INVESTIGATING CYBER ATTACKER TEAM COGNITION

Craig J. Johnson[1], Kimberly J. Ferguson-Walter[2], Robert S. Gutzwiller[1], Dakota D. Scott[3], & Nancy J. Cooke[1]
[1]Arizona State University, [2]Laboratory for Advanced Cybersecurity Research, [3]Wichita State University

Cyber attackers commonly operate in teams, which may process information collectively and thus, may be best understood when the team is treated as the unit of analysis. Future research in Oppositional Human Factors (OHF) should consider the impact of *team-influencing* and *team-level biases* and the impact that defensive interventions have on team cognition in general. Existing measurement approaches using team interactions may be well suited for studying red teams, and how OHF interventions impact cyber attackers.

INTRODUCTION

Cyber-attacks are a costly and increasingly prevalent problem. Traditional defenses focus primarily on preventing or detecting security breaches, mainly through technical solutions. Recently, the cyber defense domain has been improved by adding human factors research and engineering, which focuses on the fit between the humans doing the work and technology, further improving defender effectiveness (Gutzwiller et al., 2015; Simonson et al., 2020). However, both are defender-centered views. Oppositional Human Factors (OHF; Gutzwiller et al., 2018), in contrast, proposes that cyber defense can also be improved by selectively reversing human factors advice and recommendations to create systems that disrupt the performance of attackers. Examples of OHF include implementing cyber and psychological deception (Ferguson-Walter, Major, et al., 2019; Ferguson-Walter, Major, et al., 2021) and inducing decision-making biases in cyber attackers (Cranford et al., 2021; Ferguson-Walter et al., 2017).

OHF research has thus far focused on individuals (Ferguson-Walter, Gutzwiller, et al., 2021). However, malicious attackers (and red teamers) often work within teams. Red teamers are groups of people brought together to attempt to break into networks and hack systems and hardware and report their findings as "good guys" to the tech owners so the vulnerabilities can be fixed. Both malicious attackers and red teamers, as teams, will possess their own dynamics which cannot be fully understood by studying individuals in isolation (Klein & Kozlowski, 2000; Mathieu & Luciano, 2019). Therefore, we must extend understanding of OHF to the team level to learn how to disrupt team actions.

This paper has two main aims. First, we explain and explore why, in contrast to disrupting individual cognition, different techniques and theories are required to disrupt cyber attacker team cognition and provide some examples of OHF at the team level. Second, we lay out the potential for the application of interaction-based methods and measurements (Cooke & Gorman, 2009) as a valuable approach to studying the dynamics of cyber attacker teams. Interaction approaches will, for example, help to study and augment the OHF defensive arsenal with team-level OHF. Interaction-based measures of team cognition can be adapted by researchers to aid understanding of the impact of various interventions on teams within experimental, and eventually real-world attacker scenarios. The overall benefit will be to allow defenders to measure if—and when—OHF techniques against attacker teams are succeeding. These team-level measures can also characterize red team behavior during capture the flag (CTF) and penetration testing exercises, including interactions with technology and teammates. Finally, these interaction-based measures are often unobtrusive, and provide the potential for real-time sensing in operational settings.

Cyber Attacker Team Cognition

A team is defined as "two or more individuals who have specific roles and interact adaptively, interdependently, and dynamically toward a common and valued goal" (Roberts et al., 2021). The interactions of team members with one another and with their technology result in emergent states and behaviors that can be qualitatively different from the sum of individual properties (Mathieu & Luciano, 2019). *Team cognition* characterizes how teams perform team-level cognitive activities such as planning, decision making, situation assessment, and collective action. Team cognition is the cognitive activity that occurs on the team level, and it plays an essential role in effective teamwork and team performance (Cooke et al., 2013; Mathieu et al., 2000). Team interaction (e.g., communication) is an essential component of team cognition (Cooke et al., 2013) and can provide an avenue to observe and measure it directly. Research has explored aspects of team interaction and team cognition in cyber defense teams (Buchler et al., 2018; Deline et al., 2021; Granåsen & Andersson, 2016; McNeese et al., 2012; Rajivan & Cooke, 2018). However, there is a lack of work specifically examining team cognition in cyber attackers.

There is a growing prevalence of organized cyber attackers operating in teams (Bulanova-Hristova et al., 2016). The time and space dependencies of cyber-attack teams vary, impacting the nature of their interactions. Some attack teams may operate synchronously, but others may work asynchronously or serially, handing off tasks between individuals or groups. Similarly, some attack teams may be collocated while others are geographically distributed (Meyers et al., 2009). The variability in cyber attacker team profiles can make it difficult to predict, influence, or disrupt their behavior. A variety of methods using different inputs may be necessary to understand their collective behavior.

Many organizations utilize red teams composed of "white hat" hackers who emulate cyber attackers, provide feedback, and identify security weaknesses to improve the security posture of the organization. Red teams have also been used to understand potential attacker behaviors. Other approaches can offer opportunities for understanding cyber attackers in action. Some Capture-the-flag (CTF) exercises pit "red team" attackers against "blue team" defenders in a game-like environment that simulates certain aspects of cyber attacker–defender dynamics (Buchler et al., 2018). Experimental settings further seek to

remove some of the confounds inherent to CTF exercises while still capturing the essence of hacker behavior (Ferguson-Walter, Major, et al., 2019). These studies still tend to focus on cyber defenders, while OHF research has examined cyber attacker performance in realistic experimental settings, with a focus on the role of cognitive biases that impact behavior by disrupting attention and resulting in confirmation biases and anchoring effects (Gutzwiller et al., 2018).

Team Cognitive Biases

Scaling OHF in the form of biases from the individual to the team level is a multi-relational idea (see Figure 1). Unfortunately, very little research exists on the impact of biases within teams and at the team level. Additionally, how cognitive biases scale from the individual to the team level is not clear, and in many cases, inconsistent (Kerr et al., 1996).

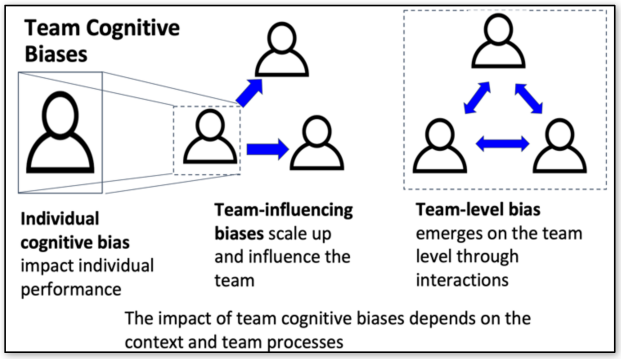

Figure 1. The multilevel relationship between *individual, team-influencing,* and *team-level* cognitive biases.

A cognitive bias is a systematic pattern of deviation from rationality in thinking that impacts a person's judgment or decision-making. An example is the *availability bias*, through which people assess that something is more likely or common because it is more easily brought to mind (Tversky & Kahneman, 1974). Traditionally, research has focused on reducing biases and decision-making errors, but OHF research examined the impact of biases on cyber attacker decision-making, intending to disrupt cyber attacker performance by inducing or intensifying biases that degrade attacker performance. For example, individual red team hackers operating as a team in a real network were found to show preliminary evidence of multiple biases, including *framing effects*, *the sunk-cost bias*, *irrational escalation*, *illusion of control*, the *take-the-best bias*, *confirmation bias*, and *anchoring bias* (Gutzwiller et al., 2018; Gutzwiller et al., 2019). Because cyber attackers commonly operate within teams, cognitive biases can cascade to impact the performance of the rest of the team. For example, a cognitive bias may influence a team member's decision (perhaps, but not necessarily, a leader), resulting in the team pursuing an erroneous course of action together. This type of bias is referred to in this work as a *team-influencing bias* (Figure 1). Other team biases may manifest exclusively on the team level referred to here as *team-level biases* (Figure 1; see Table 1 for examples).

Team-level biases are emergent from the interactions of individuals and cannot exist on the individual level. For example, the *information pooling bias*—in which people are more likely to share things everyone on the team knows and less likely to share unique information available only to them—must be described in terms that refer to at least two individuals. Rajivan and Cooke (2018) found evidence that this team-level bias degraded teamwork and reduced defender team performance in a cybersecurity task, though the impact on attacker teams remains to be seen.

Table 1. Examples of team-level cognitive biases

Team-level bias	Definition
Information pooling	People in groups are more likely to share information everyone already knows and neglect unique information (Stasser & Titus, 1985)
False consensus	The tendency to overestimate the degree of similarity between self and others (Ross et al., 1977)
Groupthink	When a collective phenomenon where group members' striving for unanimity overrides their motivation to appraise alternative courses of action (Janis, 1972)
Group escalation of commitment	The tendency for groups to continue to support a course of action in a non-rational manner despite evidence that it is failing (Jones & Roelofsma, 2000)
Group serving bias	Attributing success to the in-group while attributing in-group failure to an external factor (Taylor & Doria, 1981)
Conformity bias	Individuals are compelled to side with the group opinion when faced with an obvious incorrect response (Asch, 1952)

Despite a lack of research, an examination of transcripts from a longitudinal study involving a 3-person cyber attacker team in a realistic network penetration scenario seems to provide evidence of team-level and team-influencing biases (See Ferguson-Walter, Shade, et al., 2019 for details of the study). Labeled transcripts of team-level communications among the members (labeled A, B, and C) are provided below with bolded explanation text accompanying.

Transcript Excerpt 1 from Ferguson-Walter, Shade, et al. (2019)

C: *You want a printer?*

A: *Yes. By IP address, please, if possible.*

[C provides A with 3 IP addresses, 2 of which are decoys]

[both decoys detect intrusion attempts.]

A: *That did not work. Very disappointed. That is a very stable exploit. Very well known. Very well documented. I really thought I had something there…*

[decoy exploit attempt detected]

B: *That should have worked, right?*

A: *That's the default. I'm going to throw another SMB exploit at it.*

B: *Honeypot*

A: *It could be a honeypot. At least someone knows what they're doing…*

…

[decoy exploit attempt detected]

A: *That exploit, the way that it spent so much time trying to stage, that seemed very promising….*

A: *This is the same exploit, but the difference is the last time I tried it, I did the reverse connection…*

[decoy exploit attempt detected]

A: *Yes, I will clumsily launch hundreds of exploits.*

In the above example (Excerpt 1), the team is unaware that decoys are present on the network, and as a result, false target information was unknowingly passed from person C to person A. This results in person A spending hours trying to exploit two

(unexploitable) decoy targets, despite discussion that it was a honeypot (i.e., decoy) with C, and evidence that the repeated exploits were failing. This eventually culminated in the team executing a very high-risk "hail Mary" exploit attempt. This series of interactions provides an example of team-level *escalation of commitment* (See Table 1) and team-influencing *confirmation bias* (the tendency to interpret new evidence as confirmation of one's existing beliefs or theories).

In another example (Excerpt 2), the team had been deceptively informed that decoys were present in the exercise when they actually were not. Here, the teammates appear convinced (incorrectly) that the targets are decoys, resulting in a strategy change.

Transcript Excerpt 2 from Ferguson-Walter, Shade, et al. (2019)

A: *Yea, those are DNS servers. Yea, yea. Agree.*

C: *I am using mine to figure out a convention for your list. I don't know what that is.*

A: *Yea, I think that, if I had to guess, it would be like a room, possibly like a port number or something.*

C: *And then on your list, like when you see stuff like [random nouns]. I could go one of two ways. I would either think that those are server names*

A: *Concur.*

C: *or I could think that they are phony.*

A: *Ok, so what you are saying is that they merit more.*

C: *Let's test. Pick a handful and try to decide if they look like honeypots or if they look like real servers.*

A: *Ok. That sounds like a great idea. Let's follow that one.*

C: *Because we also have users.*

A: *Something else that I would like to do is at some point in time, let's run a full-blown scan of our subnet and then let's look at what all is out there…*

Person C has convinced A to work on trying to figure out if the machines are decoys, taking the team off-track from their actual goal to recon and exfiltrate. This is an example of the propagation of misinformation leading to a disruption to the team's current strategy and an apparent loss of team situational awareness. Team-level cognitive biases could be one of many potential ways that cyber attacker team cognition is disrupted.

Team Cognition Breakdowns

Disrupting team cognition can take other forms aside from decision-making biases. In particular, Wilson et al. (2007) found that disruptions to coordination, communication, and cooperation led to breakdowns in teamwork on the battlefield. Other work also supports that disrupting lines of communication within teams disrupts their effectiveness (Lane et al., 2019). Research on human-machine teams operating remotely piloted aircraft found poor team communications, miscalibrated trust, and overly rigid coordination dynamics resulted in poor responses to novel events (Grimm et al., 2018; Johnson et al., 2021). Some disruptions or perturbations initially impairs team performance but can also strengthen it over the long run by increasing team adaptivity (Gorman et al., 2010). A deeper understanding of teamwork within cyber-attack teams could be utilized to disrupt the cyber-attack kill chain (Hutchins et al., 2011) through the intentional disruption of the building blocks of team cognition, such as communication channels and team situation awareness.

Of course, though examples can be found, a more systematic examination of cyber attacker team cognition is needed. This requires methods and tools suited for capturing team-level phenomena within the constraints of data collection and analysis. Existing techniques for the analysis of team interactions can be adapted to this end. We provide a path forward concerning these measures in the next section.

INTERACTION-BASED MEASURES

Interaction-based measures are well suited for understanding processes that rely on interdependent coordination, and the transfer of information within the team (Cooke & Gorman, 2009). Because most team biases and breakdowns result from the propagation and synthesis of information, affect, and social cues throughout a team and the emergence of collective states and behaviors, we propose that interaction-based measures can easily be adapted to the study of team biases and team cognition breakdowns. However, interaction measures vary in form, resource requirements, inputs, and outputs. Determining which may be more useful is something we attempt in the next section.

Interaction Data Collection

Interaction measures rely on event data generated by interactions within teams, which can be quantitatively collected and analyzed. Event data includes interactions between people and technology, or interactions among teammates (Cooke & Gorman, 2009). For example, this could be keystroke data, actions in a network initiated with software, or attempts to exploit decoys within a network (Ferguson-Walter, Major, et al., 2021).

Human-human interactions in the form of communications provide another valuable source of event data, often mediated by technology in the cyber domain. Data collection for face-to-face communication differ from technology-mediated communications, but in either case the data is multidimensional, and the collection and analysis method depends on research goals, pragmatics of collection, and resources for analysis. Communication analyses typically focus on communication *content,* communication *flow*, or a combination of both (Cooke & Gorman, 2009).

Content refers to what is talked about, whereas flow refers to how information moves throughout a team, for instance by characterizing who talks to who. Several studies use content analysis to categorize words or messages according to meaning (e.g., "request," "update"). Some types of content analyses are very labor-intensive, although some automated analysis approaches do exist (Baker et al., 2021). It is more likely that quick research benefits will come from studying communications flow because it is easier to collect and analyze by avoiding content analysis. Understanding flow still yields important information about cyber teams, including some of the team-level issues described previously, as it may help understand or monitor the spread of ideas (e.g., measures of communication distribution).

Interactions can be analyzed with both *static* analyses, which do not preserve the structure of the communications over time; and *sequential* or *timing* analyses, which in part do (Cooke & Gorman, 2009; and see Table 2). In general, sequential methods rely on the order of interactions whereas

timing methods rely on the specific time of interactions. In the following sections, we provide an overview of a few promising static interaction measures that primarily use team communications as inputs including distributions, social networks, and Event Analysis of Systemic Teamwork (EAST). Other promising sequential and timing approaches can be found in Table 2.

Table 2. Example Team communication analysis techniques and references.

	Content	Flow	Content + Flow
Static	Word counts/ratios (Entin & Serfaty, 1999)	Distributions, Social networks (Wasserman & Faust, 1994)	Event Analysis of Systemic Teamwork (Stanton et al., 2018)
Sequential	LSA Lag Coherence (Gorman et al., 2004)	Relational Event Modeling (Gibson et al., 2019)	Sample Entropy of Speaker and Content (Strang et al., 2012)
Timing	Conceptual Recurrence Analysis (Tolston et al., 2019)	Recurrence Quantification Analysis (Gorman et al., 2020)	Entropy phase transition detection (Wiltshire et al., 2018)

Potential Applications

Communication Distributions. Communication distribution is a static flow measure that describes how much each teammate communicates relative to one another. These measures involve calculating individual interaction quantities (e.g., seconds talking per minute) and then typically aggregating them to provide a team-level estimate of distribution. These analyses are simple and may be useful in co-located teams when the intended recipient is not easily determined. For disrupting cyber teams, each member is a potential vector to disrupting a team; but any methods here (such as biasing one member) usually rely on that information making it to the team. Measures of communication distribution then would help find both the most 'talkative' members, but also determining whether a compromised member is actually communicating with the team. In an analysis of the team communications collected by Ferguson-Walter, Shade, et al. (2019), the more senior member of the team communicated proportionally more during 3 of the 4 sessions and team communications became increasingly heterogeneous over the course of the sessions.

Social networks. Social network analysis is another static flow measure that can provide a deeper analysis of team communication distributions and structural patterns, but requires for each message that the senders and receivers be known. In this analysis method, the network consists of *nodes* representing entities (e.g., team members) and *edges* representing relationships (e.g., interactions), which can be directed (becoming arcs, if the sender is known) or undirected (an interaction occurred but it is not clear who originated it). Social network analysis provides several aggregation measures that can characterize interactions at the individual and team level (Wasserman & Faust, 1994). Whole-network measures have been applied to understanding team effectiveness in cybersecurity. For example, Buchler et al. (2018) found a negative association between the connectedness of the whole network's face to face interactions and performance in a capture the flag exercise.

Centralization is a specific set of network analysis measures which reveals the relative dominance of one or a few nodes, and has also been explored extensively in relation to team performance (Katz et al., 2004), with some findings revealing the impact of task load and complexity. For example, one study found that surgical teams tended to have less centralized communications when surgical task complexity was increased (Barth et al., 2015). Therefore, measures of centralization may help determine whether attempts to make the attackers' tasks more difficult (via OHF methods) are working, both from an individual and a team perspective. Individual network metrics (i.e., degree centrality) could also reveal highly central team members who could serve as vectors to influence the team. Other network-based approaches that used flow data, such as Relational Event Modeling, can also provide insight into how similar phenomena evolve over time (Gibson et al., 2019).

EAST. The Event Analysis of System Teamwork (EAST) method is an extension of social network analysis which combines *flow* and *content* to characterize system-level information flow (Stanton et al., 2018). It has been used to characterize collective cognitive activities in a variety of teams. The outputs are social, information, and task networks which can be used to characterize the propagation of information throughout the team as snapshots in time. EAST is a promising approach for characterizing breakdowns in team situation awareness, and potentially for understanding how biased-laden information passes through teams (e.g., Excerpt 2). EAST has been applied to understanding team cognition in cyber defense teams (Rajivan & Cooke, 2017). However, even the abbreviated version of EAST using primarily communication transcripts can be very labor-intensive to perform (Stanton, 2014).

Future Research

A number of questions remain for future research. Deep understanding of how cyber-attack teams interact remains limited. More research is needed to understand how cyber-attack teams collaborate, and the potential avenues through which their teamwork can be effectively disrupted. Future studies should continue to explore the applications of OHF to disrupting team cognition specifically, perhaps by intervening to reduce team situation awareness or communication capacities. Future work is also needed to decipher exactly what specific characteristics of team communications (flow vs. content) correspond to specific biases and breakdowns within the content of their dynamics (static vs. sequential). Finally, in both potential experimental and real-world applications, it will be necessary to determine what sensors can be used to gather team interaction data on more realistic adversaries.

Acknowledgment

This work was funded by the Laboratory for Advanced Cybersecurity Research and the Naval Research Enterprise Internship Program.

REFERENCES

Asch, S. E. (1952). Group forces in the modification and distortion of judgments. *Social Psychology, 452*, 112-117.

Baker, A. L., Fitzhugh, S. M., Huang, L., Forster, D. E., Scharine, A., Neubauer, C., Lematta, G., Bhatti, S., Johnson, C. J., Krausman, A., Holder, E., Schaefer, K. E. & Cooke, N. J. (2021). Approaches for assessing

communication in human-autonomy teams. *Human-Intelligent Systems Integration, 3*, 99-128.

Buchler, N., Rajivan, P., Marusich, L. R., Lightner, L., & Gonzalez, C. (2018). Sociometrics and observational assessment of teaming and leadership in a cyber security defense competition. *Computers & Security*, *73*, 114-136.

Bulanova-Hristova, G., Kasper, K., Odinot, G., Verhoeven, M., Pool, R., de Poot, C., Werner, Y., & Kordell, L. (2016). *Cyber-OC-Scope and manifestations in selected EU member states*.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Jrnl. of Cognitive Eng. and Decision Making, 3*(1), 27-46.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285.

Cranford, E. A., Gonzalez, C., Aggarwal, P., Tambe, M. Cooney, S. & Lebiere, C. "Towards a cognitive theory of cyber deception." *Cognitive Science* 45, no. 7 (2021): e13013.

Deline, S., Guillet, L., Rauffet, P., & Guérin, C. (2021). Team cognition in a cyber defense context: Focus on social support behaviors. *Cognition, Technology & Work*, *23*(1), 51-63.

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors*, *41*(2), 312–325.

Ferguson-Walter, K. J., Gutzwiller, R. S., Scott, D. D., & Johnson, C. J. (2021). Oppositional human factors in cybersecurity: A preliminary analysis of affective states. *IEEE/ACM International Conference on Automated Software Engineering, Workshop on Human-centric Software Engineering and Cyber Security (HACS)*.

Ferguson-Walter, K. J., LaFon, D., & Shade, T. (2017). Friend or Faux: Deception for Cyber Defense. *Jrnl. of Information Warfare*, *16*(2), 28-42.

Ferguson-Walter, K. J., Major, M. M., Johnson, C. K., & Muhleman, D. H. (2021). Examining the Efficacy of Decoy-based and Psychological Cyber Deception. *In Proc. of 30th USENIX Security Symposium*, 1127-1144.

Ferguson-Walter, K. J., Major, M., Van Bruggen, D., Fugate, S., & Gutzwiller, R. S. (2019). The World (of CTF) is Not Enough Data: Lessons Learned from a Cyber Deception Experiment. In *Proc. of 2019 IEEE 5th Intl. Conference on Collaboration and Internet Computing*, 346-353.

Ferguson-Walter, K. J., Shade, T., Rogers, A., Trumbo, M. C. S., Nauer, K. S., Divis, K. M., Jones, A., Combs, A., & Abbott, R. G. (2019). The Tularosa study: An experimental design and implementation to quantify the effectiveness of cyber deception. In *Proc. of the 52nd Hawaii Intl. Conference on System Science (HICCS),* 7273-7281.

Gibson, C. B., Buchler, N., Hoffman, B., & La Fleur, C.G. (2019). Participation shifts explain degree distributions in a human communications network. *PloS One*, *14*(5), e0217240.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training adaptive teams. *Human Factors*, *52*(2), 295-307.

Gorman, J. C., Cooke, N. J., & Kiekel, P. A. (2004). Dynamical perspectives on team cognition. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 673-677.

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human Factors*, *62*(5), 825-860.

Granåsen, M., & Andersson, D. (2016). Measuring team effectiveness in cyber-defense exercises: A cross-disciplinary case study. *Cognition, Technology & Work*, *18*(1), 121-143.

Grimm, D., Demir, M., Gorman, J. C., & Cooke, N. J. (2018). Systems level evaluation of resilience in human-autonomy teaming under degraded conditions. In *2018 IEEE Resilience Week (RWS)*, 124-130.

Gutzwiller, R. S., Ferguson-Walter, K. J., & Fugate, S. J. (2019). Are cyber attackers thinking fast and slow? Exploratory analysis reveals evidence of decision-making biases in red teamers. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, *63*, 427-431.

Gutzwiller, R. S., Ferguson-Walter, K. J., Fugate, S., & Rogers, A. (2018). "Oh, look, a butterfly!" A framework for distracting attackers to improve cyber defense. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*, 272-276.

Gutzwiller, R. S., Fugate, S., Sawyer, B. D., & Hancock, P. A. (2015). The human factors of cyber network defense. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, *59*, 322-326.

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, *1*(1), 80.

Janis, I. L. (1972). *Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes.* Houghton, Mifflin.

Johnson, C. J., Demir, M., McNeese, N. J., Gorman, J. C., Wolff, A. T., & Cooke, N. J. (2021). The impact of training on human–autonomy team communications and trust calibration. *Human Factors*, Advance online publication.

Jones, P. E., & Roelofsma, P. H. M. P. (2000). The potential for social contextual and group biases in team decision-making: Biases, conditions and psychological mechanisms. *Ergonomics*, *43*(8), 1129-1152.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Rev.*, *103*(4), 687-791.

Klein, K. J., & Kozlowski, S. W. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*(3), 211-236.

Lane, B. R., Salmon, P. M., Cherney, A., Lacey, D., & Stanton, N. A. (2019). Using the Event Analysis of Systemic Teamwork (EAST) broken-links approach to understand vulnerabilities to disruption in a darknet market. *Ergonomics*, *62*(9), 1134–1149.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Jrnl. of Applied Psychology*, *85*(2), 273-283.

Mathieu, J. E., & Luciano, M. M. (2019). Multilevel emergence in work collectives. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 163–186). APA.

McNeese, M., Cooke, N. J., D'Amico, A., Endsley, M. R., Gonzalez, C., Roth, E., & Salas, E. (2012). Perspectives on the role of cognition in cyber security. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*, 268-271.

Meyers, C. A., Powers, S. S., & Faissol, D. M. (2009). *Taxonomies of cyber adversaries and attacks: A survey of incidents and approaches* (Report # LLNL-TR-419041). Lawrence Livermore National Lab.

Rajivan, P., & Cooke, N. (2017). Impact of team collaboration on cybersecurity situational awareness. In: Liu P., Jajodia S., Wang C. (eds) *Theory and Models for Cyber Situation Awareness*, 203-226. Springer, Cham.

Rajivan, P., & Cooke, N. J. (2018). Information-pooling bias in collaborative security incident correlation analysis. *Human Factors*, *60*(5), 626-639.

Roberts, A. P., Webster, L. V., Salmon, P. M., Flin, R., Salas, E., Cooke, N. J., Read, G. J., & Stanton, N. A. (2021). State of science: Models and methods for understanding and enhancing teams and teamwork in complex sociotechnical systems. *Ergonomics*, 1-27.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Jrnl of Experimental Social Psychology*, *13*(3), 279–301.

Simonson, R. J., Keebler, J. R., Lessmiller, M., Richards, T., & Lee, J. C. (2020). Cybersecurity teamwork: A review of current practices and suggested improvements. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, *64*, 451-455.

Stanton, N. A. (2014). Representing distributed cognition in complex systems: How a submarine returns to periscope depth. *Ergonomics*, *57*(3), 403-418.

Stanton, N. A. D., Salmon, P. D., & Walker, G. H. D. (2018). *Systems thinking in practice: Applications of the event analysis of systemic teamwork method*. CRC Press.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Jrnl. of Personality and Social Psychology, 48*(6), 1467-1478.

Strang, A. J., Horwood, S., Best, C., Funke, G. J., Knott, B. A., & Russell, S. M. (2012). Examining temporal regularity in categorical team communication using sample entropy. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, *56*, 473-477.

Taylor, D. M., & Doria, J. R. (1981). Self-serving and group-serving bias in attribution. *The Jrnl. of Social Psychology*, *113*(2), 201-211.

Tolston, M. T., Riley, M. A., Mancuso, V., Finomore, V., & Funke, G. J. (2019). Beyond frequency counts: Novel conceptual recurrence analysis metrics to index semantic coordination in team communications. *Behavior Research Methods*, *51*(1), 342-360.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. *Human Factors*, *49*(2), 243-256.

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive Science*, *42*(1), 129–167.