# Use of Expert Judgments to Inform Bayesian Models of Insider Threat Risk

Frank L. Greitzer[1]*, Justin Purl[2]†, Paul J. Sticha[2]‡, Martin C. Yu[2], and James Lee[3]
[1]PsyberAnalytix, LLC, Richland, WA USA
Frank@PsyberAnalytix.com
[2]Human Resources Research Organization, Alexandria, VA USA
Justin.purl@gmail.com, Paul.Sticha@psychinference.com, myu@humrro.org
[3]George Mason University, Fairfax, VA USA
jlee194@gmu.edu

Received: December 21, 2020; Accepted: June 13, 2021; Published: June 30, 2021

## Abstract

To promote effective detection and mitigation of insider threats, research has sought to identify, validate, and integrate cyber and behavioral (sociotechnical) indicators into comprehensive models of insider threat risk. Because validation of proposed indicators is hampered by a lack of appropriate real-world data, innovative approaches have used expert judgments as an initial step in developing and evaluating threat assessment models. For probabilistic models such as Bayesian networks, assigning probability values to posterior evidence is particularly challenging because it often relies on subjective base-rate (prior) and conditional probabilities estimates that are difficult to obtain and fraught with human errors and biases. The purpose of the present study was to test the efficacy of an expert knowledge elicitation method that does not rely on probability judgments in supporting development of probabilistic as well as non-probabilistic/risk-based predictive models of insider threat. We compared previously obtained expert judgments of threat/risk levels for a large set of indicators within a comprehensive ontology of technical and behavioral indicators of insider threats with corresponding likelihood ratio estimates that we obtained in the present study, concluding that the observed high correlation between the risk versus probability judgments demonstrates the efficacy of acquiring expert judgments of threat/risk levels as a practical alternative to the difficult and unreliable methods of acquiring conditional probability estimates from human experts. Based on these results, we created a Bayesian model of insider threat that incorporates all (∼200) individual factors specified in the ontology and compared the performance of the Bayesian and risk-based models in predicting the judgments of experts, as proxies for real data and ground truth. Results indicated that the Bayesian model performed slightly better than a risk-based model that had been proposed and examined in prior research. This research demonstrated benefits of cross-fertilization of methods used in developing non-probabilistic/risk-based and probabilistic models in the insider threat domain. Implications of these findings for advancing insider threat predictive analytics, and future research needs, are discussed.

**Keywords**: Insider threat, SOFIT ontology, expert judgments, Bayesian models, threat assessment models

*Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 12(2):3-47, June 2021
DOI:*10.22667/JOWUA.2021.06.30.003*

*Corresponding author: 651 Big Sky Dr, Richland, WA 99352, Tel: 509-539-4250

†Justin Purl is now at Google.

‡Paul J. Sticha is now at PsychInference, LLC.

# 1 Introduction

An insider is an individual at an organization or company with authorized access to or knowledge of information systems, services, and missions [36]. An *insider threat*—according to a recently updated definition by the Carnegie Mellon University Software Engineering Institute, CERT Division—is the *potential for an individual who has or had authorized access to an organization's assets to use their access, either maliciously or unintentionally, to act in a way that could negatively affect the organization* [11] (cf. [8]). Intentional, malicious insider crimes and abuses include espionage, sabotage, embezzlement, extortion, bribery, corruption, intellectual property theft, negligent use of classified data, fraud, unauthorized access to sensitive information, providing sensitive information to unauthorized recipients, and workplace violence. Unintentional insider threats are non-malicious/inadvertent acts by insiders that harm the organization or expose it to harm, such as inadvertent information leaks or responding to social engineering attacks that expose the organization to outside attacks.

Challenges facing the development and testing of insider threat mitigation/monitoring approaches include (a) an inadequate framework or infrastructure to provide systematic integration of predictive or relevant cyber and behavioral contributing factors and (b) lack of empirical data and ground truth to inform detection models. To help address these challenges, the present work reports on research to advance a knowledge base framework integrating cyber and behavioral factors that was developed and informed through expert knowledge elicitation methods; and describes an empirical study that supports the implementation of probabilistic, risk-based predictive models of insider threat—informed by expert judgments of the severity of predictive factors—as a proxy for missing ground truth information.

Guidelines and policies for establishing insider threat programs across the US government represent progress in meeting these challenges, but there is much more to be done. For example, Executive Order 13587 [48] and the National Insider Threat Policy [37] specify only minimum standards for establishing an insider threat program and for collecting and analyzing vast amounts of data to support monitoring and mitigation systems. Numerous techniques and approaches are used in research aimed to detect or predict insider threats. Sanzgiri and Dasgupta [41] distinguish nine categories of insider threat detection or mitigation techniques, including approaches that are anomaly-based, rule-based, scenario-based, risk-based models that incorporate psychological factors, risk analysis of workflow, and several defensive approaches involving decoy-based solutions, network defense, access control improvements, and process control improvements. As pointed out in [41], the increase in consequences of insider attacks over recent years has led to a greater emphasis on risk-assessment approaches, which tend to address *human factors*: i.e., behavioral/psychological characteristics that include assessment of attack-related behaviors [42], physical behavior indicators [36], psychological characteristics or personality traits [21] [22], and modeling of adversarial behavior [2]. To achieve more effective detection and mitigation using these techniques, the research community must address critical challenges in identifying, validating, and integrating cyber and behavioral (sociotechnical) indicators of insider threat risk [50] and must also seek to establish more comprehensive *behavioral + technical* (sociotechnical) models [19] [18].

Deterring insider threats requires identifying and monitoring indicators of these insider actions. Indicators are observable characteristics, capabilities, and behaviors associated with increased probability of threat. Creating an inference enterprise, or organizational unit devoted to detecting threats [7], is one way to harden an organization against insider threats. A crucial element of an inference enterprise is the application of sound methods and models to facilitate the threat assessment process. As Homoliak et al. [27] suggest, to achieve a robust insider threat program, a combination of several independent solutions should be employed—mitigation and prevention techniques provide the first line of defense, misuse-based detection supports the second line of defense, and anomaly-based detection and risk assessment methods underlie the final line of defense, where each successive solution should feed into or "alert" the next level of analysis. Risk-based assessment or prediction models have been developed using a variety of methods including statistical, rule-based, Bayesian, machine learning, and artificial intelligence methods.

The lack of sufficient quantities of real-world data available to build models and detection tools, along with ground truth to test them, was identified by the Information Security Research Council (IRC) as a major reason why insider threat was listed as #2 on the INFOSEC Hard Problems List [28], and these challenges remain to this day. Despite many incidents of insider attacks, organizations are reluctant to share data with the research community. This applies even more to behavioral data, which is scarcely available for legal, privacy and other reasons. Given these challenges, there is a need for information to support predictive modeling—outside a particular organizational context and without ground truth information on any given exploit—that organizations may build upon to produce their own useful mitigation solutions.

In the absence of empirical data, innovative approaches have tapped expert knowledge to inform insider threat models as an initial step in developing robust threat assessment tools (e.g., [7]). This requires creative ways to elicit expert judgments about individual indicators and to incorporate this knowledge into quantitative models that aggregate individual indicator threat values to assess threats of cases comprising multiple observed indicators. These predictions must then be evaluated against expert judgments of insider threat cases. For probabilistic models such as Bayesian Networks (BN), the challenges associated with obtaining reliable probability ratings are particularly acute. The purpose of this work was to describe and test the applicability of a knowledge elicitation method that may be used to support development of both probabilistic and non-probabilistic, risk-based predictive models of insider threat.

This paper is organized as follows: Section 2 reviews related research and methods used to support development and testing of risk-based insider threat models as background for the development and testing of BN models of insider threat risk. Section 3 describes the development of two such BN models. In Section 4 we describe how probabilities that inform Bayesian models of insider threat risk may be derived from an expert judgment knowledge-elicitation study; and we describe the method and results of a likelihood survey that was used to specify the relationship between expert judgments of insider threat indicator level of concern and likelihood-ratio/probability estimates that support Bayesian Model development. Section 5 documents the methods we used to derive conditional probability tables for these BN models based on the level of concern ratings, and presents the results obtained in testing the BN models predictions of expert judgments of risk obtained in an expert knowledge elicitation exercise that served as a proxy for ground truth evidence. We close the paper in Section 6 with a discussion of conclusions, limitations, and future research needs.

## 2 Background

In this section we describe the motivation and problem that stimulated the present study; we provide an overview of the knowledge elicitation approach upon which the study was based; and as a backdrop for the BN models that are described later, we summarize a knowledge base—the Sociotechnical and Organizational Factors for Insider Threat (SOFIT)—that informed the Bayesian models. Also, for comparison with these new BN models, we review related research that developed and investigated BN models in this application domain; and we describe research that investigated several risk-based insider threat predictive models based on SOFIT that informed the present study and that provide several alternative predictive models for comparison.

## 2.1 Problem

As noted in the Introduction, in the absence of empirical data, some approaches to insider threat predictive modeling have used expert judgments as an initial step in developing threat assessment tools (e.g., [7], [17]). In general, these modeling efforts define and represent insider threat indicators and collections of multiple indicators (which are observed or reported as cases), with the aim of incorporating the judgments of experts who rate the "severity" of the threat. One way to characterize these judgments is the use of probability estimates that are incorporated into a BN, a particularly useful approach to address the challenge of insufficient historical data and ground truth [9]. Elicitation of expert knowledge can combine different sources of knowledge to inform the nodes and conditional probabilities within such Bayesian belief networks. In their review of BN models in cybersecurity, Chockalingam et al. [9] determined that 11 out of 17 of the models exclusively used expert knowledge to populate conditional probabilities, while only three of the models exclusively relied on empirical data (the remaining three models used both sources of information). Among these was a model developed by Axelrad et al. [2] that used a set of 83 indicators potentially related to insider threat (e.g., psychosocial and counterproductive workplace behavior indicators), which were ranked and combined into a single risk score using a BN. Also listed in the Chockalingam et al. [9] review was a BN to predict the psychosocial risk of an individual, addressing only behavioral indicators [20]. The Axelrad et al. [2] model used both empirical data and expert knowledge; the Greitzer et al. [20] approach (which included a Bayesian Model as well as statistical/regression and artificial neural network models) relied only on expert knowledge.

The task of estimating probabilities, especially those of rare events that are typically associated with insider threats, is a difficult one that can yield biased outcomes. The use of expert knowledge to inform BN models is particularly problematic since, as noted by Kwan et al. [31], these estimates are based on subjective personal beliefs, and people have difficulty thinking in terms of conditional distributions and probabilities. Studies of human probability assessment, across a range of elicitation techniques, reveal that human estimators generally are prone to overconfidence (giving probability estimates too close to zero or one); but experts tend to underestimate their confidence [38]. Harris et al. [24] showed that probability estimates are systematically biased, such that rare events with extremely negative consequences are judged to be more likely to occur than neutral events.

Because of these reliability concerns in obtaining human-generated probability estimates, other approaches to obtain ratings of insider threat have sought to gain a sense of the "threat level" on a numeric scale. This non-probabilistic rating approach was the foundation of new models developed by Greitzer and colleagues [17] [16]. These models were informed by the development of the Sociotechnical and Organizational Factors for Insider Threat (SOFIT) ontology (e.g., [17]). Greitzer et al. [17] obtained expert judgments of individual insider threat indicators using a 0-100 "Level of Concern" rating scale; these estimates were used to test various alternative threat models for aggregating threat values of collections of indicators observed in insider threat cases. Model predictions were compared to expert judgments obtained for a collection of cases examined in a second knowledge elicitation study. The Level of Concern judgments obtained in these studies were not unlike the judgments underlying the BN model developed by Axelrad et al. [2] that sought to predict the "degree of interest" in a potentially malicious insider, although the modeling techniques differed substantially.

Because the SOFIT knowledge base and the expert knowledge elicitation approach used to develop the non-probabilistic models also served as a foundation for the methods to be described in this paper for instantiating BN probabilities, for completeness these methods are briefly reviewed here (more detailed information may be found in [17] and [16]).

## 2.2 SOFIT Knowledge Base

Research and case studies over the last two decades offer a large set of insider threat indicators that have been described by many papers with divergent foci. Based on examination of over 200 sources, including published papers, technical reports, and case studies in cybersecurity, human factors/organizational effectiveness, workplace violence, and associated taxonomies or ontologies, we compiled a structured list of indicators implemented as the SOFIT ontology. To the best of our knowledge, this is the most comprehensive insider threat indicator knowledge base that encompasses both behavioral and cyber/technical indicators. The SOFIT ontology lists indicators with labels that reflect commonly used terms, with brief definitions or descriptions, and with selected citations of source material. Compared to previous works (e.g., [52], [1], [5], [43], [12]), the SOFIT knowledge base includes not only a large number of cyber-technical indicators, but also specifies psychosocial (behavioral and psychological) indicators based on works such as [20], [47], [15], [46], [4], [10], [45], and [44].

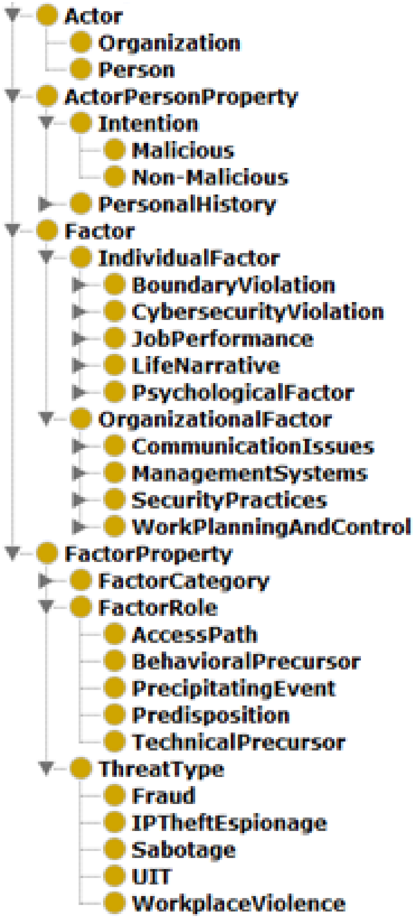


Figure 1: Upper levels of SOFIT ontology

Fig. 1 shows the main constructs (or classes) comprising the upper levels of the SOFIT knowledge base developed and reported by [17]. The taxonomy accounts for both malicious and non-malicious (unintentional) insider threats, and it distinguishes Individual Factors associated with employees (as insiders) from Organizational Factors such as problematic responses to potential threats, poor institutional policies, or security practices. It also distinguishes among five threat types. In Supplement I, we provide documentation of the entire set of SOFIT constructs, its class structure, relationships among these constructs, estimated threat values for indicators, and selected references/resources used in compiling the ontology—this includes the SOFIT ontology (.owl file) and the SOFIT taxonomic listing of this information in PDF format.

Fig. 2 expands the Individual Factor node, exhibiting classes (and associated subclasses) corresponding to the third and fourth levels of the taxonomy. For example, the subclass Boundary Violation reflects a large set of individual actions (indicators) such as Concerning Work Habits, Blurred Professional Boundaries, and Interpersonal Problems. Subclasses of these constructs are at the fifth and lower levels of the taxonomy and referred to as "observables" (not shown in the figure). The subclass Psychological Factor includes the indicator classes Dynamic State and Enduring Trait; dynamic states comprise the observables Attitude and Mental States; enduring traits include observables Personality Dimensions and Dark Triad. Each of these subclasses is further delineated with lower-level observables (not shown).

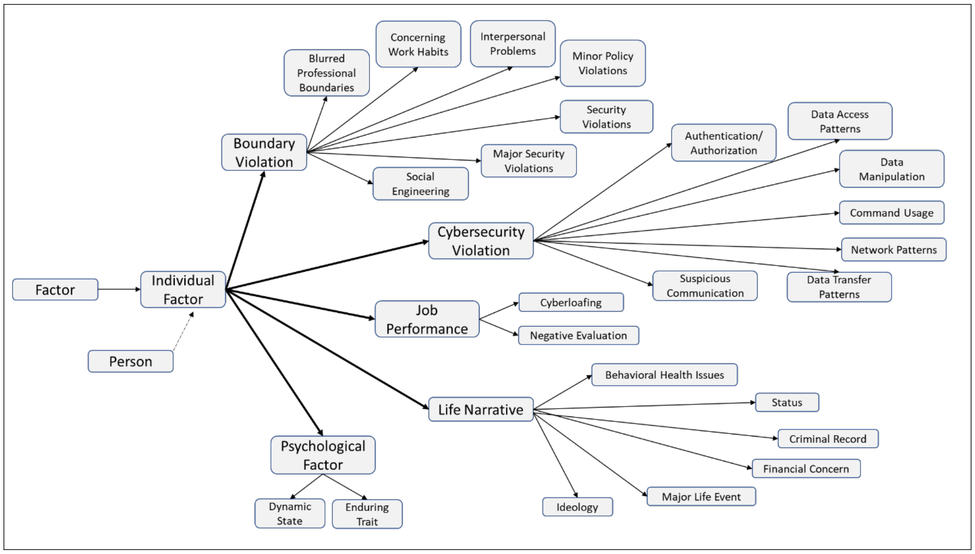


Figure 2: Individual Factor branch of SOFIT ontology (Greitzer et al., 2018)

Development of an insider threat knowledge base, particularly as implemented within an ontology, can facilitate specification of threat/risk models to meet mitigation goals. The structure of the knowledge base can inform the structure of the threat model—as will be seen, this applies both to the non-probabilistic risk models and to BN models. Development of quantitative and/or predictive models has proceeded slowly, with few attempts to implement rigorous models that integrate behavioral and technical indicators.

## 2.3 Expert Knowledge Elicitation Methods

Several Level of Concern based models, as well as the BN models discussed in Sections 3-5, were developed and tested with the aid of data acquired in expert knowledge elicitation exercises reported by [17]. To provide background and context for the evaluations reported in this paper, we briefly review the methods employed. The surveys were conducted with experts from across the research and operational communities to calibrate the relative levels of concern for SOFIT indicators. Experts assessed the extent to which individual indicators and groups of indicators contribute to perceived insider threat risk. Because the SOFIT knowledge base comprises more than 300 individual constructs, only a small fraction of all possible subsets composed of these indicators could be examined. Fortunately, any representative group of indicator combinations can provide enough breadth to allow a test for the plausibility of tested models.

As context, the experts were given a fictional backstory with instructions that expert judgments were needed to help evaluate an insider threat alerting system that analyzes data collected from diverse employee monitoring systems and alerts a security team about suspicious behavior. Thus, it was explained, the objective of the insider threat alerting system was not to apprehend a supposed insider criminal, but rather to identify cases for further review by the insider threat assessment team. For each factor, definitions and descriptions were provided to aid the expert in judging the degree of potential insider threat risk (measured by the selected "level of concern") associated with the factor. In the first stage of the task (individual factor risk judgment), our experts rated each factor using a 0-100 point scale ranging from "no concern" to "extreme concern" [17]. Fig. 3 depicts the ranges of the mean threat values for these indicators, grouped within their respective classes within the SOFIT ontology. The indicator classes are shown in the left column and examples of indicators within these classes are shown in the right column. As reported previously, the ranges plotted in the graph in the middle demonstrate clearly that individual factors vary widely in level of concern as insider threat indicators.

It is evident that in most situations, a combination of (multiple) indicators would need to be observed before an analyst would judge a case to be of extreme concern. For example, a case in which an individual is observed to work at unusual hours represents a rating of "somewhat concerning"; a case in which the individual has a big ego would generate a slightly higher rating of "concerning"; and a case exhibiting attempts to access the system against policy would generate a rating "very concerning." None of these indicators, on its own, would rise to the level of extreme concern. However, a case presenting a pattern of concerning behaviors (such as an individual who works at unusual hours, has a big ego, and attempts to access the system against policy) might generate an extremely concerning insider threat risk—i.e., a risk that would justify further monitoring and analysis of the associated person of interest. However, this simple rating task on individual indicators does not inform a model describing how multiple indicators might be combined to yield an overall risk judgment. This was the motivation for conducting the sorting tasks involving cases with multiple indicators reported in [17].

Thus, in the second stage of the study reported in [17], experts were given descriptions of 45 cases comprising 2 to 5 indicators and asked to use a sorting procedure that ultimately produced a rank-ordering of all the cases from least (i.e., not at all concerning) to most (i.e., extremely) concerning. Data obtained in this phase of the study were used to evaluate the abilities of various models to account for the case rankings, given the individual threat values (as represented in Fig. 3, above) that were estimated from the initial phase of the study. The advantage of this methodology is that all statistical evaluations of the models are based on the number of cases tested rather than the number of subject matter experts providing the ratings. Specifically, the degrees of freedom in the analysis are based on the number of cases evaluated and not the number of individuals providing ratings. One drawback of this methodology is that the error in the model stems from rater level error and case level error and statistical significance may be harder to obtain. This drawback is offset in the current situation by the lack of potential subject matter experts that can contribute to open resources on insider threat. Models described in the next subsection, as well as the new BN models presented in Section 3, were evaluated against the rank orderings that emerged from this expert knowledge elicitation study.

## 2.4 Related Insider Threat/Risk Models

Numerous insider threat assessment and detection approaches have been discussed and proposed in the literature dating back at least two decades. Here we focus on computational models that include behavioral factors in their threat assessment approach. These examples demonstrate the difficulties in estimating probabilities in the construction of quantitative models in insider threat research.

**CMO-Based Approach**. Kandias et al. [30] describe the development of an Insider Threat Prediction Model that used a multi-dimensional perspective that relies on several stages of analysis including real-time monitoring, psychometric testing, and other measures, this model applies the capability-motivation-opportunity (CMO) concept that assumes each threat requires capability, motivation, and opportunity [52]. The model produces a composite threat score, $T = M + O + C$, which sums risk across three main dimensions of Motivation, Opportunity, and Capability. The multi-stage assessment accumulates threat ratings, which are generally numeric values such as low (1-2), medium (3-4), high (5-6), that reflect indicators within several different assessment components. A user taxonomy component contributes the assessment of user access levels such as system role (novice, advanced, administrator); a psychological profiling component contributes ratings that reflect user technical sophistication, predispositions, and stress level; and a real time usage component assesses technical indicators such as system calls, intrusion detection system alerts, and honeypot alerts. The scores reflect findings in scientific research literature and are apparently derived from expert judgments by the model developers. While there was no indication that this model was evaluated against empirical, simulated, or expert judgments, this work merits mention here due to its broad scope that includes indicators related to Opportunity and Capability.

**Psychosocial Risk Modeling**. Greitzer et al. [22] developed a BN model focusing on psychosocial indicators of insider threat risk. The task of estimating prior and conditional probabilities to populate the model requires many complicated judgments involving possible combinations of the twelve indicators (the "power set" of all subsets of the twelve indicators), of which there are 4,096 possible cases. Because this was impractical, an alternative approach to derive conditional probabilities was employed that requires expert judgments for only about 3% of the total number of possible cases. They constructed 110 scenario cases comprising up to five indicators and then asked two Human Resources (HR) experts to

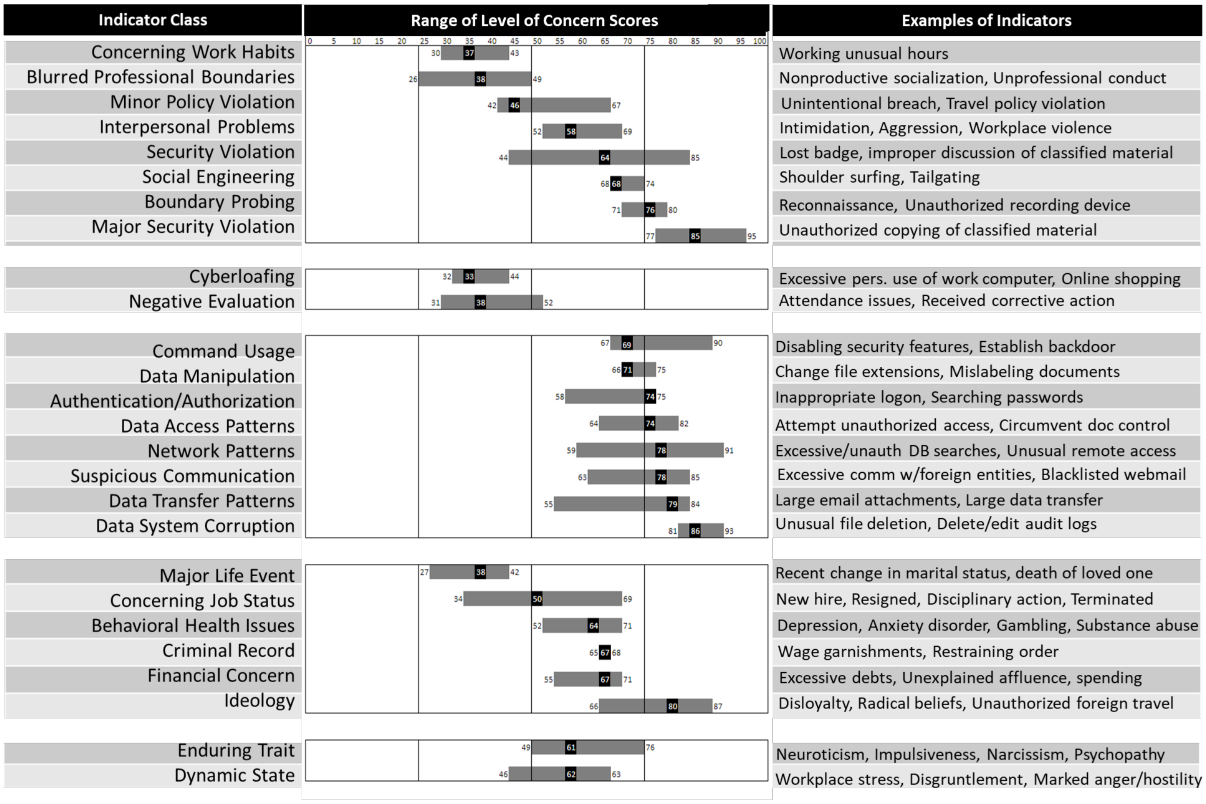


Figure 3: Level of Concern Ratings for Classes of SOFIT Individual Indicators (based on data collected by Greitzer et al. [17]).

assign insider threat risk levels to employees who would exhibit those behaviors. These judgments were used to populate the BN model, and then the model was tested in an experiment that obtained judgments of ten HR experts for 24 hypothetical cases. The participants used a sorting task to rank the cases from lowest to highest concern, and the BN model was used to predict the rankings. The BN model performance was compared with several other models including a nonlinear feedforward Artificial Neural Network (ANN) model, a linear regression model, and a simple counting model. The BN, regression, and ANN models, which differentially weigh the individual indicators, performed about equally well, with $R^2$ values accounting for roughly 60% of the variance; the counting model that assumes all indicators are equally predictive performed poorly, with an $R^2$ accounting for only 25%. The methods employed in this relatively narrowly focused study on psychosocial factors were in many ways a precursor to the approach described in [17].

**Bayesian Modeling of Behavioral Risk Factors**. Axelrad et al. [2] developed a BN model that uses psychological factors to predict interpersonal and organizational counterproductive behaviors. An initial model was developed to incorporate critical variables for predicting "degree of interest" in a potentially malicious insider. Predictor variables in the model were defined based on published findings in the research literature, with conditional probabilities informed by correlations reported in the literature (and supplemented as needed by judgments of the authors). Predictor variables included personality traits of agreeableness, neuroticism, conscientiousness, excitement seeking, perceived stress, hostility, and job satisfaction. The variables to be predicted in the model were interpersonal deviance (counterproductive behaviors directed toward an individual) and workplace deviance (counterproductive behaviors directed against the organization). A survey assessing the relationship between the predictor variables and the counterproductive behavior variables was used to further refine the BN. The updated Bayes model was then tested in a survey data to determine its ability to predict the counterproductive behavior variables from the set of psychological predictors incorporated in the model. The survey was conducted with individuals recruited using Amazon Mechanical Turk, i.e., in an environment that does not represent any particular organization. Also, the outcome variables were *proxies* for insider threat behaviors reflecting personal and organizational deviance or job satisfaction. The resultant prediction error rates showed that the performance of the revised model was better than the original model in predicting the empirical data. However, while the percentage of cases for which the model-predicted value of the counterproductive behavior variable was different from the actual value of the variable was substantially lower than chance, the prediction accuracy of the model was not high. Possible limitations in the predictiveness of the model might include the low associations between variables, the validity of the proxy measure used for counterproductive behaviors, and the fact that these are rare events.

**BAIT Framework**. Azaria et al. [3] developed the Behavioral Analysis of Insider Threat (BAIT) Framework, which defines a set of 28 features that were analyzed using Support Vector Machines (SVM) and multinomial naïve BN modeling. The framework uses behavioral cues to identify insider threats: sixteen basic features track the number of times an action was performed; twelve additional features were derived from various types of "send" actions (i.e., sending information out of the organization), "transfer" actions (e.g., save to CD, save to USB, print), or "fetch" actions (e.g., from internal database, from USB drive). A distinguishing feature of this approach is that the modeling approach was evaluated against the performance of real users who both act normally and who are charged with exfiltrating data from an organization. Thus, in contrast to other testing approaches that compare model predictions either with expert knowledge or with a synthetic dataset, this approach used real data, albeit the users did not belong to any real organization since they were participants in a game-based study implemented using Amazon Mechanical Turk. The study recruited 795 people to play a one-person game in which most were assigned the role of benign user while a very small number were malicious. Seven BAIT algorithms were compared to determine which yielded the best accuracy in identifying benign versus malicious users. The best performing algorithms had a recall of 0.6 with precision = 0.3, meaning that it could correctly identify 60% of the malicious insiders while guaranteeing that about one of three flagged suspects would indeed be a malicious insider.

**Anomaly Analysis of User Activity Monitoring Data**. Legg et al. [33] (see also [32]) developed an anomaly-based insider threat detection system that performs user and role-based profiling to identify and classify anomalies, producing multiple anomaly scores. Input data include activity logs corresponding to five activities: login, USB device, email, web, and file access. A user and role-based profiling stage uses a tree-structured profile to represent all users and roles; a diverse set of 168 features representing the five targeted activities is extracted from input data; a rule-based approach determines whether the features exhibited in a user's daily profile represent policy violations or match previously-recognized attack patterns. Key types of assessment are the features of the user's daily observations, comparisons between daily activity and previous activity, and comparisons between daily activity and previous activity for the associated role. The threat assessment generates anomaly scores for the set of features and applies weights reflecting importance of the features (if defined, otherwise these are equally weighted). Sixteen anomaly metrics (examples include *Login_anomaly, USB_insertion_anomaly, Web_anomaly, File_anomaly, Role_anomaly*) are computed based on analysis of associated sets of features. An anomaly score matrix is evaluated to make a final classification decision of threat/no-threat. Examples of decision criteria include a decision threshold for normalized anomaly scores exceeding two standard deviations of the norm; or use of the Mahalanobis distance to assess distance of an individual's observation from the rest of the distribution; or to assess the signed differences between covariances in the covariance matrix of a user's anomaly scores. Alerts are generated for anomalies whose scores exceed a threshold. The system was tested using synthetic datasets that contained ten insider threat scenarios. The best results were reported for scenario 3, whose data included 4200 daily assessments made for 12 employees over 350 days, of which 24 assessments were flagged as anomalies. These alerts identified ten out of ten cases with malicious activity, yielding a precision of 42% and recall of 100%.

**Inference Enterprise Modeling**. The Scientific advances to Continuous Insider Threat Evaluation (SCITE) program sponsored by the Intelligence Advanced Research Projects Activity (IARPA) was created to improve performance of insider threat detection and forecasting models. The research program assigned competing teams increasingly complex challenge problems that required the use of inference enterprise modeling (IEM). The competitors were given aggregated and often incomplete data from an unnamed organization with approximately 4,000 employees with the aim of identifying individuals that exhibited certain target behaviors of insider threats (e.g., excessive personal use of work computers, unusual working hours, and unusual foreign contacts). In IEM, the enterprise processes incoming data to derive indicators thought to be associated with the target behavior of interest. Missing data requires either forecasting methods or expert knowledge elicitation to infer reasonable values. Performance of the different teams competing within the SCITE program was evaluated by an independent organization, which created a baseline BN model as a reference against which the competitors' models were measured. The most successful team, led by Innovative Decisions, Inc. (IDI) [7] [6] used a Multi-model Inference Enterprise Modeling (MIEM) technical approach that integrated the results of a diverse set of modeling methods, including Bayesian networks, discrete-event threat scenarios, discrete-event activity counts, stochastic optimization, factorized stochastic optimization, Markov Chain Monte Carlo, gaussian copula, neural networks, classification trees, linear programming, logistic regression, and historical model using weighted linear opinion pools. As a general rule, these models relied on expert judgments to refine the models and/or augment missing data. The fused model produced by the MIEM multi-modeling approach yielded better performance metrics than any of the individual models. Most relevant to the present discussion is the fact that both the baseline model developed by IARPA and the models developed by the IDI team were informed by, and tested against, data that were augmented by expert judgments.

**Risk Models Based on SOFIT**. Recent research by Greitzer and colleagues [17] [16] examined various mathematical models of risk judgment to gain a better understanding of the ability to estimate or predict the level of concern for cases comprising collections of indicators, as might be observed in an organization. In contrast to the approaches described above, this research incorporated a broad knowledge base, the SOFIT ontology, as a framework for the threat assessment models. The models were tested by comparing their performance in predicting expert rankings of the level of threat for hypothetical insider threat cases. The models express the level of concern or risk score for a set of indicators such that an extremely high score prescribes further scrutiny by insider threat analysts. The variable, $R_j$, denotes the predicted level of threat or risk for person $j$ specified by a risk model. It is important to distinguish this measure from a *probability* that the case is an insider exploit, which is complicated by extremely low base rates and much more difficult to estimate from expert judgment studies. The risk score is simply an indication of concern for the case, where more concern represents an increased need to review the case further. Several alternative risk models that were developed are briefly described below (we have updated and improved some mathematical notations, but the models are the same):

- *Counting Model.* A simple Counting Model serves as a comparison to define a baseline for predicting judgments of insider threat cases. This model defines the risk value as the number of indicators observed, i.e.,

  $R_j = \sum x_{ji}$,

  where $x_{ji}$ takes the value of 1 if indicator $i$ is present for person $j$, otherwise 0. Thus, if there are $n$ indicators in a case, the risk will be $n$, irrespective of any differences in threat level for individual indicators. As noted in [20], the counting model assigns equal weights to all indicators.

- *Regression Model.* A regression model is defined by estimating empirically derived weights to predict the rankings of insider threat cases. An early use of a regression modeling approach that was focused only on psychosocial insider threat indicators was described by [22] and [20]. In this model, the risk value is

  $R_j = \sum b_i x_{ji}$,

  where $b_i$ is the regression weight for indicator $i$. This approach, which freely estimates the weight of each indicator in figuring the case rankings, is labor and computationally intensive in requiring the estimates of empirically derived weights, but it provides a reasonable upper bound for the degree of fit for quantitative models.

- *Sum-of-Risk Model.* In the sum-of-risk model, the risk for a case is the sum of the ratings of concern ($r_i$) for the individual indicators contained in the case, i.e.,

  $R_j = \sum r_i x_{ji}$,

  where the $r_i$ represents the rating of concern for indicator $i$. By adding the threat values of the observed indicators, this model recognizes their variability as revealed in the rating task. Fagade and Tryfonas [13] provide a recent example of a conceptual model using an algorithm that adds weights of observed risk factors. Similarly, the present formulation specifies an additive model that incorporates indicator weights, which are determined directly from analyst judgments. By assuming that analysts judge the combination of multiple indicators as the sum of potential risk stemming from the individual indicators, the simple sum-of-risk model provides a parsimonious solution that is at least logically consistent. The sum-of-risk model is similar to the Kandias C-M-O model [30] that sums risk over a set of indicators; the C-M-O model is broader in that it covers indicators related to capability and opportunity, but the SOFIT sum-of-risk model is deeper and represents a much larger set of indicators.

In summary, consistent with the observation reported by [9] regarding the use of expert knowledge to inform BN models in cybersecurity, the insider threat modeling approaches described in this section rely to a large degree on input from experts to augment or refine models; and most of the models have been tested against data that derived wholly or in part from expert judgments. This is mainly due to the limitations and challenges already noted that are attributed to missing data and the absence of ground truth in available data sets. Even synthetic data sets that are compiled through simulations or by augmenting real data with injected data to represent target behaviors (e.g., [35]) are informed by expert knowledge. This underlines the motivation and application for the research presented here that offers an additional method for populating the probabilities in BN models of insider threat risk.

## 3 SOFIT Bayesian Model Development

Methods of detecting or predicting insider threats must provide a means of classifying individuals within an organization into high- versus low/no-threat groups. To make this determination, monitoring processes gather information on indicators of insider threats, and analysts apply a triage process to identify the high threat individuals for further investigation [16]. To facilitate and simplify the modeling effort, we focus on the main component of the triage process that implements a predictive model or algorithm transforming monitored data into threat probabilities. Our most recent work, reported here, shows that the knowledge representation in SOFIT may be implemented in a comprehensive BN representing both cyber-technical and psychosocial-behavioral indicators.

### 3.1 Modeling Approach

Simple models of expert risk judgments may be examined to gain a better understanding of the ability to estimate or predict the level of concern for cases comprising collections of indicators, such as might be observed in an organization [30] [17] [16]. The likelihood that an analyst would recommend mitigation can be considered a worthy criterion for this insider threat triage process. Those insider threat cases that experts rank higher (in level of risk or concern) should be more likely to be acted on by an organization's threat analysts than cases that are ranked lower in risk/concern.

The basis for our decontextualized predictive models includes indicators in the SOFIT ontology and expert ratings of concern for those indicators. Any predictive model classifies cases using the probability that a case with given characteristics will be in a given class (e.g., threat, non-threat) versus the probability that it will appear in any other class. We propose that experts may internalize this threat classification probability as "level of concern." In the simplest case (i.e., the presence of one observed indicator), an analyst can articulate a level of concern that an individual demonstrating a given indicator will be a threat (which is arguably easier than estimating its threat probability). Such judgments are more challenging, indeed onerous, when estimating threats of cases comprising multiple indicators: With over 200 individual indicators in the SOFIT ontology, the task of obtaining expert judgments for combinations of just 1-5 indicators would require ratings of over 2.6 billion such cases; increasing the size of cases to 10 indicators yields a much larger number of cases (2.3 x 1016) for experts to evaluate. Therefore, it would be extremely useful if, instead of requiring expert judgments of (the desired) combinations of indicators, these ratings could be estimated using models that use only the expert judgments of individual indicator threat ratings, which may be readily captured in expert knowledge elicitation studies. Our approach to model development is based on the idea that the concern rating for any combination of indicators may be *derived* from judgments of level of concern of each indicator independently, provided by insider threat experts via expert knowledge elicitation surveys. The models may then be tested by comparing their outputs with rankings of insider threat cases obtained in expert judgment surveys, which provide a means of evaluating models in the absence of ground truth.

To apply this approach to the development and testing of probabilistic models such as BN models, a mapping is needed between probabilities comprising these models and the concern ratings produced in expert judgment surveys. Determining the best equation linking independent indicator judgments to case concern is the basis and rationale for the current study. Specifically, we conducted a study to determine the functional form of the indicator judgments and applied a BN model using the functional form and basic probability assumptions. In the remainder of this section we describe two proposed BN models. Then in Section 4 we report on the empirical study used to derive the mapping equation that relates the level-of-concern judgments to probabilities, and in Section 5 we present results of the study that tested the BN models' predictions against expert judgments of sample insider threat cases.

### 3.2 Two Bayesian Models

The probability that a case is an insider threat is the ultimate criterion for the triage process. As such, specifying a probabilistic model would be operationally convenient. Bayesian classifiers are a simple way to produce probabilities from presence/absence information. These Bayes models would need to have probability information for each indicator to facilitate the calculation of probability for the hypothesis (i.e., the criterion of a Bayes model). The work we present will provide evidence that ratings of concern can be transformed into probability information to support Bayesian modeling. The efficacy of this approach depends upon the strength of relationship between insider threat risk judgments and judgments of likelihood. To assess this relationship, we conducted an expert knowledge elicitation survey focusing on likelihood ratio judgments, described in Section 4. Once having obtained positive results showing that the necessary probabilities can be estimated from risk judgments, we could pursue the development and testing of Bayesian models (as discussed in Section 5).

Even with probability information, the structure of the BN is indeterminate. Fortunately, the SOFIT ontology contains a wealth of information on the structure of insider threat indicators. Two simple structures can be derived from SOFIT: Indicator Bayes and Behavior Bayes.

- **Indicator Bayes**. We created a simple BN model where each indicator acts independently to produce the probability of insider threat. This model has about 200 nodes defined by the set of individual SOFIT indicators.

The Indicator Bayes model is the simplest form of a Bayes model that relates each indicator individually to insider threat. However, the SOFIT ontology and prior research (e.g., [16]) suggests that indicators aggregate to form behaviors, which are then related to threat. Aggregating indicators to the behavior level requires specification of conditional probability of threat given presence of the behavior (rather than indicator). This produces a second type of Bayes model:

- **Behavior Bayes**. Here we developed a BN that adds a dependency between indicators within the same behavioral domain. Namely, indicators within a domain all contribute to a common source of insider threat probability bound by the highest probability indicator within the behavior. As such, many indicators from a common behavior will provide less than additive information to the hypothesis. We were interested in whether this more complex model, which reflects the class structure, would provide a better prediction of the case judgment data.

Since the Behavior Bayes model has more than 200 nodes that correspond to the SOFIT Individual Factors, a detailed graphical display of the model is not easy to decipher; nevertheless, the complexity of the model is evident in Fig. 4. A more detailed, readable subset of the Bayes net model – exhibiting the bottom nodes of the network for the (a) Data Manipulation sub-class (which falls under the Cybersecurity Violation class) and (b) Interpersonal Problems sub-class (which falls under the Boundary Violation class) – is shown in Fig. 5.

It is worth noting that the Indicator Bayes model is a naive Bayes classifier, which is a simple classification model that assumes all features (in this case, indicators) are independent of one another, while the Behavior Bayes model allows conditional dependencies between indicators representing the same behavior. The strong assumption of the Indicator Bayes model might be particularly questioned, since one might expect higher correlations among indicators belonging to the same class. We attempted to address this issue using the Behavior Bayes implementation, which increased model complexity. Even though the more complex model captures expected dependencies between indicators, the simpler model might still provide a good accounting for the experts' judgments. Naive Bayes models have worked well in many complex real-world situations; and there is theoretical justification why, even when strong dependencies exist among nodes in the network, naive Bayes models still may perform optimally—with such dependencies canceling each other out and having limited influence on the classification [54].

The structure of the Bayes model determines which conditional probabilities must be estimated to implement the model. In general, probabilities for each state of a node, conditional on the state of its parents, must be estimated. For example, In Figure 5(b), the probability of each of the seven types of interpersonal problem must be estimated conditional on the presence or absence of interpersonal problems. With this information, the network can be used to calculate the probability of any node in the network, given information about the state of a subset of the remaining nodes. One way to obtain these probabilities is to acquire a large data set with cases that specify the values for each node in the network, including ground truth about insider threats, or in the absence of that, expert judgments as to whether or not any given case (comprising a pattern of observed indicators) should be considered as worthy of an "alert"(i.e., the expert deems the case to reflect sufficient insider threat risk to justify its flagging for further analysis). This type of data set would be difficult to obtain and would likely come from highly sensitive operational information. Moreover, people tend to overestimate probabilities of rare events (e.g., [34]) and small probabilities generally, as examined and described by Prospect Theory [29]—such biases have been documented and explained by invoking the availability heuristic [49]. In lieu of ground truth or reliable probability estimates, we conducted a study to determine if it is feasible to map expert judgments of level of concern ($r_i$) that have been obtained in several studies to independent expert judgments of likelihood. That is, we hypothesized that level of concern ratings reflect a likelihood ratio judgment, which may be considered to be a transformation of the conditional probability. The next section describes our test of this hypothesis.

## 4 Likelihood Judgment Survey to Inform Bayes Models

We conducted a new expert knowledge elicitation survey (Likelihood Judgment Survey) to serve as a basis for translating judgments from the earlier expert knowledge surveys (described in the Background Section 2.3 and in [17]. The objective of the Likelihood Judgment survey was to test whether there is a correspondence between the level of concern judgments $r_i$, used in our research and likelihood or probability metrics used in probabilistic models (e.g., Bayesian models, such as described by [2]. The logic behind this approach is the idea that concern reflects an internal representation of the degree to which a given indicator suggests a possible insider threat. For a strong indicator of insider threat, the probability that the indicator would appear in a threat population is very high and the probability that the indicator would appear in a normal population is relatively low. Thus, the likelihood ratio ($L_i$) for this indicator, as generated from these probabilities would be high. As concern decreases, the indicator may be less likely in the threat population, more likely in the normal population, or both—which produces a lower likelihood ratio. This study tested to see if there is a correlation between the $r_i$ and $L_i$ measures. If we find that there is a functional relationship between $r_i$ and $L_i$, then it will be possible to map the ratings of level of concern to probabilities and probabilistic models. Furthermore, because the functional

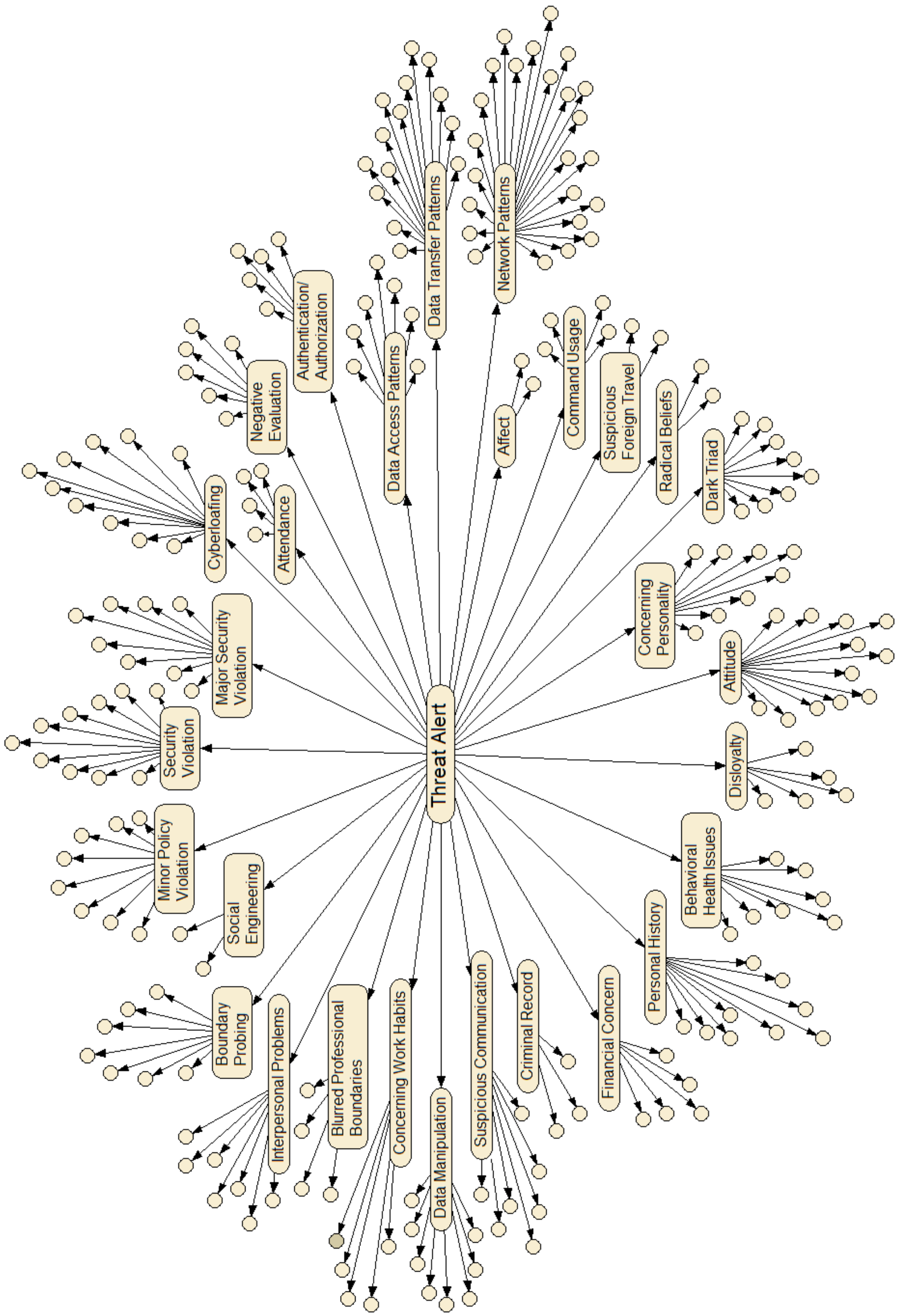


Figure 4: Behavior Bayes network model with nodes corresponding to the SOFIT Individual Factors.

form of the relationship is likely a property of the expert judgments that is consistent across indicators, this allows us to discover the functional form using a sample of indicators, and then to apply the derived function to all other indicators. Importantly, applying a function to transform level of concern ratings into probabilities creates an opportunity to implement probabilistic models making use of the level-of-concern ratings obtained in our surveys.

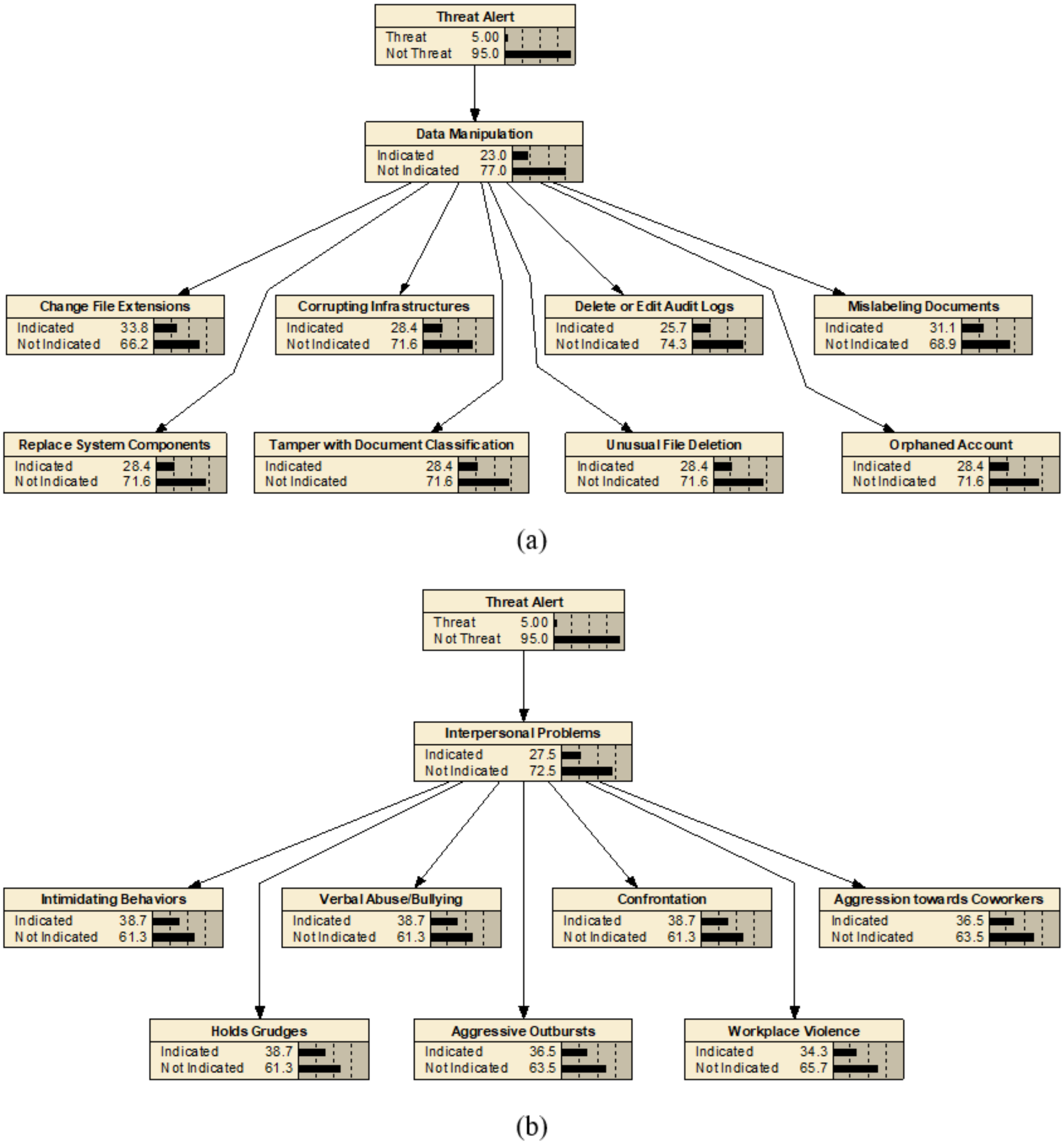


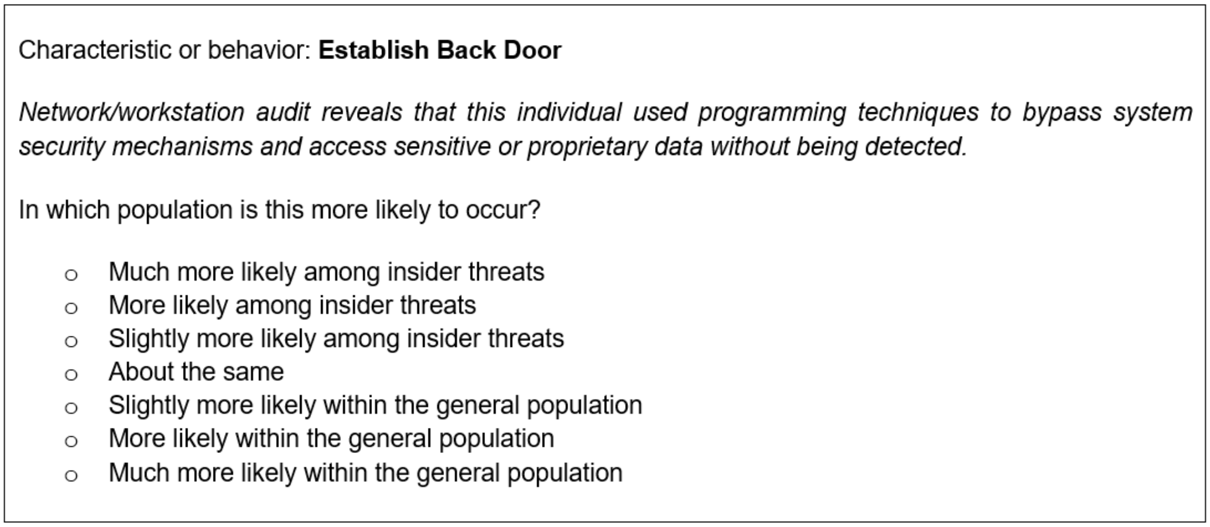
Figure 5: Detail for two portions of SOFIT Individual Factors Bayes network model showing nodes for sub-classes (a) Data Manipulation and (b) Interpersonal Problems.

## 4.1 Method and Procedure

This survey investigated whether judgments of concern (i.e., the "Level of Concern" or "risk score" metric) were related to likelihood ratio judgments. Likelihood ratio judgments were obtained through expert evaluations of the likelihood of indicators occurring in insider threat populations versus the general population. The survey was approved by the Human Resources Research Organization (HumRRO) IRB and conducted in January-February 2019. Participants were recruited using email and online solicitation of experts in research/operational communities. More than 100 such recruitment emails were sent, in addition to posting an announcement on LinkedIn.

The survey presented 25 questions that asked for a judgment comparing the likelihood of an indicator being observed for an insider threat versus a member of the general population of workers in an organization. Responses were requested using a seven-point scale ranging from "Much more likely among insider threats" to "about the same" to "Much more likely within the general population." Of the 25 questions, 20 questions concerned indicators that represent some degree of concern for insider threat. To avoid predisposing the respondent to always slant answers toward insider threat populations, we inserted five questions that represent positive characteristics that should not be expected to be associated more strongly with insider threats. The five positive observables were: (a) received a promotion, (b) received a certificate of appreciation, (c) received a significant raise/salary increase for superior performance, (d) received a bonus for superior performance, and (e) received recognition for outstanding mentoring. These positive instances could potentially ameliorate or counterbalance negative indicators in a model of insider threat; however, since our models do not attempt to address potential mitigating events, these five questions were not included in the analysis of survey results.

It was not possible to include all SOFIT indicators in the survey, since this would impose an undue burden on expert participants. We selected twenty indicators across the five SOFIT indicator classes (representing about one-tenth of the individual insider threat indicators defined in SOFIT) based on the criteria that (a) the indicator and its brief description would be readily understood without further explanation, and (b) the selected set would reflect insider threat risk factors/constructs from the full range of SOFIT indicator classes. A sample question is shown in Fig. 6. The complete Likelihood Ratio Survey is provided in Appendix A.



Characteristic or behavior: **Establish Back Door**

*Network/workstation audit reveals that this individual used programming techniques to bypass system security mechanisms and access sensitive or proprietary data without being detected.*

In which population is this more likely to occur?

○ Much more likely among insider threats
○ More likely among insider threats
○ Slightly more likely among insider threats
○ About the same
○ Slightly more likely within the general population
○ More likely within the general population
○ Much more likely within the general population

Figure 6: Excerpt from Likelihood Survey.

Table 1 lists the indicators used in the survey, grouped by SOFIT indicator class. Three to six indicators were chosen from each indicator class. In SOFIT, the number of indicators comprising each class varies widely with the greatest number defined for Cybersecurity Violations (well over 500 indicators and observables) and two to three dozen indicators defined within each of the other four classes. Considering representation as a percentage of indicators selected to indicators defined within a class, the six indicators selected from the Cybersecurity Violation class represent about 7% of the large set of cybersecurity indicators; the indicators selected for Boundary Violations and Job Performance classes represent about 10% of associated SOFIT indicators; and indicators selected for Psychological and Life Narrative classes represented 13-14% of their respective indicator classes. To interpret these percentages, it is useful to note that many indicators within SOFIT subclasses are similar—for example, the Boundary Violations class includes the subclass, Blurred Professional Boundaries, which in turn includes the indicator Interpersonal Problems, which is further broken down into observables such as Aggression and Intimidation, etc. Because such subtle variations at the lower level nodes in the hierarchy are common, the representativeness of the selected indicators is greater than what might be inferred merely from the above counts or percentages. An alternative conceptualization is to compare the overall threat values associated with the selected indicators with the overall threat associated with their respective classes: For example, the sum of threat values for the seven subclasses of Boundary Violations is 455, and the sum of the threat values for the three selected indicators within this class is 177—this represents 39% of the threat values associated with Boundary Violations. Similarly, corresponding proportions for the classes Job Performance, Psychological Factors, and Life Narrative factors range from 37% to 48%. The threat coverage percentage for the selected Cybersecurity Violations class is 62%. Regardless of whether we assess representativeness using simple counts or the more complex accounting that relies on estimated threat values, we are confident that the selected indicators provide sufficient breadth in covering SOFIT indicators to allow generalization of the findings to other indicators in the SOFIT ontology.

Table 1: SOFIT Indicators (Organized by Class) Used in Likelihood Survey.

**Boundary Violation**
- Lack of confidentiality
- Unauthorized contact with foreign nationals
- Unauthorized copying of classified material

**Job Performancen**
- Cyberloafing
- Excessive Absences
- Negative performance evaluation

**Psychological**
- Disgruntled
- Low honesty-humility
- Manipulative
- Poor time management

**Cybersecurity Violation**
- Attempts access to prohibited file sharing sites
- Delete audit logs
- Establish backdoor
- Unusual file deletion
- Receives unusually large attachments in emails
- Simultaneous use of multiple printers

**Life Narrative**
- Change in marital st
- Passed over for promotion
- Terminated
- Unexplained affluence

## 4.2 Results

When the survey was closed, we had obtained ratings from 20 experts representing over 275 years of experience collectively, 18 of whom completed the entire survey (90%). Because the questions are independent, even incomplete surveys provided ratings that could be included in the analysis. Additionally, as noted earlier, the number of subject matter experts matters less to the resulting analysis than the number of cases included in the study. Ages of respondents ranged from 30 to over 70 years old, with 50% falling in the range of 40-60 years old. Most respondents held advanced degrees of PhD (61%) and Masters (28%), and predominant fields of study included computer and information sciences, psychology, and mathematics/statistics. Respondents' background and expertise included fields of social/behavioral sciences (61%), intelligence/counterintelligence (44%), information assurance/cybersecurity (44%), computer science (33%), and personnel security (28%)—note that these percentages total more than 100% because individuals could report expertise in multiple categories. Sixty-one percent of respondents reported more than 20 years of work experience in their fields, and 28% reported between 11-20 years of work experience. Thirty-nine percent of respondents identified their positions as government contractors; 28% were in commercial/private industry; 22% were associated with academic institutions; 5% were employed in government agencies. Eleven respondents identified themselves as researchers (61%) and seven identified themselves as practitioners (39%).

To assess the results of this survey, the mean likelihood scores (Li values ranging from -3 to +3) where compared to the mean level of concern scores that were obtained in the 2018 survey (ri values ranging from 0-100). The results are shown in Fig. 7, which plots the mean likelihood scores against the mean level of concern scores. The test of the predicted direct relationship between these two measures was accomplished by computing their correlation. As hypothesized, there was a significant positive correlation between the mean ratings of concern and the mean likelihood ratio ratings: $r = 0.87$ (p<0.05); the linear $R^2$ = 0.762 indicates that this linear relationship accounted for 76% of the variance in predicting the $L_i$ values from $r_i$ values. [Note that the correlation was obtained for 19 of the 20 indicators because there was not an available $r_i$ value for the indicator, passed over for promotion]. This result provides strong support for the proposed direct relationship between these measures (i.e., level of concern scores follow the pattern of likelihood scores) and allows us to conclude that one can derive, using the relationship of likelihood ratio to probability, reasonably accurate probability estimates from the 200+ level of concern values that were obtained for the indicators examined in our prior studies.

These results provide a mapping between the level of concern scores and probabilities that may be implemented within a BN with the SOFIT indicators implemented as nodes within the network. For completeness, we tested other possible (nonlinear) relationships and found the variance accounted for to be marginally higher than the linear model. The results imply that level of concern judgments close to zero may be inflated due to difficulties conceptualizing low/no risk. However, the functional form of a quadratic relationship implied a particularly desirable characteristic (i.e., an intercept close to zero). As such, all 200+ level of concern ratings were converted to comparable likelihood ratio scores via the following quadratic formula:

(1) $L_i = 0.0003\, r_i^2 + 0.0023\, r_i + 0.0083$

The resulting ratings on a -3 to 3 scale represent the likelihood ratio judgment expected given prior expert ratings of concern. Because only concerning indicators were used in the analysis, no values for the likelihood scores were lower than zero (same probability in the insider threat and normal populations). The arbitrary scale does not accurately represent likelihood ratios that may range from zero to infinity. To extend the range of values and account for the idea that higher values on the scale likely represent greater increases in threat, the antilog transformation was applied to this scale. Finally, the resulting likelihood ratios were converted to conditional probabilities. This step is based on the fundamental property of likelihood ratios:

(2) $L_i = p(I)\ /\ (1 - p(I))$

Solving the above for $p(I)$ allows us to convert from likelihood ratio to probability via the following formula:

(3) $p(I_i|A) = L_i \;/\; (1+L_i)$

This formula, which specifies the probability that Indicator Ii is present, given the truth of the hypothesis A that the individual is an insider threat, assumes that the conditional probability table is symmetric; that is, $p(I_i|A) = 1 - p(I_i| \sim A)$. This approach represents the likelihood ratio without specifying the indicator probabilities within either the threat or non-threat populations. These probabilities reflect the relative likelihood that a case with a specific indicator represents an insider threat. This conditional probability is the foundation for the BN models defined in Section 3.2.

Appendix B lists the derived likelihood ratios and conditional probabilities for a large selection of SOFIT indicators, computed using the above equations (1) and (2) and the associated level-of-concern ratings that were previously obtained by Greitzer et al. (2018) and documented in the SOFIT ontology (provided in Supplement I).

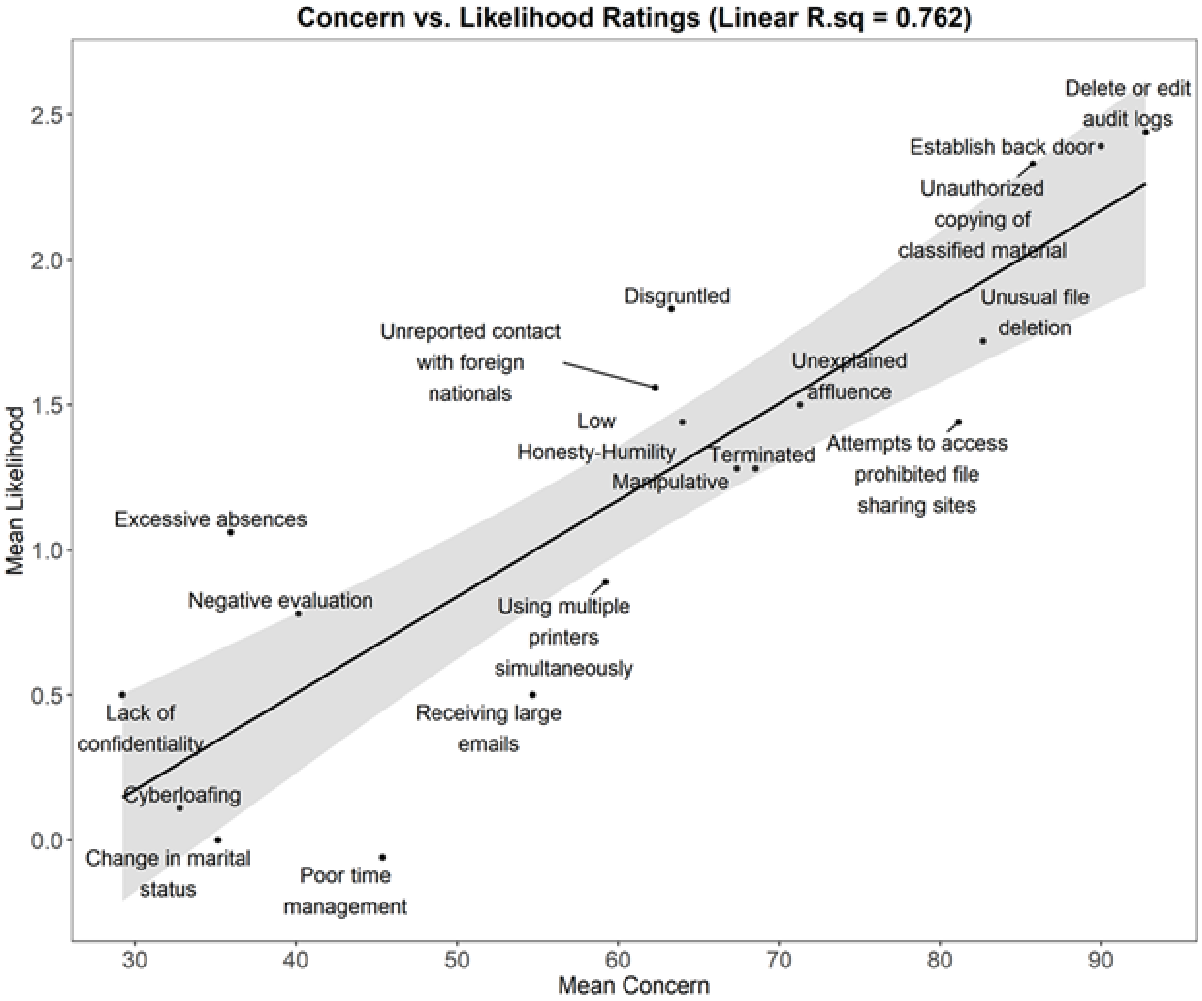


Figure 7: Scatter plot of Mean Concern and Mean Likelihood Ratings.

# 5 Specifying and Testing Bayesian Models

The probability that an employee who poses a threat also exhibits a given indicator, compared to the probability of that indicator within the population provides a fundamental assertion about the predictive strength of the indicator. Accumulating these probabilities across indicators provides an assessment of the threat potential for any employee being monitored. The potential for threat assessment with a probabilistic model is promising, and the likelihood transformation applied to level of concern ratings has provided the underlying probabilities. As a preliminary test of the validity for the probabilistic model, the threat assessment resulting from a Bayes model was used to rank order cases previously ranked by experts in the 2018 survey study [16]. Given the strong relationship between the threat level-of-concern ratings and likelihoods, it is very straightforward and feasible to compute probabilities that can populate a Bayes model that may be developed from the SOFIT ontology.

## 5.1 Implementation of Bayesian Models

In general, a BN uses observations to inform a decision by estimating the probability that a hypothesis is true, conditional on the observations. For example, we may decide to issue an "alert" if the indicators that are present in the case imply that the probability that an insider is threatening is greater than a criterion value. When we built the BN, we began by specifying the conditional probability that an indicator is present, given the truth of the hypothesis that an employee represents a threat, $(p(I_i|A)$, based on the expert judgments of level of concern, and on Equations (1) and (2). The probability of no indicator, given that the hypothesis is true is the complement of $p(I_i|A)$. Assuming the symmetry of the conditional probability table (CPT), we also set $p(I_i| \sim A)$ to be $1 - p(I_i|A)$, so that the CPT reflects the likelihood ratio. Finally, $p(\sim I_i| \sim A)$ would be equal to the complement of $p(I_i| \sim A)$, which is $p(I_i|A)$. These probabilities form a conditional probability table (or CPT), as shown in Table 2. The mathematics of the model then allows for the calculation of the probability of the truth of the hypothesis, given any combination of indicators, $p(A|I_i)$.

Table 2: Example of a Conditional Probability Table.

| | | **OBSERVED:** | |
|---|---|---|---|
| | | **Indicator *i* Present ($I_i$)** | **Indicator *i* Absent ($\sim I_i$)** |
| **HYPOTHESIS:** | **True (A)** | Conditional probability of indicator present, given hypothesis is true<br>$p(I_i \mid A)$ | Conditional probability of no indicator, given hypothesis is true<br>$p(\sim I_i \mid A)$ |
| | **False (∼A)** | Conditional probability of indicator present, given hypothesis is false<br>$p(I_i \mid \sim A)$ | Conditional probability of no indicator, given hypothesis is false<br>$p(\sim I_i \mid \sim A)$ |

Derivation of CPTs for all 200+ indicators is required to produce a fully specified belief network that reflects the information specified in the SOFIT ontology. The resulting Bayesian belief network is a framework that takes as input any specified pattern of observed indicators, and then, using the associated conditional probabilities, produces as output a marginal probability that an alert is advisable. The following procedure is followed in specifying the model:

1. Obtain ratings reflecting the level of concern (0-100) for relevant indicators
2. Use equation (1) above to convert ratings using ri to Li

3. Apply antilog transformation using base with parameter a
4. Convert Li to p(Ii|A) using equation (2), above
5. Populate each CPT with probabilities:
    a. $p(I_i|A) = p(I_i|A)$
    b. $p(\sim I_i|A) = 1 - p(I_i|A)$
    c. $p(I_i|\sim A) = 1 - p(I_i|A)$
    d. $p(\sim I_i|\sim A) = p(I_i|A)$

Additional procedures were used for the Behavior Bayes model:

6. Add parent nodes for each group of indicators specified in SOFIT (e.g., Security Violation)
7. Populate the CPT for the parent node with parameter b and 1 - b; the reverse for absence.

The base of the antilog **(a)** and the conditional probabilities for behaviors **(b)** can be varied to determine the values that optimize the prediction of the rank data. Otherwise, the selection of these values would be arbitrary. To support optimization, the Indicator Bayes model was nested within the Behavior Bayes model, where the Indicator Bayes is equal to the Behavior Bayes with behavior conditional probabilities of unity.

The antilog base **(a)** was first narrowed to a value of two by varying the antilog by units of one from two to ten. Then, values from 2.00 to 2.25 were used in increments of 0.01 to optimize the values further. The optimal value was determined to be 2.18, with a resulting optimal correlation between the Indicator Bayes and case rankings of r = 0.7297.

The behavior conditional probabilities **(b)** were specified in terms of the probability that the behavior is observed given that the case is a threat. Logically, the probability for the behavior should be equal to or greater than the likelihood for the most likely indicator for a behavior. That determined a range of values for our optimization spanning from the maximum child indicator likelihood to unity. The percentage of the distance between this maximum and unity was used as a consistent metric across behaviors. For example, if the maximum probability is .6, we evaluated .6 as 0% of the remaining range and unity as 100% of the remaining range. We tested between 0 and 100 percent of the remaining range in units of .01 percent. The optimal value was determined to be 80% of the remaining value, with a resulting optimal correlation of r = 0.7349. This value is very similar to that obtained through optimization of the Indicator Bayes model. The Behavior Bayes model, with optimized behavior conditional probabilities less than unity, provides a slight improvement over the indicator model.

## 5.2 Testing the Bayesian Models

The BN models were used to predict the rankings of insider threat cases in the sorting task described in Section 2.3 and in [16 and [16]. As described in [17], the sorting task produced a ranking of all cases ranging from 1 (for the case with highest concern) to 45 (for the case with lowest concern). This set the stage for the testing of proposed models to determine their ability to predict the rankings from individual indicator risk scores. The results are summarized in Table 3 and discussed in the following subsections.

The results demonstrate that the predictive power of a Counting model (using only the number of indicators observed) provides a logical lower bound on our measure of predictive strength ($R^2$ = 0.26); whereas a Regression model freely estimating the weight of each indicator on rank provides a logical upper bound ($R^2$ = 0.76). The Sum-of-Risk model predicted the rankings nearly twice as well as the Counting model ($R^2$ = 0.48).

The Likelihood Judgment survey demonstrated the correspondence between the threat severity (Level of Concern) metric $r_i$ and likelihood ratings $L_i$ that may be used to derive probabilistic models. Indeed, we showed that a BN model based on the SOFIT indicators, constructed using the derived probabilities,

Table 3: Variance in Case Ranks Explained by Alternative Models.

| Model[a] | $R^2$ |
|---|---|
| Counting Model | 0.26[b] |
| Sum-of-Risk Model | 0.48[b] |
| Regression Model | 0.76[b] |
| Indicator Bayes Model | 0.53 |
| Behavior Bayes Model | 0.54 |

[a] Testing of all models was based on the data from the Greitzer et al. [17] study described in Section 2.4.2. Development and testing of the Bayes models reported here extends the original analysis.
[b] Results reported in [17].

performed slightly better than the Sum of Risks model in predicting rankings of cases in the 2018 survey. More specifically, we examined two implementation of Bayes models, one (Indicator Bayes) reflecting the large set of individual indicators, and a second (Behavior Bayes) that, in addition to the individual indicators, incorporated nodes that reflect the class structure. Both models yielded predictions of case rankings that correlated well with the observed rankings, and the Behavior Bayes model, which more strongly reflected the SOFIT class structure, yielded a slightly higher correlation.

# 6 Discussion and Conclusions

## 6.1 Summary of Findings

A major challenge in developing BN models is assigning probability values to evidence, which is usually based on subjective personal beliefs. To instantiate the model, developers often rely on expert judgments of their degree of belief for various base-rate (prior) and conditional probabilities, a difficult process that is often impractical or fraught with bias or errors. The focus of the present study was on the use of an expert knowledge elicitation method that can inform this probability estimation process without relying on probability judgments.

As background for the present work, we reviewed several non-probabilistic and probabilistic insider threat assessment models. To support the development of two BN models based on SOFIT, we conducted a survey to determine if expert ratings of threat/risk level (level of concern), which have been used previously in developing threat assessment models based on SOFIT, are correlated with corresponding likelihood ratios. Having found a high correlation between the previously obtained threat ratings and the likelihood ratios observed in the present study, we were able to use this information to derive conditional probability tables and construct BN models reflecting the large set of $\sim 200$ indicators defined in the SOFIT ontology. Applying the BN models to the previously obtained case-sorting data (reported in [17] and described in the Background Section 2.4) revealed that the best Bayesian model performed slightly better than the previously-best-performing Sum-of-Risk model, accounting for 53% of the variance in the data predicting the expert judgments.

Considering the counting model as a minimum or baseline and the linear regression model as a maximum or best possible fit to our judgment data, the relatively large improvements provided by the sum-of-risk model and the Bayes models provide some support for our underlying assumptions. Specifically, classifying insider threat into indicators and providing a framework (i.e., SOFIT) for categorizing indicators contributes substantially to explaining expert judgments. Although the improvement in variance accounted for in the Behavior Bayes model was minimal, the optimal **b** parameter was less than

100% implying that accounting for the SOFIT class structure may be important for explaining expert judgments. The present study only tested one possible representation of the class structure of the ontology and allowed only one parameter to determine the overlap in explanatory power among indicators. Given these limitations, a small improvement is encouraging. Altogether the results suggest that further development of insider threat models informed by expert judgments will continue to be productive.

The research approach described in this paper is not meant to replace threat detection based on ground truth, but rather to expand the conversation to include structured elicitation of threat analysts' judgments of insider threat. Our method produces information that is well-suited for integration into data-driven models of insider threat. Without baseline data for non-threat conditional probabilities an assumption was made that the non-threat probabilities would mirror the threat probabilities. This assumption is unlikely to hold in real data and our prediction of expert judgments may have improved with and investigation into non-threat probabilities. Within the organizational context, non-threat baseline data is abundant. Simply pairing data driven baseline information with our expert judgments of conditional probabilities given threat might provide a powerful initial model for an inference enterprise. Specifically, we recommend specifying $p(I_i| \sim A)$ and $p(\sim I_i| \sim A)$ using baseline (i.e., non-threat) organizational data, and then using the likelihood ratios from SOFIT to estimate $p(I_i|A)$ and $p(\sim I_i|A)$. Our hope is to demonstrate the potential for expert judgments to augment inference enterprise models of insider threat.

## 6.2 Contributions

This study investigated the relationship between risk judgments and probability judgments to support cross-fertilization of findings and ontology details with probabilistic insider threat modeling approaches. The results demonstrate that the method of acquiring expert judgments of level of concern to support insider threat assessment may be applied not only to the development of non-probabilistic risk-based models—exemplified in the recent works by Greitzer and colleagues (e.g., [17], [16]) and potentially to related works by Kandias et al. [30]—but also to probabilistic-based BN modeling approaches (e.g., [2], [7], [20], [31], [6], [51]) that are faced with the difficulties in estimating conditional probability tables for threat indicators with very low base rates, and for which reliable probability estimates are difficult to obtain. By providing a practical alternative means of likelihood estimation, these methods may also serve to enhance and extend insider threat models that have focused only on technical indicators. More broadly, since Bayesian modeling approaches have been widely deployed to integrate expert judgments into predictive analytics for inference enterprises, the method demonstrated here for acquiring BN conditional probabilities from expert judgments that do not rely on probability estimates can help overcome biases evident in judgments of probability. Future research could compare transformed Level of Concern ratings to directly estimated probabilities to more clearly delineate the advantages of each approach.

These results also provide further support for and demonstrate the utility of integrating an array of (massive, streaming) technical data and (relatively sparse) behavioral data into a comprehensive insider threat monitoring approach. As noted in [17], insider threat risk assessment is fraught with ambiguity and uncertainty: It is difficult to distinguish malicious cyber/technical activities from normal activities. Observation of precursor, psychosocial indicators that occur weeks or months prior to the actual incident [45] [44] can help to disambiguate an uncertain threat profile. If informed by such behavioral indicators, this initial stage of the threat assessment process provides a triage analysis that produces alerts for the most concerning cases before they have a chance to cause damage [17] to achieve a proactive insider threat mitigation strategy. Therefore, the findings reported here should encourage threat assessment model and tool developers to incorporate behavioral indicators in addition to technical indicators: The methods described for deriving probability estimates from judgments of severity for insider threat indicators help to address the difficulties in populating models with probability estimates for behavioral variables. This will facilitate the development of more effective, comprehensive insider threat monitoring applications that extend beyond the typical cyber data sources ([32], [25], [23]).

The insider threat domain is cumbersome to model because of the large number of concepts and their complex interrelationships. A major contribution of this research is outlining an approach to take advantage of information contained in an ontology to specify the structure of predictive models. By starting with estimated threat/risk values of individual indicators housed in a structured ontology, we can incrementally improve the fit of the model to expert judgments. Our finding that the optimal Behavior Bayes b parameter differed from 100%, given the great potential for further optimizing the structure, strongly implies that the structure of the Bayes model has the potential to improve prediction. This finding is partially in contrast with findings that structural dependencies cancel out [54]. Refining the structural components of the SOFIT ontology through operational use is strongly recommended. Given the success of this approach to developing ontology-informed BN models that are elicited using an intuitive level-of-concern scale, we suggest that the procedure may be applied to any indicators using any sufficiently informed group of experts.

## 6.3 Limitations

The principal limitation of this research is the lack of real data and ground truth that are needed to test the insider threat models examined in the studies reported here. The lack of appropriate data sets that represent "real world" cyber activities and behaviors was identified as one of the "capability gaps" responsible for placing the insider threat problem second in a list of eight "hard problems" identified in a 2005 Information Security (INFOSEC) Research Council Hard *Problems List* [28]. While there have been reference data sets established for certain research programs, the problem persists to this day because even these data sets are extremely limited in scope. This limitation—which is evident in virtually all other efforts to develop predictive models of insider threat risk—applies especially to the lack of appropriate human-behavioral data (which would be available from diverse sources such as human resource, security, and other such repositories). The present study, therefore, is not meant to replace the development of threat detection models based on ground truth, but rather to serve an intermediate objective that uses expert judgments as proxies for empirical data and ground truth. Expert estimates of threat values for individual indicators, when informed by relationships defined in the ontology, may be used to predict threats of cases comprising collections of indicators. Tests of the models seek to assess the "fit" between the model predictions and expert judgments. The development of threat models based on this research will help to identify formulations that best match expert judgments, and which, we assume, will better position the model-development community for testing the best-performing models against real data and ground truth. This line of research would benefit from future comparative studies, if/when appropriate data become available, to validate the expert judgments obtained here against ground truth.

A second limitation of the survey studies reported here concerns the additional error variance in the estimates stemming from the limited number of respondents. The inflated error variance due to estimates being derived from a small sample is evident in the regression model accounting for only 76% of variance. As such, we caution that the performance of any model should be evaluated as it falls between the counting model and regression model rather than being compared for absolute variance accounted for. In any case, the estimates do seem sufficiently accurate to provide a proof of concept. While we contacted around 100 professionals across several relevant disciplines, the 20% return rate, although typical of survey studies, was less than desired. Nevertheless, we obtained sufficient sample sizes to produce statistically significant results. Moreover, a perusal of research literature relating to knowledge-based model development and forecasting suggests that sufficient accuracy can be obtained with 6-10 experts (e.g., [26], [53]). Such investigations typically have fewer than 20 experts—a dissertation on expert judgment [14] provided a table showing the number of experts used in 66 studies from expert knowledge elicitation literature in which 44 of the studies (two-thirds) used fewer than 20 experts—and the median number of experts was 12 (Appendix B in Forrester, 2005). Another study that used Delphi-based expert knowledge elicitation methods to estimate a BN prior probability distribution had only four experts [40], noting that the representativeness of the expert panel is more a function of its quality than its size [39]. More relevant to the present research, our sample size of 20 experts compares favorably to the similar BN development efforts ([22] [2] [30]) described in Section 2.4.1.

Thirdly, our goal has been to find the best model to represent the underlying clinical model that experts employ when evaluating insider threat. The expert clinical model may differ from the empirical model relating clinical judgment to actual threat probability. Because there has been no opportunity to test models against actual data or ground truth, our approach (like others described in the Background section) has been to use other information sources as proxies for such data—in this case, we use expert judgments acquired through expert knowledge elicitation studies. While our models reflect expert judgments rather than operational data, we believe that these models provide a key piece of the overall inference enterprise which would optimally use an empirical model to alert threats to an expert who ultimately decides the organizational action. Approaches that are based on enterprise data but lack ground truth, and models tested against synthetic/simulated data (whose creation largely relies on expert judgment) provide complementary information focused earlier in the inference enterprise workflow. We expect that any models shown to do a good job in predicting expert judgments of insider threats will be in a leading position for future validation studies that use operational data and ground truth.

## 6.4 Future Research

The present research helps to advance efforts to model and mitigate insider threats. Informed by extant research on human factors associated with insider threats, the constructs and indicators represented in the SOFIT ontology can be used to develop models to assess individual risk, as well as to inform operational risk management practices. We have postulated that a simple risk model could be informed by the structure of the SOFIT ontology and the estimated risk weights or scores obtained through expert knowledge elicitation exercises. We showed how computational models—including a probabilistic, Bayes model—could be derived and used by organizations to objectively identify cases that are deemed most likely, or to be at the highest percentile of risk, and for whom additional monitoring and analysis may be advised. This provides an objective means of focusing limited resources on the most at-risk individuals. Since the models that we examined accounted for slightly more than half of the variance in the expert judgments of insider threat risk, there is room for improvement: For example, a study could be designed to replicate the case judgment findings of Greitzer and colleagues [17] [16] with an aim to better distinguish between the Indicator and Behavior Bayes models by comparing rankings of cases that contain multiple, strong risk indicators of the same classes of behavior with those of cases containing indicators from different classes. More generally, future research may benefit from examining more complex relationships, such as those that may be reflected in meaningful patterns of indicators, either accounting for recognizable attack vectors, or possible temporal/sequential relationships, or both.

We suggest that the following lines of research will be useful in advancing these findings:

1. Continue research to solidify the foundation provided in the SOFIT ontology class structure and relationships among constructs. Conduct research to examine additional relationships among factors, such as exploration or specification of critical patterns that may improve threat detection beyond the level of performance attained by approaches focused merely on the contribution of individual indicator risk to the threat assessment process.
2. Continue research to refine/validate predictive models using actual incident/outcome data to move beyond validation approaches that use expert judgments as proxies for real data and ground truth.

3. Conduct applied research to develop and validate operationally oriented tools for assessing individual insider threats.

Another implication and potential application of our expert knowledge elicitation exercise is the possibility of using the ontology structure, along with estimated subclass risk estimates, to help determine which of several proposed indicators might be most effective. In other words, the knowledge base provided here may have further operational impact by informing the structure of data to be captured by enterprises for effective insider threat monitoring and analysis. Therefore, another suggestion for advancing this research is:

4. Conduct applied research to develop tools and metrics for evaluating the effectiveness of an organization's insider threat monitoring approach, which will foster self-assessment and facilitate the development of more mature and effective insider threat programs.

## Acknowledgments

The research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract 2016-16031400006. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. This research was reviewed and approved by the HumRRO IRB (Memorandum for the Record dated September 1, 2017). The authors wish to acknowledge the contribution of collaborators Dr. Yung-Mei Leong and Dr. D. E. (Sunny) Becker in the initial development of the SOFIT ontology; we thank Dr. Dennis Buede and Dr. David P. Brown (Innovative Decisions, Inc.) and Dr. Kathryn B. Laskey and Dr. Abbas Zaidi (George Mason University) for their support and oversight.

## References

[1] R. H. Anderson, T. Bozek, T. Longstaff, W. Meitzler, and M. Skroch. Research on mitigating the insider threat to information systems-# 2. Technical report, Rand National Defense Research Inst Santa Monica CA, 2000.

[2] E. T. Axelrad, P. J. Sticha, O. Brdiczka, and J. Shen. A bayesian network model for predicting insider threats. In *Proc. of the 2013 IEEE Security and Privacy Workshops (SPW'13), San Francisco, California, USA*, pages 82–89. IEEE, May 2013.

[3] A. Azaria, A. Richardson, S. Kraus, and V. Subrahmanian. Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data. *IEEE Transactions on Computational Social Systems*, 1(2):135–155, June 2014.

[4] S. R. Band, D. M. Cappelli, L. F. Fischer, A. P. Moore, E. D. Shaw, and R. F. Trzeciak. Comparing insider it sabotage and espionage: A model-based analysis. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST, 2006.

[5] R. C. Brackney and R. H. Anderson. Understanding the insider threat. proceedings of a march 2004 workshop. Technical report, RAND CORP SANTA MONICA CA, 2004.

[6] D. P. Brown, D. Buede, and S. D. Vermillion. Improving insider threat detection through multi-modelling/data fusion. *Procedia Computer Science*, 153:100–107, 2019.

[7] D. M. Buede, E. T. Axelrad, D. P. Brown, D. W. Hudson, K. B. Laskey, P. J. Sticha, and J. L. Thomas. Inference enterprise models: An approach to organizational performance improvement. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6):e1277, 2018.

[8] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak. *The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud)*. Addison-Wesley, 2012.

[9] S. Chockalingam, W. Pieters, A. Teixeira, and P. van Gelder. Bayesian network models in cyber security: a systematic review. In *Proc. of the 22nd Nordic Conference on Secure IT Systems (NordSec'17), Tartu, Estonia*, volume 10674 of *Lecture Notes in Computer Science*, pages 105–122. Springer, November 2017.

[10] E. Cole and S. Ring. *Insider Threat Protecting the Enterprise from Sabotage*. Syngress Publications, 2006.

[11] D. Costa. Cert definition of 'insider threat'—updated, March 2017. `https://insights.sei.cmu.edu/insider-threat/2017/03/cert-definition-of-insider-threat---updated.html` [Online; accessed on June 22, 2021].

[12] D. L. Costa, M. L. Collins, S. J. Perl, M. J. Albrethsen, G. J. Silowash, and D. L. Spooner. An ontology for insider threat indicators development and applications. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST, 2014.

[13] T. Fagade and T. Tryfonas. Malicious insider threat detection: A conceptual model. *Security and Protection of Information 2017*, pages 31–44, 2017.

[14] Y. Forrester. *The quality of expert judgment: an interdisciplinary investigation*. PhD thesis, University of Maryland, December 2005.

[15] M. Gelles. Exploring the mind of the spy, 2005. `http://www.dss.mil/search-dir/training/csg/security/Treason/Mind.htm` [Online; accessed on June 22, 2021].

[16] F. Greitzer, J. Purl, D. Becker, P. Sticha, and Y. M. Leong. Modeling expert judgments of insider threat using ontology structure: Effects of individual indicator threat value and class membership. In *Proc. of the 52nd Hawaii International Conference on System Sciences (HICSS'19), Maui, Hawaii, USA*, pages 1–10. University of Hawaii at Manoa, Association for Information Systems IEEE Computer Society Press, January 2019.

[17] F. Greitzer, J. Purl, Y. M. Leong, and D. S. Becker. Sofit: Sociotechnical and organizational factors for insider threat. In *Proc. of the 2018 IEEE Security and Privacy Workshops (SPW'18), San Francisco, California, USA*, pages 197–206. IEEE, 2018.

[18] F. L. Greitzer. Insider Threats: It's the HUMAN, Stupid! In *Proc. of the 2019 Northwest Cybersecurity Symposium (NCS'19), Richland, Washington, USA*, pages 1–8. ACM, April 2019.

[19] F. L. Greitzer and D. A. Frincke. Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation. In W. P. Christian, H. Jeffrey, G. Dieter, and B. Matt, editors, *Insider threats in cyber security*, volume 49 of *Advances in Information Security*, pages 85–113. Springer, Boston, MA, July 2010.

[20] F. L. Greitzer, L. J. Kangas, C. F. Noonan, C. R. Brown, and T. Ferryman. Psychosocial modeling of insider threat risk based on behavioral and word use analysis. *e-Service Journal: A Journal of Electronic Services in the Public and Private Sectors*, 9(1):106–138, 2013.

[21] F. L. Greitzer, L. J. Kangas, C. F. Noonan, and A. C. Dalton. Identifying at-risk employees: A behavioral model for predicting potential insider threats. Technical Report PNNL-19665, Pacific Northwest National Lab.(PNNL), Richland, Washington, USA, 2010.

[22] F. L. Greitzer, L. J. Kangas, C. F. Noonan, and A. C. Dalton. Identifying at-risk employees: A behavioral model for predicting potential insider threats. In *Proc. of the 45th Hawaii International Conference on System Sciences (HICSS'10), Maui, Hawaii, USA*, pages 1–10. IEEE, January 2010.

[23] F. L. Greitzer, J. Purl, Y. M. Leong, and P. J. Sticha. Positioning your organization to respond to insider threats. *IEEE Engineering Management Review*, 47(2):75–83, May 2019.

[24] A. J. Harris, A. Corner, and U. Hahn. Estimating the probability of negative events. *Cognition*, 110(1):51–64, January 2009.

[25] J. Henderson and N. Cavalancia. Insider threat program maturity model report. techvangelism & insider threat defense, January 2019. `https://cdn2.hubspot.net/hubfs/5260286/PDFs/Whitepapers/insider-threat-maturity-report-2019.pdf` [Online; accessed on June 22, 2021].

[26] R. M. Hogarth. A note on aggregating opinions. *Organizational behavior and human performance*, 21(1):40–46, February 1978.

[27] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa. Insight into insiders and it: A survey

of insider threat taxonomies, analysis, modeling, and countermeasures. *ACM Computing Surveys (CSUR)*, 52(2):1–40, May 2019.

[28] I. R. C. (IRC). Hard problems list, November 2005. `https://www.infosec-research.org/docs_public/20051130-IRC-HPL-FINAL.pdf` [Online; accessed on June 22, 2021].

[29] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. In M. Leonard C and Z. William T, editors, *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

[30] M. Kandias, A. Mylonas, N. Virvilis, M. Theoharidou, and D. Gritzalis. An insider threat prediction model. In *Proc. of the 7th International Conference on Trust, Privacy and Security in Digital Business (TrustBus'10), Bilbao, Spain*, volume 6264 of *Lecture Notes in Computer Science*, pages 26–37. Springer, August 2010.

[31] M. Kwan, K.-P. Chow, F. Law, and P. Lai. Reasoning about evidence using bayesian networks. In *Proc. of the 2008 IFIP International Conference on Digital Forensics (DigitalForensics'08), Kyoto, Japan*, volume 285 of *IFIP — The International Federation for Information Processing*, pages 275–289. Springer, Boston, MA, January 2008.

[32] P. A. Legg. Human-machine decision support systems for insider threat detection. In C. Palomares, K. Iván, K. Harsha, and Y. Huang, editors, *Data Analytics and Decision Support for Cybersecurity*, pages 33–53. Springer, 2017.

[33] P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese. Automated insider threat detection system using user and role-based profile assessment. *IEEE Systems Journal*, 11(2):503–512, June 2015.

[34] S. Lichtenstein, P. Slovic, B. Fischhoff, M. Layman, and B. Combs. Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory*, 4(6):551, 1978.

[35] B. Lindauer, J. Glasser, M. Rosen, K. C. Wallnau, and L. ExactData. Generating test data for insider threat detectors. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 5(2):80–94, June 2014.

[36] M. Maybury, P. Chase, B. Cheikes, D. Brackney, S. Matzner, T. Hetherington, B. Wood, C. Sibley, J. Marin, and T. Longstaff. Analysis and detection of malicious insiders. Technical report, MITRE CORP BEDFORD MA, 2005.

[37] P. Memorandum. National insider threat policy and minimum standards for executive branch insider threat programs, November 2012. `https://obamawhitehouse.archives.gov/the-press-office/2012/11/21/presidential-memorandum-national-insider-threat-policy-and-minimum-stand` [Online; accessed on June 22, 2021].

[38] M. G. Morgan, M. Henrion, and M. Small. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge university press, 1990.

[39] C. Powell. The delphi technique: myths and realities. *Journal of advanced nursing*, 41(4):376–382, February 2003.

[40] C. Rietbergen, R. H. Groenwold, H. J. Hoijtink, K. G. Moons, and I. Klugkist. Expert elicitation of study weights for bayesian analysis and meta-analysis. *Journal of Mixed Methods Research*, 10(2):14, October 2014.

[41] A. Sanzgiri and D. Dasgupta. Classification of insider threat detection techniques. In *Proc. of the 11th annual cyber and information security research conference (CISRC'16), Oak Ridge, Tennessee, USA*, pages 1–4. ACM, April 2016.

[42] E. E. Schultz. A framework for understanding and predicting insider attacks. *Computers & security*, 21(6):526–531, October 2002.

[43] S. E. I. (SEI). Analytic approaches to detect insider threats. white paper, pittsburgh, pa, sei, December 2015. `http://resources.sei.cmu.edu/asset_files/WhitePaper/2015_019_001_451069.pdf` [Online; accessed on June 22, 2021].

[44] E. Shaw and L. Sellers. Application of the critical-path method to evaluate insider risks. *Studies in Intelligence*, 59(2):1–8, June 2015.

[45] E. D. Shaw. The role of behavioral research and profiling in malicious cyber insider investigations. *Digital investigation*, 3(1):20–31, March 2006.

[46] E. D. Shaw and L. F. Fischer. Ten tales of betrayal: The threat to corporate infrastructure by information

technology insiders analysis and observations. Technical report, DEFENSE PERSONNEL SECURITY RESEARCH CENTER MONTEREY CA, 2005.

[47] E. D. Shaw, J. M. Post, and K. G. Ruby. Inside the mind of the insider. *Security Management*, 43(12):34–42, 1999.

[48] The White House. Executive order—13587, structural reforms to improve the security of classified networks and the responsible sharing and safeguarding of classified information, october 2011. `https://obamawhitehouse.archives.gov/the-press-office/2011/10/07/executive-order-13587-structural-reforms-improve-security-classified-net[Online; accessedonJune22,2021]`, October 2011.

[49] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, September 1974.

[50] L. Val. Rethinking 'red flags' - a new approach to insider threats., January 2020. `https://www.securitymagazine.com/articles/91499-rethinking-red-flags---a-new-approach-to-insider-threats`[Online; accessed on June 22, 2021].

[51] L. Wilde. A bayesian network model for predicting data breaches caused by insiders of a health care organization. Master's thesis, University of Twente, December 2016.

[52] B. Wood. An insider threat model for adversary simulation, August 2000.

[53] M. S. B. Yusoff. Abc of content validation and content validity index calculation. *Resource*, 11(2):49–54, 2019.

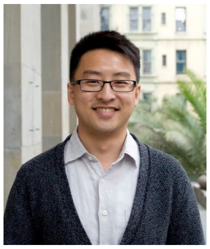
[54] Y. Zhang. Using bayesian priors to combine classifiers for adaptive filtering. In *Proc. of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'04), Sheffield, United Kingdom*, pages 345–352. ACM, July 2004.

## Author Biography

**Frank L. Greitzer** is owner and Principal Scientist of PsyberAnalytix, which performs consulting in applied cognitive and behavioral systems engineering and analysis with a focus on cybersecurity and insider threat research. Prior to founding PsyberAnalytix in 2012, Dr. Greitzer served as a Chief Scientist at the U.S. Department of Energy Pacific Northwest National Laboratory, where he performed human factors and social/behavioral science research in diverse problem domains including national security and energy systems. Dr. Greitzer received a B.S. in mathematics from Harvey Mudd College and a M.A. in psychology and Ph.D. in Mathematical Psychology from UCLA, with specialization in memory and cognition.He has published numerous scientific articles in peer-reviewed journals and has served as an Associate Editor or peer reviewer for several journals in cybersecurity and cyber psychology domains. Recently he was named Editor-in-Chief for a new journal in this field, Counter-Insider Threats Research and Practice.

**Justin Purl**, formerly a research scientist at the Human Resources Research Organization (HumRRO), has coordinated projects for focused research teams with a variety of clients in private sector, military, and educational organizations. He has managed databases consisting of over one million persons on over five thousand characteristics, leveraged advanced statistical analytic techniques to produce accurate insights, and automated reports to save time and money. Along with a team of experts, he has devoted the last three years to the advancement of insider threat research through the development and implementation of an insider threat ontology. He has also published research in the related areas of motivation and attrition. He is now at Google.

**Paul J. Sticha** is Principal of PsychInference, LLC, where he applies the results of psychological and other social science research, along with probabilistic inferential models to develop and evaluate methods to anticipate, prevent, and/or mitigate insider threats. Previously, he was Principal Scientist at the Human Resources Research Organization (HumRRO), where he worked for 36 years, focusing on the development and validation of mathematical models of human information processing, the application of probabilistic and decision analytic methods to aid individual and group problem solving, the development of computer decision support systems for planning and design, and the application of advanced technology to aid training and training management. He has contributed to several research programs on insider threats, focusing his efforts on the development of Bayesian network models to assess the risk of insider threat, based on psychological variables, and on development of methods to elicit estimates of insider behavior from experts. In addition, he worked with intelligence analysts to develop over 30 models to forecast actions of political leaders. He received a Ph.D. in Mathematical Psychology, and M.A. degrees in Mathematics and Psychology from the University of Michigan.



**Martin C. Yu** is a Research Scientist in the Strategic Human Capital Management (SHCM) program at the Human Resources Research Organization (HumRRO). He has broad experience and expertise in leading applied psychometric research projects to design, implement, and evaluate personnel assessment, selection, and development systems, and his recent work has involved coordinating research efforts to develop machine learning techniques for assessment development and evaluation. He has published research on a variety of topics including judgment and decision making, psychometrics and research methods, and prediction of performance. Yu received his PhD in Industrial-Organizational Psychology from the University of Minnesota.

**James Lee**[1] is a Systems Engineering and Operations Research Ph.D. student at George Mason University. Mr. Lee received a B.S. in statistics from the University of Michigan and a Master's in Systems Engineering at George Mason University.His research focuses on systems architecture, process management, and ontology development. He has conducted research on formalizing the inference enterprise modeling process with the use of ontologies and workflow languages. Mr. Lee implemented the SOFIT ontology.

[1] No photo is available.

## Appendix A: Likelhood Survey

The likelihood survey included initial material to acquire informed consent, describe the purpose of the study, and to identify the principal investigator. This introduction informed participants about its minimal risks, potential benefits, and confidentiality of the survey responses. It was suggested that the total time to complete the survey was approximately 20 minutes.

The survey comprised two parts. Questions 1-10 included one question to indicate informed consent (not shown) followed by nine demographic items. The rest of the questions consisted of the substantive survey items, questions 11-35.

# Demographics

Please answer the following questions about your background and experience.

2. What is your age?
   - 18-24
   - 25-29
   - 30-39
   - 40-49
   - 50-59
   - 60-69
   - 70+

3. Years of education completed beyond high school:
   - 1-2
   - 3-4
   - 5-6
   - 7-8
   - 8+

4. Highest degree achieved:
   - High School Diploma/GED
   - Associates Degree
   - Bachelor's Degree
   - Master's Degree
   - Doctoral Degree

5. Primary field of study:
   - Business Management/Marketing
   - Computer and Information Sciences
   - Engineering
   - Legal Studies
   - Mathematics and Statistics
   - Psychology
   - Public Administration and Social Services
   - Security and Protective Services
   - Other

6. Secondary field of study:
   - Business Management/Marketing
   - Computer and Information Sciences
   - Engineering
   - Legal Studies
   - Mathematics and Statistics
   - Psychology
   - Public Administration and Social Services
   - Security and Protective Services

- Other

7. Please indicate the areas in which you have background or experience (check all that apply):
   □ Computer Science
   □ Intelligence/Counterintelligence
   □ Information Assurance/Cybersecurity/Network Security
   □ Human Resources
   □ Personnel Security
   □ Social/Behavioral Sciences
   □ Other/Related Disciplines

8. Total years work experience:
   - Less than 3
   - 3-4
   - 5-10
   - 11-20
   - Greater than 20

9. What type of organization do you work for?
   - Military or DoD civilian
   - Non-DoD Government
   - Government Contractor (primary role)
   - Education
   - Other private organization
   - Other

10. Primary area of experience/expertise:
    - Researcher
    - Practitioner

## *Survey Body*

## General Instructions:

In this task, we ask you to compare the occurrence of different indicators between individuals who are insider threats versus the normal population.

INSTRUCTIONS: Each question describes a behavior or characteristic. Please indicate whether you think this characteristic or behavior is more likely or less likely to occur for individuals who are insider threats versus the general population. Seven response options are provided–please pick one:

- Much more likely among insider threats
- More likely among insider threats
- Slightly more likely among insider threats
- About the same
- Slightly more likely within the general population
- More likely within the general population
- Much more likely within the general population

When you read each characteristic or behavior, think about how likely it would be within the general population and how likely it would be among insider threats. If you think that the characteristic or behavior is more likely among insider threats, select one of the first three options. If you think that it is more likely to occur within the general population than among insider threats, then select one of the last three options. Which of these options you select depends on how much more (or less) likely you think the characteristic or behavior is among insider threats than it is within the general population.

If you do not think there is a difference in occurrence of a characteristic or behavior between insider threats and the general population, then select "About the same." For example, consider the behavior **"Establish a backdoor."** Experts would agree that this cybersecurity violation is a very suspicious insider threat indicator that would not be expected to occur within the general population; thus most would agree that it would be much more likely among insider threats, and most would select the option "Much more likely among insider threats." On the other hand, the event **"Received a Promotion"** might be considered to be more likely within the general population than among insider threats.

11. Characteristic or behavior: **Establish Back Door**
*Network/workstation audit reveals that this individual used programming techniques to bypass system security mechanisms and access sensitive or proprietary data without being detected.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

12. Characteristic or behavior: **Received a Promotion.**
*The individual has recently been promoted based on consistently exceeding performance expectations.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

13. Characteristic or behavior: **Excessive Absences**
*Management reports indicate that this individual has had excessive unexcused absences, many with little or no advance notice.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

14. Characteristic or behavior: **Delete or Edit Audit Logs**
*Irregularities in network audit logs reveal that this individual has deleted or edited audit logs, in violation of organizational policy.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

15. Characteristic or behavior: **Unauthorized Copying of Classified Material**
*Security records indicate a security incident in which the individual was reported to be making copies of classified material without authorization.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population

o Much more likely for general population

16. Characteristic or behavior: **Certificate of Appreciation**
*The employee has received a certificate of appreciation from a client for extraordinary customer service.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

17. Characteristic or behavior: **Unusual File Deletion**
*Network/host workstation audits indicate the individual has been deleting sensitive data files at a higher rate than is consistent with this individual's role.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

18. Characteristic or behavior: **Change in Marital Status**
*The individual has been making excuses for poor performance by telling the manager of marital problems, recent separation, and pending divorce.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

19. Characteristic or behavior: **Negative Evaluation**
*Performance reports by project managers and peers indicates that this individual's performance has recently fallen significantly below expectation.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same

o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

20. Characteristic or behavior: **Raise/Salary Increase**
*The employee has received a significant salary increase for sustained superior performance.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

21. Characteristic or behavior: **Unexplained Affluence**
*Management and peer reports indicate that the individual appears to be living beyond his/her means. These reports are substantiated in financial records showing several recent purchases of high value items (e.g., real estate, stocks, vehicles, or vacations).*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

22. Copy of Characteristic or behavior: **Terminated**
*The individual has been notified that they are under review for conduct or performance issues that will result in termination if not corrected.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

23. Characteristic or behavior: **Low Honesty-Humility**
*Management and peer reviews report that this individual has been characterized as greedy, self-promoting, and often given to exaggerating, lying, or deceit.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

24. Characteristic or behavior: **Disgruntled**
*Management and peer reviews describe numerous serious complaints about the behavior of this individual, indicating strong irritability and disgruntlement over company policies and management decisions that the individual considers unfair.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

25. Characteristic or behavior: **Bonus**
*The employee has received a cash bonus for outstanding performance in completing a critical project on time and within budget.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

26. Characteristic or behavior: **Passed over for Promotion**
*The individual was passed over for a promotion to manage a new, prestigious project after having lobbied to get the position for months.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population

o Much more likely for general population

27. Characteristic or behavior: **Receiving Large Emails**
*Network audit logs reveal that this individual has received emails with attachments that are much larger than the individual has received in typical prior weeks or that would be expected for this individual's role.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

28. Characteristic or behavior: **Manipulative**
*Management and peer reviews report that this individual often uses unfair or subtle means to influence others to one's own advantage.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

29. Characteristic or behavior: **Cyberloafing**
*Management and peer reports indicate that this employee often uses company resources, and company time, to do personal business: excessive browsing to non-work-related websites or unsanctioned personal use of work email.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

30. Characteristic or behavior: **Unreported Contact with Foreign Nationals**
*This individual's security clearance requires that any contact with foreign nationals must be reported to security. Security records indicate that this individual had unreported contact with foreign nationals at technical conferences.*

In which population is this more likely to occur?

o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

31. Characteristic or behavior: **Poor Time Management**
*Management reviews indicate that this individual's performance is negatively affected by poor time management skills, such as procrastination and difficulties in scheduling/planning and completing tasks.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat o Slightly more likely for insider threat
o About the same o Slightly more likely for general population
o More likely for general population o Much more likely for general population

32. Characteristic or behavior: **Recognition for Outstanding Mentoring**
*The employee has received a plaque and a cash performance award for outstanding performance in mentoring a junior member of his technical team.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

33. Characteristic or behavior: **Attempts to Access Prohibited File Sharing Sites**
*Network monitoring indicates numerous attempts to access prohibited file sharing web sites.*

In which population is this more likely to occur?
o Much more likely for insider threat
o More likely for insider threat
o Slightly more likely for insider threat
o About the same
o Slightly more likely for general population
o More likely for general population
o Much more likely for general population

34. Characteristic or behavior: **Lack of Confidentiality**
*Management and peer reviews report that this individual often engages in idle talk, spreading rumors, and gossiping about personal affairs of others.*

In which population is this more likely to occur?
o Much more likely for insider threat

- o More likely for insider threat
- o Slightly more likely for insider threat
- o About the same o Slightly more likely for general population

35. Characteristic or behavior: **Using Multiple Printers Simultaneously**
*Network audit logs reveal that this individual is using multiple printers concurrently much more than they have in typical previous weeks or that would be expected for this individual's role.*

In which population is this more likely to occur? *

- o Much more likely for insider threat
- o More likely for insider threat
- o Slightly more likely for insider threat
- o About the same
- o Slightly more likely for general population
- o More likely for general population
- o Much more likely for general population

# Appendix B: Conditional Probabilities of SOFIT Indicators

As described in Section 4.2, based on the empirical results reported in this study, the SOFIT level-of -concern ratings ri may be converted to comparable likelihood ratio ratings Li via the quadratic formula,

$L_i = 0.0003{r_i}^2 + 0.0023r_i + 0.0083$,

and the associated conditional probabilities are obtained via the formula,

$p(I_i|A) = L_i/(1+L_i)$,

which specifies the probability that Indicator Ii is present, given the truth of the hypothesis A that the individual is an insider threat.

These probabilities reflect the relative likelihood that a case with a specific indicator represents an insider threat. This conditional probability is the foundation for the BN models defined in this paper. The table below shows the converted conditional probabilities for a large set of SOFIT indicators, based on the threat rating values (listed in the SOFIT taxonomy listing in Appendix III).

**SOFIT Insider Threat Indicators with Threat Ratings, Likelihood Ratios and Conditional Probabilities**

| Classes - Indicators | Threat Rating | Likelihood Ratio | Conditional Probability |
|---|---|---|---|
| **Boundary Violations** | | | |
| Working At Unusual Hours | 43 | 0.662 | 0.398 |
| Nonproductive Socialization | 26 | 0.271 | 0.213 |
| Unprofessional Conduct | 48 | 0.810 | 0.447 |
| Aggression | 63 | 1.344 | 0.573 |
| Intimidation | 55 | 1.042 | 0.510 |
| Workplace Violence | 69 | 1.595 | 0.615 |
| Attempted Unauthorized Access | 71 | 1.684 | 0.627 |
| Unauthorized Recording Device | 80 | 2.112 | 0.679 |
| Shoulder Surfing | 68 | 1.552 | 0.608 |
| Tailgating | 74 | 1.821 | 0.646 |
| Unintentional Breach | 42 | 0.634 | 0.388 |
| Ignore Security Norms | 48 | 0.810 | 0.447 |
| Travel Policy Violation | 42 | 0.634 | 0.388 |
| Lost security badge | 44 | 0.690 | 0.408 |
| Leave Unlocked- Unattended Security Container | 64 | 1.384 | 0.581 |
| Improper Discussion Of Classified Material | 65 | 1.425 | 0.588 |
| Improper Destruction Classified Material | 63 | 1.344 | 0.573 |
| Unauthorized Disclosure Of Classified Information | 88 | 2.534 | 0.717 |
| Improper Handling Of Classified Material | 81 | 2.163 | 0.684 |
| Unlawful Removal Classified Material | 85 | 2.371 | 0.703 |
| Unauthorized Copying Of Classified Material | 86 | 2.425 | 0.708 |

| Classes - Indicators | Threat Rating | Likelihood Ratio | Conditional Probability |
|---|---|---|---|
| **Job Performance** | | | |
| Cyberloafing | 35 | 0.456 | 0.313 |
| Attendance Issues | 32 | 0.389 | 0.280 |
| Performance Issues | 40 | 0.580 | 0.367 |
| **Cybersecurity Violation** | | | |
| Attempts To Access System Against Policy | 73 | 1.775 | 0.640 |
| Finding Or Searching For Passwords | 74 | 1.821 | 0.646 |
| Attempts To Change File Permissions | 69 | 1.595 | 0.615 |
| Circumvent Document Control | 80 | 2.112 | 0.679 |
| Search Own Name | 74 | 1.821 | 0.646 |
| Excessive Unauthorized Database Searches | 78 | 2.013 | 0.668 |
| Unusual Remote Access | 80 | 2.112 | 0.679 |
| Use Of Anonymizers | 79 | 2.062 | 0.673 |
| Use Of Covert Channels | 88 | 2.534 | 0.717 |
| Compromised Machine | 79 | 2.062 | 0.673 |
| Use Of Keystroke Logger | 91 | 2.702 | 0.730 |
| Use Of Steganography | 86 | 2.425 | 0.708 |
| Uses Network Mapping Software | 84 | 2.318 | 0.699 |
| Unauthorized Wireless | 78 | 2.013 | 0.668 |
| Using Multiple Printers Simultaneously | 59 | 1.188 | 0.543 |
| Attempts To Access Prohibited File- Sharing Websites | 81 | 2.163 | 0.684 |
| Copy Large Amount Of Data Offline | 81 | 2.163 | 0.684 |
| File Sharing Or Personal Storage Websites | 76 | 1.916 | 0.657 |
| High Use Of U S B Devices | 81 | 2.163 | 0.684 |
| Large Data Transfer Outgoing | 75 | 1.868 | 0.651 |
| Disabling Security Features | 69 | 1.595 | 0.615 |
| Establish Backdoor | 90 | 2.645 | 0.726 |
| Mislabeling Documents | 75 | 1.868 | 0.651 |
| Delete Or Edit Audit Logs | 93 | 2.817 | 0.738 |
| Excessive Communication With Foreign Entities | 78 | 2.013 | 0.668 |
| Blacklisted Webmail Correspondence | 81 | 2.163 | 0.684 |
| **Life Narrative Factors** | | | |
| Restraining Order | 68 | 1.552 | 0.608 |
| Violence Outside Workplace | 67 | 1.509 | 0.601 |
| Wage Garnishments | 65 | 1.425 | 0.588 |
| Suspicious Spending Trends | 67 | 1.509 | 0.601 |
| Unexplained Affluence | 71 | 1.684 | 0.627 |
| Credit Problems | 67 | 1.509 | 0.601 |
| Excessive Debts | 68 | 1.552 | 0.608 |
| Gambling Addiction | 70 | 1.639 | 0.621 |
| Substance Abuse | 71 | 1.684 | 0.627 |
| Clinical Syndromes | 60 | 1.226 | 0.551 |
| Disloyalty | 77 | 1.964 | 0.663 |
| Radical Beliefs | 82 | 2.214 | 0.689 |

| Classes - Indicators | Threat Rating | Likelihood Ratio | Conditional Probability |
|---|---|---|---|
| New- Hire | 34 | 0.433 | 0.302 |
| Retiring | 40 | 0.580 | 0.367 |
| Resigned To Take Another Job | 52 | 0.939 | 0.484 |
| Terminated | 69 | 1.595 | 0.615 |
| Disciplinary Action | 69 | 1.595 | 0.615 |
| Passed Over For Promotion | 48 | 0.810 | 0.447 |
| Recent Birth Of Child | 27 | 0.289 | 0.224 |
| Recent Change In Marital Status | 35 | 0.456 | 0.313 |
| Recent Death Of Loved One | 42 | 0.634 | 0.388 |
| Severe Medical Condition | 40 | 0.580 | 0.367 |
| **Psychological Factors** | | | |
| Marked Anger/ Hostility | 63 | 1.344 | 0.573 |
| Other Negative Emotions | 58 | 1.151 | 0.535 |
| Mood Swings | 53 | 0.973 | 0.493 |
| Workplace Stress | 46 | 0.749 | 0.428 |
| Lack Of Motivation | 41 | 0.607 | 0.378 |
| Disengaged Socially | 42 | 0.634 | 0.388 |
| Overly Competitive | 45 | 0.719 | 0.418 |
| Disgruntlement | 66 | 1.467 | 0.595 |
| Overly Critical | 61 | 1.265 | 0.558 |
| Emotional Instability/ Neuroticism | 59 | 1.188 | 0.543 |
| Low- Conscientiousness | 50 | 0.873 | 0.466 |
| Unreliable | 49 | 0.841 | 0.457 |
| Impulsivity | 63 | 1.344 | 0.573 |
| Poor Time Management | 45 | 0.719 | 0.418 |
| Disagreeableness | 55 | 1.042 | 0.510 |
| Excitement-seeking | 61 | 1.265 | 0.558 |
| Low Honesty- Humility | 64 | 1.384 | 0.581 |
| Machiavellianism | 64 | 1.384 | 0.581 |
| Narcissism | 64 | 1.384 | 0.581 |
| Psychopathy | 64 | 1.384 | 0.581 |