Proposer's Day: Reimagining Security with Cyberpsychology-Informed Network Defense (ReSCIND) Proposers' Day

Dr. Kimberly Ferguson-Walter | Program Manager | Feb 28, 2023

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Welcome to the ReSCIND Proposers' Day!

- Thank you for your interest in this program and participation in this event
- To assure a clear broadcast stream, audio and video are disabled for meeting participants
- Comments and questions can be submitted in one of three ways:
  - Using the WebEx Chat feature, send questions to "Host"
  - To the alias (dni-iarpa-ReSCIND-proposers-day@iarpa.gov) during the meeting
  - Index cards to drop boxes in the meeting rooms or to the registration desk
- *No questions will be fielded during talks.*

Proposer's Day Goals

1. Familiarize participants with IARPA's outline of the ReSCIND program and solicit questions and feedback
2. Foster discussion of complementary capabilities among potential program participants, i.e., **teaming**
   - Teaming information: https://www.iarpa.gov/research-programs/rescind
   - An attendance list, with contact details from participants who opted to share their information, will be distributed
   - The chat feature is enabled for participants to plan future discussions associated with teaming
   - Teaming interests, capability summaries, lightning talk slides, and posters, will be posted publicly on the ReSCIND IARPA webpage until the BAA submission period closes

Please ask questions and provide feedback, this is your chance to alter the course of events.
Please talk with others, find great team members.

Disclaimers

- This presentation is provided solely for information and planning purposes
- The Proposers’ Day does not constitute a formal solicitation for proposals or proposal abstracts
- Nothing said at Proposers’ Day changes the requirements set forth in a BAA
- The BAA language supersedes anything presented or said by IARPA at Proposers’ Day
- This meeting is being recorded and will be posted for public viewing
- For those viewing the recording, email aliases and POCs may be dated, please refer to IARPA.gov for updated information

Feedback and Questions

- Questions can be submitted until 9:40am PT/12:40pm ET
- There will be a break after the contracting presentation at 9:30am PT/ 12:30pm ET
- Responses to selected questions will be broadcast at 11:00am PT/2:00pm ET, so please don't log out or close your WebEx connection
  - All programmatic, technical, and contractual questions will be captured, but not necessarily answered in this session
- Feedback about the draft technical description may be submitted to the IARPA team at dni-iarpa-ReSCIND-proposers-day@iarpa.gov
  - A new alias will be established for when the full BAA is released
- After this Proposer's Day, IARPA will review all the feedback received for a final BAA to be posted on SAM.gov

Collaborations

- **Collaboration is highly encouraged; ReSCIND is an extremely interdisciplinary endeavor**
- Lightning-Talk session at 11:30am PT/ 2:30pm ET
- Teaming Discussions (in person only) at 2:00pm PT
- Remote participants are encouraged to organize their own teaming discussions
- Capability Statements will be received and posted publicly, pending minimal review for appropriateness.
  - Capability Statements can be submitted until the BAA closes by sending to: dni-iarpa-ReSCIND-proposers-day@iarpa.gov
- Lightning Talks, Capability Statements, and Teaming Forms are for peers to explore collaborations and resources, for forming the best proposal. The government's evaluation resides *only* with the proposal.

ReSCIND Proposers' Day Agenda

Time	Topic	Speaker
8:00am-8:10am PT	Welcome, Logistics, Proposers’ Day Goals	Kimberly Ferguson-Walter, Program Manager
8:10am-8:20am PT	IARPA Overview	Robert Rahmer, Director Office of Analysis Research, IARPA
8:20am-9:10am PT	ReSCIND Program Overview	Kimberly Ferguson-Walter
9:10am-9:30am PT	Contracting Overview	Stephen Enokida, Contracting Officer
9:30am-11:00am PT	Break (Submit questions in chat or drop boxes before 9:40am PT)	
11:00am-11:30am PT	Answers to Selected Technical Questions	Kimberly Ferguson-Walter
11:30am-11:35am PT	Introductions to Lightning Talks	Kimberly Ferguson-Walter
11:35am-2:00pm PT (est.)	Lightning Talks	Potential Performers
2:00pm-3:30pm PT	Informal Teaming Discussions	In-Person Participants

Lightning Talks Agenda

Please submit questions before 9:40am PT/12:40pm ET.

Time	Speaker	Institution	In Person
11:35am-11:40am PT	Joseph Dingley	Social Machines Co	No
11:40am-11:45am PT	Merve Sahin	SAP Security Research	No
11:45am-11:50am PT	Scott Brown	University of Newcastle	No
11:50am-11:55am PT	Radu Marculescu	University of Texas, Austin	No
11:55am-12:00pm PT	David Starobinski	Boston University	No
12:00pm-12:05pm PT	Alexander Poylisher	Peraton Labs	No
12:05pm-12:10pm PT	Zak Fry	GrammaTech	No
12:10pm-12:15pm PT	Yu Huang	Vanderbilt University	No
12:15pm-12:20pm PT	Dan Thomsen	Smart Information Flow Technologies (SIFT)	No
12:20pm-12:25pm PT	Gentry Lane	Anova Intelligence	No
12:25pm-12:30pm PT	Mary Aiken	Capitol Technology University	No

Lightning Talks Agenda (Continued)

Time	Speaker	Institution	In Person
12:30pm-12:40pm PT	BREAK		
12:40pm-12:45pm PT	Frank DiGiovanni	Parallax Advanced Research	Yes
12:45pm-12:50pm PT	Prashanth Rajivan	University of Washington	Yes
12:50pm-12:55pm PT	Anthony Palladino	Draper Labs	Yes
12:55pm-1:00pm PT	Frederico Araujo	IBM (Watson Research Center)	Yes
1:00pm-1:05pm PT	Palvi Aggarwal	University of Texas, El Paso	Yes
1:05pm-1:10pm PT	Michael Sieffert	Assured Information Security	Yes
1:10pm-1:15pm PT	Michael Lundie	Applied Research Associates (ARA)	Yes
1:15pm-1:20pm PT	Noam Ben-Asher	SimSpace	Yes
1:20pm-1:25pm PT	Aaron Brown	c3.ai	Yes
1:25pm-1:30pm PT	Diego Gomez-Zara	University of Notre Dame	Yes

Lightning Talks Agenda (Continued)

Time	Speaker	Institution	In Person
1:30pm-1:35pm PT	Brenda Wiederhold	Virtual Reality Medical Center	Yes
1:35pm-1:40pm PT	Robert McGraw	RAM Labs	Yes
1:40pm-1:45pm PT	Sean Guarino	Charles River Analytics	Yes
1:45pm-1:50pm PT	Amory Bennett	Quorum Research	Yes
1:50pm-1:55pm PT	David Huberdeau	Riverside Research Institute	Yes
1:55pm-2:00pm PT	Sanjay Goel	University of Albany, SUNY	Yes
2:00pm-3:30pm PT	Informal Teaming Discussions and Poster Session		In-Person Participants

IARPA Overview

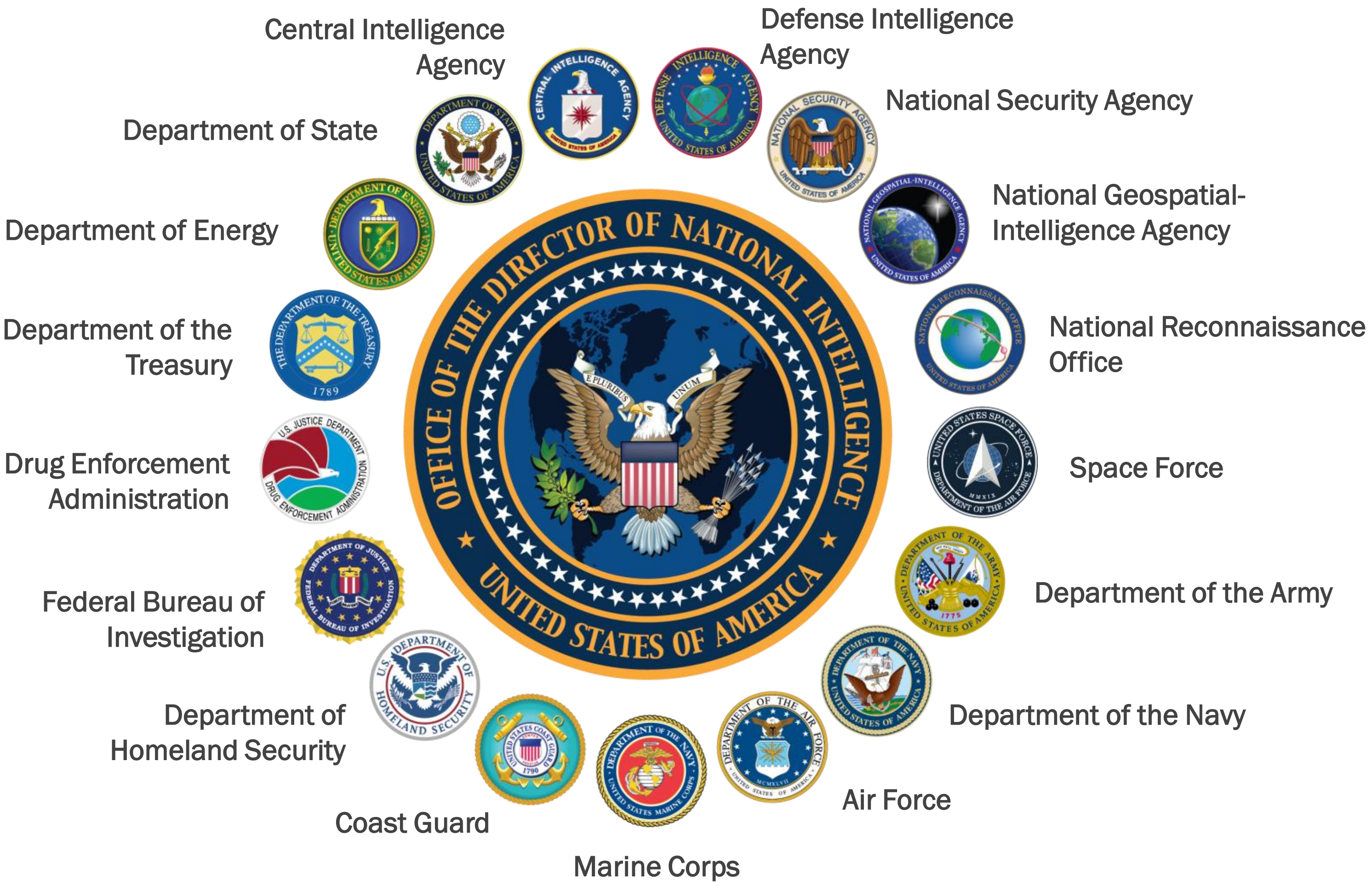
Robert Rahmer | Director, IARPA Office of Analysis | ReSCIND Proposers' Day, Feb 28, 2023

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Office of the Director of National Intelligence



Central Intelligence Agency

Defense Intelligence Agency

National Security Agency

National Geospatial-Intelligence Agency

National Reconnaissance Office

Space Force

Department of the Army

Department of the Navy

Air Force

Marine Corps

Coast Guard

Department of Homeland Security

Federal Bureau of Investigation

Drug Enforcement Administration

Department of the Treasury

Department of Energy

Department of State

IARPA Mission

IARPA envisions and leads *high-risk, high-payoff research* that delivers innovative technology *for future overwhelming intelligence advantage*

- Our problems are complex and multidisciplinary
- We emphasize technical excellence & technical truth

IARPA Method

- Bring the best minds to bear on our problems
  - Full and open competition to the greatest possible extent
  - World-class, term-limited Program Managers
- Define and execute research programs that:
  - Have goals that are clear, ambitious, credible and measurable
  - Run from three to five years
  - Publish peer-reviewed results and data, to the greatest possible extent
  - Employ independent and rigorous Test & Evaluation
  - Involve IC partners from start to finish
  - Transition new capabilities to intelligence community partners

IARPA R&D

- Technical and programmatic excellence are required
- Each program has a clearly defined and measurable end-goal
  - Intermediate milestones to measure progress are also required
  - Every program has a beginning and an end
- This approach, coupled with term-limited PM positions, ensures
  - IARPA does not “institutionalize“ programs
  - Fresh ideas and perspectives are always coming in
  - Status quo is always questioned
  - Only the best ideas are pursued, and only the best performers are funded

IARPA Snapshot

IARPA's research portfolio is diverse, including math, physics, chemistry, biology, microelectronics, neuroscience, linguistics, political science, cognitive psychology, and more.

- 70% of completed research transitions to U.S. Government partners
- 3,000+ journal articles published
- IARPA funded researchers have been awarded the Nobel Prize in Physics for quantum computing research, a MacArthur Fellowship, and a Bell prize
- IARPA serves on National Science and Technology Council (NSTC) committees and actively engages with the White House BRAIN Initiative, National Strategic Computing Initiative, and the NSTC Select Committee on Artificial Intelligence, the NSTC Subcommittee on Quantum Information Science (SCQIS), and NSTC Subcommittee on Economic and Security Implications of Quantum Science (ESIX)

How to Engage with IARPA

ENGAGE WITH US

Throughout our website you can learn more about engaging with us on our highly innovative work that is having a positive impact in the Intelligence Community and society in general. Click on any of the below links to learn more.

iarpa.gov | 301-243-1995

dni-iarpa-info@iarpa.gov

- Reach out to our Program Managers.
- Schedule a visit if you are in the DC area or invite us to visit you

Open BAAs

Broad Agency Announcements (BAAs) solicit research proposals for specific programs. Learn more about current BAA opportunities and ways to get involved...

Requests For Information

Requests for Information (RFIs) are designed to gather more information on an idea in an area in which our program managers are not fully informed...

Seedlings

Seedlings are typically 9 – 12 month research efforts that are less than $1M in cost. They are intended to address highly innovative ideas and concepts within...

Proposer's Day: Reimagining Security with Cyberpsychology-Informed Network Defense (ReSCIND) Overview

Dr. Kimberly Ferguson-Walter | Program Manager | Feb 28, 2023

OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE
UNITED STATES OF AMERICA

CELEBRATING 15 YEARS

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Technical Slides Disclaimer

- All images, references, and figures are included as illustrative examples only
- ODNI and IARPA do not endorse any product or company referenced within
- Changes have occurred since the draft technical document was released and additional changes may occur in the final released BAA

Problem Statement

- Cyber attacks are increasing in quantity and severity
- Gaps exist in cyber defense technologies and evaluation techniques
- Lack of research on the decision-making processes of cyber attackers.
- Attackers take advantage of human limitations and errors, but defenses generally do not
- Many sophisticated and persistent cyber attacks facing the IC are primarily human-driven

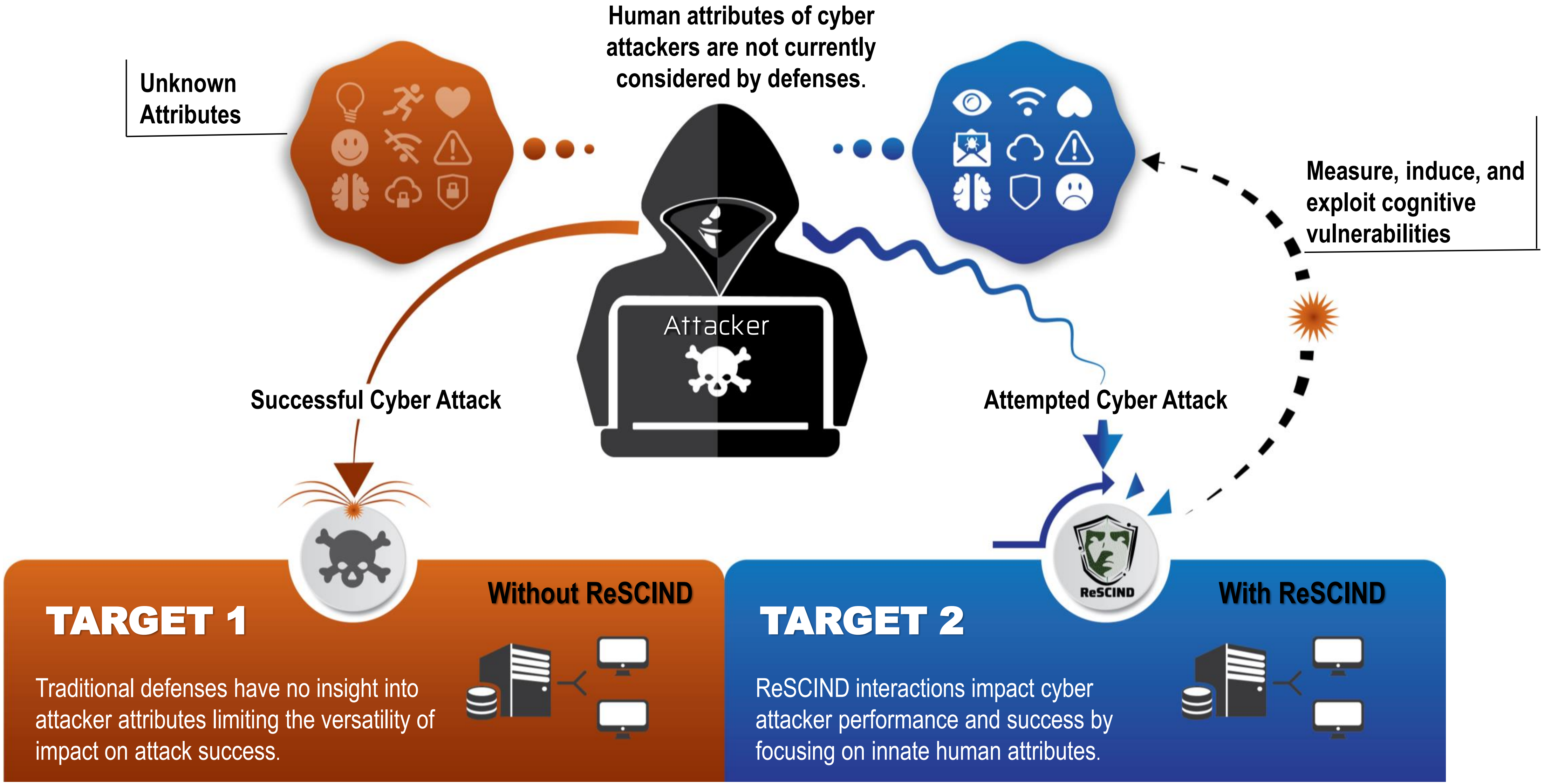
> ***"The human factor is the weakest link in cyber attacks."***

ReSCIND Program Objectives

- Shift the asymmetric nature of cyber defense to benefit defenders
- Influence and manipulate cyber attacker’s decision-making throughout the phases of a cyber attack
- Build novel cyberpsychology-informed defenses *(CyphiDs)*
- Increase the effort and resources for cyber attackers

Research & develop novel & effective CyphiDs to exploit the cognitive vulnerabilities of attackers

Solving the Problem: ReSCIND Approach



Human attributes of cyber attackers are not currently considered by defenses.

Unknown Attributes

Measure, induce, and exploit cognitive vulnerabilities

Without ReSCIND

TARGET 1

Traditional defenses have no insight into attacker attributes limiting the versatility of impact on attack success.

With ReSCIND

TARGET 2

ReSCIND interactions impact cyber attacker performance and success by focusing on innate human attributes.

Why at this Time?

- Other domains successful profit from Cognitive Vulnerabilities (CogVuls), but cyber defense lags behind
- Cyberpsychology for cyber defense is an emerging area
  - Historically focused elsewhere (e.g., online dating, cyberbullying, online gaming)
  - Behavioral scientists and cyber security researchers rarely work together
  - Cyber Deception research and technologies lay groundwork, but utilize only a few human attributes
- Cyber-relevant cognitive biases have begun to be hypothesized, but scientific groundwork still needed

Cyberpsychology Foundation

Cyberpsychology: *The scientific field that integrates human behavior and decision-making into the cyber domain, allowing us to understand, anticipate and influence attacker behavior*

Cyberpsychology-Informed Defenses (CyphiDs)

- Well-established **behavioral science constructs**
- **Scientifically rigorous**
- Establishes useful **metrics and measures**
- Quantifies **effectiveness** of methods
- Defines **research limitations**
- Understands human **cognition and decision making**
- Develops methods to **influence** cyber behavior
- **Informs** automated defense systems

Cyberpsychology Applied to Cyber Defense

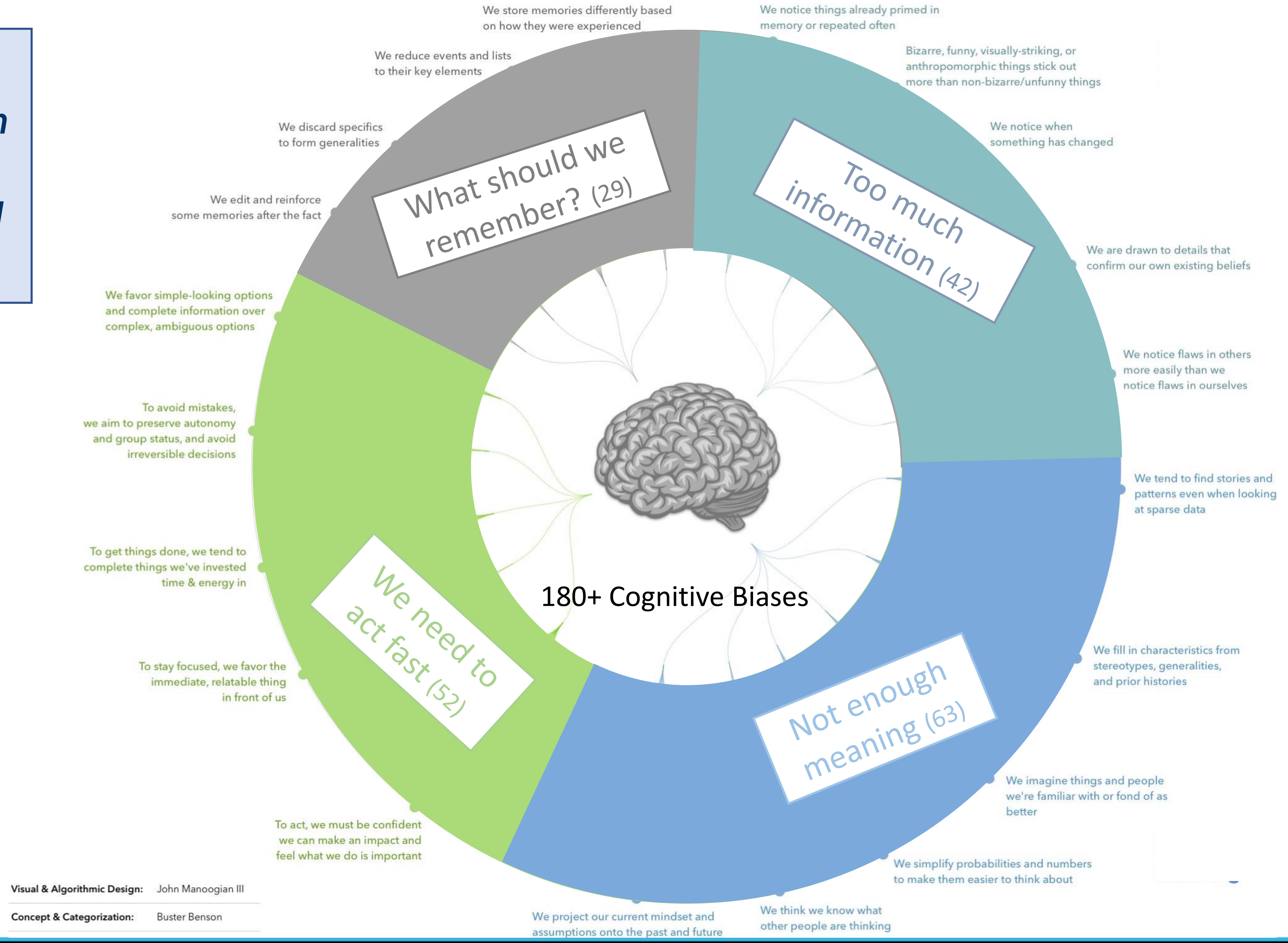
- Existing research on decision-making doesn’t easily abstract to cyber
  - Fictitious, hypothetical decision-making scenarios
  - Often students asked to role play
  - Little effort required in task

- Cyber activities are different from previous, simple studies
  - Time-constrained, multi-step decisions in diverse and complicated situations with high-impact risks and rewards
  - Existing theory must be extended into more realistic cyber decision-making scenarios

New human subjects research (HSR) required to explore dynamic cyber attack tasks with skilled human participants

State of Current Versus Needs

Most cognitive biases and human inclinations have yet to be explored for cyber defense



What should we remember? (29)

Too much information (42)

Not enough meaning (63)

We need to act fast (52)

180+ Cognitive Biases

We store memories differently based on how they were experienced

We reduce events and lists to their key elements

We discard specifics to form generalities

We edit and reinforce some memories after the fact

We notice things already primed in memory or repeated often

Bizarre, funny, visually-striking, or anthropomorphic things stick out more than non-bizarre/unfunny things

We notice when something has changed

We are drawn to details that confirm our own existing beliefs

We notice flaws in others more easily than we notice flaws in ourselves

We tend to find stories and patterns even when looking at sparse data

We fill in characteristics from stereotypes, generalities, and prior histories

We imagine things and people we're familiar with or fond of as better

We simplify probabilities and numbers to make them easier to think about

We think we know what other people are thinking

We project our current mindset and assumptions onto the past and future

To act, we must be confident we can make an impact and feel what we do is important

To stay focused, we favor the immediate, relatable thing in front of us

To get things done, we tend to complete things we've invested time & energy in

To avoid mistakes, we aim to preserve autonomy and group status, and avoid irreversible decisions

We favor simple-looking options and complete information over complex, ambiguous options

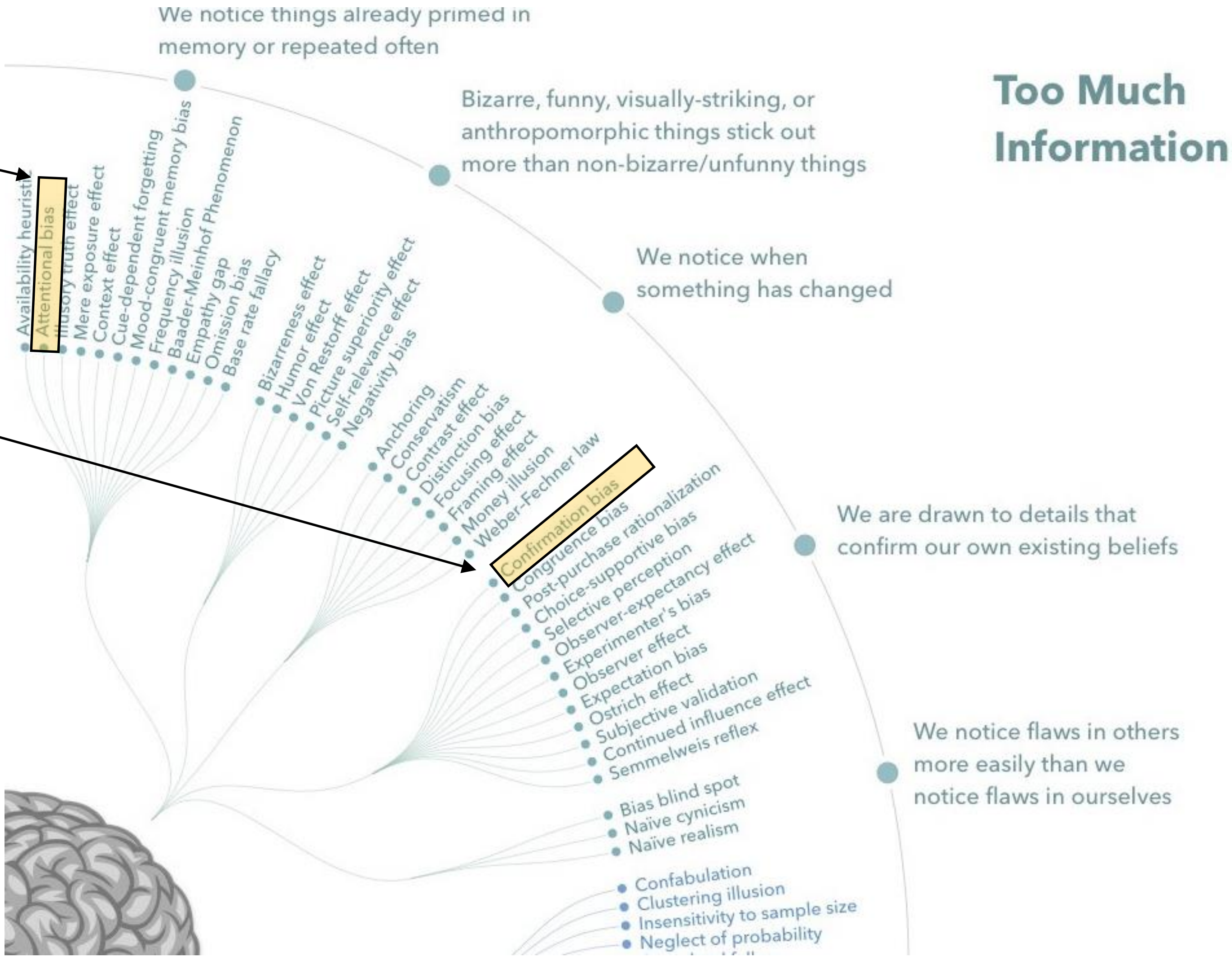
Visual & Algorithmic Design: John Manoogian III

Concept & Categorization: Buster Benson

Cyber Deception Focuses on Few Biases

- Honeypots are designed to induce **attentional tunneling** and hold an attacker’s attention.
- Decoys and honeytokens benefit from **confirmation bias,** the tendency to search for or interpret information in a way that affirms one’s preconceptions.

Many additional CogVuls could be influential in the cyber domain



Too Much Information

We notice things already primed in memory or repeated often

Availability heuristic, Attentional bias, Illusory truth effect, Mere exposure effect, Context effect, Cue-dependent forgetting, Mood-congruent memory bias, Frequency illusion, Baader-Meinhof Phenomenon, Empathy gap, Omission bias, Base rate fallacy

Bizarre, funny, visually-striking, or anthropomorphic things stick out more than non-bizarre/unfunny things

Bizarreness effect, Humor effect, Von Restorff effect, Picture superiority effect, Self-relevance effect, Negativity bias

We notice when something has changed

Anchoring, Conservatism, Contrast effect, Distinction bias, Focusing effect, Framing effect, Money illusion, Weber-Fechner law

We are drawn to details that confirm our own existing beliefs

Confirmation bias, Congruence bias, Post-purchase rationalization, Choice-supportive bias, Selective perception, Observer-expectancy effect, Experimenter's bias, Observer effect, Expectation bias, Ostrich effect, Subjective validation, Continued influence effect, Semmelweis reflex

We notice flaws in others more easily than we notice flaws in ourselves

Bias blind spot, Naïve cynicism, Naïve realism

Confabulation, Clustering illusion, Insensitivity to sample size, Neglect of probability

Achieving Defender Goals

***Cognitive Vulnerability:** Cognitive and decision-making biases, innate cognitive limitations, emotional or mental state, or physiological vulnerabilities resulting in reduced attacker success or effectiveness*

ReSCIND performers will design novel defenses spanning different categories to influence cyber attackers through **manipulation of well-established cognitive vulnerabilities.**

Attacker

Notional CogVul Categories

- Influencing Decisions
- Altered Risk Taking
- Memory Effects
- Attention Allocation
- Inducing Errors
- Other

Defender Goals

- Impeded Attack Goals
- Increased Detection
- Wasted Attack Resources
- Delayed Attacker Goals
- Increased Attacker Effort
- Other

Defender

Examples of Potentially Cyber-Relevant CogVuls

There are a plethora of unexplored cyber-relevant CogVuls that can be used against attackers.

Influencing Decisions

- Choice Overload: Too many available choices can cause difficulty making a decision.
- Sunk Cost Fallacy: Tendency to continue with a specific strategy because of prior investments, such as time or effort.
- Ambiguity Effect: Tendency to avoid options that have an unknown probability of a favorable outcome.
- Default Effect: When given a choice between several options, the tendency to favor the default one.
- Availability Heuristic: Tendency to use easily available information and ignore not easily available sources of significant information.

****List not exhaustive**

Examples of Potentially Cyber-Relevant CogVuls

Altered Risk Taking

- Peltzman Effect: Tendency to take greater risks when perceived safety increases.
- Loss Aversion: The tendency for people to strongly prefer avoiding losses more than acquiring gains.

Memory Effects

- Von Restorff Effect: Tendency for an item that stands out like a sore thumb to be more likely to be remembered than other items.
- Information Access Cost: The time, physical and mental cost of accessing information can effect powerful changes in cognitive processing strategies that subsequently affect performance.

**List not exhaustive

Examples of Potentially Cyber-Relevant CogVuls

Attention Allocation

- Attentional Tunneling: Allocation of attention to a particular channel of information or task goal, for a longer than optimal duration.
- Inattention Blindness: The failure to perceive an unexpected stimulus in plain sight, purely as a result of a lack of attention.
- Endowment Effect: The tendency for people to value something higher as soon as they own it.

Inducing Errors

- Heavy Cognitive Load: Increase in the amount of mental effort used in the working memory typically creates error or interference in the task at hand.
- Representativeness Bias: The tendency to overweight the representativeness of a piece of evidence while ignoring how often it occurs.

**List not exhaustive

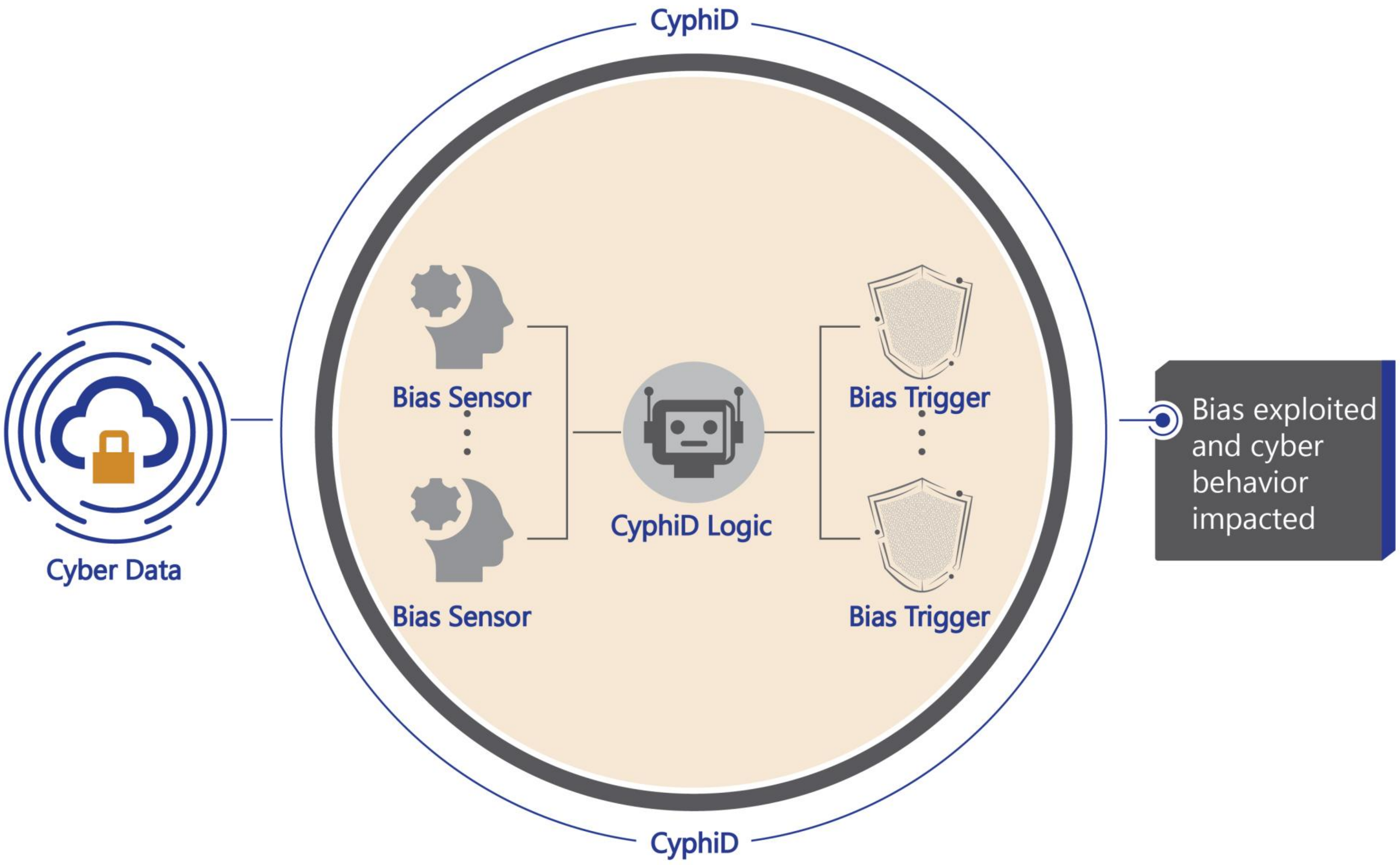
ReSCIND Program Plan

ReSCIND Program Overview

Phase 1	18 months
Phase 2	15 months
Phase 3	12 months

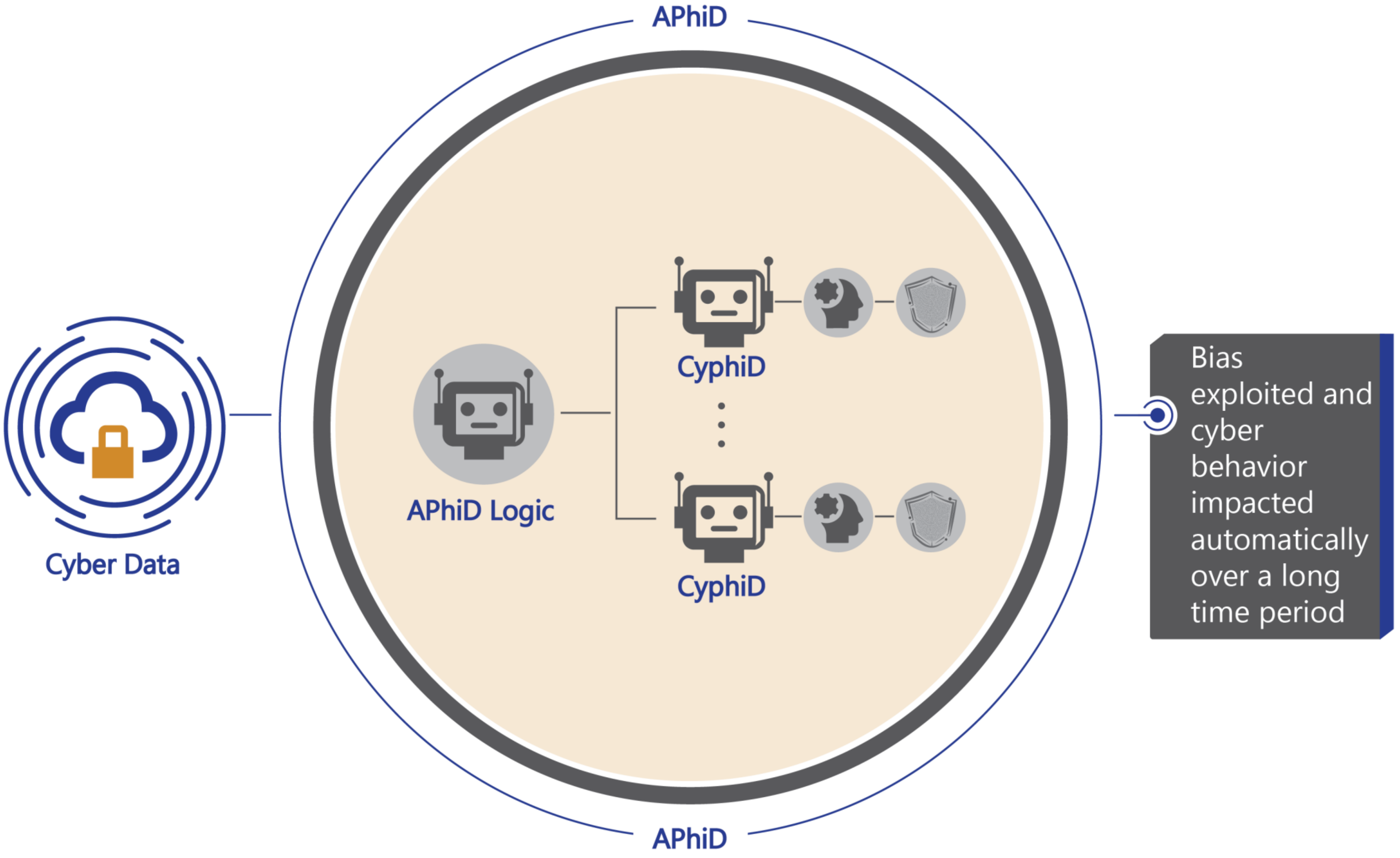
1. Identify cyber-relevant cognitive vulnerabilities (Phase 1)
   - Cognitive vulnerabilities may not be mutually exclusive; theoretically founded clusters are acceptable
   - Bias sensors measure cognitive vulnerabilities using cyber data
2. Induce changes in cyber attacker behavior/success (Phase 1 & 2)
   - Bias triggers create cyber situations that intensify/exploit the cognitive vulnerability
3. Develop Cyberpsychology-informed Defenses (CyphiDs) (Phase 2)
4. Create Cyber-specific Computational Cognitive Models (C3M) that reflect and predict attacker behavior (Phase 3)
5. Produce Adaptive Psychology-informed Defenses (APhiDs) which automate CyphiD sequence based on observables (Phase 3)

Notional Cyberpsychology-informed Defenses (CyphiDs)



Phase 2

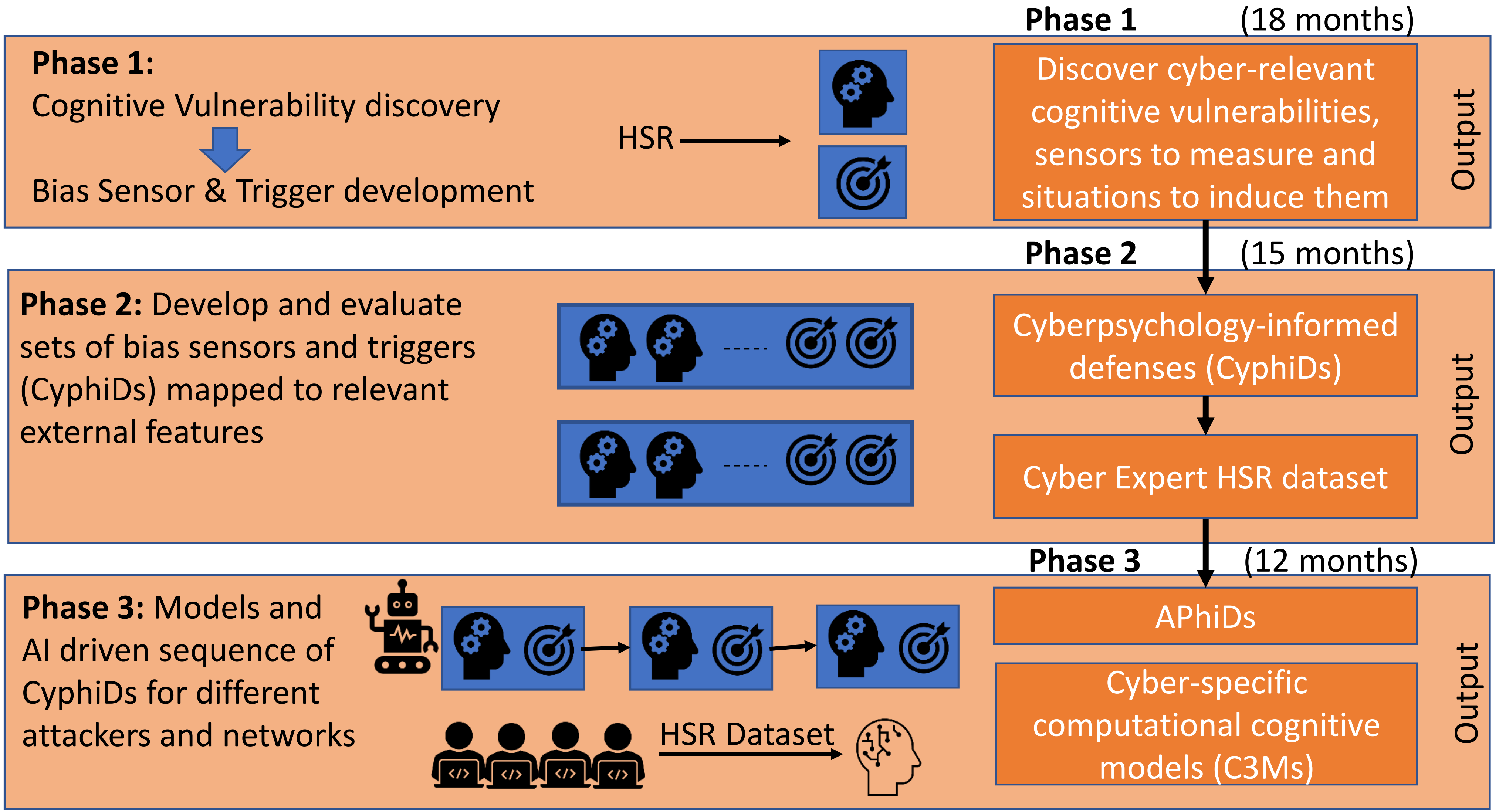
Notional Adaptive Psychology-informed Defenses (APhiDs)



Bias exploited and cyber behavior impacted automatically over a long time period

Phase 3

ReSCIND Program Plan



Phase 1 (18 months)

Phase 1:
Cognitive Vulnerability discovery

Bias Sensor & Trigger development

HSR

Discover cyber-relevant cognitive vulnerabilities, sensors to measure and situations to induce them

Output

Phase 2 (15 months)

Phase 2: Develop and evaluate sets of bias sensors and triggers (CyphiDs) mapped to relevant external features

Cyberpsychology-informed defenses (CyphiDs)

Cyber Expert HSR dataset

Output

Phase 3 (12 months)

Phase 3: Models and AI driven sequence of CyphiDs for different attackers and networks

HSR Dataset

APhiDs

Cyber-specific computational cognitive models (C3Ms)

Output

Out-of-Scope Approaches

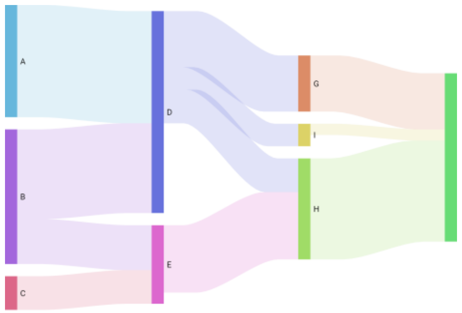
- Research without strong theoretical/experimental foundations
- Research not supporting a CyphiD
- Bias sensors requiring unobtainable cyber data
- Bias sensors/triggers solely targeting non-human attacker
- Bias triggers lacking a cyber behavioral impact
- Technologies focused solely on cyber deception or on traditional cyber defenses
- OSINT research or attacker activity prior to network access
- Reliance of live human actors
- Hardware solutions
- Techniques solely focused on intelligent gathering or attribution
- Anything involving classified data

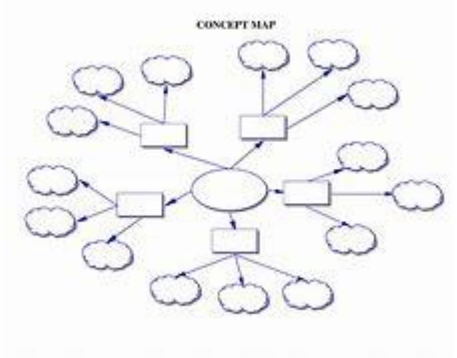
Structured Visual Representation

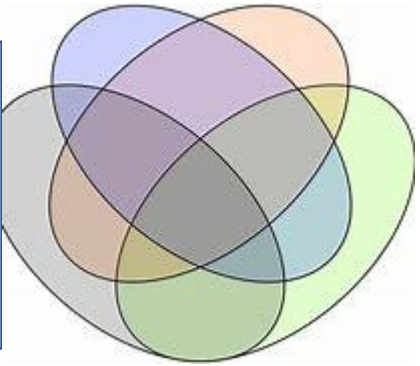
Structured visual representation: displays relationships among relevant variables

- What it looks like?
- What goes in it?
  - **External features**
    - Host & network characteristics
    - Time factors
    - Mission context
    - Situational attributes
  - **Attacker attributes**
    - Attacker behaviors
    - Individual differences
  - **Theoretical foundations**
    - Characteristics of specific cognitive vulnerabilities
  - **Phases of Cyber Kill Chain**
    - Attacker Tactics, Techniques & Procedures (TTPs)
  - Cyber behavioral impacts/defender goals
  - **Cognitive vulnerability-specific factors**
    - i.e., ambiguity, time constraints
- **How does it relate to bias sensors & triggers?**

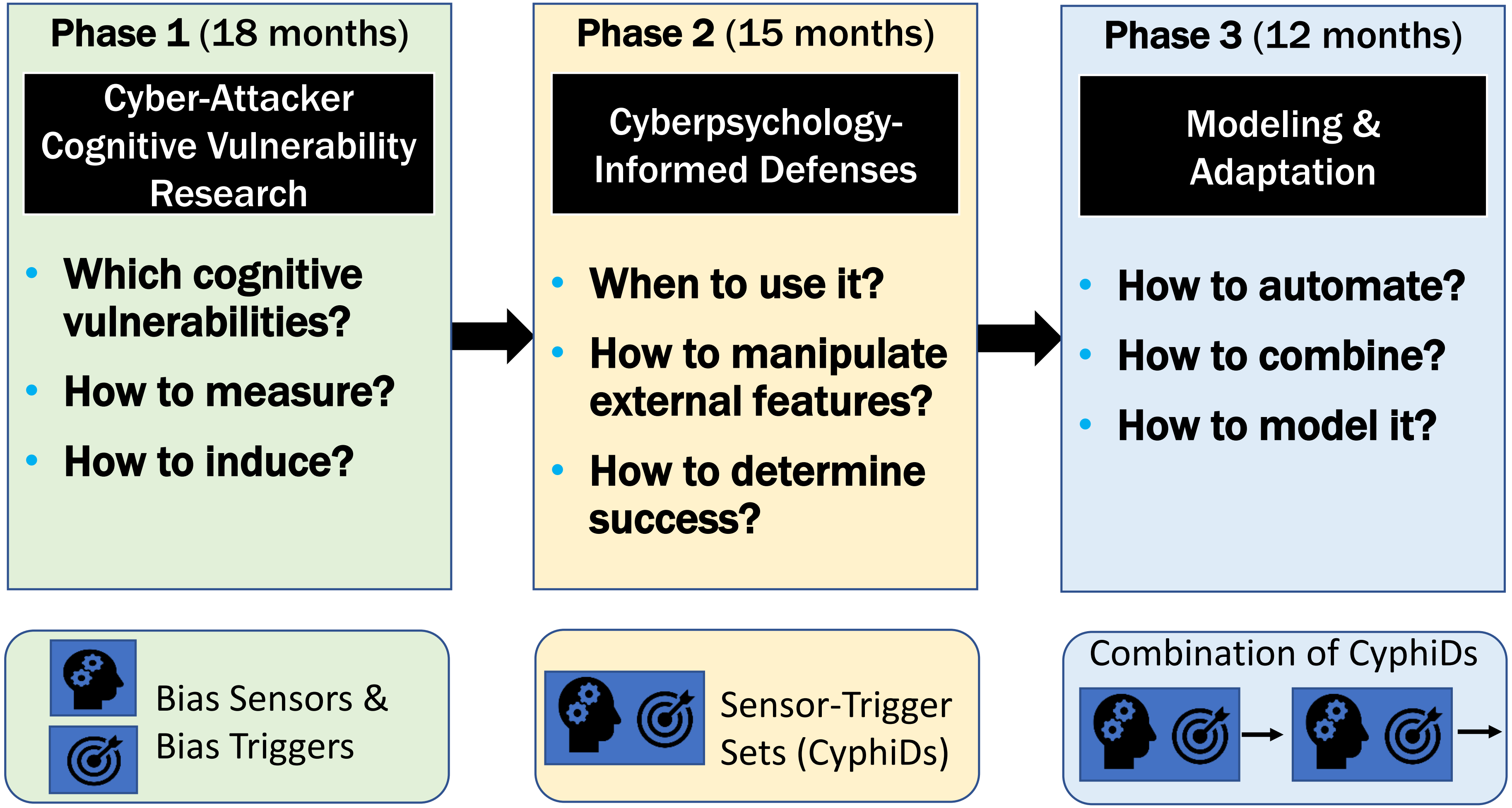
- *Working documents for performer teams*
  - *Fostering CyphiD development*
- *Integrated into a master representation by T&E*
  - *Fostering APhiD development*







ReSCIND Program Phases



Phase 1 (18 months)

Cyber-Attacker Cognitive Vulnerability Research

- **Which cognitive vulnerabilities?**
- **How to measure?**
- **How to induce?**

Bias Sensors & Bias Triggers

Phase 2 (15 months)

Cyberpsychology-Informed Defenses

- **When to use it?**
- **How to manipulate external features?**
- **How to determine success?**

Sensor-Trigger Sets (CyphiDs)

Phase 3 (12 months)

Modeling & Adaptation

- **How to automate?**
- **How to combine?**
- **How to model it?**

Combination of CyphiDs

ReSCIND Program Plan

Phase 1 (18 months)

IRB Submission Required

- Identify at least 3 additional cyber-attack relevant cognitive vulnerabilities
    - 2 mandatory biases assigned by IARPA
        - Loss aversion
        - Representativeness bias
    - Justify with execution of Human Subjects Research (HSR)
- Create bias sensors that measure to what degree each bias is present and bias triggers that induce the bias, in a cyber situation
    - Performers to provide established methodologies for bias sensor validation
- Evaluation
    - Performer experimental designs and results evaluated with a SME rubric (months 5 & 16)
    - Sensor and trigger software test for functionality (months 12 & 16)
    - Bias sensor validation to be performed by T&E (months 12 & 16)
    - Trigger effect size to be calculated as part of performer HSR (months 10 & 14)

Expectations for Responsible & Efficient Research

- Scientifically sound methods & measures are expected
- Empirically grounded theory is required
- Empirically & statistically efficient designs are encouraged
- Cyber-attack scenarios with skilled human participants
  - Performers must obtain ethics review board approval or an IRB waiver
  - Performers must ensure removal of PII
  - Datasets will be made publicly available and must be appropriately labeled and documented
- T&E will provide a subset of standardized IRB language as GFI at Phase 1 Kickoff

ReSCIND Phase 1 Program Metrics

Statistical Metrics	Phase 1 Target
External validity check	Bias sensor: within 1.5 SD of baseline
Higher effect size	Bias trigger: $d \geq 0.3$

Within 1.5 SD of baseline: Each bias sensor corresponds with the established methodology by approximately 90% (Phase 1)

Cohen's *d*: Measures how well performer solutions trigger each cognitive vulnerability; Cohen's d analog for non-parametric (Phases 1, 2, 3)

d=(M_1-M_2)/SD

Cohen's $d \geq 0.30$ = medium effect

Cohen's $d \geq 0.70$ = large effect

Qualitative Metric	Phase 1 Evaluation
Manipulation and validity check	Experimental design & findings: SME Rubric

ReSCIND Program Plan

ReSCIND

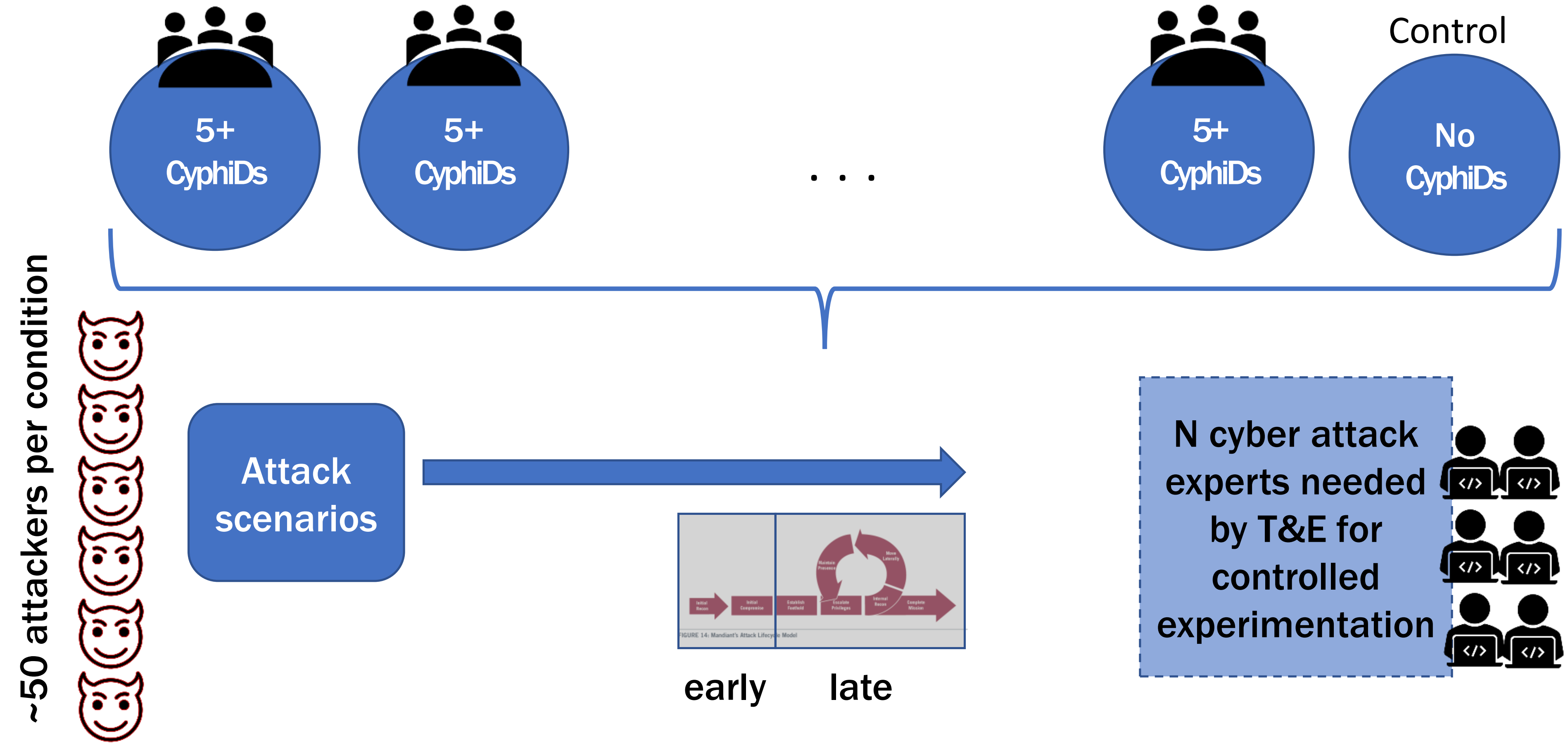
Phase 2 (15 months)

IRB Submission/Mods May Be Required

- Develop software for sensor-trigger sets (CyphiDs)
  - Interact with attacker via triggers based on observables collected by sensors
  - Create logic to link sensors to triggers
  - Both early and late phases of a cyber attack
  - Validate with self-testing (month 24 & 29)
  - Additional/improved sensors and triggers will be developed based on Phase 1 or new HSR
  - Multiple CyphiDs per CogVul are expected
    - At least 5 CyphiDs for early kill chain and 5 CyphiDs for late kill chain
- Evaluation
  - Performers will be compared across common metrics for cyber behavioral impact
  - Validation will be done by T&E with controlled HSR using expert participants (months 26 & 31)
  - Performers may request additional bias-specific metrics, data collection, etc.

Phase 2 Controlled Experimentation

Notional ReSCIND T&E: HSR Testing Plan for CyphiDs: at least 5 CyphiDs for <u>early</u> kill chain and 5 CyphiDs for <u>late</u> kill chain per Performer Team



5+ CyphiDs

5+ CyphiDs

. . .

5+ CyphiDs

Control

No CyphiDs

~50 attackers per condition

Attack scenarios

early

late

N cyber attack experts needed by T&E for controlled experimentation

ReSCIND Phase 2 Program Metrics

Cyber Behavioral Impacts	Behavioral Metrics	Phase 2 Target
Decrease Rate of Attack Success	Attack success vs. HSR control	50% ≤ baseline
Decrease Progress Towards Goal	Progress to goal vs. HSR control	50% ≤ baseline
Decrease in Time Until Detection	Time to detection vs. HSR control	50% ≤ baseline
Decrease Defender Effort Spent	Decreased defender effort vs. HSR control	50% ≤ baseline
Increase Attacker Cognitive Effort Spent	Attacker effort vs. HSR control	50% ≥ baseline
Increase Attack Resources Wasted	Attack resources wasted vs. HSR control	50% ≥ baseline
Increase Time to Task Completion	Time to task completion vs. HSR control	50% ≥ baseline
Cyber Behavioral Impacts	**Statistical Metrics**	**Phase 2 Target**
All Seven Cyber Behavioral Impacts	Higher effect size	CyphiD: $d \geq 0.5$
	Predictive power	N/A

Each CyphiD focuses on at least one cyber behavioral impact. The collection of a performer's CyphiDs should meet all targets.

Examples of Experimental Data Types

- Environmental data
- Scenario data
- Forward progress
- Alert data
- Attack data
- Host data
- User data
- Network data
- Individual measures
- Self-report data
- CyphiD data
- APhiD data

Performer teams will propose and justify any data requested in addition to what will be provided by T&E.

Examples of Experimental Data Types

Data Type	Data Example
Scenario Data	subject ID, date, day, condition, environment, daily start/end time, breaks/lunch, subject time on task, screen capture
Environment Data	subject IP, target IPs, target host configuration (e.g., OS, ports), host name, vulnerabilities
Host Data	Process logs, file touches, services, process history, file data, system & application host logs
Network Data	packet ID, pcap timestamp, destination IP, pcap size, source IP, destination IP, port, timestamp
User Data	User accounts, access logs, privilege, user files, login attempts
Attack Data	exploit timestamp, exploit name, exploit CVE, success/failure
Alert Data	signature ID, IDS alert description, CVE, severity, target IP, timestamp
Forward Progress	flags captured, data exfiltrated, lateral movement, privilege escalation
Self-Report Data	timestamp, self-reported vulnerabilities identified, self-reported exploit attempts, self-reported success/failure, Red Team Briefing
Individual Measures (HSR Data)	Bias-specific questions, Reported Cognitive State, Experience, Demographics, interviews, General Decision-Making Style Inventory (GDMSI), Indecisiveness Scale (IS), Big Five Inventory (BFI-44)
CyphiD/APhiD Data	To be included in proposal by Offerors

ReSCIND Program Plan

ReSCIND

Phase 3 (12 months)

IRB Submission Required

- Improve solutions with AI-guided adaptation (APhiDs)
  - Develop algorithms to select sequences of CyphiD defenses
  - All CyphiDs to be shared among Performer teams
  - Validate with self- testing (month 40)
- Create cyber-specific computational cognitive models (C3Ms)
  - Reflect & predict variability of cyber behavior tied to presence of each CogVul
  - Validate with self- testing using previous phase datasets

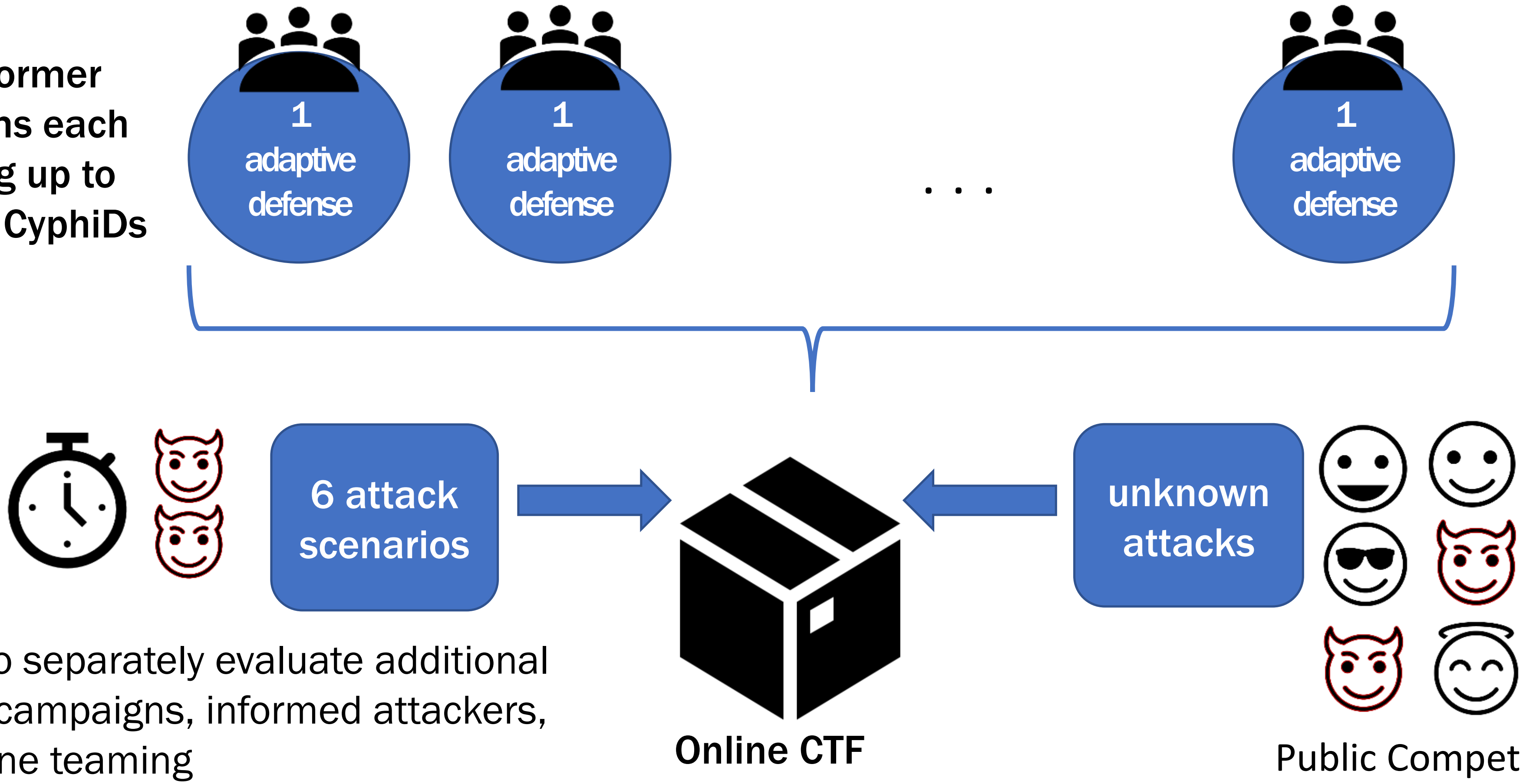
Evaluation

- Additional scenarios and use cases will be tested by T&E (month 39)
- C3Ms to be tested against existing/collected HSR data (month 44)
- APhiD validation will be done via open Capture-the-flag (CTF) prize competition (month 43)

Phase 3 Quantitative Experimentation

Notional ReSCIND T&E: HSR Testing Plan for APhiDs

Performer teams each using up to 50+ CyphiDs



T&E red teaming to separately evaluate additional use cases: longer campaigns, informed attackers, and human-machine teaming

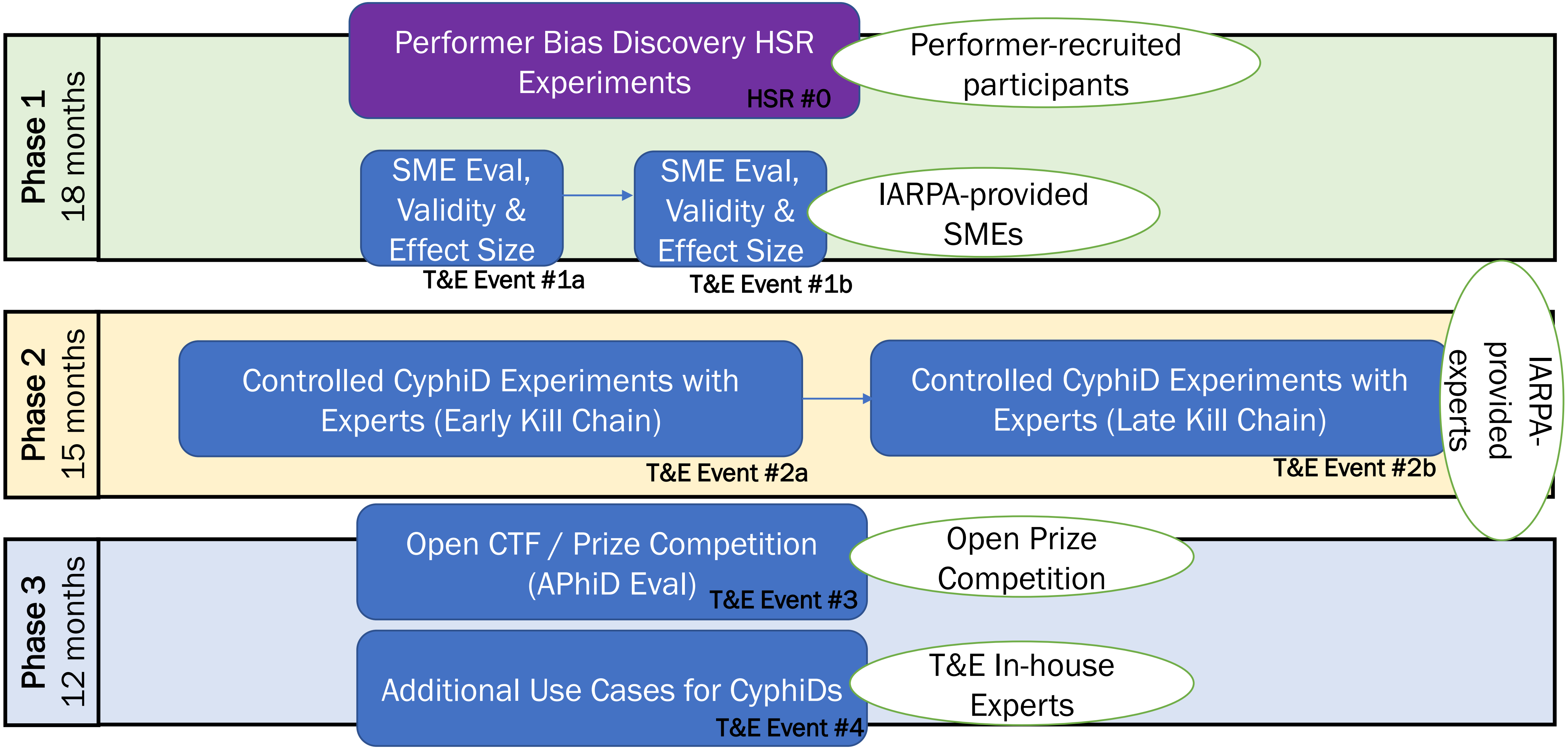
ReSCIND Phase 3 Program Metrics

Cyber Behavioral Impacts	Behavioral Metrics	Phase 2 Target	Phase 3 Target
Decrease Rate of Attack Success	Attack success vs. HSR control	50% ≤ baseline	APhiD: 10% improvement on best team's Phase 2 results for each cyber behavioral impact
Decrease Progress Towards Goal	Progress to goal vs. HSR control	50% ≤ baseline	
Decrease in Time Until Detection	Time to detection vs. HSR control	50% ≤ baseline	
Decrease Defender Effort Spent	Decreased defender effort vs. HSR control	50% ≤ baseline	
Increase Attacker Cognitive Effort Spent	Attacker effort vs. HSR control	50% ≥ baseline	
Increase Attack Resources Wasted	Attack resources wasted vs. HSR control	50% ≥ baseline	
Increase Time to Task Completion	Time to task completion vs. HSR control	50% ≥ baseline	
Cyber Behavioral Impacts	**Statistical Metrics**	**Phase 2 Target**	**Phase 3 Target**
All Seven Cyber Behavioral Impacts	Higher effect size	CyphiD: $d \geq 0.5$	APhiD: $d \geq 0.7$
	Predictive power	N/A	C3M: RMSE ≤ 0.2

The root-mean-square error (RMSE): Measures how well the model predicts real data

Testing and Evaluation (T&E)

Program Test and Evaluation Plan



Phase 1
18 months

Performer Bias Discovery HSR Experiments HSR #0

Performer-recruited participants

SME Eval, Validity & Effect Size
T&E Event #1a

SME Eval, Validity & Effect Size
T&E Event #1b

IARPA-provided SMEs

Phase 2
15 months

Controlled CyphiD Experiments with Experts (Early Kill Chain)
T&E Event #2a

Controlled CyphiD Experiments with Experts (Late Kill Chain)
T&E Event #2b

IARPA-provided experts

Phase 3
12 months

Open CTF / Prize Competition (APhiD Eval) T&E Event #3

Open Prize Competition

Additional Use Cases for CyphiDs
T&E Event #4

T&E In-house Experts

Examples Mapping Metrics to Data

- Delayed or Impeded Attacker Goals
  - Time to stated goal
  - Forward Progress
  - Protection of key terrain
- Increased Attacker Effort
  - Increased scanning behavior
  - Packet or keystroke count
- Increased Detection
  - Time until detection
  - Alerts triggered

- Persistent Effects
  - Deterrence
  - Self-doubt
- Wasted Attack Resources
  - Unsuccessful exploit attempts
  - Increased mistakes
  - Unnecessary change in strategy
- Additional Performer-Specified Metrics
  - Cognitive Vulnerability-specific

Example Defender Goals:
Deny
Delay
Degrade
Detect
Disrupt

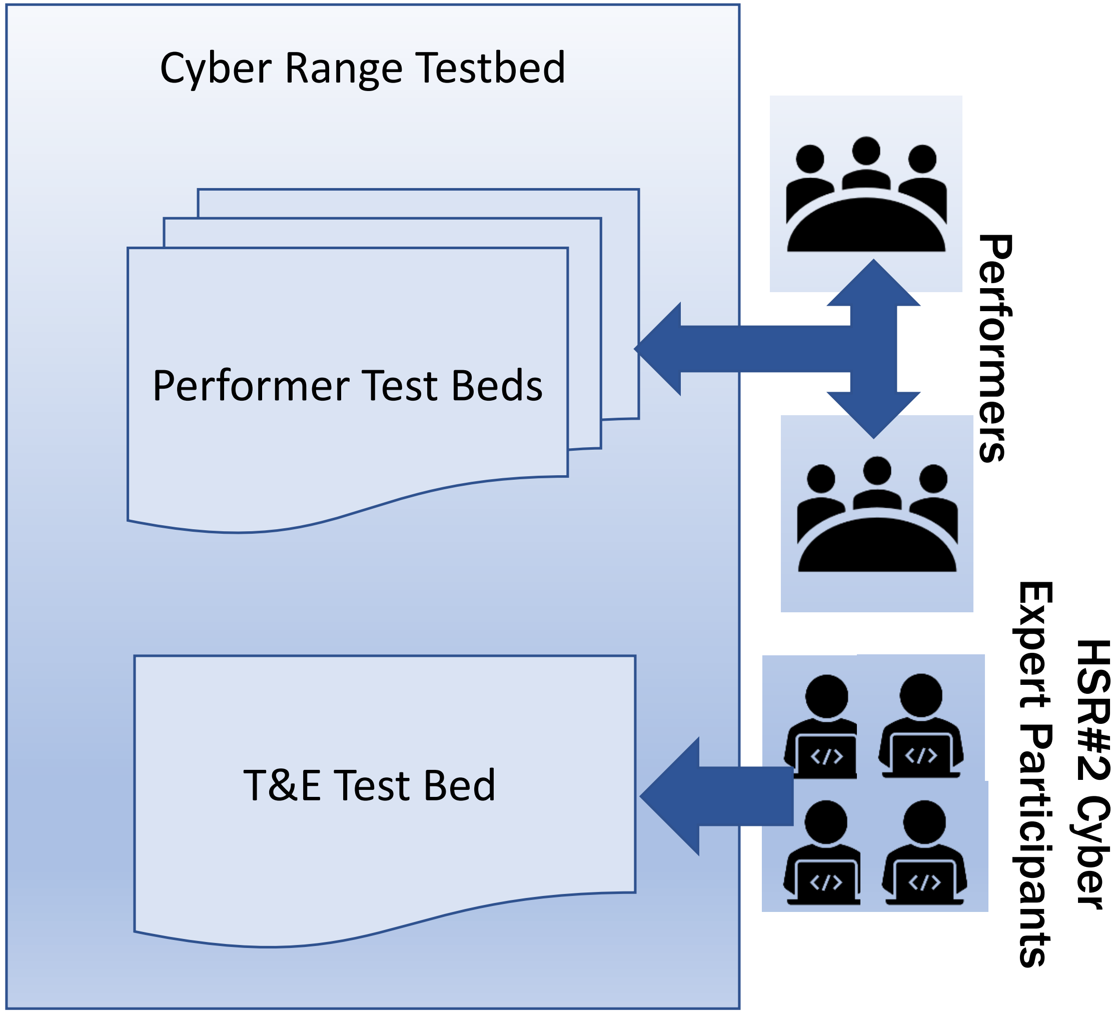
Subjective Measures:
System usability, system adoptability, system security, coverage of attack phases & TTPs

Datasets Created

- Each experiments will create a new cyber research dataset which can jumpstart new human-focused research across the community.
- Program will host all T&E datasets for future research.
- May share HSR #0 dataset independently or have them co-located with the T&E datasets.
- Unrestricted rights or (at least) government purpose rights for all data and software.

CyberVAN Cyber Range Testbed Usage

- DoD-funded T&E testbed hosted/managed by T&E Team
  - Evaluation and experimentation
- Provided independent performer testbed instances for self-testing
- Performers will not be given all details about the configuration
- Will not be supplied for performer experiments (HSR #0)
- Data collected within T&E testbed will be made publicly available
- API provided at Phase 1 kick-off



Program Schedule

	Phase 1																		Phase 2															Phase 3											
	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12	Month 13	Month 14	Month 15	Month 16	Month 17	Month 18	Month 19	Month 20	Month 21	Month 22	Month 23	Month 24	Month 25	Month 26	Month 27	Month 28	Month 29	Month 30	Month 31	Month 32	Month 33	Month 34	Month 35	Month 36	Month 37	Month 38	Month 39	Month 40	Month 41	Month 42	Month 43	Month 44	Month 45
Kickoff Meeting	O																		O															O											
IRB Milestone		Δ						Δ											Δ				Δ					Δ																	
Document Delivery			X	X				X		X	X			X	X	X	X			X				X	X			X		X			X			X		X		X		X			X
Performer Self-testing									X				X	X										X					X								X	X		X					
Software Delivery										X				X	X		X								X					X			X									X		X	
T&E Event					♦											♦										♦	♦				♦								♦	♦			♦	♦	
Site Visits					O						O					O								O						O							O					O			
Demos											Δ							Δ						Δ						Δ												Δ			Δ
PI Meetings									O									O								O							O						O						O
Final Report																		X															X												X
Monthly Status Report	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Year 1												Year 2												Year 3												Year 4								

Meeting: O Deliverable: X Evaluation: ♦ Milestone: Δ

Testing will consist of self-testing and reporting of results by performers, followed by formal testing by T&E. T&E results will be reported back to performers for iterative improvements. T&E includes both open CTF events and controlled experimentation in the cyber range testbed with skilled expert participants.

In Summary

*Utilize cyberpsychology to create novel defenses that **rescind attacker advantage** and **impose a cyber penalty***

ReSCIND

We look forward to your innovative ideas to make this happen!

Contracting Overview

Stephen Enokida | Contracting Officer | Feb 28, 2023

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Break – Last chance to submit questions is at 9:40 AM PT/ 12:40PM ET

We will start again at 11:00 AM PT/ 2:00 ET

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Addressing Submitted Questions

Dr. Kimberly Ferguson-Walter| Program Manager | Feb 28, 2023

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Lightning Talks

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Lightning Talk Overview

- Teams have 5 minutes to highlight capabilities aligning with ReSCIND interests
- Use this opportunity to fill gaps in your team
- Slides and documents will be made available on the ReSCIND website

Closeout

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Reminder on Teaming

- Participants are encouraged to find partners and collaborators; Someone might have a missing piece of your puzzle!
- Teaming and capability summaries will be accepted, with minimal review for appropriateness, and made available to the public.
  - Teaming documents and summaries can still be submitted until the BAA closes, submit to dni-iarpa-ReSCIND-proposers-day@iarpa.gov.