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

Background Concepts

- ***Cyber Deception***
  - Levels the playing field.
  - Simple to complex solutions.
- ***Cyberpsychology***
  - The scientific field that integrates human behavior and decision-making into the cyber domain, allowing us to *understand*, *anticipate* and *influence* attacker behavior.
- ***Goal***: Rigorous measures of effectiveness

The Tularosa Study

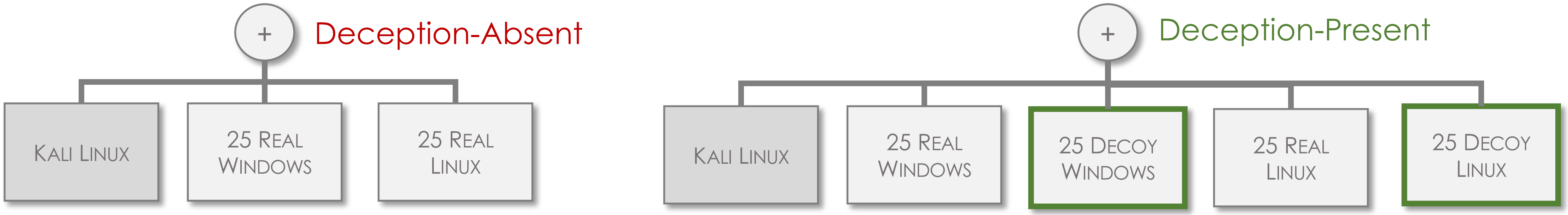
- 138 professional penetration testers (“red-teamers”)
- Full day penetration testing exercise on a test network
- Kali Linux provided *“to use for reconnaissance and system exploitation”*

> *“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****…*
> *When you learn potentially useful information about target systems on this network you will* ***immediately report this information*** *to your team”*
>
> \- Tularosa Task Instructions

K.J. Ferguson-Walter, T.B. Shade, A.V. Rogers, E.M. Niedbala, M.C. Trumbo, K. Nauer, K.M. Divis, A.P. Jones, A. Combs, R.G. Abbott.
The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception. *HICSS 2019*: 1-10.

The Tularosa Experiment

- Experimental Conditions:
  - Deception-Absent, and participants Uninformed (control condition)
  - Deception-Absent, but participants Informed
  - Deception-Present, but participants Uninformed
  - Deception-Present, and participants Informed



Deception-Absent

Kali Linux

25 Real Windows

25 Real Linux

Deception-Present

Kali Linux

25 Real Windows

25 Decoy Windows

25 Real Linux

25 Decoy Linux

"There may be deception on the network"

- Additional statement for Informed participants

Hypotheses

- **H1:** Defensive cyber tools and psychological deception impede attackers who seek to penetrate computer systems and exfiltrate information.

- **H2:** Defensive deception tools are effective even if an attacker is aware of their use.

- **H3:** Cyber deception is effective if the attacker merely believes it may be in use, even if it is not.

- **H4:** Cyber and psychological deception affects an attacker’s cognitive and emotional state.

Data Analysis

Data Sources:

- Network Traffic (PCAP)
- Intrusion Detection System (IDS) Alerts
- Host Data
  - Keylogs on attack client
- Decoy Alerts
- Screen Recordings
  - Optical Character Recognition (OCR) derived from screen recordings
- Self-report data
  - Real-time logs
  - Retrospective

Analysis Methods:

- Data were non-normal
  - Non parametric statistical tests (Chi-Square, Kruskal-Wallis test)
  - Dunn’s post hoc test with Benjamini-Hochberg correction
- Qualitative Data – reviewed by two subject matter experts

Analysis: Measures of Success

- **Internal** versus **external** validity
  - Not a Capture The Flag (CTF)

- Individual "Success"
  - Determined by each participant

- Red Teamers as proxy for hackers
  - Measured by progress mapping, attacking, and exfiltrating from the network.

With deception, the attacker's perception of success
may not reflect *true* progress toward their goals

K.J. Ferguson-Walter, M.M. Major, D.C. Van Bruggen, S.J. Fugate, R.S. Gutzwiller. **The World (of CTF) is Not Enough Data: Lessons Learning from a Cyber Deception Experiment**. *IEEE HACS 2019*.

Analysis: Measures of Success

Forward Progress

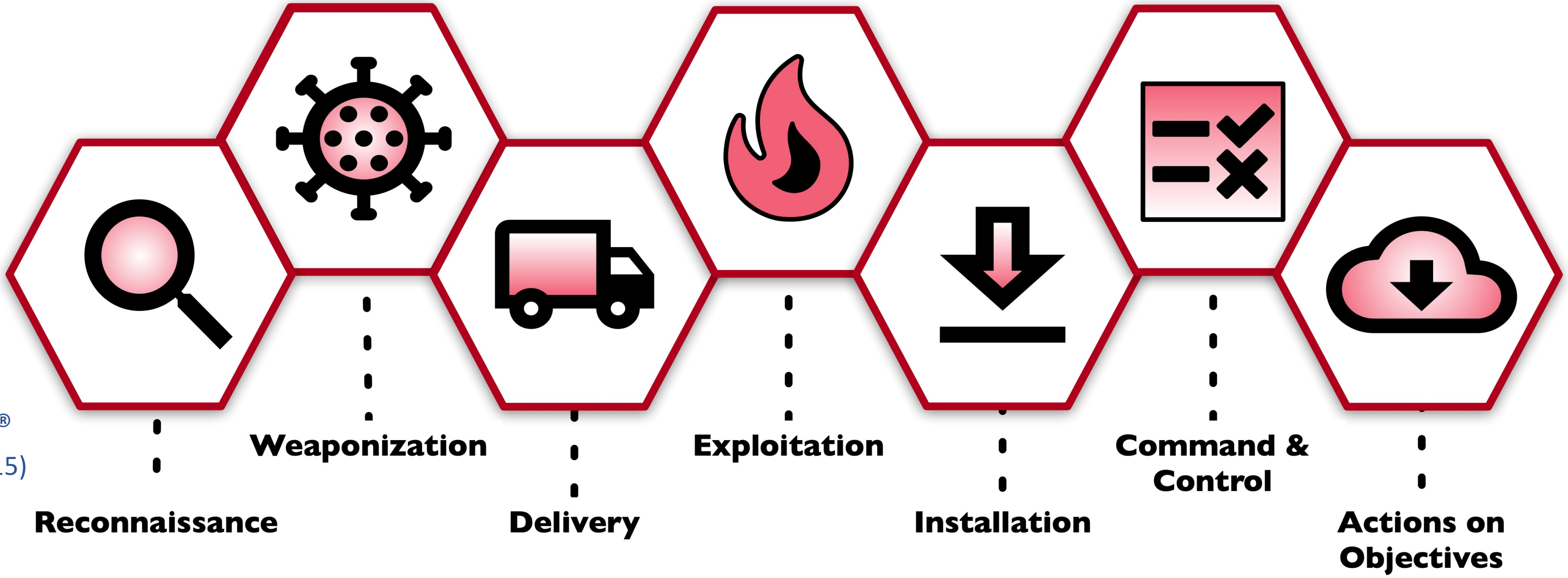
- Target Selection
- Stolen Credentials
- Use of EternalBlue Exploit
- Self-reported Exploits
- Data Exfiltration
- Keystroke Count
- Delay Effect

Wasted Resources

- Commands
- Network Traffic
- Decoy Alerts

Altered Perception

- Success/Failure
- Security Assessment



Cyber Kill Chain®
Lockheed Martin (2015)

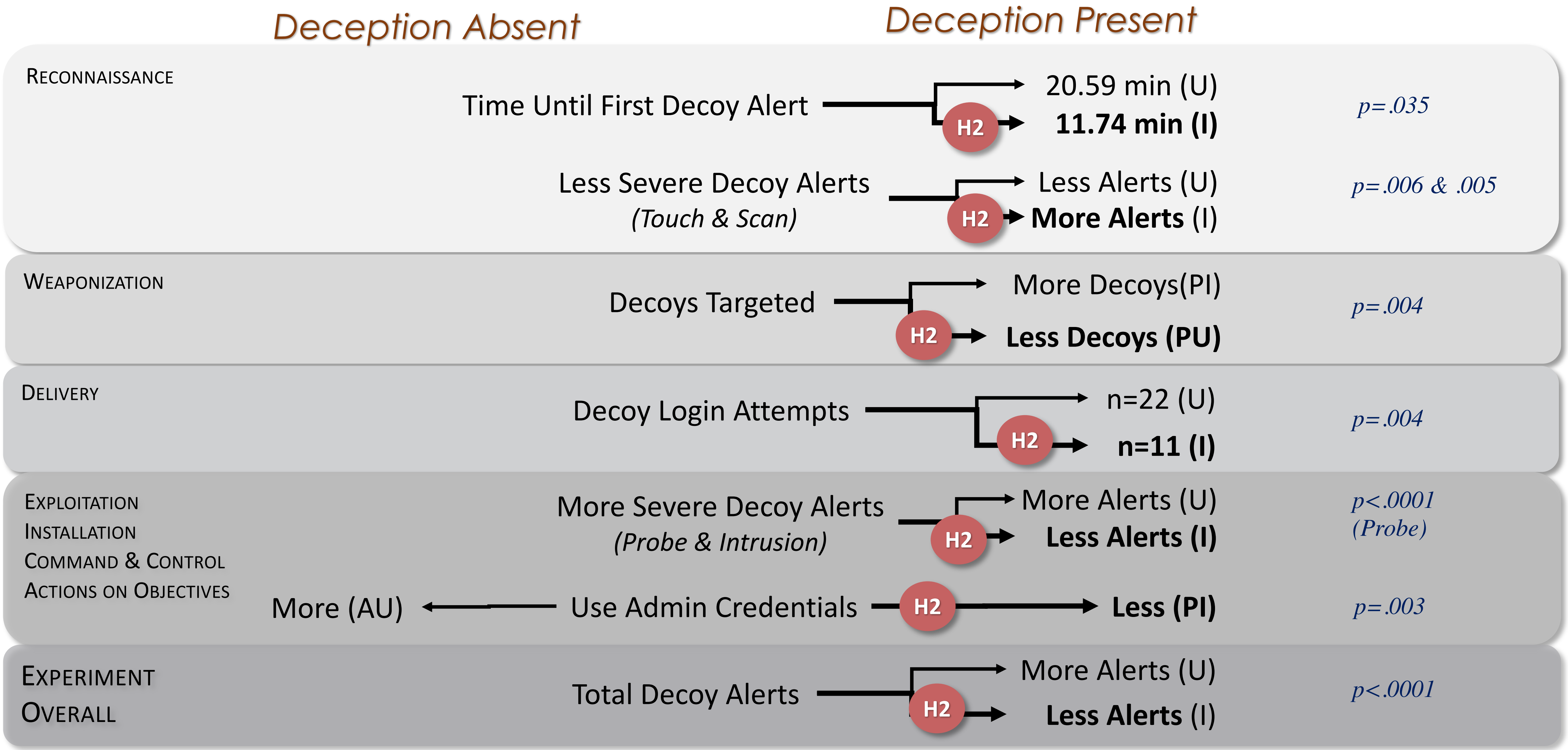
Analysis: Measures of Success

- Defender (Experimental) Success:
  - **Impeded Attacker Forward Progress** – *Strategic gains*
  - **Delayed Attacker Progress** – *Strategic gains & wasted effort*
  - **Attacker Resources Expended** – *Increased effort*
  - **Altered Attacker Perception** – *Difference between reality & deception*

Experiment Results

	Deception Absent			*Deception Present*	
Reconnaissance					
Weaponization		Decoys Targeted		More Decoys(PI)	*p=.004*
			H1	**Less Decoys (PU)**	
Delivery	Mean 17.3	EternalBlue Selected	H1	**Mean 4.6**	*p=.046*
Exploitation Installation Command & Control Actions on Objectives	Mean 3.89	EternalBlue Detected (Suricata)	H1	**Mean 1.88**	*p=.014*
	Mean 6.5	Reported Exploit Success	H1	**Mean 1.4**	*p=.011*
	More (AU)	Use Admin Credentials	H1	**Less (PI)**	*p=.003*
	Mean 3.86 files	Data Exfiltrated	H1	**Mean 1.52 files**	*p=.055*
Experiment Overall	Mean 31.98	Commands to Real Targets	H1	**Mean 22.78**	*p<.01*
	Mean 0.32 GB	Bytes to Real Targets	H1	**Mean 0.24 GB**	*p=.022*
	100% to Real	Packet Count	H1	**35% to Decoys**	

Experiment Results



Deception Absent

Deception Present

Reconnaissance

Time Until First Decoy Alert → 20.59 min (U) / H2 → **11.74 min (I)** *p=.035*

Less Severe Decoy Alerts *(Touch & Scan)* → Less Alerts (U) / H2 → **More Alerts** (I) *p=.006 & .005*

Weaponization

Decoys Targeted → More Decoys(PI) / H2 → **Less Decoys (PU)** *p=.004*

Delivery

Decoy Login Attempts → n=22 (U) / H2 → **n=11 (I)** *p=.004*

Exploitation

Installation

Command & Control

Actions on Objectives

More Severe Decoy Alerts *(Probe & Intrusion)* → More Alerts (U) / H2 → **Less Alerts (I)** *p<.0001 (Probe)*

More (AU) ← Use Admin Credentials → H2 → **Less (PI)** *p=.003*

Experiment Overall

Total Decoy Alerts → More Alerts (U) / H2 → **Less Alerts** (I) *p<.0001*

Forward Progress: EternalBlue Exploit

Weaponization:

Absent conditions loaded more of the *EternalBlue module* into Metasploit *(p = .046).*

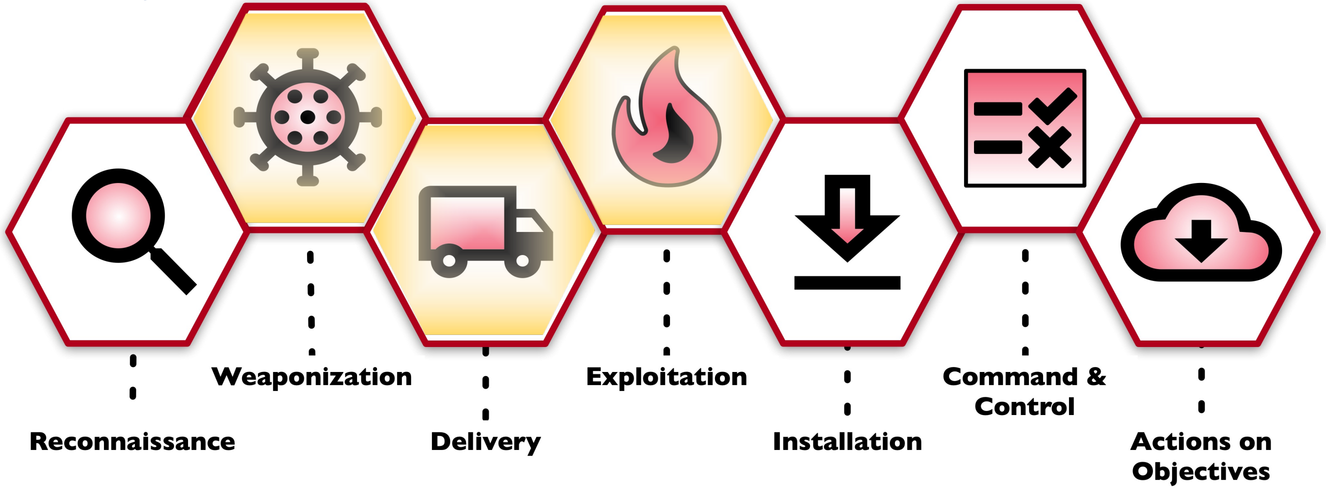
Delivery:

Absent conditions generated more *EternalBlue attempts on real targets (p = .014).*

Exploitation:

Trend of more Absent conditions reporting more EternalBlue exploit *successes (p = .076).*

✓ **H1: Presence of decoys impeded attacker forward progress**



Cyber Kill Chain®
Lockheed Martin (2015)

Forward Progress: Target Selection

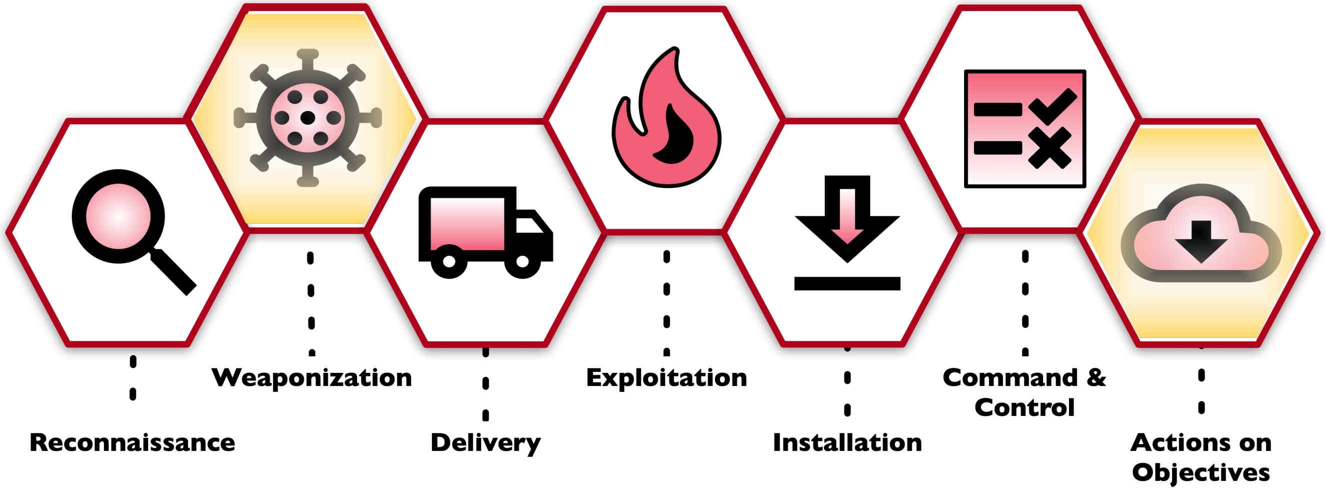
Weaponization:

Present-Informed *targeted more decoys* than Present-Uninformed (p = .004).

Privilege Escalation and Lateral Movement:

Fewer Present-Informed used *stolen domain admin credentials* than Present-Uninformed (p = .003).

- ✓ **H2: Information on deception did not impact decoy effectiveness**
- ✓ **H2: Information on deception reduced forward progress**



Cyber Kill Chain®
Lockheed Martin (2015)

Altered Perception: Success/Failure

- Data Source: End-of-day Report

- Success

Example: *“The assessment was fairly simple in terms of complexity.”*

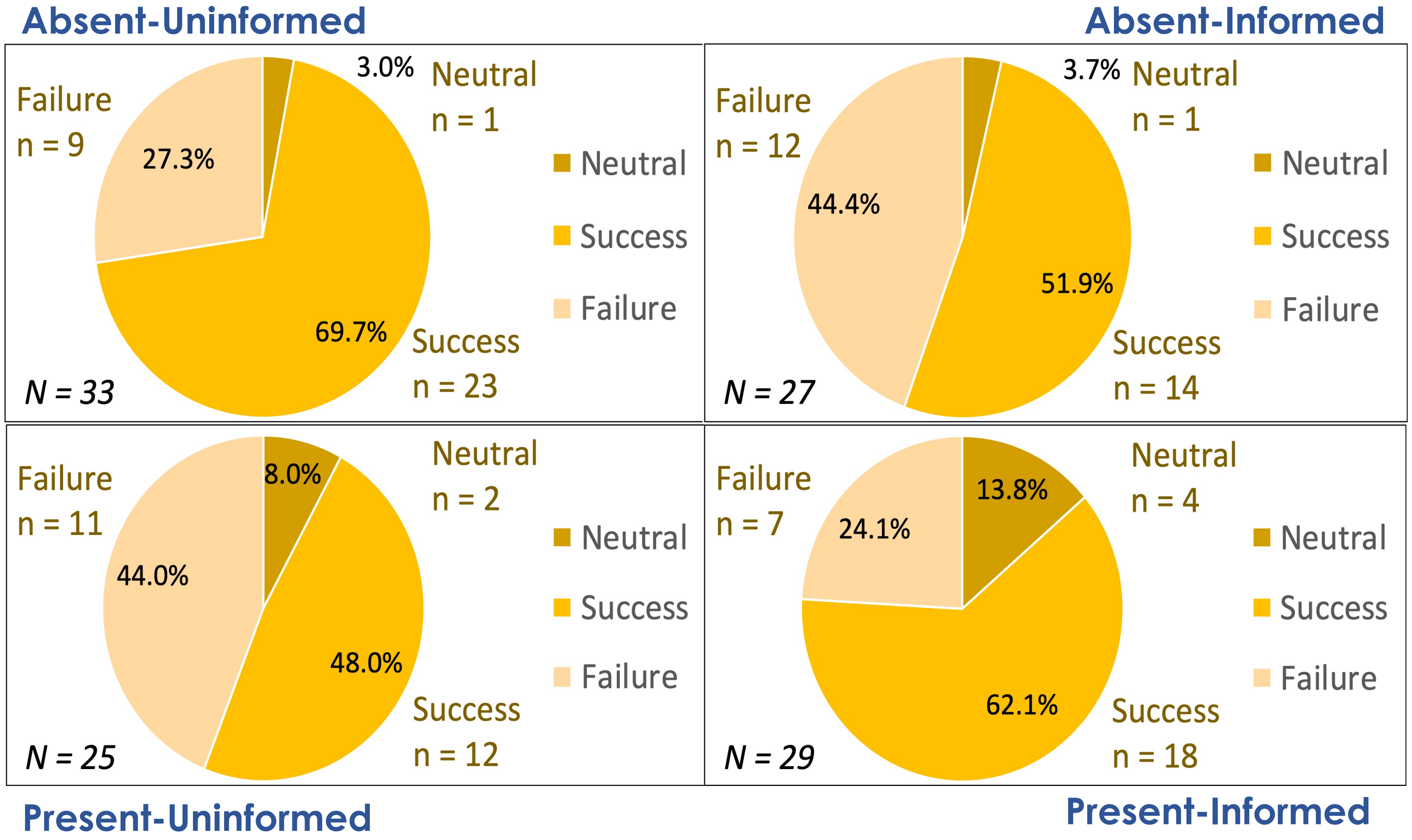
- Failure

Example: *“All of the exploits I tried to run today were not successful.”*

- Neutral

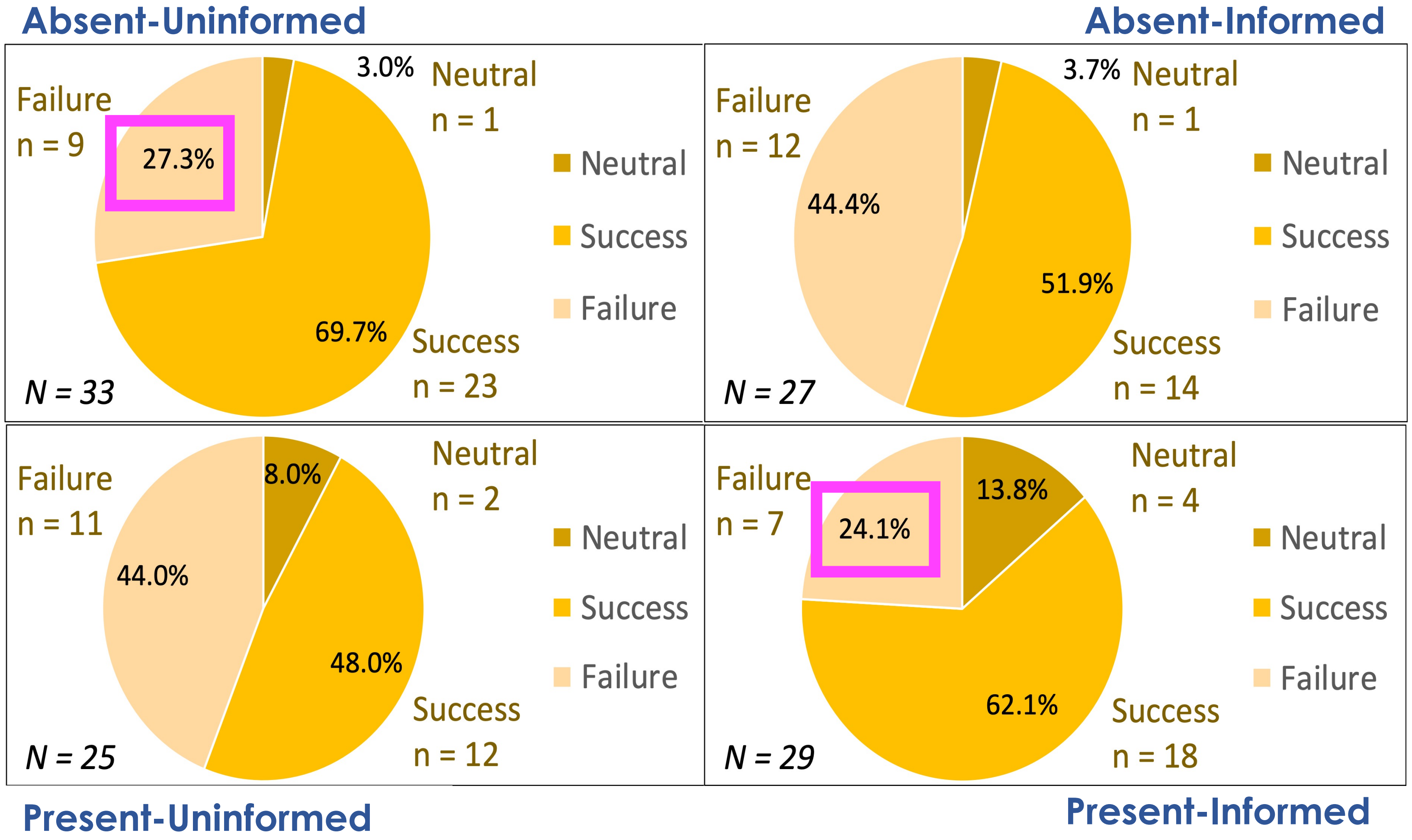
Example: *“I am extremely happy to be here. Please hire me!”*

Altered Perception: Success/Failure



- ✓ **Altered Perception**
- ✓ **H4: Reduced self-reported failures in Present-Informed condition: self-serving bias.**

Altered Perception: Success/Failure

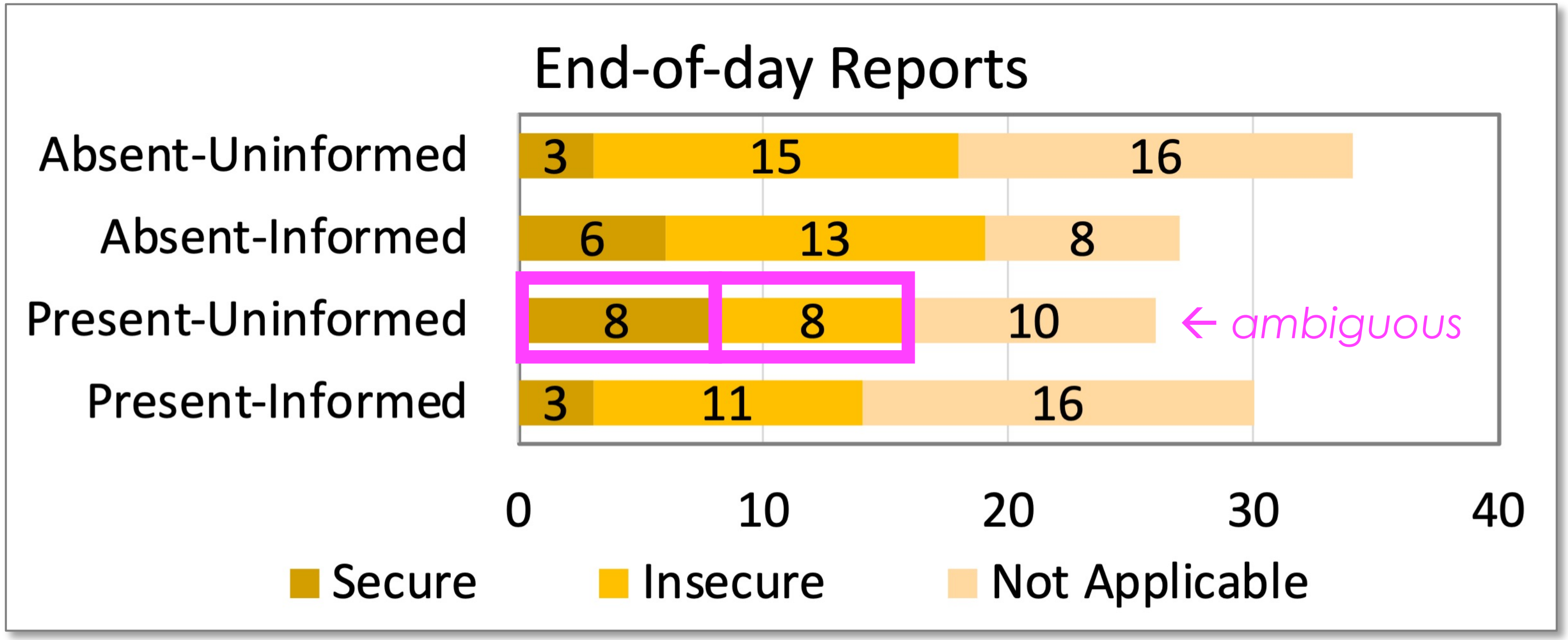


- ✓ **Altered Perception**
- ✓ **H4: Reduced self-reported failures in Present-Informed condition: self-serving bias.**

<u>Self-serving Bias:</u>
Deception provided an excuse for ***participants who no longer felt responsible for the failures***, and thus reported them less often

Altered Perception: Network Security

- More Present Uninformed described network as *secure* than Absent-Uninformed *(p=.030)*
  - *Possible Ambiguity Effect*



Ambiguity Effect: Ambiguity causes people to be unwilling to act.

Present-Uninformed had the most ambiguity.

Data Analysis Results

- **H1:** *Cyber and psychological deception impedes attackers.*
  - Participants in Deception Present conditions:

Impeded Forward Progress

- *Targeted more decoys (p=.004)*
- *Used domain admin account less (p=.003)*
- *Less Eternal Blue exploit attempts (p=.046)*
- *Reported less exploit successes (p=.011)*
- *Generated less keystrokes overall (p=.047)*
- *Exfiltrated fewer files (p=.055)*

Altered Perception

Delayed Progress

- *Sent less bytes to real targets (p=.022)*
- *Typed less commands with real IPs (p=.009)*
- *Sent 35% of packets overall to decoys*
- *Over 10 GB of network traffic sent to decoys*

Resources Expended

Data Analysis Results

- **H1:** *Cyber and psychological deception impedes attackers.*

> *"I eventually pwned everything.*
>
> ***Every. Single. Domain. Asset. Pwned.****"*
>
> \- Absent-Uninformed Participant S104

> *"There was **a lot of frustration**. . .*
>
> *I don't really think there is too much that is actually exploitable."*
>
> \- Present-Uninformed Participant S87

Data Analysis Results

- **H2:** *Cyber deception tools are effective even if an attacker is aware of their use.*
  - With Deception Present, Informed participants:

Impeded Forward Progress

Altered Perception

- *Selected more decoys as targets (p=.004)*
- *Used the domain admin account less (p=.003)*
- *Generated less late-stage decoys (p<.0001)*
- *Generated the less Eternal Blue alerts (p=.05)*

Delayed Progress

- *Took less time to trigger a decoy alert (p=.035)*
- *Took more time to select first real target (p=.072)*

Resources Expended

- *Generated more early-state decoy alerts (p<.006)*

> *"I think I **wasted a lot of time** looking for the deception."*
>
> \- Present-Informed Participant S116

Data Analysis Results

- **H3:** *Cyber deception is effective if the attacker merely believes it may be in use, even if it is not.*
  - Observational support only:

- *Mismatch between self-reports and reality*

Altered Perception

> *"This network was **filled with deception** and I spent the majority of the day going down rabbit holes that led me nowhere."*
>
> \- Absent-Informed Participant S106

> *"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."*
>
> \- Absent-Informed Participant S119

Data Analysis Results

- **H4:** *Cyber and psychological deception affects an attacker's cognitive and emotional state.*
  - Compared to the control condition:
    - *More Present-Uninformed considered network secure (p=.03)*
    - *Fewer Present-Informed reported failure on cyber task (self-serving bias)*

Altered Perception

> *"I did not find any aspects of the network that were frustrating or confusing.* ***Everything seemed relatively straight-forward****."*
>
> - Absent-Uninformed Participant S138

> *"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."*
>
> - Present-Uninformed Participant S87

Conclusions & Future Work

- **Human decision-making** is critical but often overlooked.
- **Decoys are effective** technique to impede, detect, & delay cyber attacks.
- **Deception** is part of the cyber arms race.
- **Cyber** and **psychological** deception _together_ have the greatest impact.
- Follow-on work:
  - ***Cognitive biases*** *relevant to cyber operations.*

Contact: Kimberly.j.ferguson-walter.civ@mail.mil

Research Directorate
Advancing Intelligence Through Science

Naval Information
Warfare Center
PACIFIC

ASU