Examining the Efficacy of Decoy-based and Psychological Cyber Deception

Kimberly J. Ferguson-Walter
Laboratory for Advanced Cybersecurity Research

Maxine M. Major
Naval Information Warfare Center, Pacific

Chelsea K. Johnson
Arizona State University

Daniel H. Muhleman
Naval Information Warfare Center, Pacific

Abstract

The threat of cyber attacks is a growing concern across the world, leading to an increasing need for sophisticated cyber defense techniques. Attackers often rely on direct observation of cyber environments. This reliance provides opportunities for defenders to affect attacker perception and behavior by plying the powerful tools of defensive cyber deception. In this paper we analyze data from a controlled experiment designed to understand how defensive deception, both cyber and psychological, affects attackers [16]. Over 130 professional red teamers participated in a network penetration test in which both the presence and explicit mention of deceptive defensive techniques were controlled. While a detailed description of the experimental design and execution along with preliminary results related to red teamer characteristics has been published, it did not address any of the main hypotheses. Granted access to the cyber and self-report data collected from the experiment, this publication begins to address theses hypotheses by investigating the effectiveness of decoy systems for cyber defense through comparison of various measures of participant forward progress across the four experimental conditions. Results presented in this paper support a new finding that the combination of the presence of decoys and information that deception is present has the greatest impact on cyber attack behavior, when compared to a control condition in which no deception was used.

1 Introduction

Cyber deception is a growing area of research in cyber defense, which considers the human aspects of an attacker in order to impede attacks and improve security [26, 46]. The goal is to use deception to better understand and influence an attacker that has already infiltrated a network, and ultimately to delay, deter, and disrupt an attack.

While the efficacy of deceptive technologies has been hypothesized for decades, controlled experiments to measure the impact on attacker behavior are relatively scant. A primary goal of our research is to provide a scientific assessment of the effectiveness of one cyber deception technology— a decoy system, which places numerous "decoy" assets on a network interspersed with the real assets as a defensive measure. These decoys can be configured to appear more or less vulnerable than, or identical to the real assets in network scans.

Bell and Whaley's highly accepted taxonomy of deception in kinetic military operations employ the term *dissimulation* for hiding the real, and *simulation* for showing the false [6]. Dissimulation includes *masking* the real so it appears not to exist, *repackaging* the real as something else, and *dazzling* to distract from the real. Simulation includes *mimicking* something true, *inventing* a new reality, and *decoying* by signaling a common truth but then changing to something different. Numerous extensions and examples of this taxonomy have been applied to cyber security [26].

A second goal of our research is to examine the effects of cyberpsychology-based techniques. Cyberpsychology is the scientific field that integrates human behavior and decision-making into the cyber domain allowing us to understand, anticipate and influence attacker behavior [36]. This can be done, for example, by leveraging an understanding of decision-making biases by taking actions that motivate an attacker to respond in specific ways that enhance a defender's ability to detect, identify, understand, and thus defend against said attacker. Furthermore, it is known that that cognitive limitations require people to form mental representations, or models of the world based on their personal knowledge and experience [32]. In this research, by providing information to the human actor (e.g., the possibility of cyber deception technology), their mental model of the target network was influenced, which in turn, affects cyber attack behavior.

In this publication we present new analyses performed on a subset of data provided from the Tularosa Study [16]. In the Tularosa Study, over 130 professional red teamers participated in a network penetration test which controlled for both the actual presence and the explicit mention of deception. The Tularosa Study leveraged two types of deception: 1) **decoys**—a cyber deception technology mostly utilizing *daz-*

zling, *mimicking*, and *inventing*, (the "Present" condition), and 2) **cyberpsychology** methods, where participants were told ("Informed") that deception might exist on the target network, regardless of whether or not it was really in use. The control condition did not use deception or cyberpsychology methods ("Absent" and "Uninformed"). See Table 1.

While a detailed description of the experimental design and execution along with preliminary results related to red teamer characteristics and cognition has been published [16], the study not yet address the main hypotheses presented and did not consider results drawn from the cyber data. Granted access to the cyber and self-report data collected from the experiment, this publication begins to address these hypotheses by investigating the effectiveness of decoy systems and cyberpsychology methods for cyber defense through comparison of various measures of participant forward progress across the four experimental conditions. Key features of the original study are summarized in Section 4.

Specifically, our analysis investigates these hypotheses, with an emphasis on hypotheses $H1$ and $H2$:

- **Hypothesis** $H1$: Defensive cyber tools and psychological deception impede attackers who seek to penetrate computer systems and exfiltrate information.
- **Hypothesis** $H2$: Defensive deception tools are effective even if an attacker is aware of their use.
- **Hypothesis** $H3$: Cyber deception is effective if the attacker merely believes it may be in use, even if it is not.
- **Hypothesis** $H4$: Cyber and psychological deception affects an attacker's cognitive and emotional state.

Experimental Conditions	
AU	Decoys Absent; Uninformed (Control)
AI	Decoys Absent; Informed
PU	Decoys Present; Uninformed
PI	Decoys Present; Informed
Groups Compared for Hypotheses	
H1	Control (AU) versus Experimental (AI, PU, PI)
H2	Uninformed (AU, PU) versus Informed (PI)
H3	Control (AU) versus Psychological Deception (AI)
H4	Control (AU) versus Experimental (AI, PU, PI)

Table 1: **Conditions and Comparisons:** Participants were randomly assigned to one of the four conditions. Analyses compared Absent versus Present and Informed versus Uninformed conditions in addition to pairwise comparisons.

This work helps to fill a critical gap in computer security research—rigorous data analysis to better understand cyber operators [5]. While several researchers have declared the need for user studies and datasets to enable cyber security research [43, 47], few have focused on cyber operators and those that do often rely solely on qualitative interviews [34, 41, 54]. For this analysis, we seek to combine the richness of qualitative data with quantitative cyber task-relevant data. See Appendix A for a table summarizing meaningful results to date, with previously published results denoted with the † symbol.

2 Related Work

The new millennium introduced new technological advancements and consequently, a new type of criminal. As such, initial investigations began to define and discover cyber adversaries, leading to what is known as modern cyber warfare. In this section, we briefly present the historical review that motivated our research. We consider the sociotechnical system in which technology and human agents create a synergistic interaction and the importance of this system to cyber defense.

In 1998, The Defense Advanced Research Projects Agency's Information Assurance (DARPA-IA) program focused on profiling cyber terrorists [49] and quantifying the impact of defense mechanisms [48]. The program assumed the cyber terrorist to be sophisticated, highly resourced, intelligent, able to influence product life cycles, risk-averse, and have specified targets. Then in 2001, Cohen et al. [8] tested human subjects to analyze attacker behavior in vulnerable systems. In addition, Tinnel et al. [55], created a Cyberwar Playbook suggesting defensive strategies, including deception, for use during an attack. Over the following decades cyber-attacks increased, and defense research and development attempted to keep pace. There exist numerous meta-analyses and surveys to introduce, define, and critique the many defensive advancements. More recently, innovative strategies have surfaced to introduce deception technology as a fruitful tactic to supplement tradition perimeter security defenses. For example, Pawlick et al. [42], provide a taxonomy based on game-theoretic defensive deception, providing six principles: perturbation, moving target defense, obfuscation, mixing, honey-x, and attacker engagement. Han et al., present a classification survey of the current application of deception techniques and discuss four categories: goal, unit of deception, layer of deception, and deployment of deception [25].

Innovative strategies are necessary to tip the scales in favor of defenders because today, network defenses have "reached the limits of what traditional defenses. . . can do" (p. iii) [26]. It is no longer enough to rely upon perimeter security [19]. Research must also focus on the human agents, which are vulnerable to decision-making biases [23] and exhibit other non-rational behavior [20]. One way to accomplish this goal is to employ a deceptive strategy to trigger these biases and influence decision-making.

In a situation where attackers can only know what they perceive, decoys and other cyber deception techniques become a powerful defensive tool. In 2002, Michael [37] suggested exploration of "software decoys [that] employ deception tech-

niques" (p. 957). However, research on the impact of cyber deception has been inconclusive, and reports on the effect of knowledge of its presence have been contradictory. For example, Fraunholz et al. [19], surveyed deception technology research, and reported the unknown presence of deception is significantly more effective than if it is known. In contrast, Yuill et al. [58] expand the early work of Heuer [27] to computer security, theorizing that when an attacker has knowledge of the presence of deception, decision-making biases cause the attacker to "see deception [even] where it does not exist" and to see benign anomalies as traps (p. 6).

We aim to clarify the psychological and technological impact of deception to delay and deter attacker behavior, regardless of whether attackers are aware of these deceptive strategies or not. We hope these basic research findings will be applied to the field, thus identifying the ideal manner to harness and induce human decision-making limitations and susceptibilities in cyber attackers, leading to the future application of novel defense strategies and technologies.

3 Motivation

From the initial definition of cyber terrorist to the determination for a need for innovative cyber defense techniques like cyber deception, we draw upon this foundational research to help answer the call for "experiments that are designed to study a focused hypothesis" [48] (p. 28). The Tularosa Study and the data analyses presented in this paper focus on evaluating the efficacy of decoy systems and psychological deception with expert human subjects. Most cyber deception experiments tend not to have rigorous experimental control or a large enough sample size of participants that generalize to the desired population. For example, participants in cyber deception studies with large participant pools often use unknown parties from the Internet [38, 40, 57] which makes it difficult to control for internal validity—the extent to which a cause-and-effect relationship established by a study cannot be explained by other factors. In this case, a trade-off seems to be made: sacrificing internal validity for high ecological validity—a form of external validity that is focused on how well the results generalize to real-world settings. This trade-off has generated interesting research results in studies such as those that have deployed honeypots on the Internet to characterize attacker behaviors [40, 45].

An alternative strategy seen in experimental cyber deception research is to design controlled studies with simplified tasks using students pursuing technology-related degrees [2, 8, 45] or other non-experts such as recruiting through Amazon Mechanical Turk [4, 9]. However, while internal validity can be easier supported in these settings, they tend to lack external validity, since the task performed and participants often do not generalize well to real-world attack scenarios. These studies have helped answer basic research questions and provide insights needed to shape future work focused on the sophisticated adversaries these defenses are employed to deceive.

In contrast, the Moonraker Study [51] was a controlled experiment designed to assess host-based cyber deception, which used "computer specialists" as participants performing cyber tasks. Likewise, the Tularosa Study focused on skilled participants in a controlled cyber experiment; for a detailed exposé on how controlled research experiments, designed to balance both internal and external validity, compare to more common cyber events such as capture the flag (CTF) see [15]. Controlled experimentation—devised to support internal validity—is defined as an experiment in which a group is used as a standard comparison condition (no variable manipulation) to other groups in treatment conditions (variable manipulation) [52]. Ecological validity was addressed by utilizing the closest analogous group to malicious cyber adversaries available for scientific testing — red teamers — and to bring in a large number of participants in hopes of providing the statistical power and reliability to detect measurable effects. Our contribution with this publication is the evaluation of a subset of previously unanalyzed cyber data from a large human subjects research (HSR) study that employed decoys as the deception technology and cyberpsychology methods, where participant conditions were aware or unaware of the presence of deceptive defenses.

4 Tularosa Study

The Tularosa Study was designed and conducted to understand how defensive deception, both cyber and psychological, affects cyber attackers [16], based on earlier pilot studies [14]. In this empirical study, *cyber deception* refers to the presence of a decoy system and *psychological deception* refers to providing information about the presence of deception on the network, which is hypothesized to influence the attacker's mental model and thus their behavior (See Table 1). Over 17 sessions, 139 experts participated in a full-day network penetration test. The total number of participants who were included for this analysis are as follows[1]: 35 for Absent-Uninformed, 28 for Absent-Informed, 30 for Present-Uninformed, and 30 for Present-Informed, for a total of 123 professional red teamers.

Professional demographics were collected for each participant including 1) level of expertise, 2) involvement in each phase of penetration testing engagement, 3) type of typical engagement, and 4) years of experience. See Appendix C for descriptive statistics. In addition to the abundant host and network data collected, a battery of questionnaires, e.g., demographics, personality; and cognitive tasks, e.g., fluid intelligence, working memory; task-specific questionnaires (TSQ)

[1] All five participants from the first session were excluded due to data collection issues. Another session of ten was excluded due to a late start which caused a reduction in the allotted cyber task time. One participant did not fit the selection criteria and was excluded. In the case of data collection glitches, affected participant data were excluded from that particular analysis.

and qualitative self-report data were also collected. More details can be found in the original publication [16] and accompanying online appendix [17].

Participants were recruited via a contracting process for qualified experts and were compensated for their participation. Approval was received on the experimental design from all relevant institutional ethics review boards (IRB)[2]. No personal identifying information (PII) was collected and all data was anonymized. No cyber task performance or HSR information was provided back to any of the participants' employers.

Participants worked independently and were presented individual identical copies of the simulated target network with 25 real Windows and 25 real Linux systems representing a variety of operating systems, patch levels, and services. The simulated network, for participants with deception present, included additional lightweight, virtualized decoys: 25 Windows and 25 decoy Linux variants. Each decoy responded to scans almost the same as their real counterparts, returning similar open/closed/filtered ports, banners, and services running. However they did not respond to any other activity, e.g., exploits launched against these non-interactive decoys always failed. This design resembles real world settings in which there are numerous reasons that exploits launched against real hosts may fail, such as: human error, faulty exploit, vulnerability not exploitable, false positive reported by vulnerability scanner, defenses are stopping the attack, etc.

The decoy system implemented *dazzling* by adding many targets to the network, distracting attackers from real assets (compared to the Absent condition) and slowing scans. It implemented *mimicking* by including some decoys that looked just like real assets during scanning (but all exploit attempts failed) and *inventing* by also including decoys that appeared different from the real assets, manifesting additional (unexploitable) vulnerabilities. These are baseline features of decoy systems common across commercial products [1, 18, 22, 29, 56]. Many commercial systems also implement high-interaction deceptive features, such as emulating or permitting exploit successes on decoys [1, 18, 29, 56]. The Tularosa Study emulated only the most generic features of decoy systems to better generalize results, with the assumption that these extra features will make the decoys more effective.

Participants were provided with a Kali Linux (Release 2017.1) attack workstation, which provided a robust environment for penetration testing with over 600 pre-installed security related tools [50], and its own offline software repository with 65 GB of additional tools and software that could be easily installed by participants. Research staff arranged ahead to include any publicly available tools requested by participants on all attack clients. The attack clients were fully instrumented for data collection. A separate internet-connected Windows workstation was provided for internet searches, self-reports and questionnaires. Time on task was monitored and cataloged, including breaks and lunch. Controls to support internal validity included directing participants not to discuss the task with each other, dividers between workstations, regulating time on task, and stationing a proctor in the room.

Participants were provided with instructions that included the following narrative for motivation on a task objective and to thoroughly report their findings:

> You represent an APT group attempting to gather information.... You have achieved an initial foothold on the company network, and now must discover as much as you can about potentially valuable targets on the network. You will conduct recon on the network and locate vulnerable services, misconfigurations, and working exploits.... Your objective is to collect as much relevant information about the target network as you can in the allotted time without compromising future network operations. [17]

Participants in Present conditions had an extra paragraph simply stating: *"There may be deception on the network."* Further instructions providing ground rules indicated all participants were to use the Kali laptop for *"reconnaissance and system exploitation"* and, to elicit data collection, specified that the internet-connected laptop was for research and real-time reporting to fictitious teammates *"when you learn potentially useful information about target systems."*

In the cyber domain, there are multiple—and sometimes competing—indicators of success or failure. Were participants stealthy in their approach? Did they exfiltrate important information? Were they able to identify and avoid decoy systems? Did they perform their tasks quickly? How this success is defined and measured for analyses is addressed in Section 6.1.

5 Data Sources

This section describes the subsets of data from the Tularosa Study that were used for the analyses presented in this publication—not all data collected were relevant for evaluating forward progress. More details about experimental design, the simulated network, and the data collected can be found in the original Tularosa Study publication [16] and online appendix [17]. Limitations to the study are discussed in detail in previous publications [15, 16], with additional comments on limitations relevant to our analyses discussed in Section 7.1.

Keylog data. Keystrokes were recorded from each participants' Kali Linux workstation, which included terminal commands, custom attack scripts, and notes to self. The keylog data did not contain web searches, self-reporting data, or survey data, as this was handled by the Windows workstation. These keylog data were searched for keywords indicating

[2]The IRB determined that the portion of the tasks that aligned with normal red team activity are not HSR and could be included in contracted work. The portion that collected data about the participants is HSR and was voluntary. Over 95% opted into the HSR portion of the study.

tools, attacks, and targets, and was vital for analyzing participant activity.

Network traffic. Network packets (PCAP) were recorded for each participant, and were useful to reveal a participant's interest or level of effort toward real or decoy systems based on traffic to certain IP addresses.

Real-time self-report data. Participants were instructed to log their thoughts and strategies (with relevant IP addresses) into the *Mattermost* chat client on the internet-connected Windows workstation throughout the cyber task. Participants were asked to freely report all *"potentially useful information about target systems on this network"*, with the backstory that a follow-on cyber team would use this information to continue the attack campaign in the future. These time-stamped reports recorded the participants' changes in belief and approach, such as the value of a target or a stated strategy.

Retrospective self-report data. An *End-of-Day Report* was also required to summarize information attained via the exercise which would be of use for a fictitious future team sent to compromise the same fictional target. These data were used to identify self-reported general task Failure/Success, and security assessment of the network.

Screen Capture. As a supplemental data source to support ground truth for participant activities, Optical Character Recognition (OCR) software was trained to extract text from the participants' attack client screen recordings. This text was validated by a cyber expert then used to discover ground truth for participants launching attacks, failure/success messages, and proof of file exfiltration from compromised targets.

Intrusion Detection Alerts There were no live Intrusion Detection Systems (IDS) on the network during the cyber task. However, to retroactively discover attacks potentially detected by an IDS we replayed PCAPs through the Suricata IDS utility. This does not include some attacks attempted by more skilled or stealthy participants. IDS alert data includes the number and frequency of IDS alerts, time to the first alert, and whether the target was real or a decoy.

Decoy alerts. The decoy system had its own alerting capability with four alert types. *Touch alerts* were triggered when any packet is sent to a decoy. *Scan alerts* were triggered when a participant scanned multiple decoy IPs within a short time period. *Probe alerts* were generated when a single decoy IP was probed for additional information with multiple packets sent in a short time frame; many exploits also triggered a probe alert. A *logon attempt alert* was triggered in response to an interactive logon attempt (e.g., RDP, SSH) from a participant or an exploit. These alerts were timestamped, and provided time of first interaction with a decoy.

6 Data Analysis Results

Prior to running our analysis on quantitative data, we performed standard outlier removal, removing data that were three standard deviations away from the mean. Data reported here were non-normal, therefore non-parametric statistical tests (Chi-Square, Kruskal-Wallis test) were performed to compare groups, followed by a Dunn's post hoc test with a Benjamini-Hochberg correction to compare specific pairs, where noted. All pairwise comparisons in Table 1 based on Hypotheses were performed, but non-significant findings are not reported due to space limitations. Where necessary, qualitative data were analyzed according to accepted practices for case study research [53]. This procedure entails four steps: 1) Two experts reviewed the participant reports, 2) noted each occurrence of the target data, 3) tallied the ratings, and 4) tested the inter-rater reliability with Cohen's Kappa statistic. Inter-rater reliability is the level of agreement between raters whereby agreement due to chance is factored out. A rate of $\kappa \geq .80$ is considered to be a sufficient level of agreement.

6.1 Measures of Success

The Tularosa experimental design did not provide a specific end-goal, nor explicit flags to exfiltrate, but rather allowed participants to self-determine what was reportable, revealing what they perceived to be of significance. In this respect, the Tularosa Study differed from a typical CTF exercise, where flags take the form of computer files containing a specified keyword hidden throughout the system. Thus "success" is pre-determined by the designers of the exercise. In contrast, real-world network exploitation requires subjective valuation of objectives and risk of exposure. While flags would have made success easier to measure across all the participants, it would have changed their motivation and behavior, thereby altering exactly what the study sought to measure. Moreover, the instructions attempted to encourage stealthy behavior, without explicitly demanding it by giving cues in the narrative. For example, telling the participants they represented an Advanced Persistent Threat (APT), often known for their stealth, hoping to not dramatically alter the natural behavior of each participant. More of the motivation behind these design decisions can be found in prior publications [15, 16].

There is no universally adopted metric to define "success" for offensive cyber actors or defenders. Attacker success is subjective, depends on the motivation behind the attack, and may depend on specific defender attributes. Success from an attacker's perspective can be examined by their progress in mapping, attacking, and exfiltrating from the network. However, in a deception scenario, the attacker's perception of success may not reflect true progress toward their goals (discussed in Section 6.1.3). While there were no human defenders in the Tularosa Study, defender success can be evaluated by the effects of the deception and measured by attacker resources wasted and altered perception caused to the attacker. The following sections discuss measures of perceived and actual success from both the attacker and defender perspectives: metrics for forward progress, effort wasted on decoys, and altered perception.

6.1.1 Forward Progress

Forward progress for a cyber actor can be measured by their strategic gains as they progress through the target network. This could include escalating privileges, compromising an increasing number of targets, accessing more valuable targets, or reaching a desired end goal. Similarly, a deceptive defender could gain an advantage by leading an attacker to believe they are achieving these same "successes" on decoy targets instead.

The industry-standard framework we used to classify each participant's activities is the Cyber Kill Chain® model [28], which describes a series of seven high-level stages for a cyber attack. The stages include: (1) Reconnaissance, (2) Weaponization, (3) Delivery, (4) Exploitation, (5) Installation, (6) Command and Control (a.k.a. C2), and (7) Actions on Objectives. Not all of these stages are observable to real-world defenders. However, the Tularosa Study was uniquely positioned to provide visibility into cyber actors' strategies and beliefs through collection of self-reports and real-time attack preparation through log files and screen recordings. Several of the following analyses indicate which stage of the Cyber Kill Chain a participant's actions indicated as well as the participant's awareness of possible attack vectors (even failed ones). These results are compiled in Appendix D.

Target Selection. Metasploit—a command line attack tool used to craft and launch cyber exploits—was the most used tool across the participants. Keylog mentions of Metasploit's `set RHOST [IP]` command, which established the intended target, was used as a proxy for target selection. We tallied each real and decoy target for every instance a participant identified a specific IP address (not a range of addresses) as the target of a Metasploit attack. Analysis indicated a significant difference with the Present-Informed condition targeting more decoys (mean = 3.42), (H(1) = 8.51, $p = 0.004$) than the Present-Uninformed condition, (mean = 1.7). This supports Hypothesis $H2$ and is consistent with results showing effectiveness is not degraded by true information given about the presence of deception on the network. Furthermore, since targeting the decoys does not help an attacker make forward progress, this is also consistent with Hypothesis $H1$. While there was no significant difference in the number of real hosts targeted between conditions, there was a trend (H(1) = 3.22, $p = .072$) of the Uninformed participants taking less time (mean = 128.6 min) to target their first real machine via `RHOST` than Informed participants (mean = 174.6). This indicates that Informed participants were delayed in the Weaponization stage of the Cyber Kill Chain, which is consistent with Hypothesis $H2$ that information on deception does not reduce the effectiveness of the decoys.

Stolen Credentials. A fictitious admin user account had been used to set up the real host machine. Participants who gained access to at least one of the real hosts typically attempted to use this account for privilege escalation and lateral movement, indicating progression toward later stages of the Cyber Kill Chain. A chi-square test of independence was performed to examine the relation between condition and the presence of the admin user name in the participant's keylog data. The relation between these variables was significant ($\chi^2 = 4.48$, $p = .0034$). There were more Absent-Uninformed participants who attempted to leverage the *lcooper* admin account than in the Present-Informed condition. This supports Hypotheses $H1$ and $H2$, and is consistent with the combination of presence and the information of deception hampering the participants' discovery and use of the admin account.

EternalBlue Exploit. The most commonly identified vulnerability reported by participants was EternalBlue, an attack on the Microsoft implementation of the Server Message Block (SMB) protocol. Often referenced by its Microsoft security bulletin identifier *MS17-010* [39], EternalBlue was a well-publicized exploit at the time of this study and was intentionally left unpatched on several of the real and decoy targets on the simulated network. As such, it is a useful metric to measure success between the conditions.

When we examine the number of mentions of *EternalBlue* or *MS17-010* in the participants' keylogs we do not see a significant difference. This is consistent across conditions, with participants initially discovering the vulnerability and searching for the exploit during the Reconnaissance and early Weaponization stages. However, when they advance further down the Kill Chain differences begin to emerge. There were significantly more (H(1) = 3.97, $p = .046$) uses of EternalBlue, as measured by OCR text matching the loading of the EternalBlue module into Metasploit, by participants in Absent conditions (mean=17.3) than Present conditions (mean = 4.6). This is the last part of the Weaponization stage of the Kill Chain, indicating the attack selection, and demonstrates further progression of Absent conditions, supporting Hypothesis $H1$. Participants' attempts to use the Eternal Blue exploit against real machines were collected by replaying PCAP traffic through community rules for the Suricata IDS, representing the Delivery stage of the Cyber Kill Chain. We see a significant difference of Absent conditions (mean = 3.89) generating more EternalBlue IDS alerts than Present conditions (mean = 1.88), (H(3) = 7.07, $p = .014$). This supports Hypothesis $H1$ because it indicates a decrease in forward progress in Present conditions. Furthermore, the Dunn's post hoc test demonstrated a significant difference between the four conditions (AU, mean = 3.87; AI, mean = 3.82; PU, mean = 2.90; PI, mean = 0.87) as the Present-Informed condition made the least forward progress ($p = .05$), supporting Hypothesis $H2$.

The total number of EternalBlue Success and Failure Messages detected by OCR are provided in Table 2. While there is no statistical difference between conditions due to the relatively small number of occurrences, the control condition (AU) had more than twice the success messages as any condition that received an experimental manipulation. We also note that the Present-Uninformed condition had nearly three times as many failure messages as any other condition. This

	Success Messages	Failure Messages
Absent-Uninformed (N=34)	235 (n=14)	1121 (n=19)
Absent-Informed (N=28)	95 (n=7)	1092 (n=12)
Present-Uninformed (N=29)	93 (n=10)	3015 (n=16)
Present-Informed (N=29)	107 (n=7)	1098 (n=16)

Table 2: **Impeded Forward Progress:** Number of EternalBlue failure and success messages detected by OCR.

is thought to be because without any information about the deception that is present, participants in this condition are more likely to repeatedly retry a failed exploit attempt, often blaming themselves, or their attack tools for the error [14]. This behavior is consistent with the confirmation bias, where decision makers tend to misinterpret ambiguous evidence as confirming their current assumption that other factors are to blame [44].

Self-Reported Exploits. A cyber expert examined the time-stamped Mattermost messages and selected those that described an exploit or vulnerability, including EternalBlue and all others, e.g., VNC vulnerabilities, pass-the-hash (using psexec), and labeled them for reporting 1) identification of a vulnerability that could be exploited, 2) exploit success or 3) exploit failure. While participants varied in the verbosity of their reporting, it was primarily the difference between conditions that mattered. An analysis of the number of exploit successes reported by participants indicated a significant difference between Absent (mean = 6.5 exploits) and Present (mean = 1.4 exploits) conditions at the Exploitation stage of the Cyber Kill Chain (AU, mean = 8.4; AI, mean = 4.8; PU, mean = 2.3; PI, mean = 1.9), (H(3) = 6.48, $p = .011$). This reveals that participants in the Absent condition reported significantly more exploit successes— across all exploits—than those in the Present condition. This is consistent with the OCR findings reported in Table 2 and supports Hypothesis $H1$ that the presence of cyber deception impedes forward progress. There was no statistically significant difference measured regarding the number of vulnerabilities identified or the number of exploit failures reported.

Data Exfiltration. Successful exfiltration of information from compromised targets is a clear indicator of forward progress, and typically occurs in the final stage of the Cyber Kill Chain (Actions on Objectives), but often needs to be achieved on multiple assets in order for the attacker to fully accomplish their objectives. OCR was used to search participants' screen recordings for status message text indicating the successful download of critical system file types such as: domain controller files, domain user hashes or credentials, local user profiles, Windows registry, and PowerShell scripts. Descriptive statistics for these exfiltrated files are provided in Appendix B. The Absent condition (mean = 3.86 files) had more than twice the number of participants (n=13) with evidence of valuable files exfilitrated than the Present condition (n = 6, mean = 1.52 files). This trend of more participants in the condition without decoys reaching the final stage of the Cyber Kill Chain approaches significance (H(1) = 3.68, $p = .055$) and is consistent with Hypothesis $H1$, that the presence of decoys impedes forward progress. Some participants exfiltrated data using a diversity of techniques, including specific attack tools i.e., `credential_collector` which are not accounted for in these results.

Keystroke Count. One coarse proxy for forward progress which revealed significant differences is the number of keystrokes a participant typed on the attack client during the course of achieving their objectives. There are limitations to this measurement, as the quantity of keystrokes does not necessarily correlate to the quality of actions taken. Keystroke counts do not take into account think-time, or that some participants might have been more productive and efficient while also typing less, e.g. keyboard shortcuts. Keystrokes may also include some non-attack activity such as note-taking. However, keystrokes may be a reasonable data source to measure the impacts of deception across conditions in that we expected to see a correlation between the number of keystrokes and forward progress, since attackers cannot progress very far without interacting with the attack client. Analysis[3] indicated a significant difference (H(1) = 3.96, $p = .047$) with participants in Absent conditions having a higher keystroke count (mean = 10564.48 keys) than those in Present conditions (mean = 8733.58 keys) which is consistent with Hypothesis $H1$ that decoys impede and delay forward progress.

Delay Effect. Any measured delay of a cyber attack that occurs within deceptive conditions can often be attributed to the cyber deception technique. One metric to measure delay is the amount of time spent before attacking begins. Decoy alerts can detect and log attacks launched against the decoys, but only exist in Present conditions. These alerts are a realistic and preferred alerting metric for defenders because there are few false-positives, by design, since no legitimate users or services would be interacting with a decoy. Delay was measured from the start of the experiment until the first decoy alert (of any type) was triggered. The Present-Uninformed condition (mean = 20.59 minutes) took a significantly longer time to initiate an interaction with a decoy than the Present-Informed condition (mean = 11.74 minutes) (H(1) = 4.44, $p = .035$). While it might be assumed that information about deception can delay an attacker by making them think twice about what to do first, this result indicates otherwise. These data support two possibilities: 1) knowledge about the deception made it less effective at delaying attacks, or 2) participants eagerly hunted for the deception because of the disclosure of information on deception. When considering all other supporting evidence for Hypothesis $H2$, we assert that the latter is the

[3]Sixteen additional outliers across conditions were identified by subject matter experts and removed from this analysis due to low number of keystrokes captured from data collection errors.

case, and instead of being cautious, participants tended to act more quickly to seek out the source of the cyber deception. This is likely because information about the deception was vague. Unlike the pilot study where participants were specifically told to expect decoy systems which slowed down their initial actions [14], in the Tularosa Study participants were merely informed that *deception may be present* which appears to have caused participants to want to quickly seek out evidence of said deception.

This change in behavior can benefit defenders, as less cautious behavior on the attacker's part can lead to faster detection and mitigation by the defender. We also note that because these were not real world situations with consequences, some participants may have felt less of a need for caution. Examples in self-report data such as Present-Informed Participant S116's statement *"I think I wasted a lot of time looking for the deception"* confirm that some informed participants hunted for deception. In line with Bell & Whaley's taxonomy, the decoys used *mimicking* and *inventing* to reveal the false in Present conditions, which caused attackers to waste resources. The decision-making bias related to this situation is the sunk cost fallacy: the tendency to continue with a specific strategy because of prior investments, such as money, time or effort [3]. Attackers succumb to the sunk cost fallacy because they continue with a course of action that is wasteful, when another less costly option or course of action is available. If this is the case, this could support the creation of novel cyber deception tactics that focus on shaping attacker beliefs and behaviors to further delay and impede attacks.

6.1.2 Attacker Resources Expended

The limited amount of time available and level of effort spent attacking decoys can be used to measure decoy effectiveness and success for the deceptive defender. The deceptive hosts logged all network interaction and each decoy interaction can be considered a wasted effort, delaying forward progress.

IP-Containing Commands. To investigate the question of increased effort expended in deception conditions, we considered the total number of commands typed by the participant which contained an IP address. This includes all keystrokes, including the Metasploit-specific `set RHOST <IP>` commands discussed in Section 6.1.1, as well as any other IP-addressed attacks, scripts, and notes taken by the participant mentioning these targets. The total number of real and decoy machines targeted by each participant in each condition was tallied. There were no statistically significant differences in the number of decoys targeted across Present conditions. However, results indicated a statistically significant difference in the number of real machines, revealing that fewer real machines were targeted in Present conditions (mean = 22.78) than Absent conditions (mean = 31.98), (H(1)) = 4.58, $p < .01$). This supports Hypothesis $H1$ that the presence of decoys impeded forward progress and protected real machines from attack. This also helps build a case for the technical effectiveness of decoys for cyber defense. A total of 710 commands included decoy IP addresses (52% of all commands that contained IPs in Present conditions). This is only the minimum number of commands that were wasted on decoys since other related commands were likely entered before and after the command containing the IP address.

Byte and Packet Count. Using PCAP data, we considered the total byte count of all the packets sent from each participant's host to all targets in the network. Results indicated that significantly fewer bytes were sent to real machines in Present conditions (mean = 0.241 GB), than in Absent conditions (mean = 0.321 GB), (H(1) = 5.28, $p = .022$). The total number of bytes sent from participants to decoys totaled over 10 GB, and most, if not all of this traffic, was a waste of effort and resources. This supports Hypothesis $H1$ that the presence of decoys wasted attacker time and resources, and thus impeded forward progress and further displayed the technical effectiveness of decoys for defense.

We also noted increased variance σ^2—a measurement of how spread the data are from the mean—specifically in the Present-Uninformed condition in this and other data types (AU mean = .4858 GB, $\sigma^2 = .3658$ GB; AI mean = .5222 GB, $\sigma^2 = .3829$ GB; PU mean = .9641 GB, $\sigma^2 = 1.467$ GB; PI mean = .6955 GB, $\sigma^2 = .7820$ GB). This could be caused by the presence of decoys and exacerbated by a lack of information about the deception. This was also seen observationally; deception caused some cyber attackers to become more cautious and work slower, but had the opposite effect on others who became less cautious, perhaps due to frustration. Regardless, a change in behavior was evident, both observationally and in the cyber data. The increase in variance can be interpreted as indicative of chaos injected into the performance of particularly when unknown deception was present.

There was also an increase in the number of assets that could have been targeted in Present conditions (the 50 real hosts *and* 50 additional decoys); any scan of the full subnet would naturally scan more assets. This is a feature of the design of a decoy system and a critical component of its effectiveness. In Present conditions, 35% of the packets were sent to decoys—another indicator of wasted attacker time and resources, and increased risk of discovery. There was no statistical difference in the number of packets sent to decoys when comparing between the Present conditions.

Decoy Alerts. We analyzed all decoy-generated alerts: (*touch, scan, probe, and login attempt*). Since these alerts are only generated by decoys, we only compared Present conditions. We found significant differences across all alert types. Moreover, every participant in a Present condition triggered a decoy alert prior to reporting any successful exploitation of real machines. This gives further evidence that the decoy alerts have utility above what a standard IDS can supply. Compared to the Present-Uninformed condition, (mean = 12676.5 alerts), the Present-Informed condition had signifi-

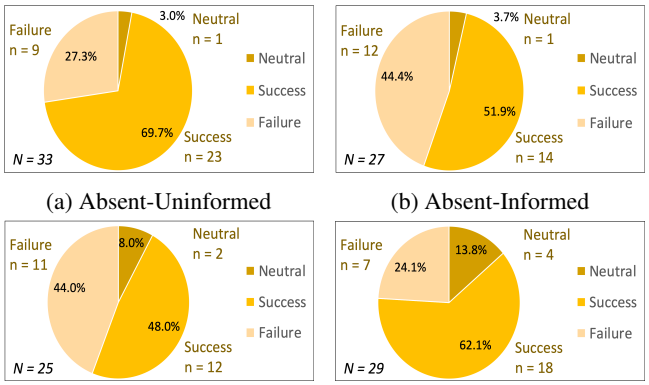
cantly *more* touch alerts (mean = 17903.33 alerts), (H(1) = 7.68, $p = .006$) and scan alerts (PU mean = 411.88; PI mean = 542.5), (H(1) = 7.91, $p = .005$), but *fewer* probe alerts (PU mean = 891.54; PI mean = 711.30), (H(1) = 36.3, $p < .0001$). *Probe* alerts, which alerted when multiple packets were sent to one decoy in a short time period and detected more targeted scanning behavior as well as several attempted exploits (Reconnaissance/Delivery/Exploitation), and *login attempt* alerts (Delivery) are triggered later in the Kill Chain than *touch* and *scan* alerts (Reconnaissance), and indicate further progress. Therefore, Present-Informed participants triggered more minor decoy alerts and fewer critical decoy alerts which supports Hypothesis $H2$ that the information of deception actually reduced their forward progress.

Login attempt alerts are generated by decoys after an interactive login attempt (e.g., SSH, RDP) made directly by the attacker or by an exploit being used. The Present-Uninformed condition had twice as many participants (n=22) who had at least one interactive login attempt on a decoy than the Present-Informed condition (n=11). The relation between these variables was significant ($\chi^2 = 8.15, p = .0043$). Participants in the Present-Informed condition were less likely than those in the Present-Uninformed condition to attempt an interactive logon manually, or with an exploit, to a decoy. Since both Present conditions continued to trigger decoy alerts throughout the cyber event, we see no evidence that the Present-Informed condition avoided logging on to decoys due to the information provided. We purport that this is a supplementary indication of impeded forward progress by participants in the Present-Informed condition, which aligned with other results supporting Hypotheses $H1$ and $H2$. Combining the alerts, we find the Present-Informed condition (mean = 19667.70 alerts) had significantly more total decoy alerts overall than the Present-Uninformed condition (mean = 12090.48 alerts), (H(1) = 18.6, $p < .0001$).

6.1.3 Altered Perception

The Tularosa Study was designed to measure both the psychological effects of cyber deception and the tangible effects of the use of cyberpsychology methods. The most pronounced psychological effect was the observable difference between reality and the altered perception caused by the deception.

Retrospective Success/Failure. To further evaluate attacker success, we examined the end-of-day report requested from all participants upon completion of the cyber task. As one measure of perceived success, we labeled these retrospective reports during post-processing as *Success* if the participant discussed more self-perceived successes, e.g., *"The assessment was fairly simple in terms of complexity"*, as *Failure* if the participant discussed more self-perceived failures e.g., *"All of the exploits I tried to run today were not successful"*, and as *Neutral* if the briefing did not discuss failures/successes or discussed an equal number (less common).



(a) Absent-Uninformed (b) Absent-Informed

(c) Present-Uninformed (d) Present-Informed

Figure 1: **Altered Perception**: Whether participants reported more failures or successes tended to correspond to their experimental condition. Reduced failures reported by the Present-Informed condition is consistent with the self-serving bias.

See Figure 1 for descriptive statistics. We see a reduced number of reported failures in the Absent-Uninformed condition (where no deception was used; 27.3%; see Figure 1a) which is unsurprising since they had the most forward progress by our measures (See Section 6.1.1). Interestingly, we also see reduced failures reported by the Present-Informed condition (24.1%; see Figure 1d) which had the least forward progress. We theorize this is because the combination of being informed of the deception and the deception being present acted as an excuse for the participants who no longer felt responsible for the failures and therefore reported failures less often. This behavior is consistent with the self-serving bias the tendency to claim more personal responsibility for successes than failures. This is particularly the case when evaluating ambiguous information [35]. The apparently altered perception displayed in the Present-Informed condition is consistent with Hypothesis $H4$ that cyber and psychological deception affect the cognitive and emotional state of an attacker. Previous results support this hypothesis in that confusion was significantly increased in both Informed and Present conditions [16]—a seemingly negative emotional effect. In this case, when minimizing the feeling of failure, the effect has a more positive feeling. Regardless of the polarity, the ability to elicit change is key. It has been suggested that if basic research can identify how to harness and induce these effects by triggering innate cognitive biases in cyber attackers, that this could lead to game-changing novel new defenses beyond, but related to, cyber deception techniques [24]. Our results take the first step by demonstrating a decision-making bias triggered by an experimental manipulation in a cyber attacker.

Security Assessment. We further evaluated the end-of-day reports for participants' assessments as to the security posture of the network. This is typically an expected component of a final report after doing a cyber assessment on a

network. A report was labeled ($\kappa = .81$) as *Secure* if the participant described the network as secure, e.g., found zero vulnerabilities, exploited nothing successfully, made statements about the network/hosts being well-secured, etc. and *Insecure* if the participant described the network as insecure, e.g., gained access to the domain controller, obtained admin credentials, exploited multiple hosts, made statements about the network/hosts being insecure, etc. Reports which fit neither category were labeled as *Not Applicable*. There were significantly more participants in the Present-Uninformed condition than the Absent-Uninformed condition who reported the network as Secure ($\chi^2) = 4.30$, $p = .030$), supporting Hypothesis $H4$. Interestingly, in the Present-Uninformed condition the number of participants describing the network as *Secure* is equal to the number describing it as *Insecure*, with the next closest condition (Absent-Informed) having less than half the amount of Secure as Insecure assessments. As described above, in conditions where decoys were present, they impeded progress and delayed participants; previous results also indicated increased confusion(see Appendix A). For the participants in the Present-Uninformed condition, most of them had no explanation for the cause of these difficulties, leading to increased ambiguity. The ambiguity effect [10], a well-researched decision-making bias, could explain the behavior displayed in Figure 2, and would be consistent with Hypothesis $H4$. The ambiguity effect suggests that ambiguity causes people to be unwilling to act. If future studies can confirm that the ambiguity effect is triggered by employing unknown cyber deception techniques, this method could lead to a delay, disruption or deterrence of cyber attack behavior—a win for defenders.

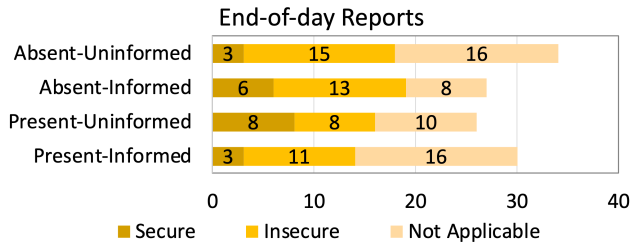


Figure 2: **Altered Perception:** Number of participants describing the network as Secure and Insecure is equal in the Present-Uninformed condition only. The ambiguity effect decision-making bias could explain this behavior.

Psychological Deception. Any perceived deception by the Absent-Informed condition is a clear example of a mismatch between perception and reality. Instances of blame being placed on the non-existent deception are exemplified in the data. Absent-Informed Participant S106 reported: *"This network was filled with deception and I spent the majority of the day going down rabbit holes that led me nowhere."* and Absent-Informed Participant S119 reported: *"I believe there were very good defense barriers and successful deception put into place in the network which didn't allow me to obtain an exploit today."* Outcomes of this study suggest that future experiments designed to assess the effect of psychological deception (when no cyber deception is present) should utilize a real network, or ensure that the simulated network has enough realistic messiness, mistakes, imperfect users, and policy mismatches, such that *real* things can be misattributed to deception. The simulated network for the Tularosa Study did not include these features, and thus we believe this is why there is only observational support rather than statistical findings supporting Hypothesis $H3$.

Decoy Interactions. We reiterate the *login attempt* decoy alert findings (PI, mean = 117.6; PU, mean = 460.5) as another example of altered perception, since attacking or attempting to log on to a decoy implies belief that it was a real, valuable asset. In Present conditions, participants often perceived the decoys as real, vulnerable machines. Unsurprisingly, we see this even more in the Present-Uninformed condition, where due to confirmation bias—a tendency to search for or interpret information in a way that confirms one's preconceptions—participants had little reason to question the veracity of machines. Others have noted similar findings [24, 51], including the pilot studies which used one-on-one observation and interviews noting that "the subjects verbally expressed confusion during the scenario and questioned their tools, their skills, and themselves rather than the authenticity of the network" [14]. This is thought to be why Present-Uninformed participants seemed more likely to perseverate on a particular machine they perceived as exploitable, even after many failed attempts. This may suggest that the Present-Informed participants are more likely to give up sooner and abandon their task, which could also benefit cyber defenders. While this is consistent with current findings, more investigation is warranted.

7 Discussion

Our work contributes to the understanding, measurement, and deployment of decoy systems and cyberpsychology methods to improve cyber defense. In this publication we performed data analysis to examine the effectiveness of a low-interaction decoy system, with consideration of attacker awareness of deception, and discussed results indicating that the combination of the presence of deception and information that deception is being used can impede attacker forward progress, increase detectability, and alter attacker perception. While much work remains to learn how to improve the use of cyber deception for cyber defense and focus on application of research results, we've investigated the following hypotheses:

- **Hypothesis $H1$: Defensive cyber tools and psychological deception impede attackers who seek to penetrate computer systems and exfiltrate information.** A difference in performance on the cyber task when

decoys were present was demonstrated by a consistent impedance in forward progress throughout the Cyber Kill Chain. In the Reconnaissance stage, Present conditions targeted significantly more decoys via *RHOST* than Absent conditions. In the Weaponization stage, Present conditions attempted significantly fewer EternalBlue exploit attacks as measured by OCR detecting loading of the module into Metasploit. For the Delivery stage, Present conditions had significantly fewer EternalBlue exploits detected by the IDS. In the Exploitation stage, Present conditions had significantly fewer self-reported exploit successes (for all exploits). For the last stage, Actions on Objectives, Present conditions exfiltrated significantly fewer data files. These results are consistent with delay or disruption in forward progress and support the hypothesis that cyber deception tools impede attackers. This is also supported observationally by contrasting self-reports from Absent-Uninformed Participant S104: *"I eventually pwned everything. Every. Single. Domain. Asset. Pwned."* and Present-Uninformed Participant S87: *"There was a lot of frustration... I don't really think there is too much that is actually exploitable."*

- **Hypothesis $H2$: Defensive deception tools are effective even if an attacker is aware of their use.** We reported a difference in performance on the cyber task between conditions when participants were informed about deception and decoys were present. Previous results indicated that even with knowledge of the deception, participants in the Informed condition reported significantly more confusion [16]. Our new analysis indicates they also triggered more total decoy alerts than the Present-Uninformed (PU) condition. Present-Informed participants also triggered the first decoy alert, indicating more aggressive initial behavior, but had statistically fewer high-severity decoy alerts which would be triggered later in the kill chain, indicating less forward progress. Even with knowledge of deception being present, participants in the Present-Informed condition wasted more effort (targeting significantly more decoys via `RHOST` than PU), and more resources (significantly fewer commands and bytes sent to real machines than Absent conditions) for less gain (fewer EternalBlue exploits detected than AU), at a delayed rate (slower to target first real machine via `RHOST`). In general, we found that the Present-Informed condition had the most affected behavior across many measurements consistent with the idea that a combination of information about and presence of deception can provide the best defense. This is counter to common thinking that deception tactics must remain hidden to be effective [19], and provides experimental support to what had previously only been theorized [58].

- **Hypothesis $H3$: Cyber deception is effective even if the attacker merely believes it may be in use, even when it is not.** We analyzed performance on the cyber task for the psychological deception condition where participants were informed that deception may be present when there were no decoys. Analysis noted no statistically significant findings. However, there was supporting evidence in the self-reports of participants in the Absent-Informed condition i.e., blaming failures on the non-existent deception. We believe that additional experiments with more real-world network, user, and system details that more accurately mimic the natural messiness of cyber space are needed to address this hypothesis. We argue that this messiness is precisely what is needed to provide the plausible deniability and uncertainty that make the psychological deception effective, as demonstrated in the related pilot studies, which were held on an operational, rather than a simulated, network [14].

- **Hypothesis $H4$: Cyber and psychological deception affects an attacker's cognitive and emotional state.** We reported various cognitive effects and altered perceptions in the experimental conditions compared to the control group. Significantly more participants in the Present-Uninformed condition assessed the network as Secure versus Insecure in their end-of-day report when compared to the control condition (AU). Moreover, fewer participants in the Present-Informed condition reported cyber task failures in end-of-day reporting than those in the Absent-Uninformed condition, which could indicate that being informed of the deception made participants no longer feel responsible for the failures. This was one example of several decision-making biases that were identified as potentially being triggered by the experimental manipulations. This hypothesis is also supported observationally in participant self-reports by contrasting the statements Absent-Uninformed Participant S138: *"I did not find any aspects of the network that were frustrating or confusing. Everything seemed relatively straight-forward."* and Present-Uninformed Participant S87: *"The results were extremely frustrating and somewhat confusing. I believe that several of the boxes that I tried to exploit were vulnerable to the exploit and payload that I threw at them."* We also provided examples of decision-making biases exhibited by the participants including: sunk cost fallacy, confirmation bias, self-serving bias, and ambiguity effect. This foreshadows an additional hypothesis that will be addressed in future work: cognitive biases are prevalent in cyber attacker behaviors and can be intensified to disrupt cyber attacks.

Additionally, our empirical assessment demonstrated the technical utility of decoy systems in the following ways:

- In conditions where decoys were present, every participant triggered a decoy alert prior to any successful exploitation of real machines.

- In conditions where decoys were present, 52% of all commands containing IP addresses contained decoy IPs and 35% of the packets sent were targeting decoy IPs.
- In conditions where decoys were present, more IDS alerts triggered on decoys than on real machines, demonstrating wasted effort; the number of alerts on real machines were reduced by about half, when compared to Absent conditions.

In summary, our data analysis provides empirical evidence that not only are cyber deception techniques, like decoy systems, effective for impeding cyber attacks, but it may be more effective if the attacker is aware of the presence of deception.

7.1 Study Limitations

The Tularosa Study was a novel attempt to measure adversarial activity and cognition in a deceptive cyber environment which attempted to balance external and internal validity considerations [15, 16], however like all experiments, limitations remain. Those most relevant to the data analysis presented in this paper are detailed below.

Attack Behavior Complexity. In a controlled experiment, an ideal situation is one where all participants experience an identical environment where any differences are tightly controlled. However, this limits the ecological validity of allowing participants to act in a manner consistent with real world behavior. Providing participants with choices can reduce the internal validity, at the expense of increased ecological validity. The tools and techniques utilized by each participant varied drastically, even for similar objectives, such as discovering the domain admin hash or exfiltrating files. Individual data sources are limited in terms of what they can reveal about a participant's cyber activity. For example, keylog cannot fully describe graphical user interface (GUI) activity, network traffic cannot reveal what took place on an encrypted channel, and OCR cannot easily piece together a timeline with multiple attacks occurring simultaneously. To score participant successes and determine ground truth, multiple data sources were scored on performed objectives such as the common use of EternalBlue. With over 1611 GB of data, the use of human experts to label the data could rival the experiment itself in scope. Hybrid approximation approaches that leverage automation were utilized, and future work will continue to refine these analyses.

Simulated Network. The goal of this basic science is to provide a foundation from which to build solutions that may be applied in the real world. As discussed in Section 3, controlled experimentation is crucial to investigate the best strategies for application. Testing within a laboratory setting is a trade-off between ecological validity and the requirements of foundational science. This study used a simulated network to ensure participants were presented the same assets and attack vectors. This was a calculated design trade-off between repeatability and realism. Participants were also given less time for this type of cyber task than they would take in a real-life scenario (over 50% indicated a week or more for typical engagement length [16]). Moreover, attackers' behavior is likely to change if they face real consequences, with realistic motivation to avoid the deceptive elements. Additionally, while it is clear that red teamers are not the same population as unethical hackers, specific differences in their behaviors and reactions have not yet been well documented.

Self-Report Data. Qualitative data adds a richness to quantitative metrics. As such, there are limitations in self-report data that may affect internal validity. Conclusions drawn from self-reported data were limited to what participants willingly transcribed, and do not tell a complete picture of a participant's frame of mind throughout the study. The variance in real-time reporting ranged from zero to 304 Mattermost messages in a day with 13% of participants providing zero. However, only 4% of participants skipped the of end-of-day report. While it's possible that personality differences could affect the quantity or quality, there were no detected differences across conditions in the conducted personality assessment [33]. Self-report data reflected the attacker's recollection and perception of their actions. Therefore, this data used in isolation is limited and may not represent the reality of tasks accomplished or actions taken.

Low-Fidelity Decoys. The decoy-based deception evaluated in this research was non-interactive and non-exploitable. Comparable decoy systems, as well as higher fidelity, interactive, exploitable decoys are commercially available. We believe these higher fidelity decoys should be more difficult to identify and avoid, and can be fine-tuned to waste additional time and resources. Thus, they will have a greater effect on attacker behavior. This suggests our results may provide a lower baseline for the efficacy of decoy-based deception. A future study is needed to focus on these more sophisticated decoy systems. Nevertheless, in this study, only one expert participant (S116) correctly identified any of the low-fidelity decoys as deceptive machines. However, as noted above, the condensed time provided for the cyber task was unrealistic, so the ecological validity of this finding is limited. While Present-Informed Participant S116 had some successes identifying decoys, the cyberpscyhology methods were still effective, as evident by misidentification of several real machines as fake. S116 reported: *"[being informed] really screwed with my head. Every machine I came across seemed like it could be a honeypot."*

7.2 Generalizability of Results

Additional studies are needed to evaluate how these results may generalize to other deception technologies. This study utilized technology that included the deception techniques of *dazzling*, *mimicking*, and *inventing* and the results may generalize beyond decoy systems to other technologies that incorporate these techniques. We also investigated cyberpsy-

chology methods as psychological deception, where true or false information about the presence of deception was provided to participants. In the latter case, this can also be viewed as an instance of *decoying* based on the Bell & Whaley taxonomy, where a signal was given that deception would be used, but in reality the decoys were never present for the Absent-Informed condition. While our analysis examined a breadth of different kinds of deception, additional studies are required, especially for *masking*, *repackaging* and *decoying*.

To further validate our findings, we highlight parallels to the Moonraker Study—a controlled experiment designed to assess host-based cyber deception using virtual machine introspection to hook system events and intercept predefined shell commands to return predefined output [51]. Participants were all unaware of the deception, so Hypotheses $H2$ and $H3$ do not apply. A variety of deception techniques were implemented in response to six different Techniques, Tactics, and Procedures (TTPs) which participants needed to execute in sequence to succeed at the specified task. While the Tularosa Study instead focused on network-based deception and included conditions with participants explicitly made aware of the deception, there are some congruent results which further support the potential generalizability of the findings presented in this publication.

The Moonraker Study indicated that the Absent condition had significantly more participants who successfully completed the cyber task, which demonstrates impeded forward progress of the Present condition. When looking at the proportion of successful TTP commands, the Absent condition had significantly more, indicating wasted resources in the Present condition. Furthermore, for participants who completed the task, those in the Present condition took significantly more time to do so, indicating delay caused by the deception. These findings correspond to results presented in Section 6.1 and help provide support for Hypothesis $H1$ across multiple cyber deception techniques and technologies. Interestingly, while the experience level of the populations differed, similar statistically significant differences in personalities were noted between the populations of the two studies and a baseline population. Both studies indicated significantly more confusion reported by participants in the Present condition, providing support for Hypothesis $H4$ across multiple cyber deception techniques and technologies. Future studies are still required to further replicate and investigate under what circumstances results apply.

Frustration, as well as other stressors like fatigue and increased cognitive workload, has been shown to reduce effectiveness in cyber operators through qualitative studies done at the National Security Agency using the Cyber Operations Stress Survey (COSS) [11]. While confusion and surprise were not included in COSS, it appears these indicators of altered perception have similar effects as discussed in Section 6.1.3. Moreover, these findings, considered with results presented to date from both the Moonraker and Tularosa Studies supporting Hypotheses $H1$ and $H4$, are consistent with the idea that exacerbating feelings of confusion and surprise negatively impact cyber performance. These data can further be used to examine the cognitive and emotional effects of cyber deception in future work, and how these deliberate additional stressors can impact the effectiveness of cyber attackers.

8 Conclusions and Future Work

The data analysis results presented in this paper are consistent with the theory that suspicion by an attacker that deceptive defenses are in place can increase their effect on cyber attack behavior and improve defensive posture. However, future work is still needed. The amount, the method, and the timing with which information about the deceptive defenses is given requires further examination. Providing too many details, such as which commercial decoy system is deployed, on which subnets, and what configuration each decoy is using, will likely make the systems less effective. Even without detailed information, some APTs will likely devise methods for differentiating or avoiding decoys on networks of interest. Cyber security is an arms race, and cyber deception does not change that. Security best-practices and behavior-based security hygiene will always be a critical, but insufficient component of cyber security. This study demonstrates that an effective deception solution has the potential to force attackers to waste time, resources, and mental effort and perhaps trigger early-warnings on zero day attacks for which typical network defenses are unprepared. Even if one APT finds a way to avoid the effects of deception, these defenses can still help protect a network against other attackers.

Future work includes further exploration of previously posed additional hypothesis $H5$ [12]: cognitive biases are prevalent in cyber attacker behaviors and can be intensified to disrupt cyber attacks. To address this hypothesis we will perform a detailed examination of cognitive biases observed in the Tularosa data. Additionally we are creating new experiments [30] specifically focused on measuring and triggering cognitive biases relevant to cyber operations [31].

Furthermore, in order to improve the effectiveness of cyber deception we will use these and future experimental findings to inform utility scores, reward functions, and models to advance artificial intelligence for adaptive decoy systems [7, 13, 21]. We plan to continue to work with experts in cyber operations to enhance understanding of attacker and defender decision-making and improve reasoning and decision-making models to better account for realistic human-behavior. Finally, we will explore how to better leverage large existing CTF-style events to better collect new useful data to help fuel the research community.

Acknowledgments

Portions of this article are based on the lead author's doctoral dissertation [12]. Constructive guidance on statistics and data analysis was provided by advisors: Prof. Brian Levine, Prof. David Jensen, Prof. Robert Gutzwiller and Dr. Dana LaFon. Many collaborators provided input and data wrangling help including: Dr. Sunny Fugate, Dr. Temmie Shade, Andrew Rogers, Mary Berlage, Rob Bruno, and Tiffany Lee. We also want to thank the anonymous USENIX reviewers and our shepherd, Laura Tinnel, for their helpful feedback.

References

[1] Acalvio. ShadowPlex™. https://www.acalvio.com/product/, (Accessed = 2019-10-11).

[2] P. Aggarwal, C. Gonzalez, and V. Dutt. *HackIT: A Real-Time Simulation Tool for Studying Real-World Cyberattacks in the Laboratory*, pages 949–959. Springer International Publishing, Cham, 2020.

[3] H. Arkes and C. Blumer. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes,*, 35:124–140, 1985.

[4] A. Basak, J. Černý, M. Gutierrez, S. Curtis, C. Kamhoua, D. Jones, B. Bošanský, and C. Kiekintveld. An initial study of targeted personality models in the flipit game. In *Decision and Game Theory for Security*, pages 623–636. Springer International Publishing, 2018.

[5] A. Baset and T. Denning. A data-driven reflection on 36 years of security and privacy research. In *Proceedings of the 12th USENIX Conference on Cyber Security Experimentation and Test (CSET)*, USA, 2019.

[6] J.B. Bell and B. Whaley. *Cheating and Deception*. St. Martin's Press, 1st edition, 1982.

[7] M. Bilinski, K.J. Ferguson-Walter, S.J. Fugate, R. Gabrys, J. Mauger, and B.J. Souza. You only lie twice: A multi-round cyber deception game of questionable veracity. *Conference on Decision and Game Theory for Security*, October 2019.

[8] F. Cohen, I. Marin, J. Sappington, C. Stewart, and E. Thomas. Red teaming experiments with deception technologies. *IA Newsletter*, 2001.

[9] E. A. Cranford, C. Gonzalez, P. Aggarwal, S. Cooney, M. Tambe, and C. Lebiere. Toward personalized deceptive signaling for cyber defense using cognitive models. In *Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, January 2020.

[10] S.P. Curley, J.F. Yates, and R.A. Abrams. Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes*, 38(2):230–256, 1986.

[11] J. Dykstra and C.L. Paul. Cyber Operations Stress Survey (COSS): Studying fatigue, frustration, and cognitive workload in cybersecurity operations. *USENIX Workshop on Cyber Security Experimentation and Test (CSET)*, 2018.

[12] K.J. Ferguson-Walter. *An Empirical Assessment of the Effectiveness of Deception for Cyber Defense*. PhD thesis, University of Massachusetts Amherst, Feb 2020.

[13] K.J. Ferguson-Walter, S.J. Fugate, J. Mauger, and M.M. Major. Game theory for adaptive defensive cyber security. *ACM Hot Topics in the Science of Security Symposium (HotSoS)*, March 2019.

[14] K.J. Ferguson-Walter, D.S. LaFon, and T.B. Shade. Friend or Faux: Deception for Cyber Defense. *Journal of Information Warfare*, 16(2):28–42, 2017.

[15] K.J. Ferguson-Walter, M.M. Major, D.C. Van Bruggen, S.J. Fugate, and R.S. Gutzwiller. The world of CTF is not enough data: Lessons learning from a cyber deception experiment. *In Proceedings of First IEEE Workshop on Human Aspects of Cyber Security (HACS)*, 2019.

[16] K.J. Ferguson-Walter, T.B. Shade, A.V. Rogers, E.M. Niedbala, M.C. Trumbo, K. Nauer, K. Divis, A.P. Jones, A. Combs, and R.G. Abbott. The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception. In *Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, 2019.

[17] K.J. Ferguson-Walter, T.B. Shade, A.V. Rogers, M.C. Trumbo, K. Nauer, K. Divis, A.P. Jones, A. Combs, and R.G. Abbott. Appendix to The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception, 2019. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/TularosaAppendix.pdf.

[18] Fidelis Cybersecurity. Fidelis Deception®. https://www.fidelissecurity.com/products/deception/, (Accessed = 2019-10-11).

[19] D. Fraunholz, S.D. Anton, C. Lipps, D. Reti, D. Krohmer, F. Pohl, M. Tammen, and H.D. Schotten. Demystifying deception technology: A survey. arXiv, 2018.

[20] D. Fraunholz, F. Pohl, and H.D. Schotten. Towards basic design principles for high-and-medium-interaction honeypots. In *16th European Conference on Cyber Warfare and Security (EECWS 2017)*, 2017.

[21] S.J. Fugate and K.J. Ferguson-Walter. Artificial intelligence and game theory models for defending critical networks with cyber deception. *AI Magazine*, 40(1):49–62, Mar 2019.

[22] Galois. CyberChaff™. https://galois.com/project/cyberchaff/, (Accessed = 2019-10-11).

[23] S. Goel, K. J. Williams, and E. Dincelli. Got phished? internet security and human vulnerability. *Journal of the Association for Information Systems*, 18:2, 2017.

[24] R.S. Gutzwiller, K.J. Ferguson-Walter, S.J. Fugate, and A.V. Rogers. 'Oh, Look, A butterfly!' A framework for distracting attackers to improve cyber defense. In *Human Factors and Ergonomics Society (HFES)*, 2018.

[25] N. Han, X.and Kheir and D. Balzarotti. Deception techniques in computer security: A research perspective. *ACM Computing Surveys*, 51(4), July 2018.

[26] K.E. Heckman, F.J. Stech, R.K. Thomas, Be. Schmoker, and A.W. Tsow. *Cyber Denial, Deception and Counter Deception: A Framework for Supporting Active Cyber Defense*. Advances in Information Security. Springer International Publishing, 2015.

[27] R.J. Heuer. Cognitive factors in deception and counter deception. In *Strategic Military Deception*. Pergamon Press Inc, 1981.

[28] E.M. Hutchins, M.J. Cloppert, and R.M. Amin. 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, 2011.

[29] Illusive Networks. Illusive platform. https://www.illusivenetworks.com/technology/platform/, (Accessed = 2019-10-11).

[30] C.K. Johnson. *Decision-Making Biases in Cybersecurity: Measuring the Impact of the Sunk Cost Fallacy to Delay and Disrupt Attacker Behavior*. PhD thesis, Arizona State University, 2021. (Manuscript in preparation).

[31] C.K. Johnson, R.S. Gutzwiller, K.J. Ferguson-Walter, and S.J. Fugate. A cyber-relevant table of decision making biases and their definitions. ResearchGate, 2020.

[32] P.N. Johnson-Laird. *Mental Models: Towards a Cognitive Science of Language, Inferences, and Consciousness*. Harvard University Press, 1983.

[33] A.P. Jones and M.C. Trumbo. Personal Communication, November 2019. Sandia National Laboratories.

[34] F.B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, and G. Ahn. Matched and mismatched SOCs: A qualitative study on security operations center issues. In *ACM Conference on Computer and Communications (CCS)*, 2019.

[35] E.A. Krusemark, W.K. Campbell, and B.A. Clementz. Attributions, deception, and event related potentials: An investigation of the self-serving bias. *Psychophysiology*, 45(4):511–515, July 2008.

[36] J. McAlaney, L.A. Frumkin, and V. Benson. *Psychological and Behavioral Examinations in Cyber Security*. IGI Global, 2018.

[37] J.B. Michael. On the response policy of software decoys: Conducting software-based deception in the cyber battlespace. In *26th Annual International Computer Software and Applications*, pages 957–962, 2002.

[38] J.B. Michael, N.C. Rowe, H.S. Rothstein, T.C. Wingfield, M. Auguston, and D. Drusinsky. Phase II report on intelligent software decoys: intelligent software decoy tools for cyber counterintelligence and security countermeasures. *Technical Report NPS-CS-04-001*, 2004.

[39] Microsoft. Microsoft Security Bulletin MS17-010 - Critical, 2017. https://docs.microsoft.com/en-us/security-updates/securitybulletins/2017/ms17-010.

[40] V. Nicomette, M. Kaâniche, E. Alata, and M. Herrb. Set-up and deployment of a high-interaction honeypot: experiment and lessons learned. *Journal in Computer Virology*, 7(2):143–157, May 2011.

[41] S.E. Parkin, K. Krol, I. Becker, and M.A. Sasse. Applying cognitive control modes to identify security fatigue hotspots. In *USENIX Conference on Usable Privacy and Security*, 2016.

[42] J. Pawlick, E. Colbert, and Q. Zhu. A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. arXiv, 2017.

[43] S.L. Pfleeger and D.D. Caputo. Leveraging behavioral science to mitigate cyber security risk. *Computer Security*, 31(4):597–611, June 2012.

[44] M. Rabin and J.L. Schrag. First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1):37–82, 1999.

[45] N.C. Rowe, E.J. Custy, and B.T. Duong. Defending cyberspace with fake honeypots. *Journal of Computers*, 2(2):25–36, April 2007.

[46] N.C. Rowe and J. Rrushi. *Introduction to Cyberdeception*. Springer International Publishing, 2016.

[47] M.B. Salem and S.J. Stolfo. On the design and execution of cyber-security user studies: Methodology, challenges, and lessons learned. In *USENIX Workshop on Cyber Security Experimentation and Test (CSET)*, 2011.

[48] G. Schudel and B. Wood. Adversary work factor as a metric for information assurance. In *New Security Paradigms Workshop*, page 23–30, New York, NY, 2000.

[49] G. Schudel and B. Wood. Modeling behavior of the cyber-terrorist. In *Workshop on Research on Mitigating the Insider Threat to Information Systems - #2*. RAND, Aug 2000.

[50] Offensive Security. Kali Linux, 2017.

[51] T.B. Shade, A.V. Rogers, K.J. Ferguson-Walter, S.B. Elsen, D. Fayette, and K.E. Heckman. The Moonraker Study: An Experimental Evaluation of Host-Based Deception. In *Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, January 2020.

[52] W. R. Shadish, T. D. Cook, and D. T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Wadsworth Cengage Learning, 2002.

[53] R. Stake. *The Art of Case Study Research*. Sage, Thousand Oaks, CA, 1995.

[54] S.C. Sundaramurthy, A. Bardas, J. Case, X. Ou, M.Wesch, J. McHugh, and S. Rajagopalan. A human capital model for mitigating security analyst burnout. In *USENIX Symposium on Usable Privacy and Security*, 2015.

[55] L. Tinnel, O. S. Saydjari, and D. Farrell. Cyberwar strategy and tactics an analysis of cyber goals , strategies, tactics, and techniques. In *IEEE Workshop on Information Assurance*, June 2002.

[56] TrapX Security. DeceptionGrid™. https://trapx.com/product/, (Accessed = 2019-10-11).

[57] G. Wagener, R. State, A. Dulaunoy, and T. Engel. Self Adaptive High Interaction Honeypots Driven by Game Theory. In *International Symposium on Stabilization, Safety, and Security of Distributed Systems*, 2009.

[58] J. Yuill, D. Denning, and F. Feer. Using Deception to Hide Things from Hackers. *Journal Of Information Warfare*, 5(3):26–40, 2006.

A Summary of Statistical Analysis Results

For reference, a concise collection of findings supportive of our hypotheses to date can be seen in Table 3.

Hypothesis H1: *Cyber and psychological deception impedes attackers.*				
Metric	**Data Source**	**Lower Mean**	**Higher Mean**	**p-value**
⋄Decoy Target Selection	Keylog/RHOST	Present-Uninformed	Present-Informed	p = .004**
⋄Stolen Credentials	Keylog	Present-Informed	Absent-Uninformed	p = .003**
Eternal Blue Attempted	OCR	Present	Absent	p = .046*
Eternal Blue Detected	IDS logs	Present	Absent	p = .014*
Reported Exploit Successes	Mattermost	Present	Absent	p = .011*
Data Exfiltration	OCR	Present	Absent	p = .055
Keystroke Count	Keylog	Present	Absent	p = .047*
Commands with real hosts	Keylog	Present	Absent	p < .01**
Bytes to real IPs	Network Capture	Present	Absent	p = .022*
Hypothesis H2: *Cyber deception tools are effective even if an attacker is aware of their use.*				
⋄Decoy Target Selection	Keylog/RHOST	Present-Uninformed	Present-Informed	p = .004**
Time to First Real Target	Keylog/RHOST	Informed	Uninformed	p = .072
⋄Stolen Credentials	Keylog	Present-Informed	Absent-Uninformed	p = .003**
Eternal Blue Detected	IDS logs	Present-Informed	Absent-Uninformed	p = .050*
Time to First Decoy Alert	Decoy Alerts	Present-Informed	Present-Uninformed	p = .035*
Less-severe Decoy Alerts	Decoy Alerts	Present-Uninformed	Present-Informed	p < .006*
More-severe Decoy Alerts	Decoy Alerts	Present-Informed	Present-Uninformed	p < .0001***
Total decoy alerts	Decoy alerts	Present-Informed	Present-Uninformed	p < .0001***
Decoy Login Attempts	Decoy alerts	Present-Informed	Present-Uninformed	p = .004**
†Reported Confusion	TSQ (Likert Scale)	Uninformed	Informed	p = .044*
Hypothesis H4: *Cyber, and psychological, deception affects an attacker's cognitive and emotional state.*				
Security Assessment	End-of-day report	Absent-Uninformed	Present-Informed	p = .03*
†Reported Confusion	TSQ (Likert Scale)	Absent	Present	p = .011*
†Reported Surprise	TSQ (Likert Scale)	Uninformed	Informed	p = .044*
†Suspicion of Deception	TSQ (Labeled)	Absent-Uninformed	Present-Informed	p = .009**

Table 3: **Summary of findings:** Significant differences are indicated as $***p < .001$, $**p < .01$, $*p < .05$.
† denotes analyses from the previous published analyses [16]. ⋄ denotes analyses which support both Hypotheses 1 and 2.

B Data Exfiltration

Exfiltrated File Type	`user.csv`	`ntds.dit`	Domain Controller Files	PowerShell `*.ps`	`NTUSER .dat`	Registry `*.reg`	`mscache`	`local_admins .csv`	**Total**
Absent-Uninformed (N=35)	0	12 (n=1)	0	0	63 (n=2)	74 (n=5)	31 (n=4)	0	**180 (n=9)**
Absent-Informed (N=28)	2 (n=1)	0	11 (n=1)	8 (n=1)	23 (n=1)	19 (n=3)	0	0	**63 (n=4)**
Present-Uninformed (N=30)	0	17 (n=1)	0	0	0	33 (n=1)	8 (n=1)	2 (n=1)	**60 (n=4)**
Present-Informed (N=30)	5 (n=1)	1 (n=1)	6 (n=1)	19 (n=1)	0	0	0	0	**31 (n=2)**

Table 4: **Exfiltration**: Counts of valuable files exfiltrated, as identified by OCR. The number of unique participants who acquired that file type is denoted by n. Participants in the Absent-Uninformed (control) condition had the most exfiltration success.

C Population Level of Expertise

For reference, Table 5 depicts the level of expertise from participants recruited for the Tularosa Study and provides descriptive statistics [33].

		Rating		
Question	Sub-Category	N	Mean	Stdev
Level of Expertise (1 = novice, 5 = expert)	Cyber Security	128	3.64	0.93
	Network penetration	128	2.92	1.08
	Host penetration	128	2.93	1.10
	Network reconnaissance	128	3.39	1.12
	Incident response	128	2.79	1.15
	Generalized defense practice	127	3.38	1.16
	Network protocol reverse engineering	128	2.02	1.05
	Binary reverse engineering	128	1.77	0.99
Involvement in each phase of engagement (1 = least, 5 = most)	Reconnaissance	128	3.38	1.35
	Weaponization	129	2.74	1.36
	Delivery of weaponized bundle	129	2.69	1.36
	Exploitation	128	3.00	1.33
	Installation of malware	126	2.73	1.38
	Command and control channel for remote manipulation	128	2.78	1.44
	Actions on objectives	126	3.27	1.32
Match to typical engagement (1 = least, 5 = most)	Compliance testing	129	3.02	1.52
	Blue team training	129	2.53	1.35
	Demonstrate the needs for increased security investments	125	3.32	1.33
	Whiteboarding/gaming/tabletop exercises	129	2.65	1.27
	Post-attack remediation effort	128	2.84	1.26
	Vulnerability analysis	128	2.62	1.32
	Security architecture review	129	3.27	1.28
	Persistent adversary emulation	129	2.59	1.46
Years of Experience	Cyber Security	128	7.87	5.61
	Network penetration	128	4.26	3.91
	Host penetration	128	4.11	3.74
	Network reconnaissance	128	5.04	4.06
	Incident response	126	3.79	4.29
	Generalized defense practice	129	6.90	6.27
	Network protocol reverse engineering	128	1.82	2.67
	Binary reverse engineering	128	1.51	2.24

Table 5: **Population Expertise:** Self-reported level of expertise for each skill set measured for participants in the Tularosa Study [16]. The highest mean level of expertise and involvement reported was in network reconnaissance, which was the most relevant skill for the decoy-based deception used in the Tularosa Study.

D Tularosa Results on Cyber Kill Chain

The following graphic summarizes the main forward progress research results from this publication along with its corresponding stage of the Cyber Kill Chain [28] and specifies whether it was supportive of Hypotheses $H1$ and/or $H2$.

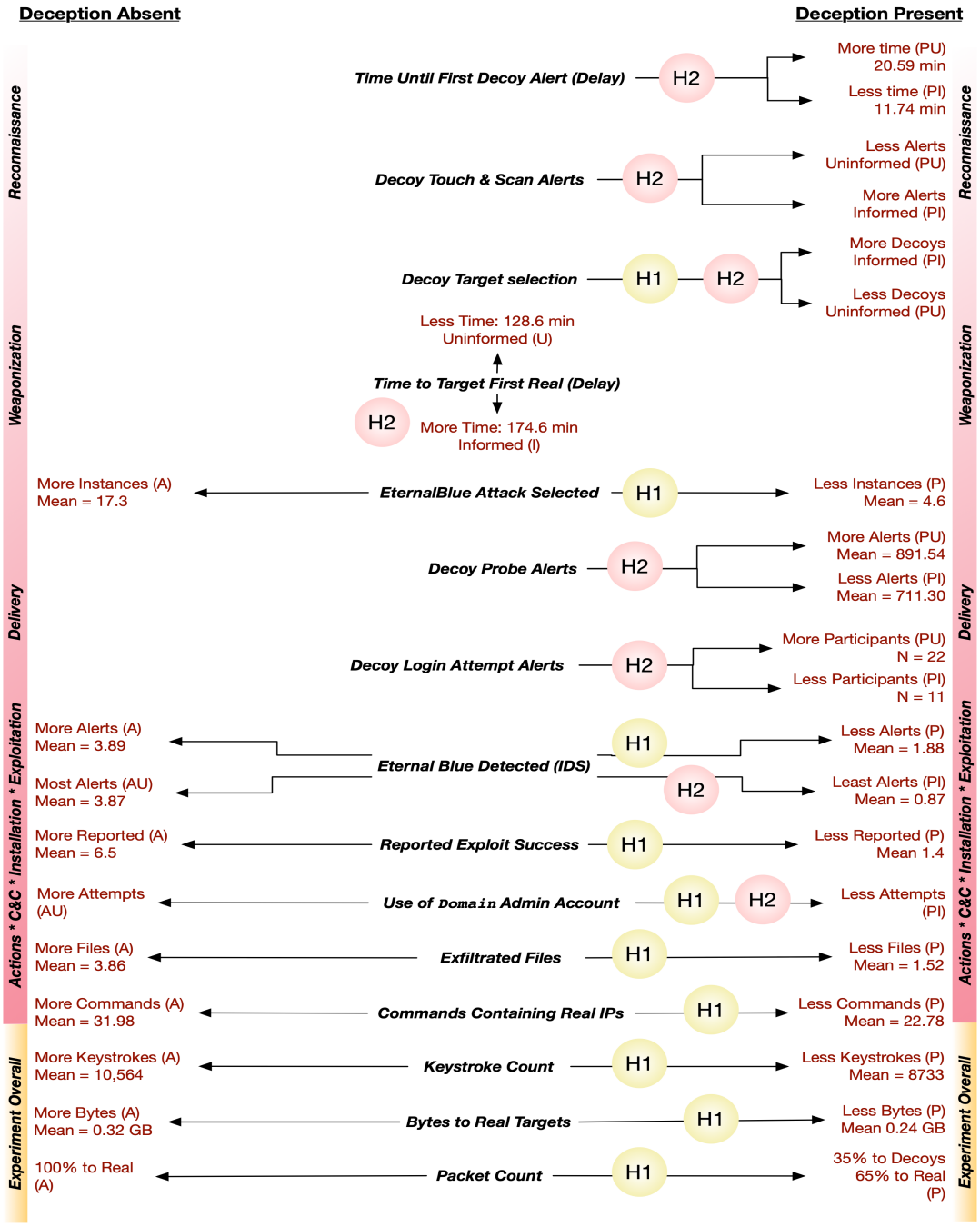


Figure 3: **Forward Progress Results Summary:** Tularosa data analysis results are displayed roughly aligned to the Cyber Kill Chain to illustrate delayed and impeded forward progress caused to a cyber attacker through use of defensive deception. Acronyms correspond to experimental conditions: Decoys Absent (A), Decoys Present (P), Informed of Deception (I), Not Informed ("Uninformed") of Deception (U). Bolded arrows and text indicate Hypothesis-supporting data for the impact of deception. See corresponding Appendix A for a table summarizing the statistical results of all meaningful findings to date.