	84	33	21.07, 0.0001	35.1	21.4	69.50, <0.0001	76.4	41.3	70.941, <0.0001	13	35	$F = 4.616$, <0.05
Requests for help	24.5	37.7	NS	10	10	NS	30	56	4.75, 0.03	16.5	27.4	59.57, <0.0001	17	35.6	7.846, 0.006	13	28	4.075, <0.05
Multiple means of communication	3.8	25	10.405, 0.001	5	14	4.47, <0.04	2	23	6.67, <0.01	*	*	*	6.6	39.4	51.966, <0.0001	21.3	95	68.227, <0.001
Antagonistic communication	17–32	3.8–9.4	<0.026	15	6	4.10, 0.04	35–65	12–35	<0.01	12.1	6.7	22.92, <0.001	70.8	43.3	22.3, <0.001	13	8.2	NS
Multiple contacts and targets	49.1	77.4	14.525, 0.002	18	37	8.20, 0.004	2.2	7.1	$T = -2.76$ <0.008	17.7–33.2	22.5–39.7	<0.001	6–43	22–58	<0.015	31.1	36	NS

*No data available.

4.0 Using the Frames in Action Platform for Protective Intelligence

Given the focus on violent intent, the *Frames in Action* platform can be used effectively to model a subject's intent to approach and harm a protectee from the subject's abnormal communications to the protectee. The list below outlines the sequence of steps and resources needed to apply the *Frames in Action* platform to the protective intelligence domain.

A. Creation of a dataset of abnormal communications to protectees, in which each communication is marked as to whether it escalated in a problematic approach. Such a dataset can be seeded with the document collections used in previous studies on problematic approaches to U.S. Congress members [21,22,23]. Additional data could be collected with the help of the USSS National Threat Assessment Center and other law enforcement organizations, and through resources such as the National Archive of Criminal Justice Data (http://www.icpsr.umich.edu/NACJD) and the National Criminal Justice Reference Center (http://www.ncjrs.gov/index.html).

B. Use of insights of existing studies on abnormal communications and approaches [18-23] to tune the information extraction and content analysis capabilities available in the *Frames in Action* platform to the protective intelligent domain. The code books (list of factors and their explanation) developed in these studies (e.g. see [24]) may prove to be particularly valuable in this pursuit.

C. Process the data set developed in step A with the calibrated information extraction and content analysis capabilities from step B to correlate categories of verbal behavior with the occurrence, or lack thereof, of a problematic approach.

D. Use the categorical correlations obtained in step C to train computer models for Proactive Protective Intelligence.

1. Categorical correlations will first be used as evidence from which to infer and validate a decision-three model capable of assessing the likelihood of abnormal communications to a protectee to escalate into problematic approaches. This model will serve as *watch-and-warn* capability, e.g. monitoring likelihood changes that incoming abnormal communications may have on the occurrence problematic approaches.

2. The decision-tree model from step D.1 will be used as ground data to develop and test a system dynamics model providing time series simulations of how abnormal communications to a protectee may develop into problematic approaches. This model will serve as a decision-making support capability by enabling *what-if* reasoning through the artificial manipulation of model parameter to explore plausible developments and test intervention strategies.

5.0 References

[1] Kessler, R. (2009) *In the President's Secret Service*. Crown Publishers, New York.

[2] Meloy JR, James DV, Mullen PE, Pathé MT, Farnham FR, Preston LF, Darnley BJ (2010) Factors Associated with Escalation and Problematic Approaches Toward Public Figures. J Forensic Sci. 2010 Oct 6.

[3] Sanfilippo, A., A.J. Cowell, S. Tratz, A. Boek, A.K. Cowell, C Posse, and L. Pouchard (2007) Content Analysis for Proactive Intelligence: Marshaling Frame Evidence. *Proceedings of the 22nd AAAI Conference*. AAAI Press, Menlo Park, CA.

[4] Sanfilippo, A., L. Franklin, S. Tratz, G. Danielson, N. Mileson, R. Riensche, and L. McGrath (2008) Automating Frame Analysis. In H. Liu, J. Salerno, and M. Young (eds.) *Social Computing, Behavioral Modeling, and Prediction*, pp. 239-248. Springer, NY.

[5] Sanfilippo, A., L. McGrath, and P. Whitney (Forthcoming) Violent Frames in Action. *Dynamics of Asymmetric Conflict. Special issue on Using Linguistic Content Analysis to Understand Islamist Terrorists.* L.Conway, K. Conway and A. Smith (eds).

[6] Goffman, E. (1974) *Frame Analysis: An Essay on the Organization of Experience*. Harper and Row, London, UK.

[7] Benford, D. and R. Snow (2000) Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology*, Vol. 26, pp. 611-639.

[8] Entman, R. (2004) *Projections of power: Framing news, public opinion, and U.S. foreign policy*. University of Chicago Press, Chicago, IL.

[9] Gamson, W. (1992) *Talking Politics*. Boston: Cambridge University Press, New York, NY.

[10] Bandura, A. (1999). Moral disengagement in the perpetration of in humanities. *Personality and Social Psychology Review, Special Issue on Evil and Violence*, 3, 193-209.

[11] Tetlock, Ph., O. Kristel, B. Elson, M. Green, and J. Lerner (2000) The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals. *Journal of Personality and Social Psychology*, 78 (2000): 853-870.
[12] Navarro, J. (2009) Unmasking Terrorists - Two Critical Characteristics! Key Signs Which Point to Potential Terrorist Activity. *Psychology Today*, December 31, 2009.
[13] Feldman, R, J. Sanger (2007) *The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data*. Cambridge, England: Cambridge University Press.
[14] Krippendorff, K. (2004) *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage.
[15] Smith, A. G., Suedfeld, P., Conway, L. G. III, & Winter, D. G. (2008). The Language of violence: Distinguishing terrorist from non-terrorist groups using thematic content analysis. *Dynamics of Asymmetric Conflict, 1*, 142-163.
[16] Murthy S. (1998). Automatic construction of decision trees from data: A multidisciplinary survey. *Data Mining and Knowledge Discovery*, 2(4):345-389, 1998.
[17] Forrester, Jay W. (1968) *Principles of Systems*, (2nd ed.). Waltham, MA: Pegasus Communications.
[18] James DV, Meloy JR, Mullen PE, Path_ MT, Preston LF, Darnley B, et al. (In Press) Abnormal attentions towards the British Royal Family: factors associated with attention and escalation. J Am Acad Psychiatry Law.
[19] Dietz P, Matthews D, Van Duyne C, Martell D, Parry C, Stewart T, et al. (1991) Threatening and otherwise inappropriate letters to Hollywood celebrities. J Forensic Sci 1991;36:185–209.
[20] Meloy JR, Mohandie K, Green M (2008) A forensic investigation of those who stalk celebrities. In: Meloy JR, Sheridan L, Hoffmann J, editors. Stalking, threatening, and attacking public figures: a psychological and behavioral analysis. New York, NY: Oxford University Press, 2008; 37–54.
[21] Dietz P, Matthews D, Martell D, Stewart T, Hrouda D, Warren J. (1991) Threatening and otherwise inappropriate letters to members of the United States Congress. J Forensic Sci 1991; 36:1445–68.
[22] Scalora MJ, Baumgartner JV, Callaway D, Zimmerman W, Hatch-Maillette MA, Covell CN, et al. (2002) An epidemiological assessment of problematic contacts to members of Congress. J Forensic Sci 2002; 47:1360–4.
[23] Scalora MJ, Baumgartner JV, Zimmerman W, Callaway D, Hatch-Maillette MA, Covell CN, et al. (2002) Risk factors for approach behavior toward the U.S. Congress. J Threat Assess 2002; 2:35–55.
[24] Dietz, Park Elliot, and Daniel A. Martell (2004) Mentally Disordered Offenders in Pursuit of Celebrities and Politicians. *Inter-university Consortium for Political and Social Research 6007*, available through http://www.icpsr.umich.edu.

ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Human Factors/Behavioral Sciences Division under Inter-Agency Agreement HSHQDC-07-X-00793, made to the Oak Ridge Institute for Science and Education (ORISE), and by the Technosocial Predictive Analytics Initiative, an ongoing research program at the Pacific Northwest National Laboratory (http://predictiveanalytics.pnl.gov). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or ORISE.