Exploring Predictive Analytic Threat Assessment Models Built Upon the SOFIT Insider Threat Ontology

2024 International Symposium on Insider Threat Mitigation

March 5-7, 2024

Brussels, Belgium

Frank L. Greitzer, PhD

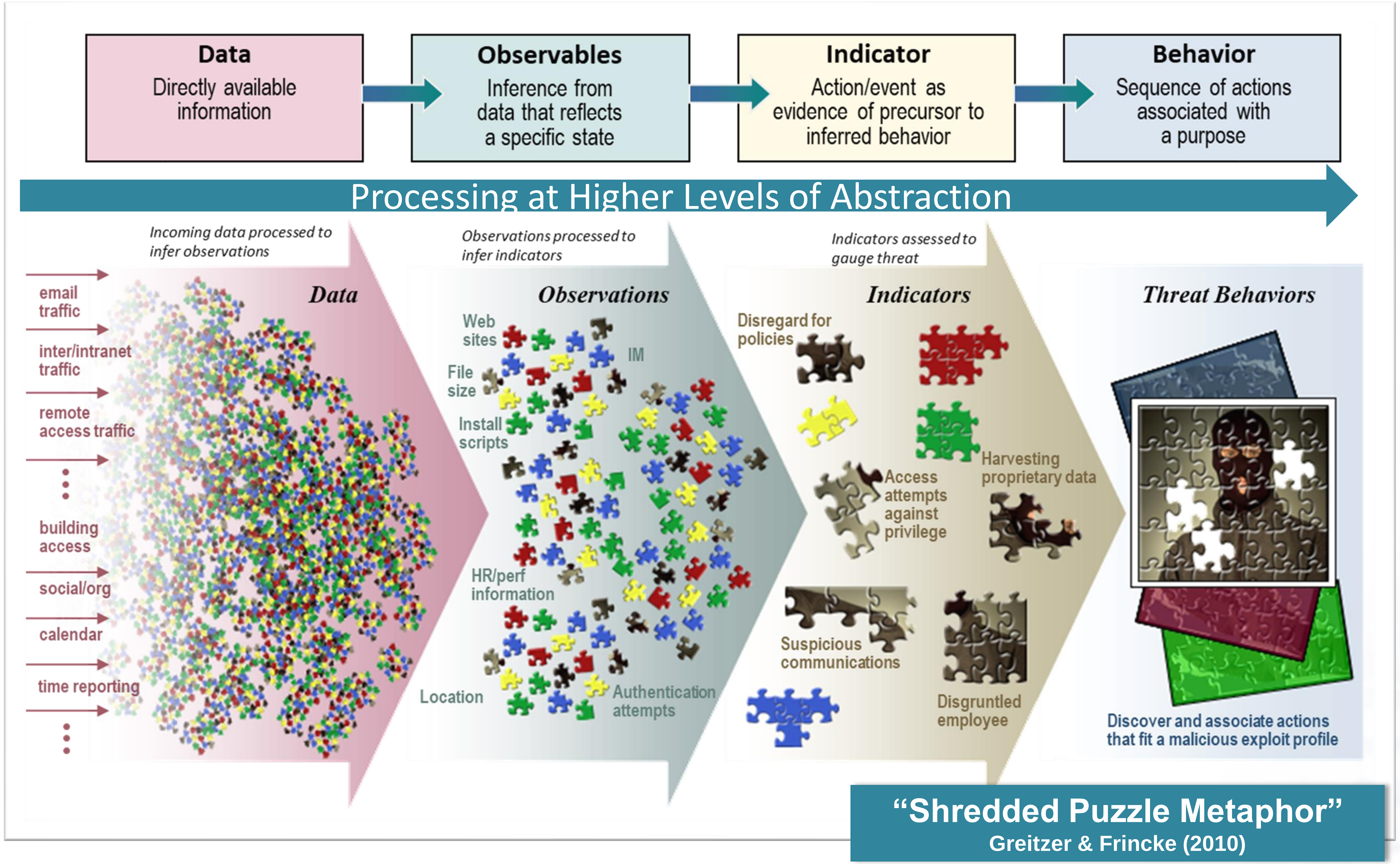
Principal Scientist, PsyberAnalytix

Chief Behavioral Scientist, COGILITY

TOPICS

- Conceptual Model and Insider Threat Indicator Knowledge Base
- Potential Risk Indicators (PRIs)
- Using Expert Judgments to estimate PRI “weights”
- Limitations in traditional predictive models
- Applying a hierarchical/pattern-based model
- Conclusions and Path Forward

CONCEPTUAL PREDICTIVE CLASSIFICATION MODEL



Data
Directly available information

Observables
Inference from data that reflects a specific state

Indicator
Action/event as evidence of precursor to inferred behavior

Behavior
Sequence of actions associated with a purpose

Processing at Higher Levels of Abstraction

Incoming data processed to infer observations

Observations processed to infer indicators

Indicators assessed to gauge threat

Data

email traffic

inter/intranet traffic

remote access traffic

⋮

building access

social/org

calendar

time reporting

⋮

Observations

Web sites

IM

File size

Install scripts

HR/perf information

Location

Authentication attempts

Indicators

Disregard for policies

Access attempts against privilege

Harvesting proprietary data

Suspicious communications

Disgruntled employee

Threat Behaviors

Discover and associate actions that fit a malicious exploit profile

"Shredded Puzzle Metaphor"
Greitzer & Frincke (2010)

- Data processed to identify "**observables**"
- Observables analyzed to recognize **Potential Risk Indicators (PRIs)**
- PRIs analyzed to recognize **behavioral patterns** relating to insider risks

Development of PRI ontology: **Sociotechnical and Organizational Factors for Insider Threat (SOFIT)**

POTENTIAL RISK INDICATORS (PRI)s

accessing classified information without need-to-know

violating security practices

disciplinary actions

poor time management

illness

lack of attention

high workload/cognitive load

persistent lateness

lack of knowledge, awareness, training

exhibiting incidents of physical violence

ties to foreign defense contractor

manipulation or destruction of sensitive information

possessing illegal weapons

engaged in criminal activity

frequent personal travel

lying to investigators

extremist views

emotional problems

disgruntlement

unexplained affluence

associating with extremist group

misuse of U.S. Government information systems

failure to comply with regulations for reporting foreign contacts or foreign travel

narcissism

access via other users' credentials

declining work performance

passed over for promotion

dismissal

demotion

failure to return company property

frequent, unreported contact with foreign persons

past untruthfulness

financial concerns – excessive debts

depression

anxiety

sleep disturbances

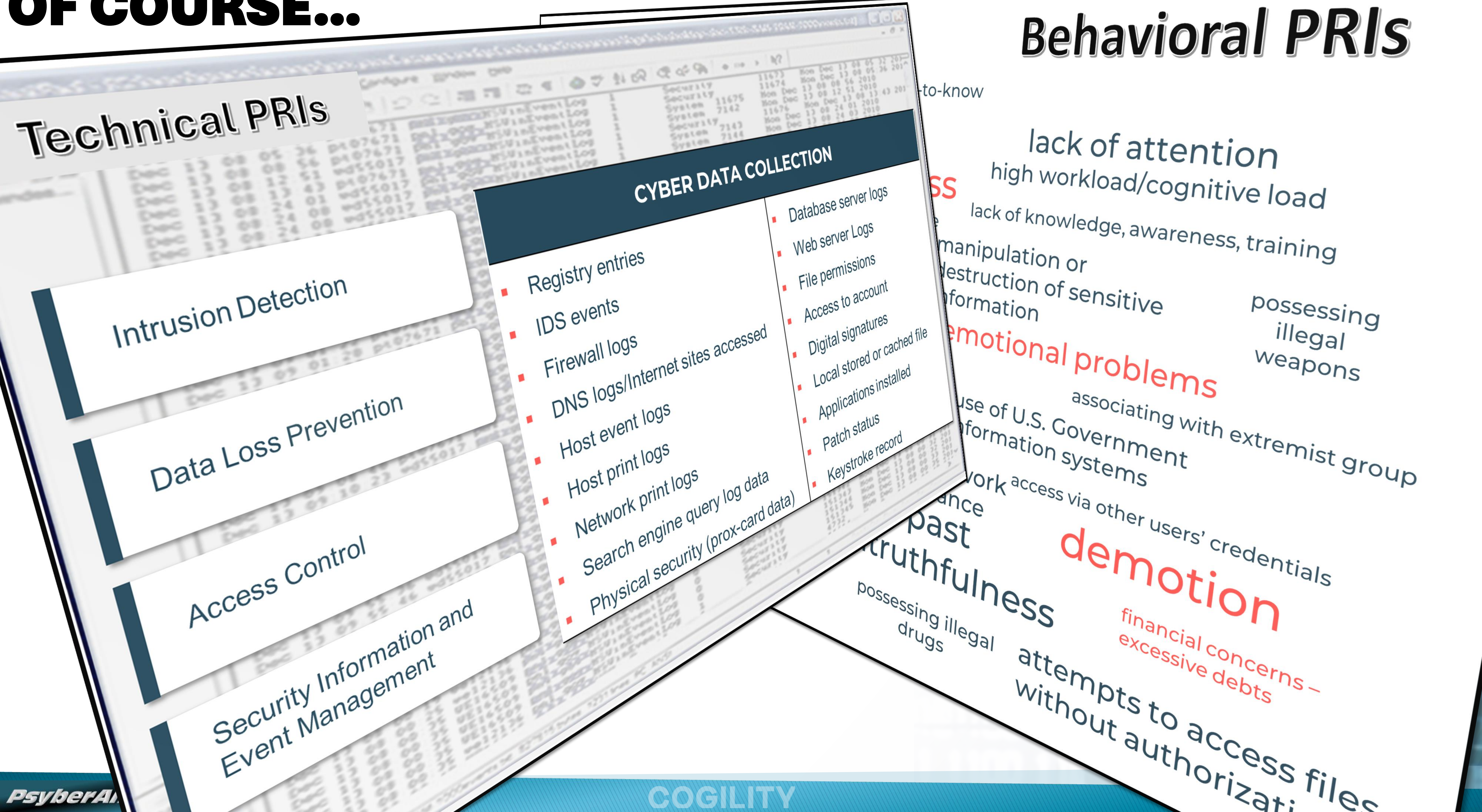
mental health counseling

possessing illegal drugs

attempts to access files without authorization

expressing ill will toward U.S. Government

OF COURSE...



Technical PRIs

- Intrusion Detection
- Data Loss Prevention
- Access Control
- Security Information and Event Management

CYBER DATA COLLECTION

- Registry entries
- IDS events
- Firewall logs
- DNS logs/Internet sites accessed
- Host event logs
- Host print logs
- Network print logs
- Search engine query log data
- Physical security (prox-card data)
- Database server logs
- Web server Logs
- File permissions
- Access to account
- Digital signatures
- Local stored or cached file
- Applications installed
- Patch status
- Keystroke record

Behavioral PRIs

-to-know

lack of attention

high workload/cognitive load

ss

lack of knowledge, awareness, training

manipulation or destruction of sensitive information

possessing illegal weapons

emotional problems

associating with extremist group

use of U.S. Government information systems

work access via other users' credentials

ance

past

truthfulness

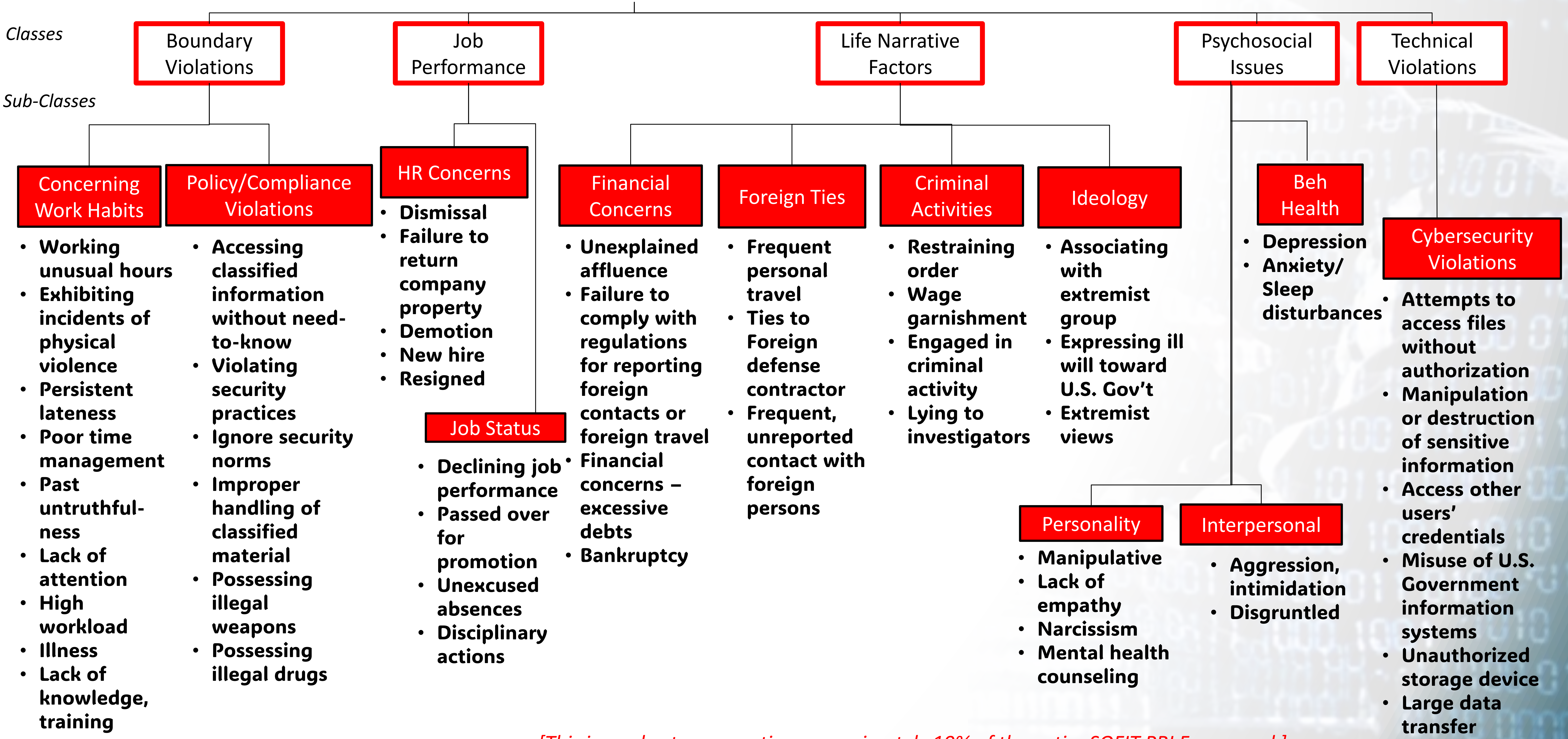
demotion

possessing illegal drugs

financial concerns – excessive debts

attempts to access files without authorizati

SOFIT PRI KNOWLEDGE BASE



Classes

Sub-Classes

Boundary Violations

Concerning Work Habits

- Working unusual hours
- Exhibiting incidents of physical violence
- Persistent lateness
- Poor time management
- Past untruthful-ness
- Lack of attention
- High workload
- Illness
- Lack of knowledge, training

Policy/Compliance Violations

- Accessing classified information without need-to-know
- Violating security practices
- Ignore security norms
- Improper handling of classified material
- Possessing illegal weapons
- Possessing illegal drugs

Job Performance

HR Concerns

- Dismissal
- Failure to return company property
- Demotion
- New hire
- Resigned

Job Status

- Declining job performance
- Passed over for promotion
- Unexcused absences
- Disciplinary actions

Life Narrative Factors

Financial Concerns

- Unexplained affluence
- Failure to comply with regulations for reporting foreign contacts or foreign travel
- Financial concerns – excessive debts
- Bankruptcy

Foreign Ties

- Frequent personal travel
- Ties to Foreign defense contractor
- Frequent, unreported contact with foreign persons

Criminal Activities

- Restraining order
- Wage garnishment
- Engaged in criminal activity
- Lying to investigators

Ideology

- Associating with extremist group
- Expressing ill will toward U.S. Gov't
- Extremist views

Psychosocial Issues

Beh Health

- Depression
- Anxiety/ Sleep disturbances

Personality

- Manipulative
- Lack of empathy
- Narcissism
- Mental health counseling

Interpersonal

- Aggression, intimidation
- Disgruntled

Technical Violations

Cybersecurity Violations

- Attempts to access files without authorization
- Manipulation or destruction of sensitive information
- Access other users' credentials
- Misuse of U.S. Government information systems
- Unauthorized storage device
- Large data transfer

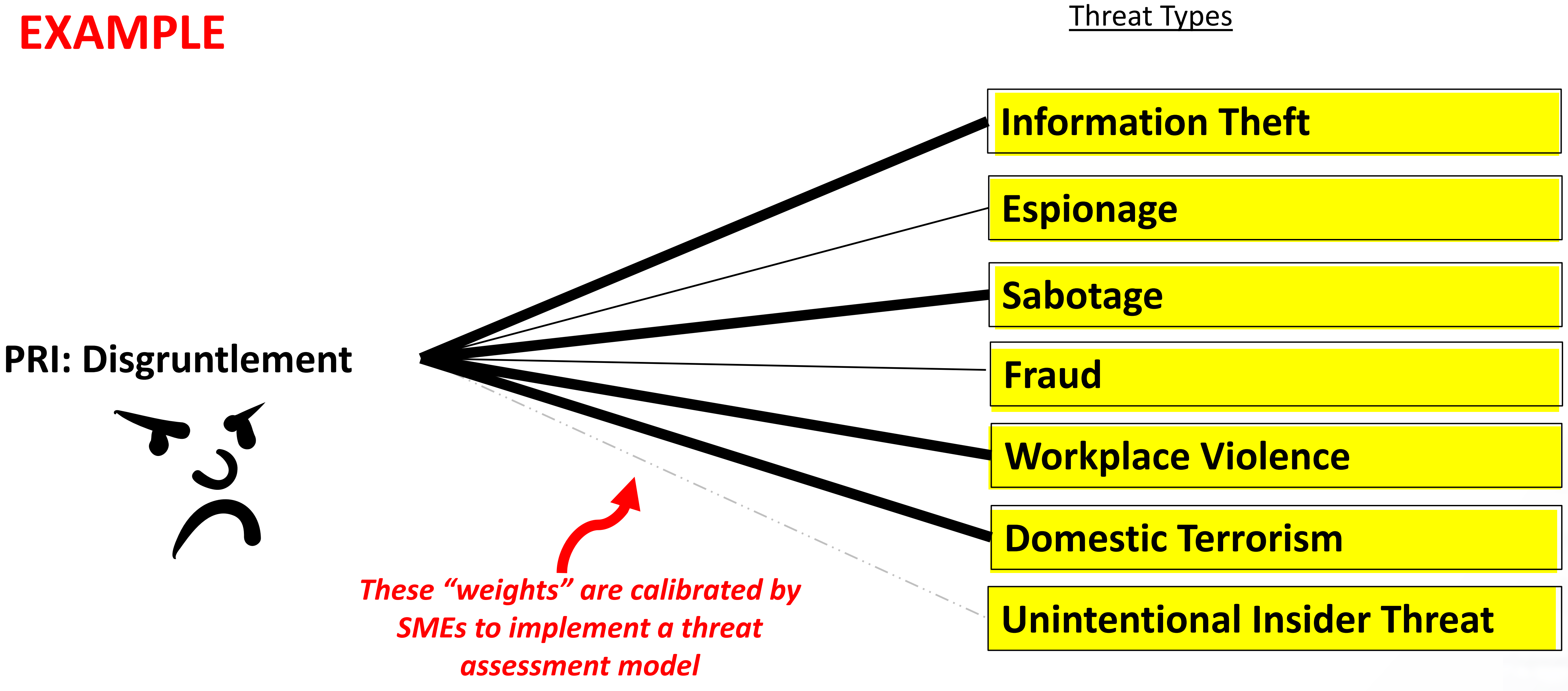
[This is a subset representing approximately 10% of the entire SOFIT PRI Framework]

PRI "CALIBRATION"

Estimating Strength of Association between a PRI and a Threat Behavior

- Each PRI is mapped to relevant Threat Types
- Strength of association may be thought of as a "weight" or "probability"
  Higher weight means that the observation of a PRI significantly increases the likelihood that the Behavior is present

EXAMPLE



HIGH Association with Information Theft

MODERATE Association with Espionage

HIGH Association with Sabotage

MODERATE Association with Fraud

HIGH Association with WorkPlace Violence

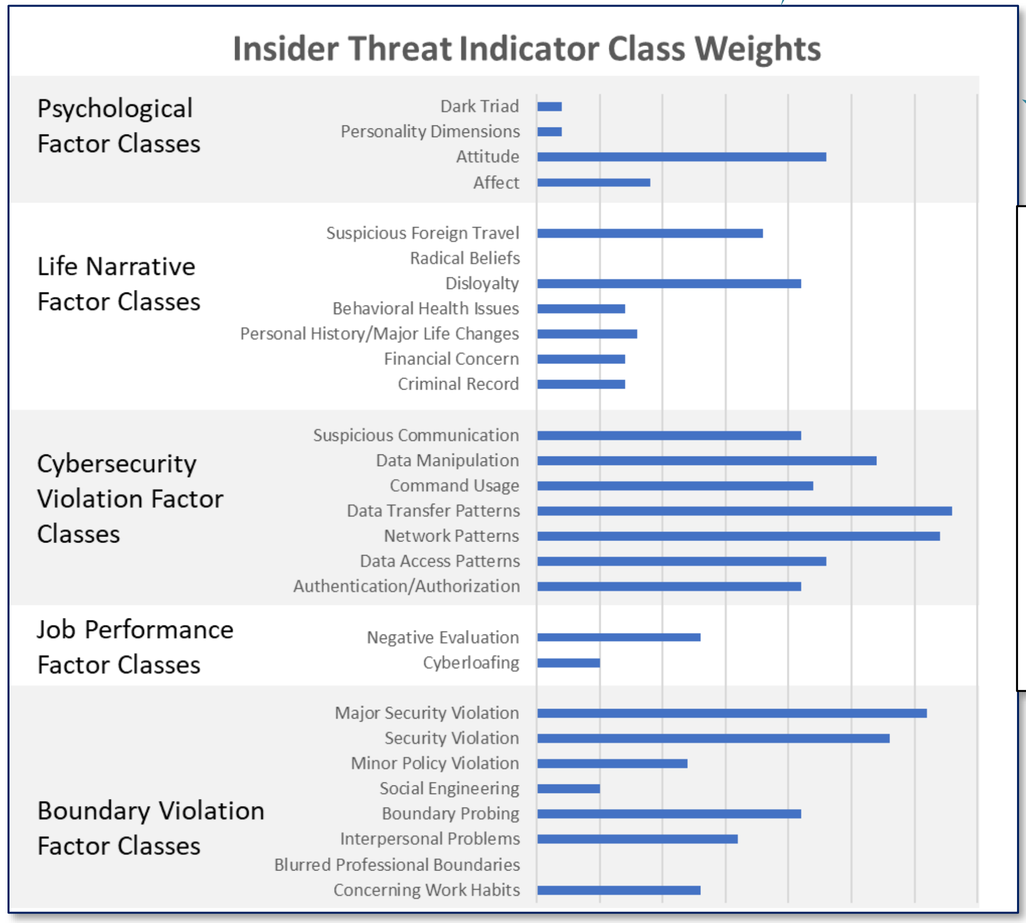
HIGH Association with Domestic Terrorism

LITTLE/NO Association with Unintentional Insider threat

These "weights" are calibrated by SMEs to implement a threat assessment model

WHAT WE'VE LEARNED FROM PRI CALIBRATION STUDIES

1. PRIs vary in their strength/ association with insider threat behaviors



Threat/Behavior Types

- Exfiltration/theft
- Sabotage
- Workplace Violence
- Fraud
- Unintentional Insider Threat
- Espionage
- Suicidal Ideation

Greitzer et al. (2018)

2. It's difficult to get reliable PRI "weight" estimates!

When we ask our analysts/experts to provide judgments about PRI weights or severity or likelihood, what are they really thinking?

We don't know!

*We use terms like **PRI risk, probability, weight, severity** interchangeably. But in our calibration exercises, our SMEs may be thinking about these weights in different ways.*

*For our probabilistic models, we need to devise an expert knowledge elicitation method that encourages experts to have the same **mindset**—i.e., focus on probability/ likelihood interpretation. I'm currently using a calibration method that acquires **Likelihood Ratio** estimates.*

3. PRIs vary in the span of time during which they influence judgments of insider threat

Greitzer et al. (2022)

Transient impact: Failed login attempt after a password was changed.

Stable impact: Narcissism – many psychological factors, and especially personality traits, are very stable over many years

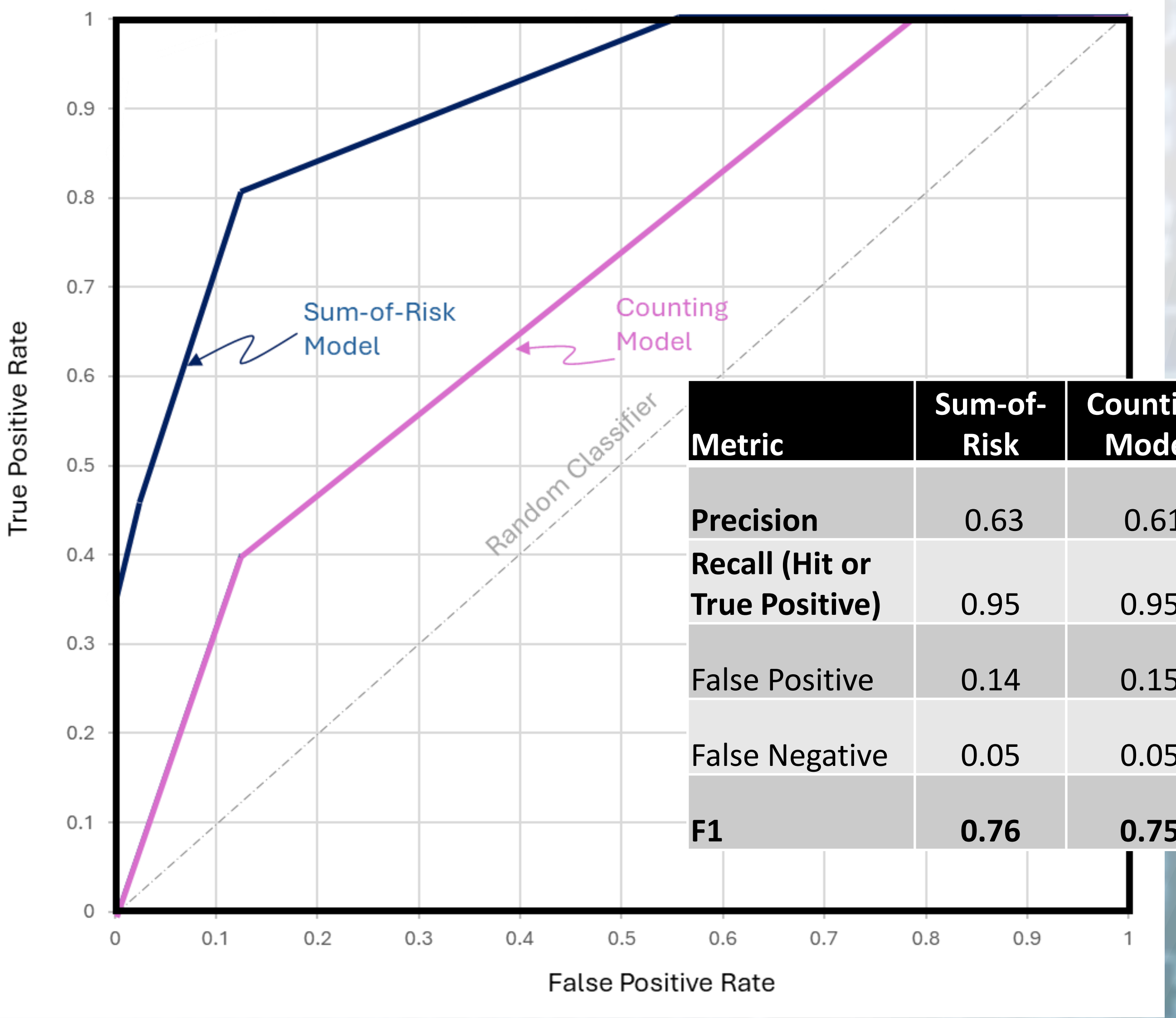
COMPUTING INSIDER RISKS BASED ON EXPERT JUDGMENTS OF PRI “WEIGHTS”

- My research investigated computational models predicting expert judgments of insider threat risk based on expert judgments of PRI “**level of concern**”
  - Counting Model
  - Sum-of-Risk Model
  - Probabilistic models (e.g. Bayesian networks)

- Performance metrics:
  - **Receiver Operating Characteristic (ROC)**
  - **Precision, Recall, False Positives, False Negatives, F1 score**
- Results indicate that these models exhibit modest predictive value, accounting for 50-60% of variance in predicting expert judgments

Based on Greitzer et al. (2018) data…

ROC Curves



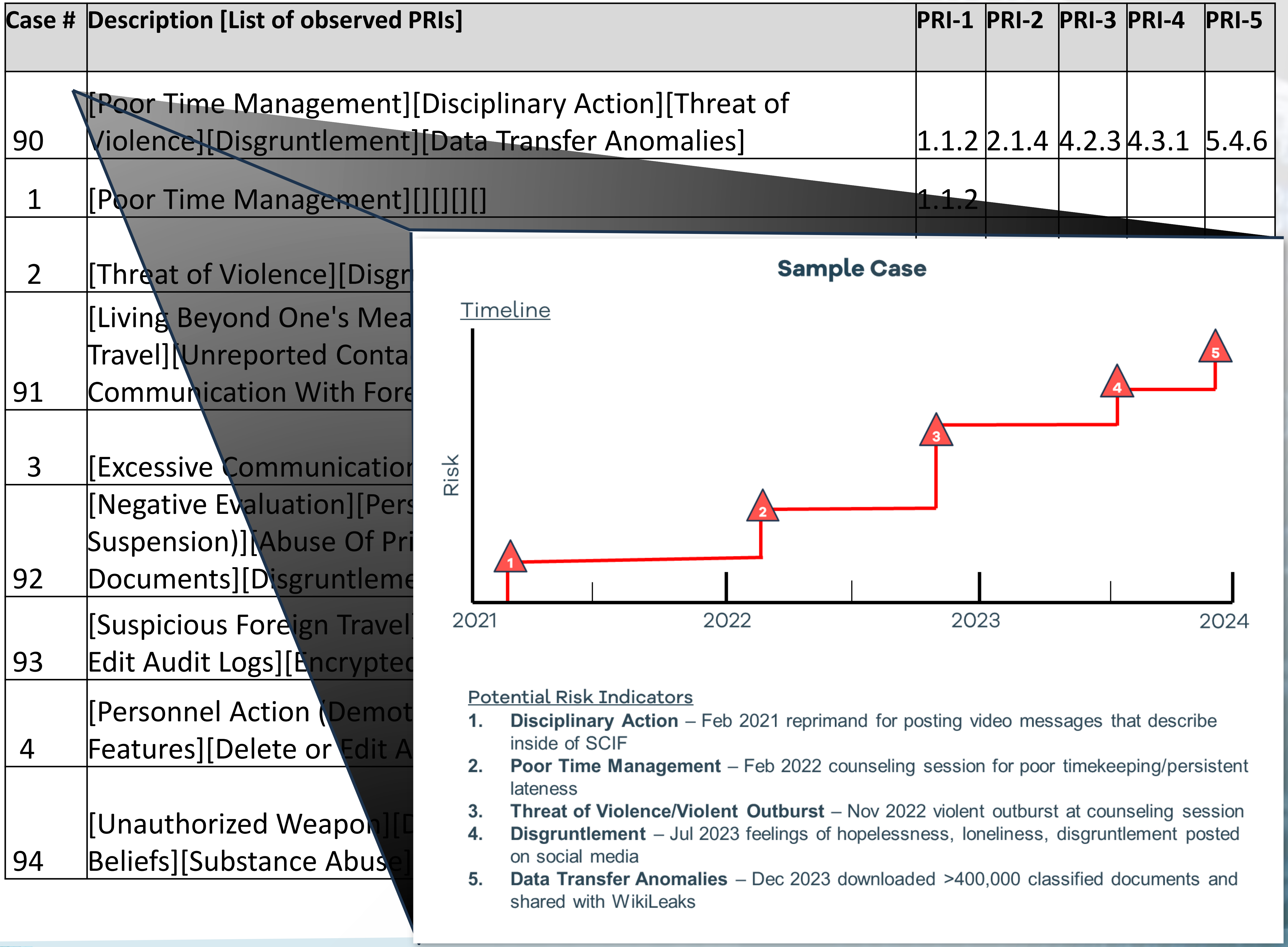
Metric	Sum-of-Risk	Counting Model
Precision	0.63	0.61
Recall (Hit or True Positive)	0.95	0.95
False Positive	0.14	0.15
False Negative	0.05	0.05
F1	**0.76**	**0.75**

- *Precision: Out of all the cases predicted to be threats, what percentage was a TRUE threat?*
- *Recall: Out of all the TRUE threats, what percentage was predicted to be threats?*
- *F1 = harmonic mean of precision and recall*

- ***Possible limitations:***
  - This early work obtained SME judgments of PRI “weights” for a generic “Insider Threat” instead of specific threat behaviors—we know PRIs contribute differentially to threat behaviors
  - Expert judgments may be conflating multiple aspects of PRIs, including probability and severity

RECENT TESTS OF MODELS USING SYNTHETIC DATA

- Informal study with a new set of synthetic data
  - 100 cases created
  - 1-5 PRIs chosen from SOFIT ontology
- Expert classified cases as “threat” vs “no-threat”
- Used new “Likelihood Ratio” method to estimate PRI weights (probabilities) for individual threat behaviors
- Applied and tested different threat models:
  - Counting Model
  - Sum-of-Risk Model
  - COGYNT Model



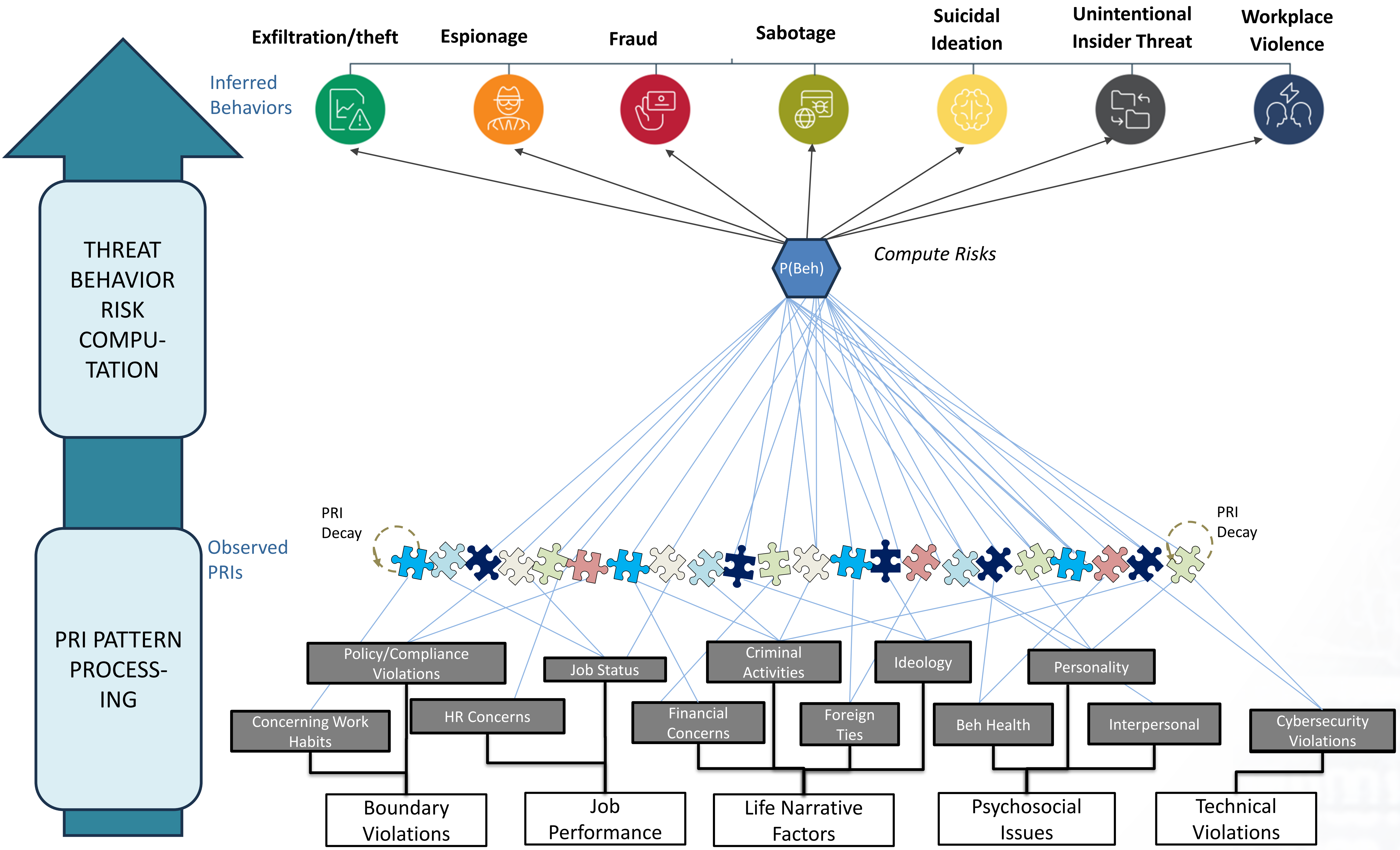
Case #	Description [List of observed PRIs]	PRI-1	PRI-2	PRI-3	PRI-4	PRI-5
90	[Poor Time Management][Disciplinary Action][Threat of Violence][Disgruntlement][Data Transfer Anomalies]	1.1.2	2.1.4	4.2.3	4.3.1	5.4.6
1	[Poor Time Management][][][][]	1.1.2				
2	[Threat of Violence][Disgr					
91	[Living Beyond One's Mea Travel][Unreported Conta Communication With Fore					
3	[Excessive Communication					
92	[Negative Evaluation][Pers Suspension)][Abuse Of Pri Documents][Disgruntleme					
93	[Suspicious Foreign Travel] Edit Audit Logs][Encrypted					
4	[Personnel Action (Demot Features][Delete or Edit A					
94	[Unauthorized Weapon][[Beliefs][Substance Abuse]					

Sample Case

Potential Risk Indicators

1. **Disciplinary Action** – Feb 2021 reprimand for posting video messages that describe inside of SCIF
2. **Poor Time Management** – Feb 2022 counseling session for poor timekeeping/persistent lateness
3. **Threat of Violence/Violent Outburst** – Nov 2022 violent outburst at counseling session
4. **Disgruntlement** – Jul 2023 feelings of hopelessness, loneliness, disgruntlement posted on social media
5. **Data Transfer Anomalies** – Dec 2023 downloaded >400,000 classified documents and shared with WikiLeaks

COGYNT HIERARCHICAL COMPLEX EVENT PROCESSING MODEL



Display final RISK scores

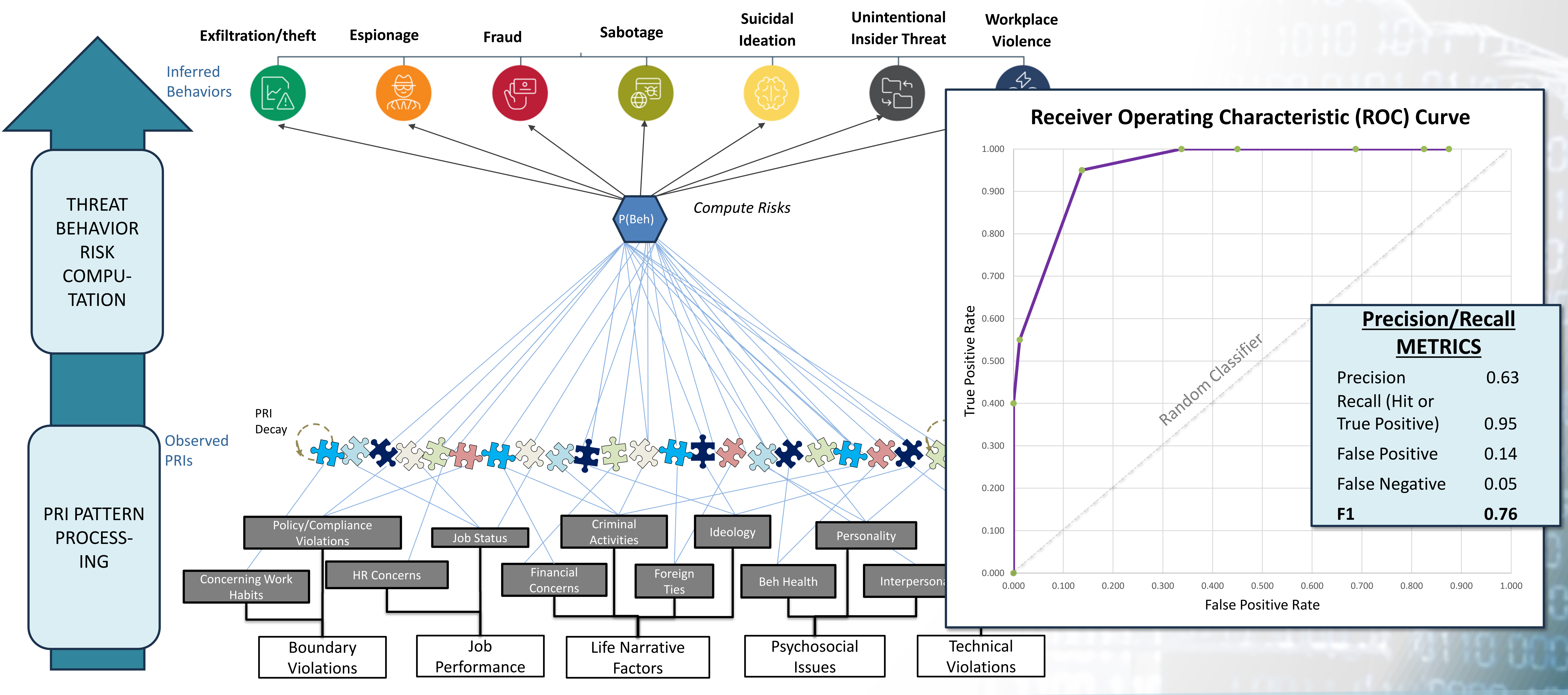
Compute risks by aggregating probabilities of observed PRIs

PRIs are "calibrated" for "weights" or probabilities associated with Behaviors

Pattern Processing of "observables" to Recognize PRIs

Incorporates SOFIT ONTOLOGY

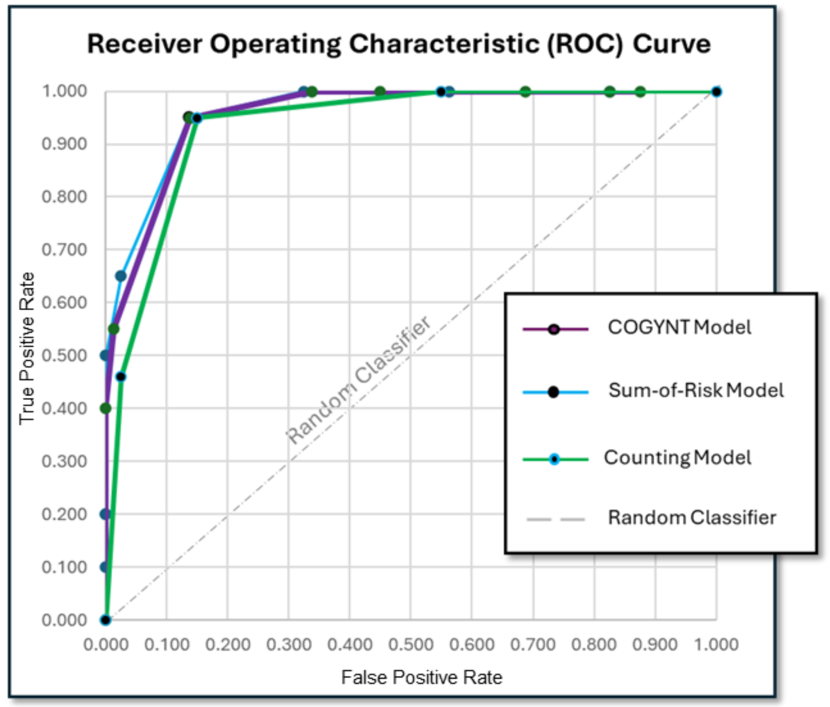
COGYNT HIERARCHICAL COMPLEX EVENT PROCESSING MODEL



Precision/Recall METRICS

Precision	0.63
Recall (Hit or True Positive)	0.95
False Positive	0.14
False Negative	0.05
F1	**0.76**

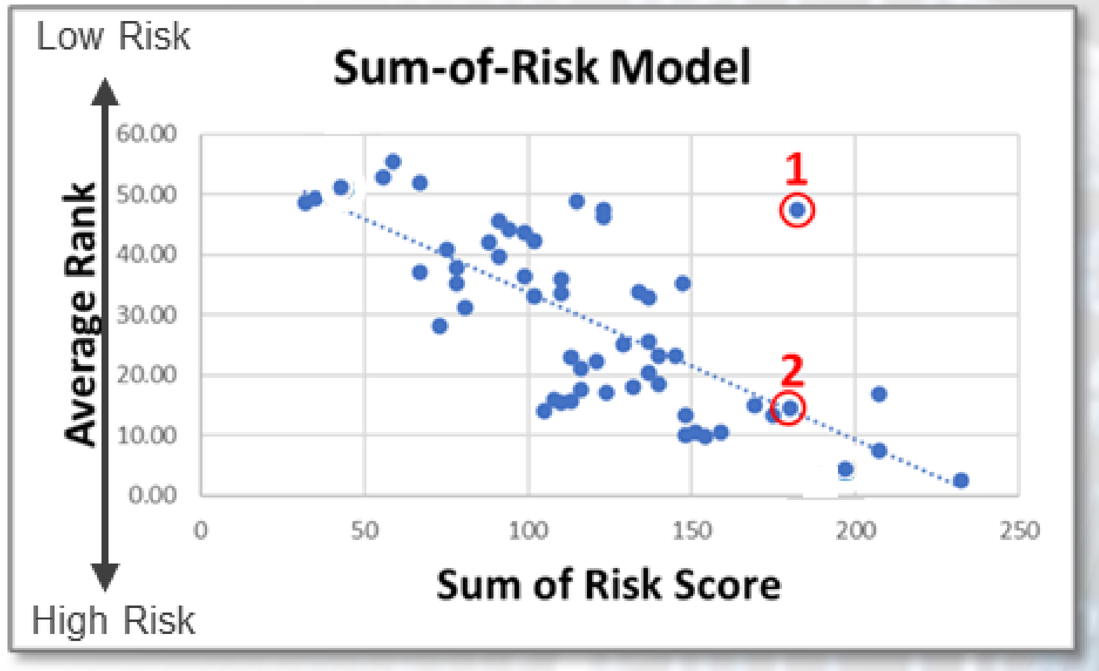
EACH OF THESE MODELS EXHIBIT PERFORMANCE LIMITATIONS



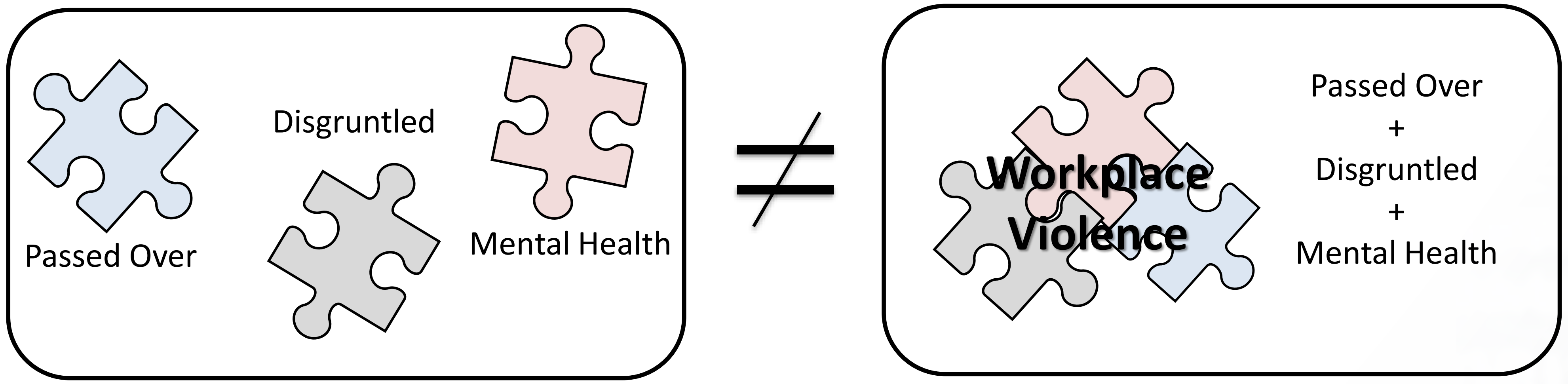
Precision ~ 0.6
Recall ~ 0.95
F1 ~ 0.75

Possible Reason: PRIs interact! They do not always contribute independently to risk

- Most computational risk modeling approaches assume that PRIs contribute independently to risk
- Research suggests that certain combinations of PRIs (***PATTERNS***) yield expert judgments of threat that are not consistent with this "independence" assumption.



Greitzer & Purl (2022)



"The whole is not equal to the sum of its parts."

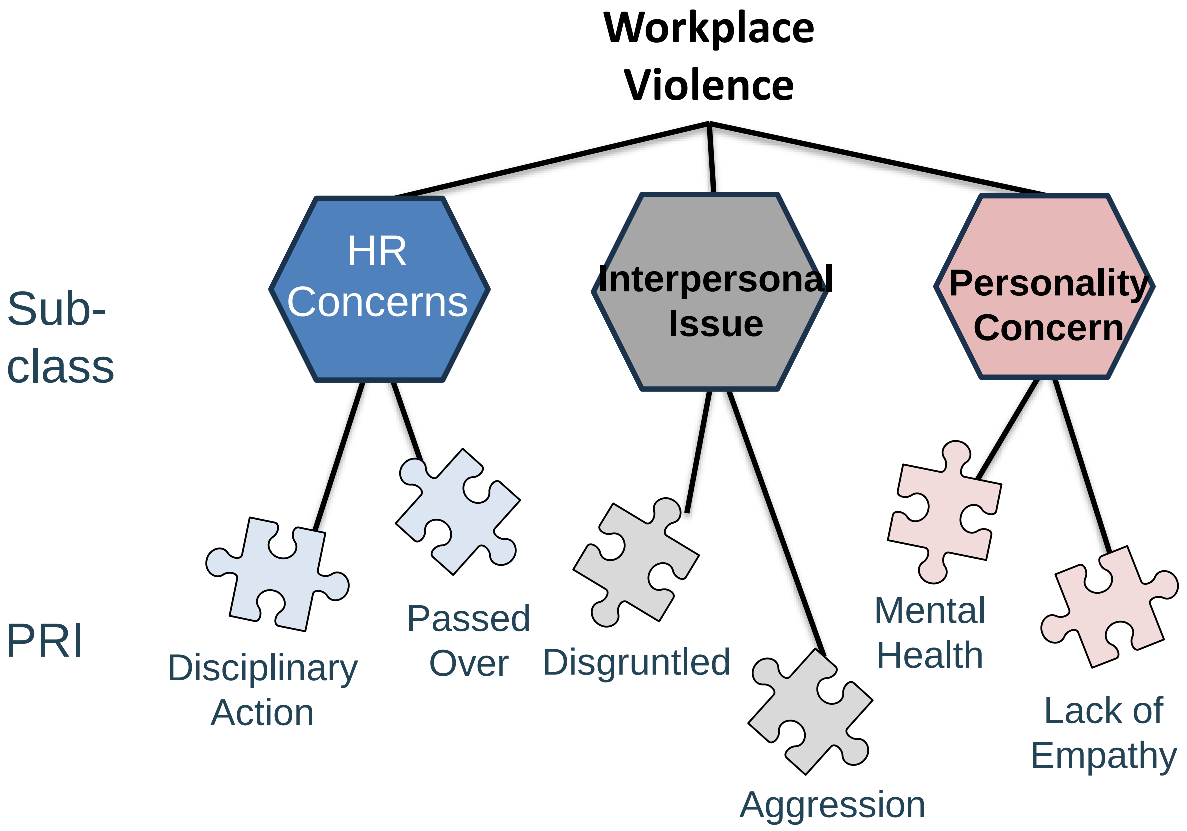
We need to account for PRI Patterns…

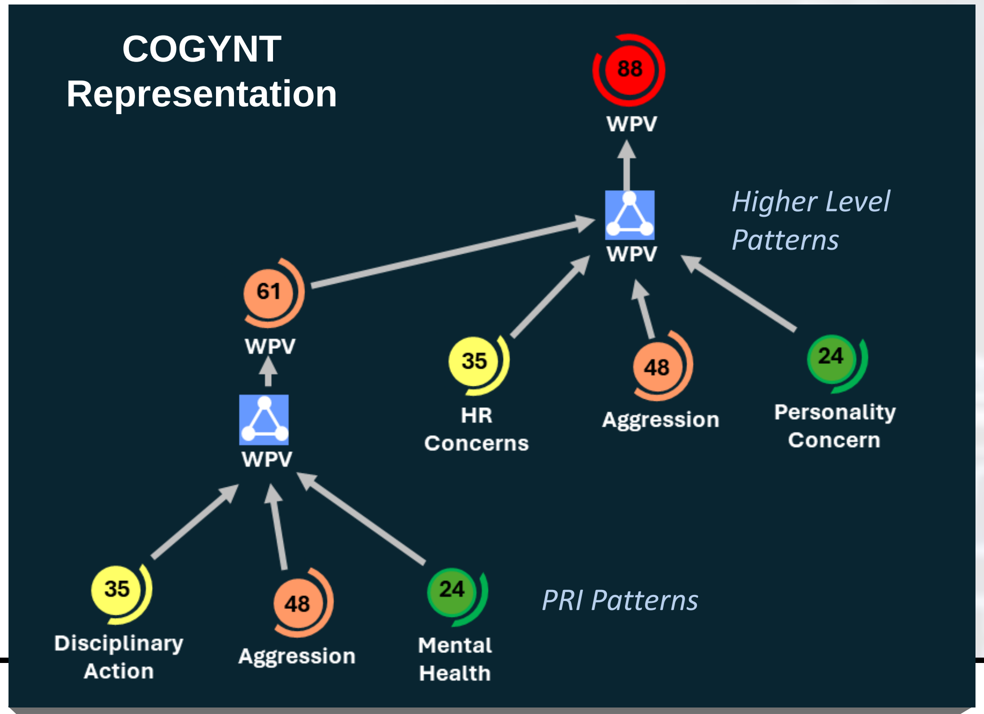
PATTERN PROCESSING APPROACH

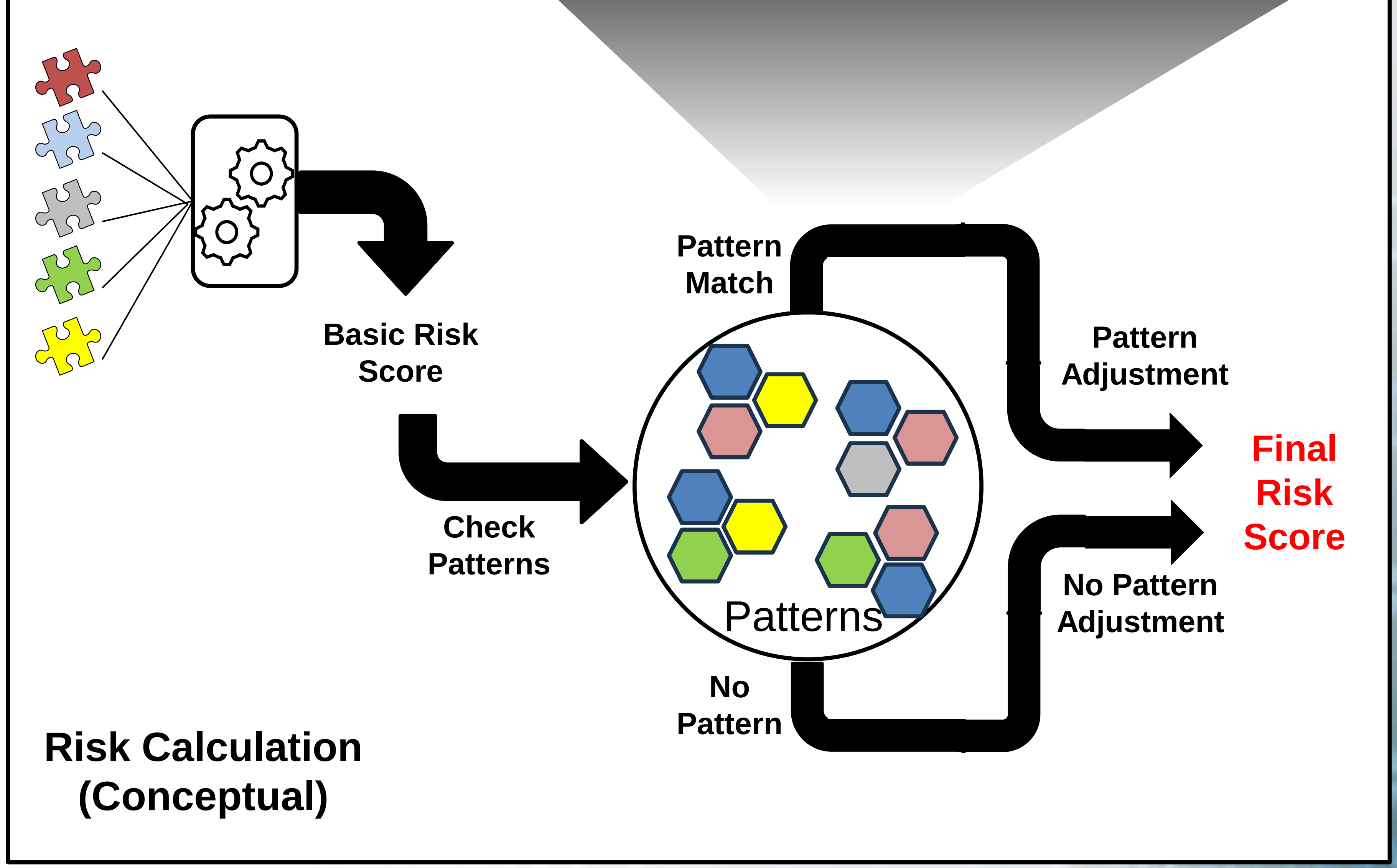
- *Bottom-Up* Examine all possible patterns…
  Power Set Limitation: 2^N patterns!

  With a set of 100 PRIs, the number of patterns is 2^{100} = 1,267,650,600,228,229,401,496,703,205,376
  Even if we limit patterns to at most 5 PRIs, the number of combinations (patterns) is 79,375,495!

- ***Top-Down***: Define patterns for behaviors based on PRI sub-classes

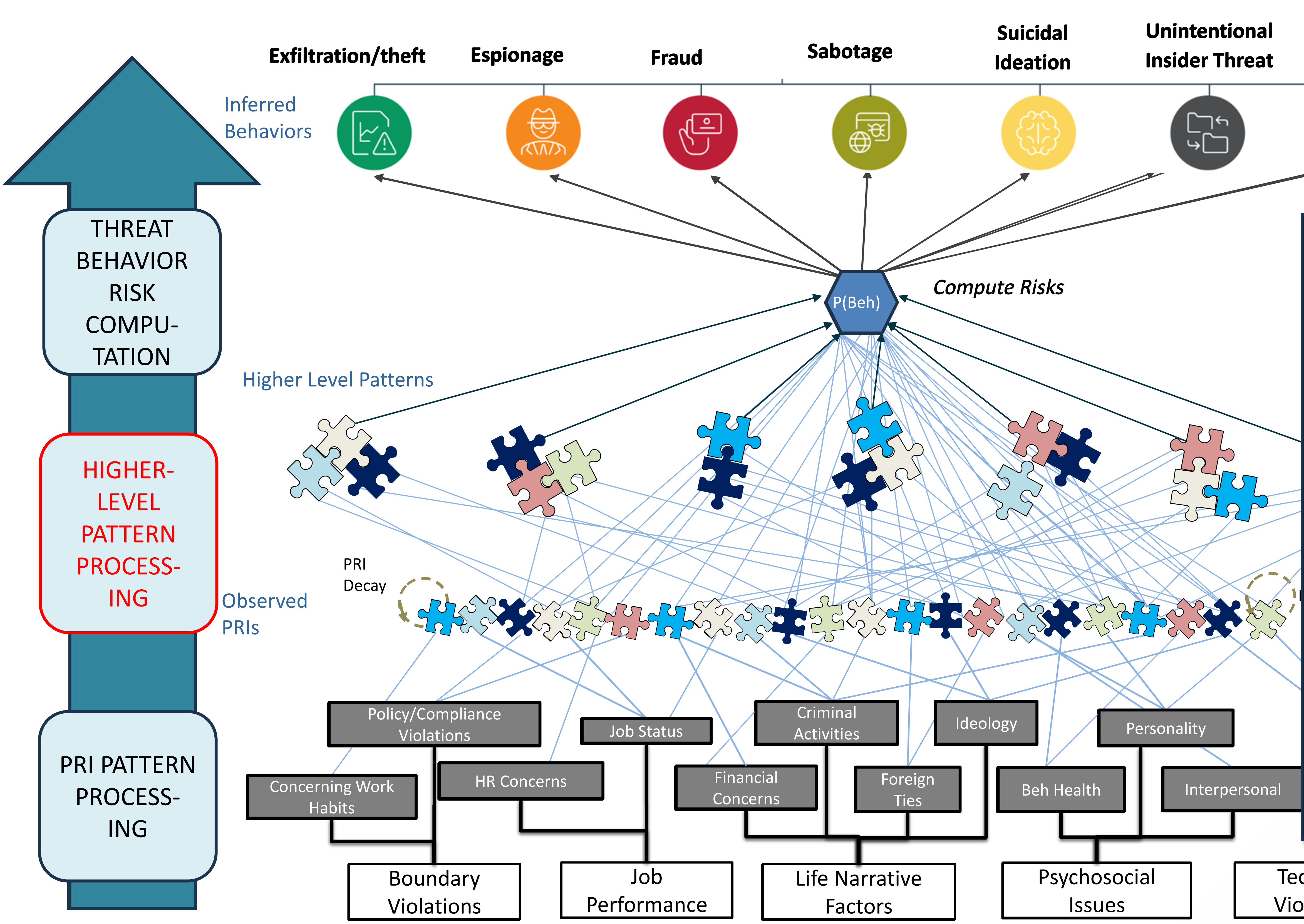


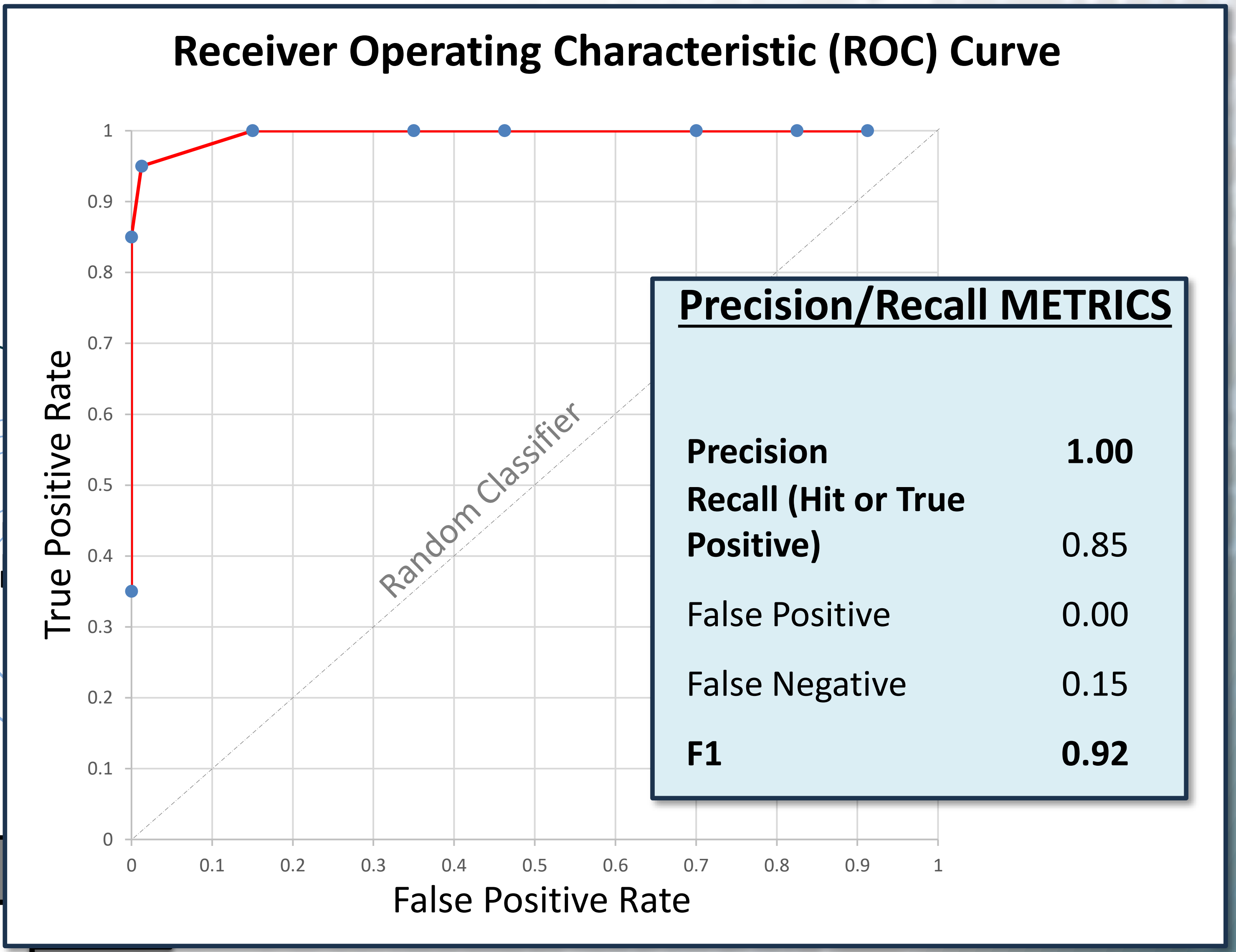




COGYNT Enhanced HIERARCHICAL COMPLEX EVENT PROCESSING

Cogynt Enhanced Model processes patterns at a higher level of abstraction

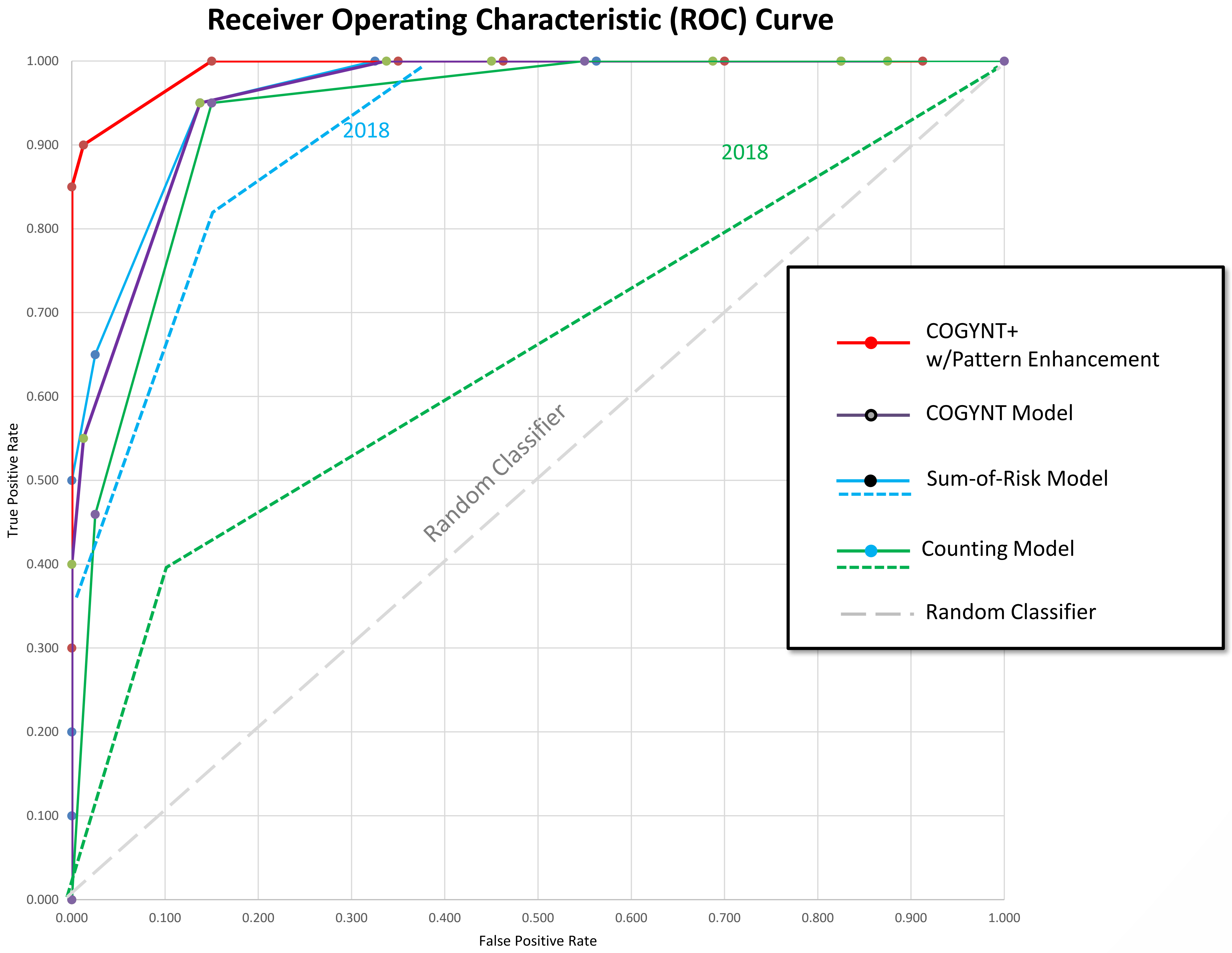




Precision/Recall METRICS

Precision	**1.00**
Recall (Hit or True Positive)	0.85
False Positive	0.00
False Negative	0.15
F1	**0.92**

INCREMENTAL IMPROVEMENTS



Model Comparisons

Metric	COGYNT+	COGYNT (Basic)	Sum-of-Risk	Counting Model
Precision	1.00	0.63	0.63	0.61
Recall (Hit or True Positive)	0.85	0.95	0.95	0.95
False Positive	0.00	0.14	0.14	0.15
False Negative	0.15	0.05	0.05	0.05
F1	**0.92**	**0.76**	**0.76**	**0.75**

- *Precision: Out of all the cases predicted to be threats, what percentage was a TRUE threat?*
- *Recall: Out of all the TRUE threats, what percentage was predicted to be threats?*

$$F1 = \text{harmonic mean of precision and recall} = \frac{2 \text{ x } (\text{Precision x Recall})}{\text{Precision} + \text{Recall}}$$

CONCLUSIONS

What we’ve learned:

- SOFIT PRI ontology provides a solid framework for characterizing and cataloguing risk indicators and contributing factors for insider threat
- PRIs vary in their degree of association with different insider threat behavior types
- PRIs vary in their spans of influence on risk judgments —models may apply different “rates of decay”
- Estimating PRI “weights” or probabilities requires a careful expert knowledge elicitation methodology to avoid “contamination” by different mindsets
- Most predictive models assume that PRIs do not “interact” – that they independently contribute to risk judgments. This lack of pattern processing may limit the effectiveness of predictive models that fail to capture complex PRI patterns, relationships, and interactions
- The enhanced Cogynt model provides a more robust threat assessment paradigm that reflects the complex hierarchical structure used by expert analysts when solving this problem
- These insights and associated research efforts have produced continual improvements.

PATH FORWARD

- There is a strong synergy between the hierarchical nature of the SOFIT PRI knowledge base and the Hierarchical Complex Event Processing (HCEP) capability of Cogility's COGYNT continuous intelligence platform
- Ongoing research with Cogility has led to enhancements in our threat assessment approach that exploit the pattern-based/HCEP processing capabilities of the COGYNT model – enabling us to develop models that reflect more complex PRI patterns, relationships, and interactions
- We're continuing to develop and test these advanced concepts:
  - Refining PRI hierarchical structure
  - Studying PRI calibration methods
  - Testing and evaluating PRI decay models
  - Defining, implementing, and testing pattern processing at higher levels of abstraction

References

Greitzer, FL, R Kliner, & S Chan. (2022). Temporal effects of contributing factors in insider risk assessment: Insider Threat indicator decay characteristics. *ACSAC WRIT Workshop*, December 2022, Austin, TX. https://www.acsac.org/2022/workshops/writ/WRIT_2022_paper_4765-Greitzer.pdf

Greitzer, FL, & J Purl. (2022). The dynamic nature of insider threat indicators. *Springer Nature Computer Science*, *3(102).* https://doi.org/10.1007/s42979-021-00990-1.

Greitzer, FL, J Purl, YM Leong & DE Becker. (2018). SOFIT: Sociotechnical and Organizational Factors for Insider Threat. IEEE Security and Privacy Workshops (SPW), Workshop on Research for Insider Threat (WRIT), San Francisco, CA, May 24, 2018, pp. 197-206. DOI: 10.1109/SPW.2018.00035 http://conferences.computer.org/sp/2018/Resources/spw/2018/SOFITSociotechnicalandOrganizationalFact.pdf

Greitzer FL & DA Frincke. (2010). Combining traditional cyber security audit data with psychosocial data: Towards predictive modeling for insider threat mitigation. In *Insider Threats in Cyber Security*, CW Probst, J Hunter, D Gollmann & M Bishop (Eds.), pp. 85-113. Springer, New York. https://link.springer.com/chapter/10.1007/978-1-4419-7133-3_5

Questions?

Thank you for your attention

Frank L. Greitzer, PhD

Frank@PsyberAnalytix.com

fgreitzer@cogility.com

For more information about COGYNT, please contact:

Stuart Booth, sbooth@cogility.com

Andrew Davis, adavis@cogility.com

www.cogility.com