A Systematic Protocol for Retrospective CPF Analysis of Cybersecurity Incidents: Methodology and Demonstration Cases

A PREPRINT

Giuseppe Canale, CISSP
Independent Researcher
g.canale@cpf3.org
URL: cpf3.org
ORCID: 0009-0007-3263-6897

September 30, 2025

Abstract

We present a systematic protocol for retrospective interpretive analysis of cybersecurity incidents through the Cybersecurity Psychology Framework (CPF). While prospective validation requires longitudinal studies, structured coding of publicly documented breaches offers insights into patterns of psychological factors appearing in incident narratives. This paper establishes a rigorous methodology for interpreting human factors in incident reports, mapping them to CPF's 100 indicators across 10 psychological categories, and identifying potential intervention opportunities. We demonstrate the protocol through interpretive analysis of three major incidents: the 2020 Twitter compromise, the 2021 Colonial Pipeline ransomware attack, and the 2020 SolarWinds supply chain breach, using publicly available documentation. Our protocol addresses inter-coder agreement, bias mitigation, and ethical considerations while providing a replicable framework for security researchers. **While retrospective analysis of constructed narratives cannot establish causal relationships or generate predictive metrics**, this structured interpretive approach enables identification of patterns in how psychological factors are documented in breach reports, supporting hypothesis generation for prospective validation studies and providing a heuristic framework for security culture assessment. This work establishes retrospective CPF analysis as a valid qualitative research method, bridging the gap between psychoanalytic theory and cybersecurity practice.

Keywords: cybersecurity methodology, incident analysis, psychological vulnerabilities, CPF protocol, retrospective analysis, human factors, qualitative methods

1 Introduction

The Cybersecurity Psychology Framework (CPF) proposes that pre-cognitive psychological states create exploitable vulnerabilities in organizational security [3]. However, prospective validation through longitudinal studies faces practical barriers: long timescales, participant recruitment challenges, and ethical constraints on creating vulnerable conditions. Retrospective interpretive analysis of documented incidents offers an alternative pathway for exploring CPF's applicability and generating hypotheses for future validation.

This paper establishes a systematic protocol for retrospective CPF analysis that:

1. Provides rigorous methodology for coding publicly available incident reports
2. Maps human factors to specific CPF indicators with defined reliability measures
3. Demonstrates practical application through three detailed case studies
4. Establishes standards for replication and comparative studies
5. Addresses ethical and methodological challenges in retrospective research

1.1 The Need for Methodological Rigor

Retrospective analysis faces inherent challenges:

Hindsight Bias: Post-hoc analysis may artificially inflate apparent predictability [5].

Incomplete Information: Public reports omit sensitive organizational details.

Narrative Reconstruction: Incident narratives are constructed artifacts reflecting organizational, legal, and political motivations rather than objective records [13]. What we analyze is narrative construction choices, not direct access to psychological states.

Selection Bias: Only certain incidents receive detailed public documentation.

Interpretive Nature: CPF coding is fundamentally qualitative interpretation, not quantitative measurement. Severity ratings reflect narrative emphasis on psychological factors, not objective vulnerability levels.

Our protocol addresses these challenges through explicit bias mitigation strategies, inter-coder reliability measures, and transparent limitation acknowledgment.

1.2 Epistemological Position

Critical clarification: This protocol produces *structured interpretations* of incident narratives, not *measurements* of organizational psychological states. The severity ratings represent coder judgments about how strongly psychological factors appear in available documentation, constrained by:

- What report authors chose to emphasize
- Legal and organizational filtering of information
- Post-hoc reconstruction of events
- Limited access to actual organizational dynamics

Therefore, CPF severity rating sums reflect "documented emphasis on psychological factors in breach narratives" rather than "actual psychological vulnerability levels." This distinction is maintained throughout the paper.

1.3 Contribution

This paper makes four primary contributions:

1. **Systematic coding protocol** for mapping incident narratives to CPF indicators with explicit interpretive procedures
2. **Reliability framework** including inter-coder agreement measures and calibration procedures
3. **Demonstration analyses** of three major incidents showing protocol application
4. **Research standards** enabling replication and comparative qualitative studies

The protocol enables researchers to conduct structured CPF interpretations without requiring organizational access or prospective data collection, supporting hypothesis generation for future validation efforts.

2 The CPF Framework: Complete Taxonomy

[Note: This section remains identical to original - full taxonomy of 100 indicators across 10 categories. Content omitted here for brevity but unchanged in actual revision.]

3 The Retrospective Analysis Protocol

3.1 Protocol Overview

The protocol comprises six phases:

1. **Incident Selection**: Identifying analyzable incidents with sufficient documentation
2. **Document Collection**: Gathering all available public sources
3. **Narrative Extraction**: Systematically extracting human factor mentions
4. **CPF Coding**: Mapping extracted factors to specific indicators through interpretive judgment
5. **Severity Rating**: Assigning interpretive ratings (Green/Yellow/Red) based on narrative evidence strength
6. **Analysis and Interpretation**: Identifying patterns within methodological constraints

3.2 Phase 1: Incident Selection Criteria

[Content remains identical to original]

3.3 Phase 2: Systematic Document Collection

[Content remains identical to original]

3.4 Phase 3: Narrative Extraction Protocol

[Content remains identical to original]

3.5 Phase 4: CPF Indicator Mapping

Map each extracted segment to specific CPF indicators through structured interpretive judgment.

Mapping Rules:

1. Each extracted segment may map to multiple indicators
2. Assign primary indicator (strongest interpretive match) and secondary indicators (contributory)
3. Use indicator definitions precisely - avoid forcing fits
4. When ambiguous, code conservatively
5. Document rationale for non-obvious mappings
6. **Acknowledge interpretive nature:** Mappings represent coder judgment about which psychological constructs best explain documented behaviors, not objective determination of psychological causation

[Rest of section content remains identical to original]

3.6 Phase 5: Severity Rating Assignment

Assign severity rating to each indicator based on narrative evidence strength.

Rating Definitions:

GREEN (0): Minimal Documentation

- No evidence of this vulnerability in available narratives
- OR explicit narrative evidence of organizational resilience
- OR effective controls documented
- *Note: Absence in narrative ≠ absence in reality*

YELLOW (1): Moderate Narrative Emphasis

- Implied or indirect narrative evidence
- Single occurrence documented
- Vulnerability mentioned but mitigating factors also present

- *Note: Rating reflects documentation emphasis, not actual severity*

RED (2): Strong Narrative Emphasis

- Explicit documentation in multiple sources
- Multiple occurrences or described as systematic pattern
- Narratively linked to breach causation
- No effective mitigation mentioned in available reports
- *Note: Reflects how strongly factor appears in narratives; subject to hindsight bias*

Critical Limitation: These ratings quantify *narrative emphasis* not *ground truth severity*. A factor rated GREEN might have been critically important but undocumented; a factor rated RED might be narratively over-emphasized due to hindsight bias or liability management.

Rating Process:

For each indicator mapped:

1. Review all evidence segments mapped to this indicator 2. Assess strength and clarity of *narrative* evidence 3. Consider organizational context as documented 4. Assign rating with documented justification 5. Flag ambiguous cases for discussion

Inter-Coder Reliability:

To ensure interpretive consistency:

- Minimum two independent coders
- Blind coding (coders don't see each other's work initially)
- Calculate Cohen's Kappa for agreement
- Resolve discrepancies through consensus discussion
- Maintain audit trail of resolution reasoning

Target reliability: Kappa greater than 0.70 (substantial agreement)

3.7 Phase 6: Analysis and Interpretation

Descriptive Analysis:

- Frequency distribution of coded indicators
- Severity rating profile across categories
- Pattern identification (which indicators co-occur in narratives)
- Timeline analysis (vulnerability progression as documented)

Interpretive Analysis:

- Identify primary vulnerability themes in narratives

- Examine convergence patterns (Category 10)
- Map documented vulnerabilities to attack chain
- Identify narratively-evident missed intervention opportunities

Limitation Acknowledgment:

Every analysis must explicitly address:

- Information completeness and source reliability
- Hindsight bias potential in retrospective interpretation
- Alternative explanations for documented patterns
- Generalizability constraints
- Distinction between narrative construction and organizational reality

[Bias Mitigation section remains identical to original]

4 Demonstration Case 1: Twitter Compromise (July 2020)

[Note: Case content remains substantively identical, but with systematic language changes throughout. Example excerpts:]

4.1 Severity Summary: Twitter Compromise

Table 1: CPF Interpretive Coding: Twitter Compromise

Category	**RED**	**YELLOW**	**GREEN**	**Rating Sum**
Authority [1.x]	3	1	6	7/20
Temporal [2.x]	1	2	7	4/20
Social Influence [3.x]	1	1	8	3/20
Affective [4.x]	1	1	8	3/20
Cognitive Load [5.x]	1	1	8	3/20
Group Dynamics [6.x]	2	0	8	4/20
Stress Response [7.x]	1	1	8	3/20
Unconscious [8.x]	0	1	9	1/20
AI Bias [9.x]	0	0	10	0/20
Convergent [10.x]	2	0	8	4/20
Total	**12**	**8**	**80**	**32/200**

Primary Narrative Pattern: Authority-based vulnerabilities dominate documented factors (7/20), followed by convergent state indicators (4/20) and group dynamics (4/20).

Key Interpretive Finding: Twitter breach narratives emphasize psychological convergence rather than technical failure—authority impersonation exploited during documented high-stress pandemic period with diffused responsibility and acknowledged inadequate training.

[Similar revisions applied to Colonial Pipeline and SolarWinds sections]

5 Cross-Case Comparative Analysis

5.1 Narrative Pattern Comparison

Table 2: Comparative CPF Severity Rating Sums Across Three Incidents

Category	**Twitter 2020**	**Colonial 2021**	**SolarWinds 2020**
Authority [1.x]	7	6	4
Temporal [2.x]	4	6	6
Social Influence [3.x]	3	2	1
Affective [4.x]	3	3	3
Cognitive Load [5.x]	3	2	4
Group Dynamics [6.x]	4	6	8
Stress Response [7.x]	3	4	1
Unconscious [8.x]	1	3	5
AI Bias [9.x]	0	0	1
Convergent [10.x]	4	6	8
Total Rating Sum	**32**	**38**	**41**
RED Count	**12**	**18**	**18**

5.2 Pattern Insights

Common Narrative Emphases Across All Cases:

- **Temporal vulnerabilities [2.x]:** All incident narratives emphasize temporal discounting—present operational demands prioritized over future security needs
- **Convergent states [10.x]:** All case narratives describe "perfect storm" conditions with multiple vulnerabilities aligning
- **Group dynamics [6.x]:** Collective processes (splitting, diffusion of responsibility, basic assumptions) documented in all cases

Incident-Specific Narrative Patterns:

[Content continues with similar language adjustments throughout]

5.3 Rating Sum Pattern Observation

The three cases show increasing CPF severity rating sums (32, 38, 41), which in these specific cases appear to track with:

- Documented attack sophistication (social engineering to supply chain compromise)
- Impact scope (single company to critical infrastructure to global supply chain)
- Detection difficulty as reported (hours to days to months)
- Recovery complexity as described (account resets to operational restart to supply chain remediation)

Critical caveat: With only three cases, this is an observed pattern, not a validated relationship. Whether CPF severity rating sums have any predictive relationship to breach outcomes requires prospective validation with large sample sizes. This observation generates a hypothesis for future testing, not a confirmed finding.

6 Methodological Validation

6.1 Inter-Coder Reliability Assessment

Two independent coders analyzed all three cases using the protocol. Results:

Table 3: Inter-Coder Reliability Metrics

Case	Cohen's Kappa	Category Agreement	Severity Agreement	Exact Match
Twitter	0.83	91%	87%	76%
Colonial	0.79	88%	84%	71%
SolarWinds	0.76	85%	81%	68%
Overall	**0.79**	**88%**	**84%**	**72%**

All Kappa values exceed 0.70 threshold for "substantial agreement," validating protocol's interpretive consistency. SolarWinds showed slightly lower agreement due to greater interpretive complexity and less explicit documentation.

Common Discrepancies:

- Category [8.x] (Unconscious Processes): Most disagreement due to highly inferential nature—these indicators require deep interpretation of limited evidence
- RED vs YELLOW severity: Close cases required consensus discussion
- Secondary indicator selection: High agreement on primary, variability on secondary

[Resolution process section remains identical]

6.2 Construct Validity Considerations

Convergent Consistency:

CPF interpretations align with expert post-mortem emphases:

- Twitter: Official reports emphasized social engineering and pandemic stress—consistent with Authority [1.x] and Stress [7.x] coding patterns
- Colonial: Experts highlighted organizational culture issues—consistent with Group Dynamics [6.x] coding patterns
- SolarWinds: Analysts emphasized trust and complexity—consistent with Affective [4.3] and Cognitive [5.x] coding patterns

Important limitation: This represents convergent *narrative emphasis* not independent validation. CPF coders read the same reports as post-mortem analysts, so alignment may reflect shared source material rather than CPF capturing independent psychological reality.

Discriminant Patterns:

CPF identifies psychological factors as distinct narrative themes from technical factors:

- SolarWinds technical sophistication was high, but documented psychological vulnerabilities (weak passwords, ignored warnings) appear as independent narrative elements
- Colonial technical deficiency (no MFA) was narratively explained through psychological factors (temporal discounting), suggesting psychological interpretation of technical gaps

6.3 Limitations and Bias Acknowledgment

Fundamental Epistemological Limitations:

Narrative vs. Reality: We analyze *constructed narratives about incidents*, not *incidents themselves*. Incident reports serve organizational, legal, and political functions. They are filtered, sanitized, and shaped by:

- Liability concerns (minimizing negligence, shifting blame)
- Regulatory requirements (emphasizing compliance efforts)
- Public relations objectives (protecting reputation)
- Incomplete information (sensitive details omitted)
- Hindsight reconstruction (post-hoc sense-making)

Therefore, **CPF severity ratings measure "how strongly psychological factors are emphasized in breach narratives" not "actual psychological vulnerability levels in organizations."** A critically important psychological factor may be entirely absent from public narratives if organizations chose not to disclose it. Conversely, minor factors may be overemphasized to explain failures.

Information Asymmetry: Public reports omit sensitive organizational details. Our interpretations are necessarily incomplete and potentially systematically biased toward information organizations were willing to disclose.

Hindsight Bias: Despite mitigation efforts, post-hoc analysis inevitably incorporates outcome knowledge. Coders know breach occurred, potentially inflating perceived vulnerability in narratives. We cannot definitively distinguish "factors that actually caused breach" from "factors that narrators emphasized post-hoc to explain breach."

Selection Bias: Only incidents with detailed public documentation are analyzable. This biases toward:

- High-profile breaches of large organizations
- Incidents triggering regulatory disclosure requirements
- Cases with legal proceedings creating public records
- Organizations prioritizing transparency

Incidents without public documentation may have entirely different psychological patterns.

Cultural Context: All three cases involve US-based organizations in 2020-2021. Cross-cultural and temporal generalizability unknown.

Category-Specific Limitations:

Category 8 (Unconscious Processes): These indicators require deep psychoanalytic interpretation that public documents cannot reliably support. Coding [8.1] Shadow Projection or [8.6] Defense Mechanisms from incident reports involves substantial interpretive inference. Alternative explanations (rational but incorrect decisions, resource constraints, technical limitations) often equally plausible. These ratings have lowest inter-coder reliability and highest uncertainty.

Weighting Arbitrariness: The choice of RED=2, YELLOW=1, GREEN=0 is methodologically arbitrary. We have not established that RED indicators are exactly twice as important as YELLOW, nor that different categories contribute equally to risk. The summed "rating totals" (e.g., 32/200) should be interpreted as ordinal rankings within this study, not interval or ratio scales with meaningful quantitative properties.

Linear Summation Assumption: Adding severity ratings across categories assumes independent, additive contributions. In reality, psychological vulnerabilities likely interact non-linearly. The total rating sums are heuristic summary statistics, not validated risk metrics.

Case-Specific Limitations:

Twitter: Limited detail about pre-pandemic organizational state makes temporal vulnerability interpretation partially speculative. Work-from-home impacts are inferred from general pandemic context, not Twitter-specific documentation.

Colonial Pipeline: Congressional testimony potentially influenced by liability concerns and may present sanitized narrative emphasizing "organizational culture issues" over individual accountability. Security-operations split may be narrat ively exaggerated.

SolarWinds: Ongoing investigations and litigation at time of public reporting may have limited disclosure. The "solarwinds123" password detail, while widely reported, may not reflect full authentication failure context. Attribution of decisions to "executive" versus other levels involves interpretation.

6.4 Researcher Positionality Statement

Reflexivity Acknowledgment: The author created the CPF framework being applied in this protocol, representing potential confirmation bias. As framework creator, there is intellectual investment in demonstrating CPF's applicability and utility. Mitigation strategies employed:

- Independent second coder with no prior CPF familiarity
- Blind coding procedures where feasible
- Explicit search for disconfirming evidence (GREEN ratings when resilience documented)
- Conservative coding when ambiguous
- Transparent documentation of interpretive reasoning

However, author's deep familiarity with CPF constructs inevitably shapes how narratives are interpreted. Alternative psychological frameworks applied to the same incidents might yield different patterns.

6.5 Ethical Considerations

[Content remains identical to original]

7 Discussion

7.1 Protocol Utility

The retrospective analysis protocol demonstrates:

Feasibility: Three major incidents coded using only public information, producing detailed structured interpretations.

Consistency: Inter-coder agreement (Kappa = 0.79) indicates protocol can be applied with substantial consistency across analysts.

Coherence: CPF interpretations align with expert post-mortem narrative emphases while providing additional psychological depth.

Heuristic Value: Protocol identifies potential intervention opportunities evident in narratives.

What This Protocol Does NOT Demonstrate:

- That CPF "measures" objective psychological vulnerabilities
- That CPF severity rating sums predict breach probability or severity
- That psychological factors identified were actual causes (vs. post-hoc narrative explanations)
- That CPF is superior to alternative psychological frameworks

The protocol's value lies in providing *structured interpretive consistency* for qualitative analysis, not quantitative measurement.

7.2 Theoretical Implications

Pre-Cognitive Vulnerabilities Documented:

All three case narratives describe psychological vulnerabilities operating below documented conscious awareness:

- Twitter employees reportedly unaware of stress-induced judgment degradation
- Colonial Pipeline leadership narratively portrayed as unaware of temporal discounting bias
- SolarWinds organization described as unaware of collective defense mechanisms

This narrative pattern is consistent with CPF's core thesis that pre-cognitive processes influence security failures, generating hypotheses for prospective validation.

Convergence Theme Apparent:

All case narratives emphasize "perfect storm" conditions (Category [10.x]) where multiple vulnerabilities aligned. Single vulnerabilities alone presented as insufficient—convergence described as enabling catastrophic outcomes.

Psychoanalytic Concepts Narratively Present:

Kleinian splitting (Colonial security vs operations), Bionian basic assumptions (SolarWinds dependency on trust), Jungian shadow projection (external threat emphasis) all appear as interpretable patterns in cybersecurity incident narratives, suggesting potential applicability of psychoanalytic frameworks.

Group Dynamics Narrative Emphasis:

Categories [6.x] and [10.x] consistently receive high ratings across cases, suggesting incident narratives emphasize organizational-level psychology over individual psychology in explaining major breaches.

7.3 Practical Applications

Proactive Heuristic Assessment:

Organizations could adapt protocol for self-assessment:

1. Conduct CPF-informed organizational observation
2. Identify categories with potential vulnerabilities
3. Consider interventions addressing psychological patterns
4. Reassess periodically

Important caveat: Without prospective validation, we cannot claim such assessments predict breach probability. They provide structured reflection on organizational psychology, not risk quantification.

Security Architecture Informed by Psychology:

Protocol suggests design considerations:

- **Authority vulnerabilities:** Design verification mechanisms resilient to authority pressure
- **Temporal vulnerabilities:** Build security architectures functional under deadline pressure
- **Convergent states:** Monitor for multiple simultaneous stressors

Incident Response Enhancement:

Retrospective protocol potentially applicable during incident response:

- Rapid CPF-informed reflection on psychological factors potentially affecting response
- Tailor communication considering organizational psychological context
- Post-incident CPF interpretation to identify narrative patterns

Security Awareness Evolution:

Move toward targeted interventions based on CPF patterns:

- Organizations showing authority vulnerability patterns: Question-authority training, anonymous reporting

- Organizations showing temporal patterns: Decision-making under pressure training
- Organizations showing group dynamic patterns: Team-based exercises, organizational psychology consultation

7.4 Research Directions

Large-Scale Retrospective Studies:

Apply protocol to 50-100 incidents to:

- Identify statistical patterns across incident types
- Examine whether rating sum patterns replicate
- Create sector-specific CPF narrative pattern benchmarks
- Test convergence theory prevalence quantitatively

Prospective Validation—Critical Next Step:

- Longitudinal studies: CPF-informed assessment of organizations, track incident rates over 24-36 months
- Test hypothesis: Organizations with identified CPF patterns experience elevated breach probability
- Control for technical security maturity to isolate psychological factors
- **This is required to move from "narrative interpretation" to "predictive assessment"**

Intervention Studies:

Test CPF-informed interventions:

- Randomized trials: CPF-targeted vs generic security awareness
- Measure incident rate reduction and near-miss detection improvements
- Identify most effective intervention types per vulnerability category

Cross-Cultural Validation:

Extend protocol to non-Western contexts:

- Test applicability in collectivist vs individualist cultures
- Examine how power distance affects Authority category manifestation
- Develop culturally-adapted indicator definitions where necessary

Alternative Framework Comparison:

Apply competing psychological frameworks to same incidents:

- Organizational behavior theory

- Human factors engineering (HFACS, STAMP)
- Safety science (Resilience Engineering, Safety-II)
- Cognitive systems engineering

Compare frameworks on interpretive coherence, intervention suggestions, inter-coder reliability. Assess CPF's unique contributions versus overlapping insights.

Automated Coding Development:

Develop NLP tools to:

- Automatically extract human factor segments from incident reports
- Suggest preliminary CPF mappings for human review
- Enable rapid large-scale retrospective studies
- Test whether machine coding achieves comparable reliability to human coding

This would make large-scale retrospective studies feasible.

7.5 Integration with Existing Frameworks

CPF provides psychological interpretive layer complementing technical frameworks:

Table 4: CPF Integration with Security Frameworks

Framework	Technical Focus	CPF Contribution
NIST CSF	Identify, Protect, Detect, Respond	Add psychological vulnerability layer
ISO 27001	Controls catalog	Human factor risk interpretation
MITRE ATT&CK	Adversary tactics	Psychological attack surface mapping
FAIR	Quantitative risk	Psychological loss event frequency factors
Zero Trust	Technical verification	Trust psychology considerations

Example Integration - NIST CSF:

Identify: Add CPF-informed organizational reflection to asset inventory phase

Protect: Design controls considering psychological vulnerabilities identified through CPF lens

Detect: Monitor for convergent psychological states (high stress + authority pressure + time pressure) alongside technical indicators

Respond: Tailor incident response considering organizational psychological context suggested by CPF

Recover: Address psychological factors in recovery to reduce recurrence risk

CPF doesn't replace these frameworks—it adds interpretive psychological dimension.

8 Conclusion

This paper establishes a systematic protocol for retrospective interpretive CPF analysis of cybersecurity incidents, providing a methodologically rigorous approach to qualitative coding of human factors in breach narratives.

Key Contributions:

Methodological Contribution: The six-phase protocol (selection, collection, extraction, mapping, severity rating, interpretation) provides replicable framework for structured interpretation of human factors in incident reports with substantial inter-coder reliability (Kappa = 0.79).

Empirical Patterns: All three analyzed incidents exhibited multiple RED-coded indicators across CPF categories, with severity rating sums ranging from 32-41. These cases showed consistent narrative emphasis on psychological factors alongside technical failures.

Pattern Identification: Temporal vulnerabilities [2.x], Group Dynamics [6.x], and Convergent States [10.x] received high ratings across cases, suggesting these categories represent common themes in breach narratives regardless of specific incident type. Whether this pattern replicates in larger samples requires further research.

Theoretical Coherence: Psychoanalytic concepts (splitting, projection, basic assumptions, defense mechanisms) and cognitive biases (temporal discounting, authority deference) appear as interpretable patterns in real-world breach documentation, supporting continued exploration of CPF's theoretical foundation through larger studies.

Heuristic Utility: Protocol identified narratively-documented missed intervention opportunities in each case—suggesting prospective CPF-informed assessment might enable preventive action, pending validation studies.

Limitations Transparently Acknowledged: Retrospective interpretation of constructed narratives faces inherent challenges (hindsight bias, information asymmetry, narrative filtering, arbitrary weighting) that explicit mitigation addresses but cannot eliminate. **Severity rating sums represent narrative emphasis patterns, not validated risk metrics or measurements of actual organizational psychological states.** Prospective validation with large samples remains essential to establish whether CPF patterns have predictive validity for actual breach outcomes.

The retrospective analysis protocol serves three functions:

Research Tool: Enables systematic qualitative study of psychological factor patterns in documented breaches, building interpretive evidence base for CPF exploration and generating hypotheses for prospective testing.

Learning Mechanism: Organizations can adapt protocol to interpret their own past incidents or study public cases to reflect on psychological vulnerability patterns, using it as structured discussion framework rather than measurement tool.

Foundation for Prospective Methods: Protocol developed here informs prospective CPF assessment methodology development, though prospective application requires additional validation before claiming predictive utility.

Critical Next Steps Required for Validation:

1. **Large-scale retrospective studies** (N=50-100 incidents) to test whether patterns observed in these three cases replicate across diverse incidents, or whether they reflect case selection artifacts

2. **Prospective longitudinal validation** to test whether organizations exhibiting CPF patterns at baseline actually experience elevated breach rates over 24-36 months—**this is essential to move from interpretive heuristic to predictive tool**
3. **Intervention effectiveness testing** to determine if CPF-informed interventions actually reduce incidents compared to generic security awareness or alternative frameworks
4. **Alternative framework comparison** to assess CPF's unique contributions versus overlapping insights with established organizational psychology, human factors, and safety science approaches
5. **Independent validation** by researchers not invested in CPF's success to mitigate confirmation bias inherent in creator-led validation

The protocol presented here makes retrospective qualitative research immediately feasible using publicly available information, generating hypotheses and demonstrating interpretive consistency. However, it does not validate CPF's predictive utility, causal explanations, or practical superiority to alternatives.

Core Insight from This Work: Major breach narratives consistently emphasize human and organizational psychological factors alongside technical failures. Temporal discounting, authority dynamics, group processes, and convergent conditions appear as recurring themes. Whether these narrative patterns reflect actual causal mechanisms operating pre-breach, represent post-hoc sense-making and blame diffusion, or combine both elements requires prospective validation that distinguishes narrative construction from organizational reality.

The field faces a clear empirical path forward: (1) apply this protocol to larger incident samples to identify robust narrative patterns and test replication, (2) develop prospective assessment methods that don't rely on post-breach narratives, (3) conduct longitudinal studies testing whether CPF patterns predict actual breach outcomes, (4) run intervention trials testing whether CPF-informed approaches outperform alternatives. Only after completing this validation pathway can CPF transition from interpretive heuristic to validated assessment tool with demonstrated predictive utility.

This protocol provides the methodological foundation for that transition while maintaining epistemological honesty about what retrospective interpretation of constructed narratives can and cannot establish. We offer a systematic way to think about psychological factors in cybersecurity incidents—not a measurement system, not a validated predictor, but a structured interpretive lens that may prove valuable pending rigorous prospective validation.

Data Availability Statement

The complete retrospective analysis protocol is described in this paper and can be applied using publicly available incident documentation. Detailed coding sheets and inter-rater reliability calculations for the three demonstration cases are available from the author upon reasonable request. Researchers interested in applying this protocol may contact the author for guidance and access to training materials.

Acknowledgments

The author thanks the organizations that prioritized transparency in their incident disclosures, enabling learning across the security community. Special thanks to the independent coder who

participated in reliability assessment and provided valuable critiques of the framework. Thanks also to reviewers who identified epistemological issues in earlier drafts, strengthening the final methodology.

Conflicts of Interest

The author is the creator of the CPF framework being applied, representing substantial potential for intellectual confirmation bias and professional investment in demonstrating CPF's utility. To mitigate this fundamental conflict, independent coding, blind procedures, and this extensive limitations section were employed. However, creator-led validation remains methodologically problematic. Independent replication by skeptical researchers is strongly encouraged. No financial conflicts exist.

Funding

This research received no external funding.

A Complete Indicator Reference

For researcher convenience, we provide the complete CPF taxonomy with brief definitions. Full definitions available in the CPF Foundation Paper [3].

Category 1: Authority-Based Vulnerabilities [1.x]

[1.1] Unquestioning authority compliance; [1.2] Responsibility diffusion; [1.3] Impersonation susceptibility; [1.4] Convenience bypass; [1.5] Fear-based compliance; [1.6] Reporting gradient; [1.7] Technical authority deference; [1.8] Executive exceptions; [1.9] Authority social proof; [1.10] Crisis authority escalation

Category 2: Temporal Vulnerabilities [2.x]

[2.1] Urgency bypass; [2.2] Time pressure degradation; [2.3] Deadline risk acceptance; [2.4] Present bias; [2.5] Hyperbolic discounting; [2.6] Exhaustion patterns; [2.7] Time-of-day windows; [2.8] Weekend/holiday lapses; [2.9] Shift change windows; [2.10] Temporal consistency pressure

Category 3: Social Influence Vulnerabilities [3.x]

[3.1] Reciprocity exploitation; [3.2] Commitment escalation; [3.3] Social proof; [3.4] Liking override; [3.5] Scarcity decisions; [3.6] Unity exploitation; [3.7] Peer pressure; [3.8] Conformity to insecure norms; [3.9] Identity threats; [3.10] Reputation conflicts

Category 4: Affective Vulnerabilities [4.x]

[4.1] Fear paralysis; [4.2] Anger risk-taking; [4.3] Trust transference; [4.4] Legacy attachment; [4.5] Shame hiding; [4.6] Guilt overcompliance; [4.7] Anxiety mistakes; [4.8] Depression negligence; [4.9] Euphoria carelessness; [4.10] Emotional contagion

Category 5: Cognitive Overload Vulnerabilities [5.x]

[5.1] Alert fatigue; [5.2] Decision fatigue; [5.3] Information overload; [5.4] Multitasking degradation; [5.5] Context switching; [5.6] Cognitive tunneling; [5.7] Working memory overflow; [5.8] Attention residue; [5.9] Complexity errors; [5.10] Mental model confusion

Category 6: Group Dynamic Vulnerabilities [6.x]

[6.1] Groupthink; [6.2] Risky shift; [6.3] Responsibility diffusion; [6.4] Social loafing; [6.5] Bystander effect; [6.6] Dependency assumptions; [6.7] Fight-flight postures; [6.8] Pairing fantasies; [6.9] Organizational splitting; [6.10] Collective defenses

Category 7: Stress Response Vulnerabilities [7.x]

[7.1] Acute stress impairment; [7.2] Chronic burnout; [7.3] Fight aggression; [7.4] Flight avoidance; [7.5] Freeze paralysis; [7.6] Fawn overcompliance; [7.7] Tunnel vision; [7.8] Memory impairment; [7.9] Stress contagion; [7.10] Recovery vulnerabilities

Category 8: Unconscious Process Vulnerabilities [8.x]

[8.1] Shadow projection; [8.2] Threat identification; [8.3] Repetition compulsion; [8.4] Transference; [8.5] Countertransference; [8.6] Defense mechanisms; [8.7] Symbolic equation; [8.8] Archetypal triggers; [8.9] Collective unconscious; [8.10] Dream logic

Note on Category 8: These indicators require substantial psychoanalytic interpretation and should be considered speculative when coded from public documents. Inter-coder reliability is lowest for this category. Consider flagging as "low inferential confidence" in applications.

Category 9: AI-Specific Bias Vulnerabilities [9.x]

[9.1] Anthropomorphization; [9.2] Automation bias; [9.3] Algorithm aversion; [9.4] AI authority; [9.5] Uncanny valley; [9.6] Opacity trust; [9.7] Hallucination acceptance; [9.8] Team dysfunction; [9.9] Emotional manipulation; [9.10] Fairness blindness

Category 10: Critical Convergent States [10.x]

[10.1] Perfect storm; [10.2] Cascade triggers; [10.3] Tipping points; [10.4] Swiss cheese alignment; [10.5] Black swan blindness; [10.6] Gray rhino denial; [10.7] Complexity catastrophe; [10.8] Emergence unpredictability; [10.9] Coupling failures; [10.10] Hysteresis gaps

Note on Category 10: Convergence emphasis may reflect narrative construction conventions rather than objective convergence patterns. Use cautiously.

B Coding Training Materials

For researchers adopting this protocol, we provide condensed training guidance.

Coder Qualifications:

Minimum requirements:

- Cybersecurity domain knowledge (CISSP, CEH, or equivalent experience)
- Familiarity with psychological concepts (undergraduate psychology or equivalent reading)
- Experience analyzing technical documentation
- Understanding of qualitative coding methodology

Recommended qualifications:

- Graduate training in psychology, organizational behavior, or related field
- Experience with qualitative coding methodologies (grounded theory, thematic analysis)
- Familiarity with psychoanalytic theory
- Training in reflexivity and bias awareness

Training Process:

Phase 1 (4 hours): Study CPF taxonomy, indicator definitions, theoretical foundations, epistemological limitations

Phase 2 (4 hours): Practice extraction on training case with answer key, discuss interpretive decisions

Phase 3 (4 hours): Practice mapping and severity rating with answer key, calibrate RED/YELLOW/GREEN thresholds

Phase 4 (4 hours): Code test case independently, compare with expert coding, discuss discrepancies and alternative interpretations

Certification: Achieve Kappa greater than 0.75 with expert coder on test case before independent coding

Common Training Challenges:

Over-interpretation: Coders tend to infer beyond evidence, especially for Category 8. Emphasize coding only what is explicitly or reasonably implied by documentation. When uncertain, code conservatively.

Confirmation bias: Coders may see vulnerabilities everywhere once trained on CPF. Actively search for and code GREEN when resilience documented. Challenge your initial interpretations—could alternative explanations fit equally well?

Category confusion: Indicators span multiple categories and boundaries are fuzzy. Use primary-secondary distinction. Accept that some segments legitimately map to multiple categories.

Severity calibration: RED threshold varies significantly by coder background. Require explicit narrative evidence for RED ratings. When in doubt between RED and YELLOW, choose YELLOW and document reasoning.

Unconscious Process indicators (Category 8): These require most interpretation and have lowest reliability. Train coders to acknowledge high uncertainty, consider multiple alternative explanations, and potentially skip these indicators when evidence is thin.

Narrative vs. reality confusion: Coders may forget they're analyzing narratives, not reality. Regularly remind: "This is what the report emphasizes, which may or may not reflect what actually happened."

Avoiding reflexivity: Coders resist acknowledging their own interpretive role. Build in regular reflexivity exercises: "How might my background/assumptions be shaping this interpretation?"

C Blockchain Timestamp

This protocol and demonstration analyses have been timestamped on the blockchain for intellectual property protection and reproducibility verification:

- **Platform**: OpenTimestamps.org
- **Document Hash**: [To be generated at final publication]
- **Block Height**: [To be recorded at final publication]
- **Bitcoin Transaction**: [To be recorded at final publication]
- **Timestamp**: [To be recorded at final publication]

Verification: https://opentimestamps.org

References

[1] Bion, W. R. (1961). *Experiences in groups*. London: Tavistock Publications.

[2] Bowlby, J. (1969). *Attachment and Loss: Vol. 1. Attachment*. New York: Basic Books.

[3] Canale, G. (2025). The Cybersecurity Psychology Framework: A Pre-Cognitive Vulnerability Assessment Model. Available at: https://cpf3.org

[4] Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York: Collins.

[5] Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288-299.

[6] Jung, C. G. (1969). *The Archetypes and the Collective Unconscious*. Princeton: Princeton University Press.

[7] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

[8] Kernberg, O. (1998). *Ideology, conflict, and leadership in groups and organizations*. New Haven: Yale University Press.

[9] Klein, M. (1946). Notes on some schizoid mechanisms. *International Journal of Psychoanalysis*, 27, 99-110.

[10] Milgram, S. (1974). *Obedience to authority*. New York: Harper and Row.

[11] Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review*, 63(2), 81-97.

[12] Selye, H. (1956). *The stress of life*. New York: McGraw-Hill.

[13] Woods, D. D., & Cook, R. I. (2010). Nine steps to move forward from error. *Cognition, Technology and Work*, 4(2), 137-144.