Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

MODELING THE EMERGENCE OF INSIDER THREAT VULNERABILITIES

Ignacio J. Martinez-Moyano

Decision and Risk Analysis Group
Argonne National Laboratory
Argonne, IL 60439-4832, U.S.A.

Eliot H. Rich

School of Business
University at Albany
Albany, NY 12222, U.S.A.

Stephen H. Conrad

Technical Staff
Sandia National Laboratories
Albuquerque, NM, U.S.A.

David F. Andersen

Rockefeller College
University at Albany
Albany, NY 12222, U.S.A.

ABSTRACT

In this paper, we present insights generated by modeling the emergence of insider threat vulnerabilities in organizations. In our model, we integrate concepts from social judgment theory, signal detection theory, and the cognitive psychology of memory and belief formation. With this model, we investigate the emergence of vulnerabilities (especially that are insider-driven) in complex systems characterized by high levels of feedback complexity, multiple actors, and the presence of uncertainty in the judgment and decision processes. We use the system dynamics method of computer simulation to investigate the consequences caused by changes to the model's assumptions. We find that the emergence of vulnerability can be an endogenous process and that leverage points to reduce this vulnerability involve improvement in information acquisition, information management, and the training of personnel in judgment and decision-making techniques.

1 INTRODUCTION

The existence of insider threats in organizations has only recently been documented (Keeney and Kowalski 2005, for excellent examples see Randazzo et al. 2004). The emergence of insider threats is difficult to identify and difficult to document because insiders have intimate knowledge of internal control and security systems, allowing them to cover their tracks and disguise their attacks as innocent mistakes. Additionally, some organizations, especially those in the telecommunications, banking, and finance sectors, do not document or disclose information about past instances of insider attacks so as not to reveal that their systems are vulnerable. Under such circumstances, learning about the emergence of insider threats becomes even more difficult. For example, in the information technology sector, reported incidents (attacks and threats) and identified vulnerabilities have experienced exponential growth in the last decade (see Figure 1, with data from <www.cert.org/stats/cert_stats.html>). However, not much is known about the specifics of the problems reported or of the circumstances that allowed these companies to become vulnerable to attack.

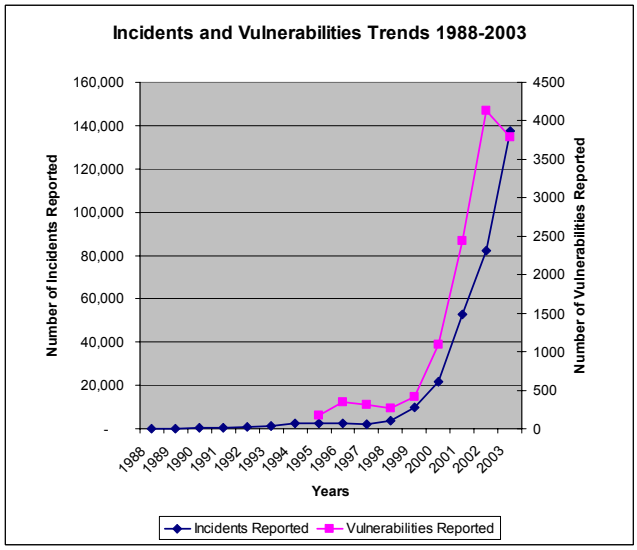

Figure 1: Incidents and Vulnerabilities Trends

CERT, a center of Internet security expertise, located at the Software Engineering Institute at CMU, reports

about 22,716 vulnerabilities (from 1995-2005) and 319,992 incidents (1988-2003). Many of these reported incidents and vulnerabilities are caused by insiders. In this paper, insiders are people who had, at some point in time, legitimate access to the system that was compromised (Randazzo et al. 2004).

In general, insider attacks/threats are those executed by a current or former employee or contractor that intentionally exceeds or misuses an authorized level of access to networks, systems, data, or resources to harm individuals and/or an organization (Keeney and Kowalski 2005).

Andersen et al. (2004), using a group model building approach with security experts and security modelers, generated the *dynamic trigger hypothesis* about insider threat emergence based on the notion that interacting feedback mechanisms in organizations have the potential to create the conditions for insiders to become malignant and cause harm. The hypothesis was further developed by Rich et al. (2005), and a preliminary formal model was developed by Martinez-Moyano et al. (2005). This paper reports further analysis and exploration using the theory and its associated simulation model.

The *dynamic trigger hypothesis* (see Figure 2) proposes prototypical feedback mechanisms that interact to create the necessary structural conditions for insider vulnerabilities to emerge. The main feedback processes identified are: the detection trap *R1*, the trust trap *R2*, and the unobserved emboldening trap *R3*.

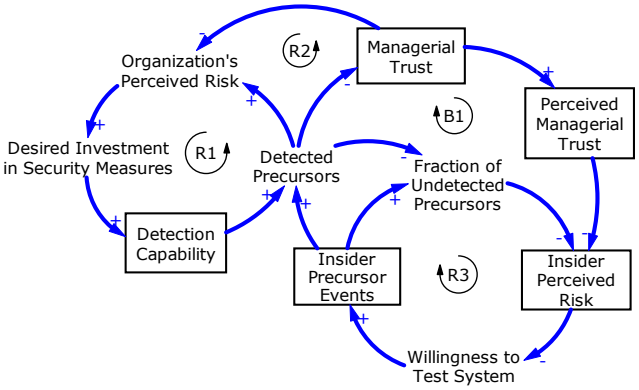

Figure 2: Dynamic Trigger Hypothesis

In organizations, detection capability determines the number of detected precursors over time. As the organization detects precursors, the organization's perceived risk increases, incrementing the desired investment in security measures and leading to a very secure environment (see cycle R1 in Figure 2). However, if the organization has poor detection technology, this results in a limited number of precursors noticed. When no threats are noticed, the organization's perceived risk falls, decreasing the desired investment in security measures eroding even further its detection capacity. This behavior becomes a vicious cycle of eroding security capability leading to increased vulnerability via an enhanced false sense of security as nothing bad is noticed.

Detection capabilities of organizations interact in a very strong way with the level of managerial trust. As managers notice that the organization is in danger by identifying precursors, their level of trust in their workers declines, increasing the overall perceived risk. When perceived risk increases, the desired investments in security rise, lifting the detection capabilities and creating the circumstances for more precursors to be detected. Finally, again, as detected precursors are noticed, managers lower their trust even further, closing a cycle of protectiveness in the organization (see cycle R2 in Figure 2). However, when managerial trust is high, organizational perceived risk is low, decreasing desired investments in security preventing future detection of precursors. As very few precursors are detected, managerial trust is reinforced and a vicious cycle settles in.

Under conditions of poor detection capacity, only a very small fraction of the launched precursors are detected. Insiders sense that the risk of getting caught is low, increasing their willingness to continue testing the system and eventually lunch a harmful attack (see cycle R3 in Figure 2).

The three traps presented—detection, trust, and unobserved emboldening—create a complex system of interactions that makes the identification and prevention of insider threats/attacks a very difficult endeavor (Andersen et al. 2004).

The exploration of the emergence of insider threats has also been characterized as a learning problem (Martinez-Moyano et al. 2006a, Martinez-Moyano et al., forthcoming, Martinez-Moyano et al. 2006b). This learning problem is especially difficult due to behavioral and cognitive traits of individuals involved in the process, high levels of uncertainty, incomplete and imperfect information, incomplete and delayed feedback, and low base rates. In order to address these elements of the problem, and recognizing its inherent dynamic and feedback-rich nature, we created a model that integrates social judgment theory (Brunswik 1943, Hammond 1996, Hammond and Stewart 2001), signal detection theory (Green and Swets 1966, Swets 1973), and learning theories from psychology (Erev 1998, Klayman 1984) using the system dynamics approach (Forrester 1961, Richardson and Pugh 1989, Sterman 2000).

We developed a computer simulation model using Vensim® software from Ventana Systems to learn more about the emergence of the vulnerability problem and about which interventions can lower the organizational risk to insider attacks.

2 THEORETICAL BASE

In our model, we use constructs from social judgment theory, signal detection theory, and psychological learning theories to capture the judgment, the decision, and learning processes that are present in the selection-detection problem of identification of insider activity in complex systems.

Social judgment theory (SJT) evolved from Egon Brunswik's (1943, 1956) probabilistic functionalist psychology and work on multiple correlation and regression-based statistical analysis (Hammond 1996, Hammond and Stewart 2001, Hammond et al. 1975).

Decomposition of judgment is one important aspect of social judgment theory. The decomposition of judgment is achieved by identifying the main elements of the judgment process and by analyzing their interactions. In social judgment theory, the lens model is used as a way to represent the relationships that exist between information cues, the phenomenon being judged (also identified as distal variable as it represents a variable that is not directly observable or knowable), and the judgment (see Figure 3).

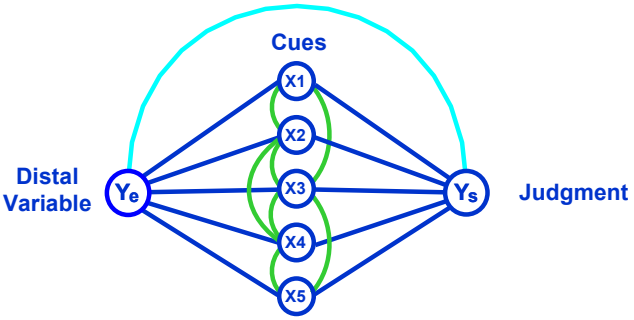

Figure 3: Lens Model of Judgment

The lens model uses information cues (in Figure 3 see X_1, …, X_5) as predictors of the distal variable—or the environment (Y_e)—when combined in a specific way.

The environment is modeled using a weighted additive linear combination of the information cues complemented by bias and error parameters. Equations 1 and 2 represent the general model for both the environment and the judgment:

$$Y = \hat{Y} + e, \qquad (1)$$

where

Y is the distal variable, or the judgment of the distal variable, (Y_s or Y_e in Figure 3),

$\hat{Y}$ is an estimate of Y, and

e is an indicator of either the inherent unpredictability of the environment or the degree of reliability of the judge.

$$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + k\,. \qquad (2)$$

where

$\hat{Y}$ is an estimate of Y,

b_n is the weight of information cue n on the distal variable or the judgment,

X_n is the information cue n, and

k is a bias term.

Identifying insider malicious activity can be seen as a case of the prototypical selection-detection problem. The selection-detection problem is typically characterized by the same underlying structure in which the base rate, the level of uncertainty, and the determination of the decision threshold determine decision outcomes.

Typically, the selection-detection problem implies identifying elements that belong to a group (positive distribution) when mixed with others (noise). Using signal detection theory (Green and Swets 1966, Swets 1973), the decomposition of the possible outcomes of the decision process is achieved: true positives, true negatives, false positives, and false negatives (see Figure 4).

		Judgment	
		Positive	Negative (Noise)
Actual Condition	Positive	True Positive	False Negative
	Negative (Noise)	False Positive	True Negative

Figure 4: Signal Detection Theory Outcomes

Social judgment theory, which allows the decomposition of the judgment process, coupled with signal detection theory (Green and Swets 1966, Swets 1973), which provides a mechanism to decompose outcomes of the decision process, present a unique framework to study and understand detection activities in complex systems as in the case of the detection of insider threat/attack activities.

3 STRUCTURE OF THE MODEL

Based on data from more than 200 cases of discovered malicious insider activity in organizations, we formulated a model of the case of long-term fraud (Randazzo et al. 2004). The formal model includes the formulation of a judgment process to characterize the likelihood of an insider threat being generated, a decision process that compares the judgment with the decision threshold to deter-

mine if action is necessary and a learning process that adjusts and updates the decision threshold depending on the mix of outcomes obtained. In Figure 5, below, we present a schematic representation of the basic feedback structure of the formal model (for more details, see Rich et al. 2005).

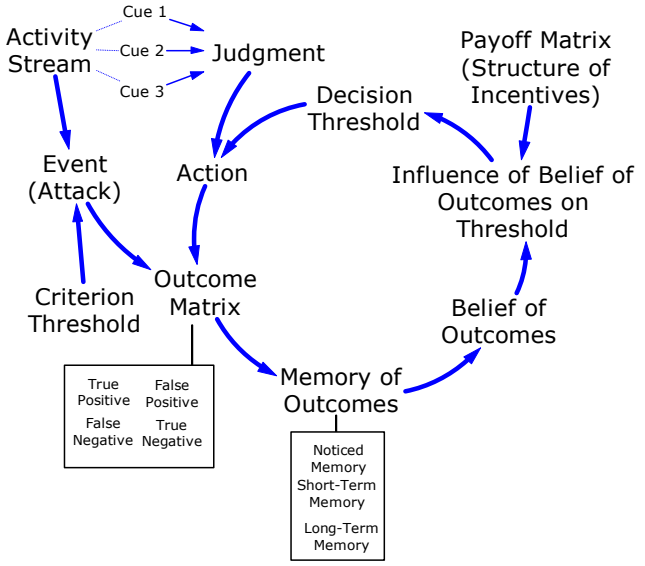

Figure 5: Model Structure

First, judgments are generated as a function of the combination of information cues from the environment according to an organizing principle that represents the cognitive process that humans follow in order to create composite indicators of variables that are not evident or not knowable in advance (e.g., trustworthiness of a person, likelihood of a people to commit a crime, probability of changes in the weather, need for surgery, etc.).

Second, decisions to act (or not) are made by means of comparing the judgments generated to the decision threshold specified for that problem. The decision threshold represents the definition of what a problem is and, specifically, when action is granted. For example, if the judgment score is 3.0 and the decision threshold is 4.0, then no action is granted as the level of judgment is lower than the threshold. However, as no judgment technology is perfect, there can be errors in this process. Two general sources of error exist in this process. The first source of error is in the judgment process. Judges can determine that the judgment score is 3.0 when in reality it is 4.1 due to imperfect knowledge of the relationship between the cues and the phenomenon or due to inconsistency in the application of their organizing principle. The second source of error is the determination of the decision threshold. Decision makers can determine erroneously that the threshold is 4.0 when the optimal threshold (the one that minimizes cost and errors) is different than that.

Third, using signal detection theory (Green and Swets 1966, Swets 1973), the decomposition of the possible outcomes of the decision process is achieved: true positives, true negatives, false positives, and false negatives (see Figure 4).

Fourth, adjustments to the decision threshold level are determined using a learning process that updates the decision threshold depending on the mix of outcomes obtained, the accuracy of the recording of these outcomes, and the beliefs formed about these outcomes. Depending on the belief formation process, adjustments to the decision threshold can be generated independent of actual results of the process (the simple case is represented in the model by providing the decision makers with perfect records of results and accurate belief formation mechanisms).

In addition to what is shown in Figure 5, the model captures several other feedback paths including effects from production pressure, productivity effects, precursor generation and identification, specifics of the mechanisms of memory and belief formation, and business dynamics.

The model produces behavior consistent with the data analyzed. In the cases studied, insiders produced precursors that allowed them to gain confidence in that they were going to be successful in attacking the organization without getting caught. We parameterized the simulation model to create a base run consistent with this behavior. In the base run we simulate an organization in which an insider probes the system until it has enough confidence to generate attacks. In the base run the organization is subject to an exogenous external attack base rate and endogenously generates an internal attack base rate depending on organizational and behavioral characteristics captured in different parameters of the model.

In Figure 6 (line 4), we show the increasing behavior of the insider's perceived probability of success as the number of undetected precursors grow. When the perceived probability of success exceeds the insider's attack mode threshold, the insider goes into attack mode stopping the production of precursors to start producing actual attacks.

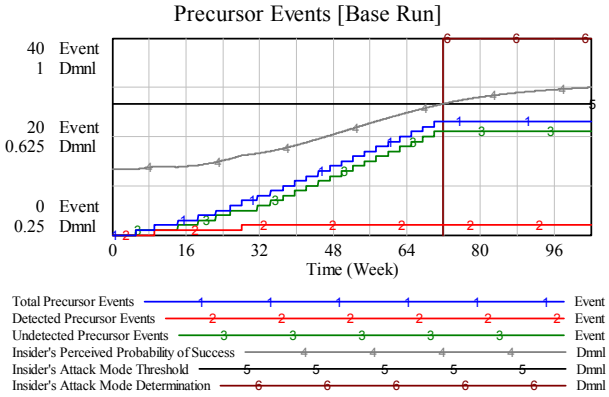

Figure 6: Model Behavior (Precursor Events)

Defenders in the organization change the level of their decision threshold by learning about the outcomes of their decisions (reinforcement learning). When nothing bad is noticed (either because nothing is actually happening or they are not capable of noticing it), defenders grow complacent as their perceived risk declines (see Figure 7, line 1) generating an increasingly unsecured environment in the organization. Once the insider starts attacking, after having produced enough successful precursors to be confident in that a successful attack is likely, defenders in the organization learn about the level of security and drop their decision threshold to increase security. These dynamics continue over time as defenders learn how to set the decision threshold in an optimal level that minimizes error and its associated costs.

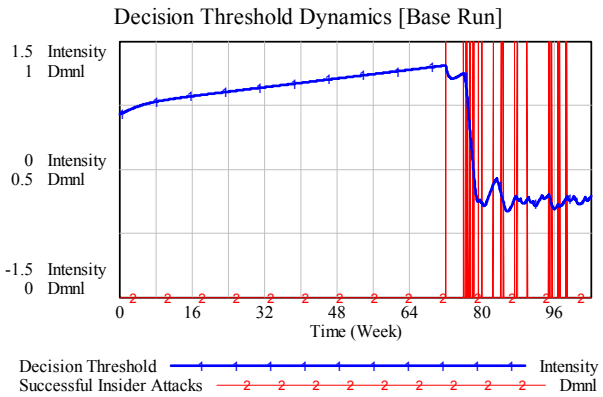

Figure 7: Model Behavior (Decision Threshold)

For further explanation about the model and for an in-depth exploration of the structure and analysis of the results, see other reports generated (Martinez-Moyano et al., forthcoming, Martinez-Moyano et al. 2006b).

4 LESSONS LEARNED

Through the modeling process, and through extensive simulation, we identified several insights and formalized them into three main lessons.

4.1 First Lesson: Framing How We Think is Crucial in Identifying Insider-Threat Vulnerabilities—Look for Whole System Effects

Insider threats and attacks, indeed most classes of threats/attacks, are as much the results of behavior and policy dynamics as they are the results of technical security issues. Technology can provide some protection, but it takes human behaviors to activate technical solutions. Workers must make choices to add a protective patch, to open or not open an attachment, or to respect and implement policies associated with automatic screening and detection systems.

Asymmetric security goals and information use within organizations will always exist: balancing security and operational effectiveness. Front line workers are responsible for production and view security in a systematically different way than security officers or those with primary responsibility for security. Understanding these differences, not as aberrations or security breaches, but as normal responses to work roles and pressures, will create more secure systems.

Additionally, recognizing the existence of competing values is crucial as many vulnerabilities arise because workers and managers have different views about how to weigh and value competing outcomes.

There are four possible outcomes for both internal and external intrusions:

1. Intercepted Attacks (true positives)—Intrusions that have been detected and defended against create big gains by avoiding losses.
2. Normal Transactions (true negatives)—Benign activity that takes place without expensive detection and defense activity.
3. False Alarms (false positives)—Normal activity made more expensive by unnecessary detection and defensive activities creates small expenses (most of the time).
4. Undefended Attacks (false negatives)—This is the worst possible outcome of a decision—attacks from internal or external sources that are not detected nor defended against often impose large losses on the organization.

The dynamics of the system influence the conditions that allow insider threats to emerge. Small forces accumulate over time creating large breaches of security in the end. Insider vulnerabilities arise from dynamic processes and are best combated by policies that recognize system level dynamics.

The impact of unobserved and often unobservable outcomes (as in the case in which no action is taken: belonging to the negative distribution) is important. Front line workers and security officers alike often lack information about the outcomes of their own activities. Furthermore, some important outcomes are deliberately hidden from view (such as insider precursor events). Learning to think correctly about unobservable events can help to identify vulnerabilities. Simulation models can help by simulating conditions of ignorance and full knowledge to provide insights about unobservable forces in the system.

4.2 Second Lesson: Everyday Human Cognitive and Behavioral Traits Create Unanticipated Vulnerabilities

Security technology is necessary to prevent vulnerabilities. However, everyday human cognitive and behavioral traits are major engines of vulnerability generation, even when all of the technical aspects of security have been considered. Our simulation model handles this by keeping the level of security technologies constant throughout all of the simulation runs. In the model, behavioral and policy differences between the simulated behavioral conditions are the only determinants of detection performance. Additionally, since the base rate of external attacks in our model remains constant throughout the simulated time, we assume that simulated defensive technology is getting better all the time, keeping up with increasingly sophisticated external attackers.

Judgment processes are crucial in the identification of insider vulnerabilities. Front line workers routinely make judgments about what may be both insider and outsider threats to an organization. Workers will inevitably make imperfect judgments because they are looking at several work factors and/or because they have less than perfect information.

Additionally, memory and belief formation play a key role in the development of vulnerability. Often, workers' judgments are based on beliefs that in turn are created by imperfect memory systems (either human or organizational). Because humans selectively pay attention to past events (hence are not perfect recorders of "the truth") additional vulnerabilities may arise or, alternatively, overly costly defense mechanisms may be put in place.

Decisions about how much security is enough are based on human judgments and perceptions about the four possible outcomes of attacks as well as human perceptions of how these outcomes should be valued. How workers and security officers make trade-offs between protecting against risk and valuing production can and will lead to vulnerabilities or overly costly defense mechanisms.

4.3 Third Lesson: It is Crucial to Create Smart Organizational Policies and Work Processes

While the costs and risks associated with vulnerabilities arise from human cognitive and behavioral traits, organizations can reduce vulnerability by carefully crafting organizational policies and training workers in various ways. These policies and this training must be undertaken in addition to technical security measures.

Our simulation model can be configured as a "management flight simulator" in which the user can systematically vary a number of pre-specified policies to see what might be their overall impact on various aspects (financial, risk, vulnerability avoidance, etc.). By systematically manipulating various policies in the simulator, users can learn lessons about which of the proposed policies are most effective. The policies in the model include the following:

1. Training workers to respect decision thresholds set by security officers.
2. Training security officers to use better information and risk factors in setting decision thresholds.
3. Training workers to more accurately weigh factors that predict insider threats.
4. Training workers to more accurately weigh factors that predict outsider threats.
5. Getting and using better information on insider threats.
6. Getting and using better information on outsider threats.

5 FURTHER RESEARCH

Further research in this area includes the development of experiments to identify specific information cues that different actors in the system look at when checking for insider activity. These information cues may change depending on the organizational level and on the type of organization. Additionally, empirical work may also include gathering data to identify the validity of the judgment and decision model in specific organizations and infrastructures.

The work reported here was developed using one specific type of insider threat: the long-term fraud case. Additional prototypical cases must be modeled, compared, and contrasted to continue the process of identifying a generic underlying structure for the emergence of vulnerability to insider attacks. CERT reports at least two more prototypical cases: the sabotage case and the information-theft case (also identified as espionage). The establishment of collaborative efforts with agencies and individuals pursuing this type of understanding will be useful in expanding the model in this direction.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Homeland Security.

REFERENCES

Andersen, D.F., D. Cappelli, J.J. Gonzalez, M. Mojtahedzadeh, A. Moore, E. Rich, J.M. Sarriegui, T.J. Shimeall, J. Stanton, E. Weaver, and A. Zagonel. 2004. Preliminary System Dynamics Maps of the Insider Cyber-threat Problem. In *Proceedings of the 22nd International Conference of the System Dynamics Society*, Oxford, UK.

Brunswik, E. 1943. Organismic achievement and environmental probability. *Psychological Review*, *50*. pp. 255-272.

Brunswik, E. 1956. *Perception and the representative design of psychological experiments*. University of California Press, Berkeley, CA.

Erev, I. 1998. Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*, *105*. pp. 280-298.

Forrester, J.W. 1961. *Industrial Dynamics*. Productivity Press, Cambridge MA.

Green, D.M. and J.A. Swets. 1966. *Signal Detection Theory and Psychophysics*. John Wiley, New York.

Hammond, K.R. 1996. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford University Press, New York.

Hammond, K.R. and T.R. Stewart (eds.) 2001. *The Essential Brunswik: Beginnings, Explications, Applications*. Oxford University Press, Oxford.

Hammond, K.R., T.R. Stewart, B. Brehmer, and D.O. Steinmann. 1975. Social judgment theory. In Kaplan, M.F. and Schwarts, S. eds. *Human judgment and decision processes* Academic Press, New York, pp. 271-312.

Keeney, M.M. and E.F. Kowalski. 2005. Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors, CERT.

Klayman, J. 1984. Learning from Feedback in Probabilistic Environments. *Acta Psychologica*, *56*. pp. 81-92.

Martinez-Moyano, I.J., E.H. Rich, and S.H. Conrad. 2006a. Exploring the Detection Process: Integrating Judgment and Outcome Decomposition. *Lecture Notes in Computer Science*, *Volume 3975*. pp. 701-703.

Martinez-Moyano, I.J., E. H. Rich, S.H. Conrad, D.F. Andersen, and T.R. Stewart. Forthcoming. A Behavioral Theory of Insider-Threat Risks: A System Dynamics Approach. *Transactions on Modeling and Computer Simulation (TOMACS)*. 1-36.

Martinez-Moyano, I.J., E.H. Rich, S.H. Conrad, T. Stewart, and D.F. Andersen. 2006b. Integrating Judgment and Outcome Decomposition: Exploring Outcome-based Learning Dynamics *International Conference of the System Dynamics Society*, System Dynamics Society, Nijmegen, The Netherlands.

Randazzo, M.R., M.M. Keeney, E.F. Kowalski, D.M. Cappelli, and A.P. Moore. 2004. Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector, U.S. Secret Service and CERT Coordination Center / Software Engineering Institute, 25.

Rich, E., I.J. Martinez-Moyano, S. Conrad, A.P. Moore, D.M. Cappelli, T.J. Shimeall, D.F Andersen, J.J. Gonzalez, R.J. Ellison, H.F. Lipson, D.A. Mundie, J.M. Sarriegui, A. Sawicka, T.R. Stewart, J.M. Torres, E.A. Weaver, J. Wiik, and A.A. Zagonel. 2005. Simulating Insider Cyber-Threat Risks: A Model-Based Case and a Case-Based Model *International Conference of the System Dynamics Society*, System Dynamics Society, Cambridge, MA.

Richardson, G.P. and A.L. Pugh. 1989. *Introduction to System Dynamics Modeling*. Pegasus Communications, Inc., Waltham.

Sterman, J.D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin McGraw-Hill, Boston MA.

Swets, J.A. 1973. The relative operating characteristic in psychology. *Science*, *182* 4116. pp. 990-1000.

AUTHOR BIOGRAPHIES

IGNACIO J. MARTINEZ-MOYANO is a Judgment and Decision Modeling Scientist at the Decision and Risk Analysis Group of the Decision and Information Sciences Division at Argonne National Laboratory. His research work focuses on the development of mathematical models of physical and psychological phenomena using numerical simulation to explore the dynamic implications of their built-in assumptions, specifically, behavioral modeling of judgment and decision-making processes and the use of the system dynamics approach in critical infrastructure modeling. He also holds an appointment as Adjunct Research Associate at the Center for Policy Research of the University at Albany, State University of New York. His e-mail address is <imartinez@anl.gov>.

ELIOT H. RICH is an Assistant Professor at the School of Business of the University at Albany. Dr. Rich holds a Ph.D. in Information Science from the University at Albany (1992) and a Masters in Public Policy from Harvard University (1981). He teaches in the areas of information policy, knowledge management, networking, and information security. His e-mail address is <e.rich@albany.edu>.

STEPHEN H. CONRAD is a Principal Member of the Technical Staff at Sandia National Laboratories. His research work includes the modeling of interdependencies in critical infrastructure. His e-mail address is <shconra@sandia.gov>.

DAVID F. ANDERSEN is a Distinguished Service Professor of Public Administration and Information Science at the Rockefeller College of Public Affairs and Policy, University at Albany, State University of New York. He is a founding fellow at the Center for Technology in Government. Dr. Andersen holds a BA in Mathematics and Social Sciences from Dartmouth College and a Ph.D. in Management from the MIT Sloan School of Management. His e-mail address is <david.andersen@albany.edu>.