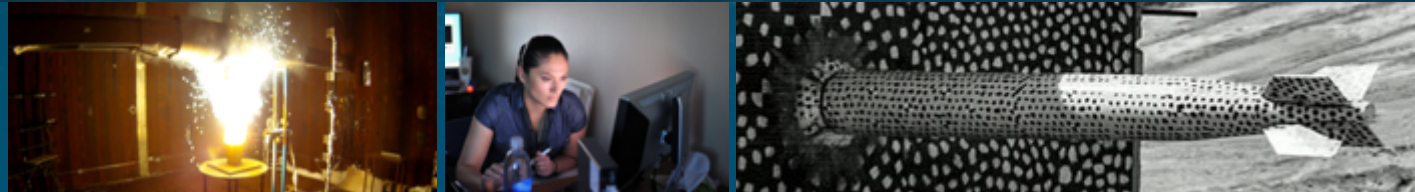




CHARACTERIZING HUMAN PERFORMANCE: DETECTING TARGETS AT HIGH FALSE ALARM RATES



Ann Speed, Jason Wheeler, John Russell, Fred Oppel, Danielle N. Sanchez, Austin Silva, and Anna Chavez



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Why worry about measuring human performance?

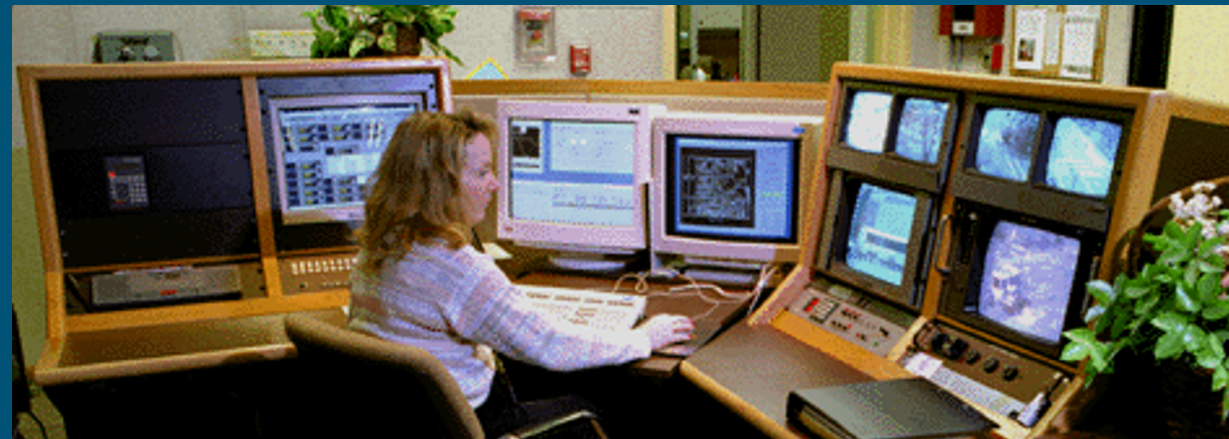
Our insight into drivers of our behavior
and cognition is limited



Human performance in low-probability, high-consequence threat environments – The Prevalence Effect



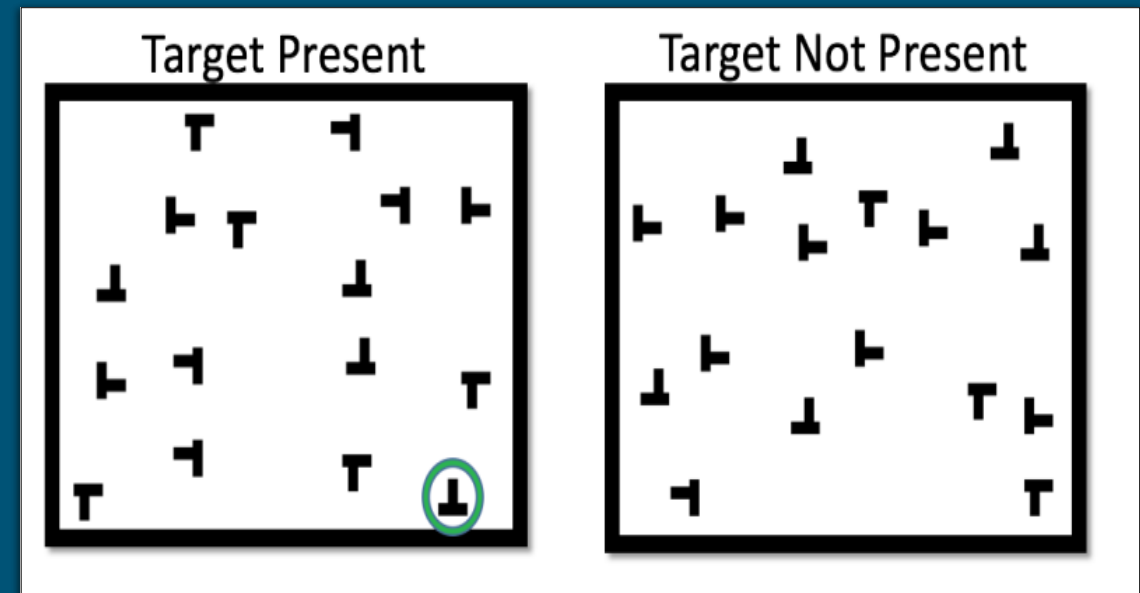
- Characteristic of many national security problems
 - Physical security CAS operators
 - Border/airport screening (*only 0.0005% of bags screened by TSA contained guns in 2018*)
 - Cybersecurity analysts
 - Treaty verification inspectors
- Analysts develop a “no threat” bias with high false alarms
- If only shown alarms for actual attacks, may never actually see an alarm
- We see this in the laboratory, but not often studied in applied environments (TSA is an exception)



Experiment 1: Domain-general study



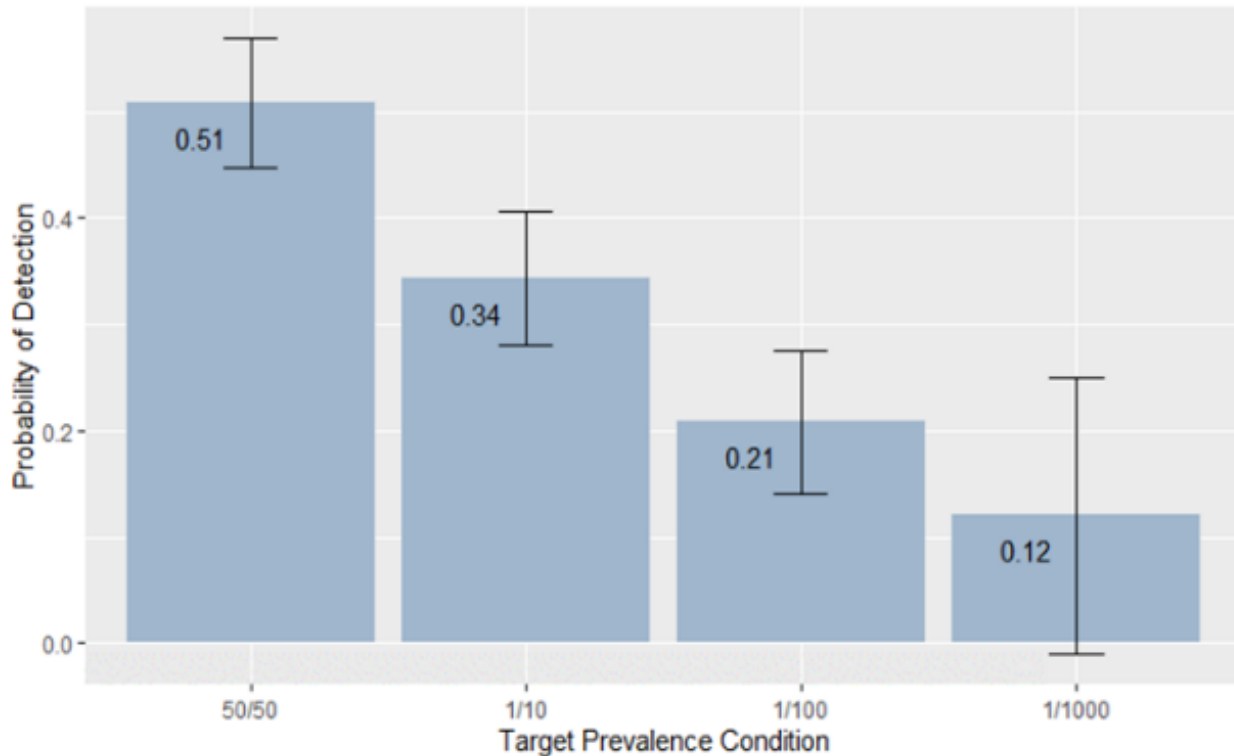
- Academic researchers studied visual search accuracy with target prevalence rates as low as 0.078%, observing detection rates of <27%
- Generalizability to national security missions challenging, environmental and operational details matter a lot
- Four conditions:
 - 500/1000 are targets
 - 100 / 1000 are targets
 - 10/1000 are targets
 - Only the last trial is a target
- Subjects search for perfect Ts amidst Ls
- Between-subjects design
- 1000 trials (approximately 2 hours) per subject
- Final trial always contains a target



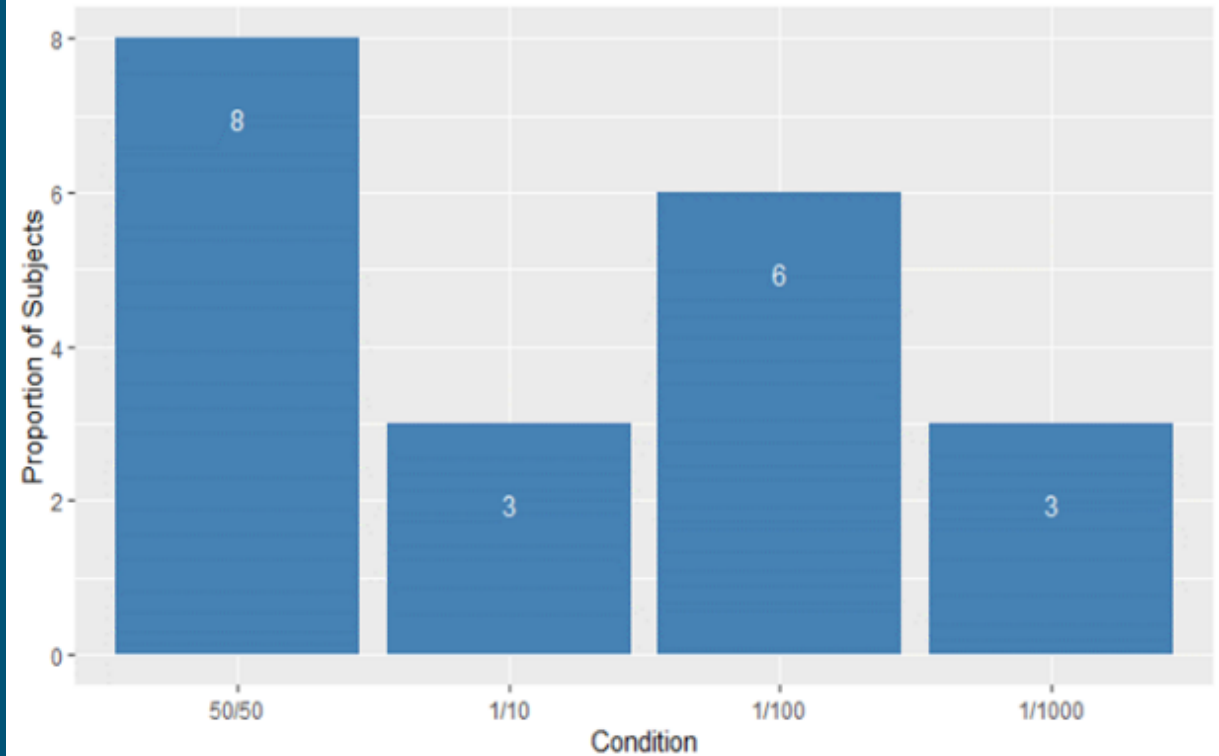
Results: Domain-general study



Average Probability of Detection by Condition

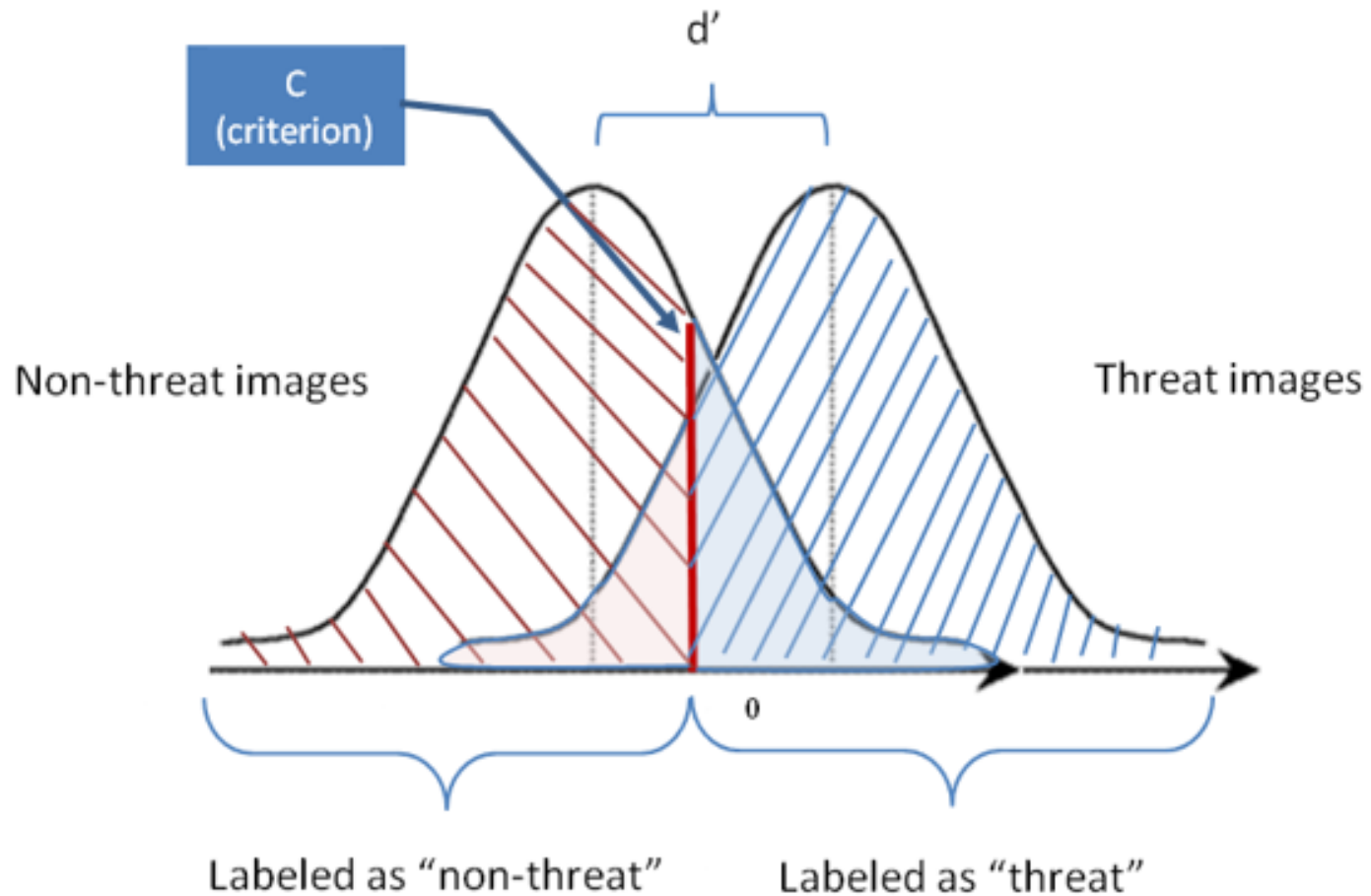


Number of Subjects Detecting the Target on the Last Trial by Condition

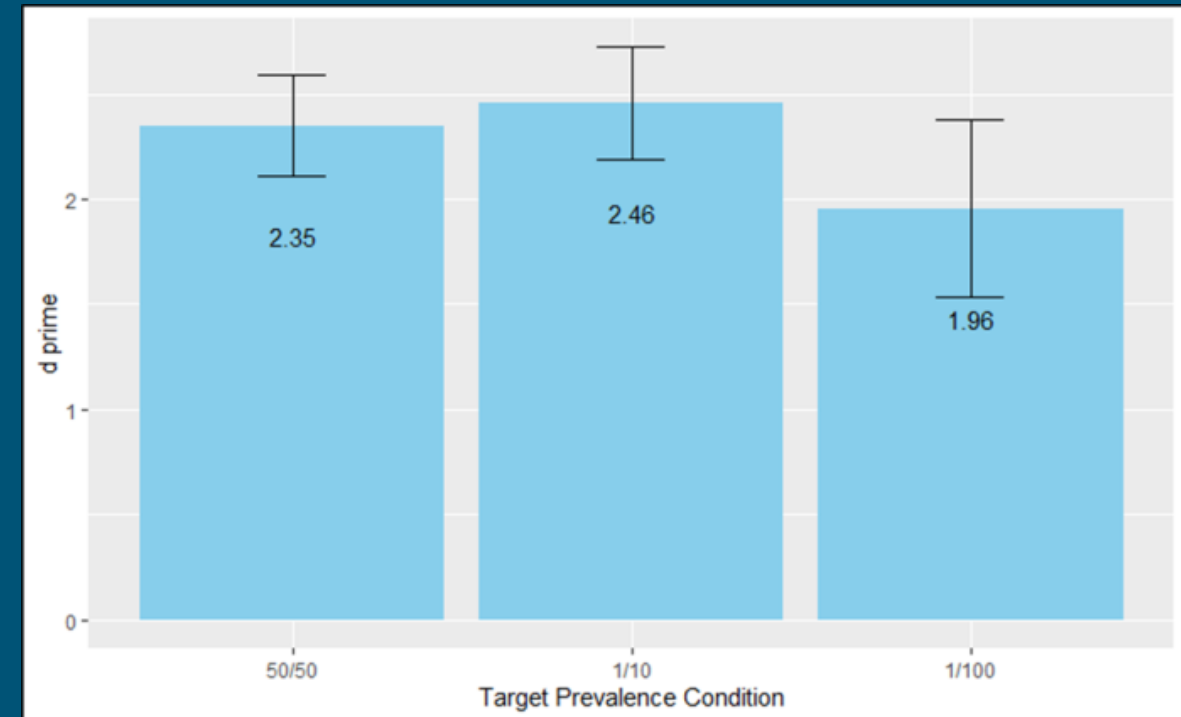
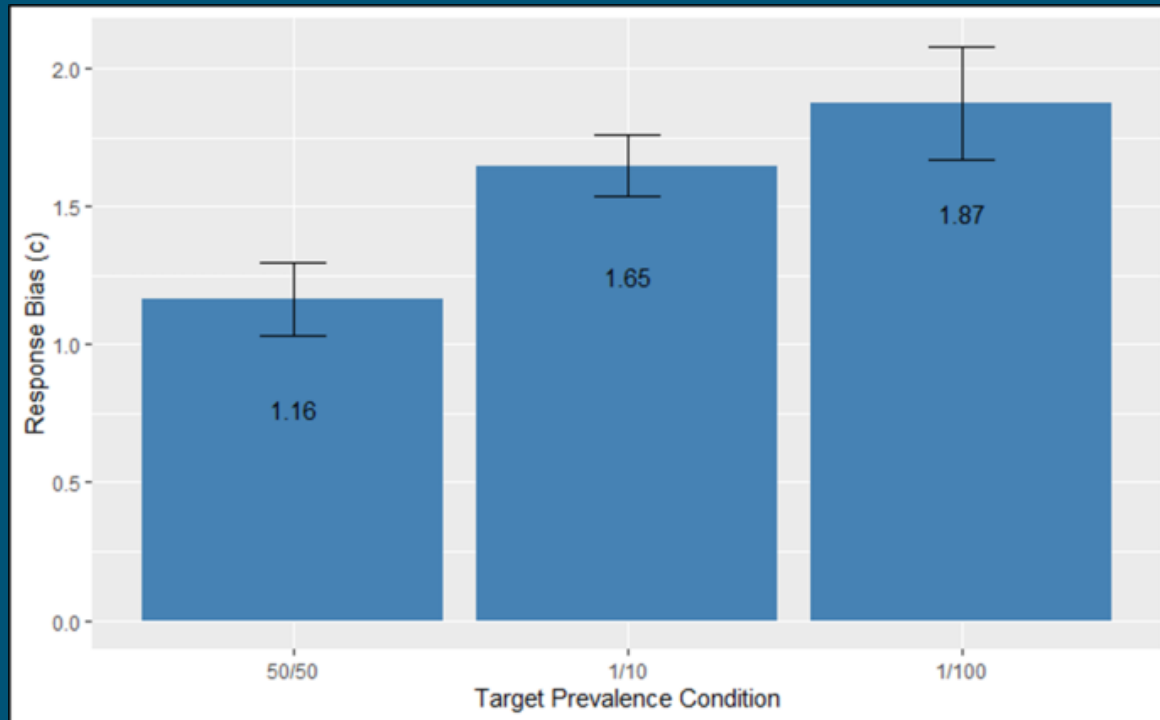


Probability of detection is as expected, but we did not expect the pattern of detection for the final trial.

Critical metrics: d' and response bias



Domain General Study: Response bias and d' as a function of target prevalence condition



Response bias. Larger positive values indicate a bias towards saying “no threat.”

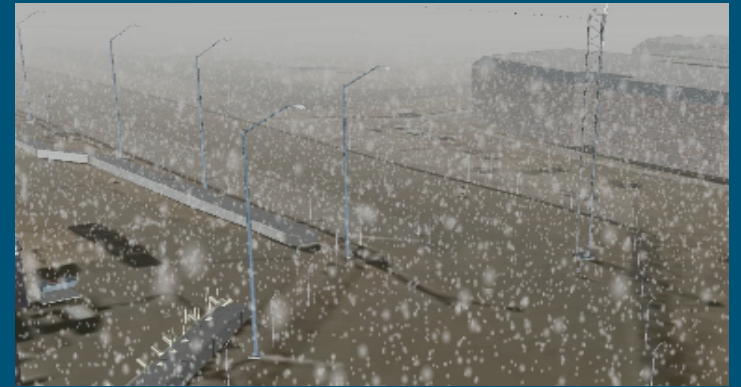
D' as a function of target prevalence condition.

Bias and d' for the 1/1000 condition was not calculated because this metric requires there be more than one target.

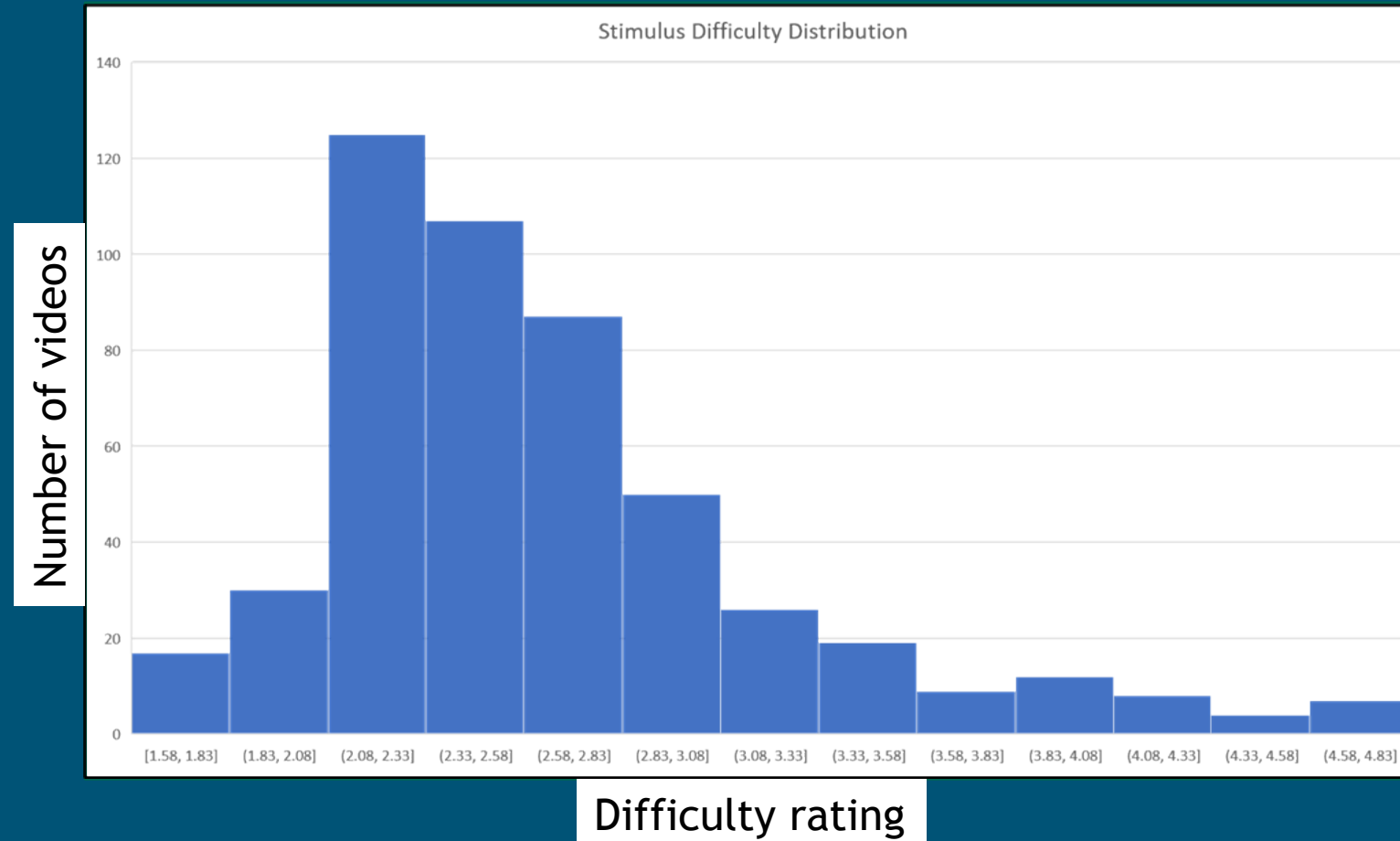
Experiment 2: Domain Specific Study of notional physical facility



- Same conditions applied to attackers on a notional alarmed perimeter, simulated in a Sandia-developed mod/sim environment called Dante
 - 500/1000 videos contain threats
 - 100 / 1000 videos contain threats
 - 10/1000 videos contain threats
 - Only the last video contains a threat
- Only nuisance alarms - proportion of causes (birds, weather, animals)
- Attack scenarios rated independently for difficulty
- Both 5sec pre/post video of alarm and a “live” video of facility available
- Overhead view of facility with alarming sector and direction of camera
- Goal was to use physical security professionals - only could recruit from Sandia general population



Distribution of independent difficulty ratings for Experiment 2 stimuli



Distribution of difficulty ratings for the attack videos used in the physical facility study. Lower scores indicate the attacker is easier to identify

Experiment 2: Domain Specific Study of notional physical facility



MRCOG NM, Atlantic Group, MRCOG NW, Savannah Huston, Maxap Powered by Esri

Start 

Ready...

14:31:03	Begin practice scenarios...
14:31:37	No threat assessed.
14:31:50	Threat assessed.
14:32:05	No threat assessed.

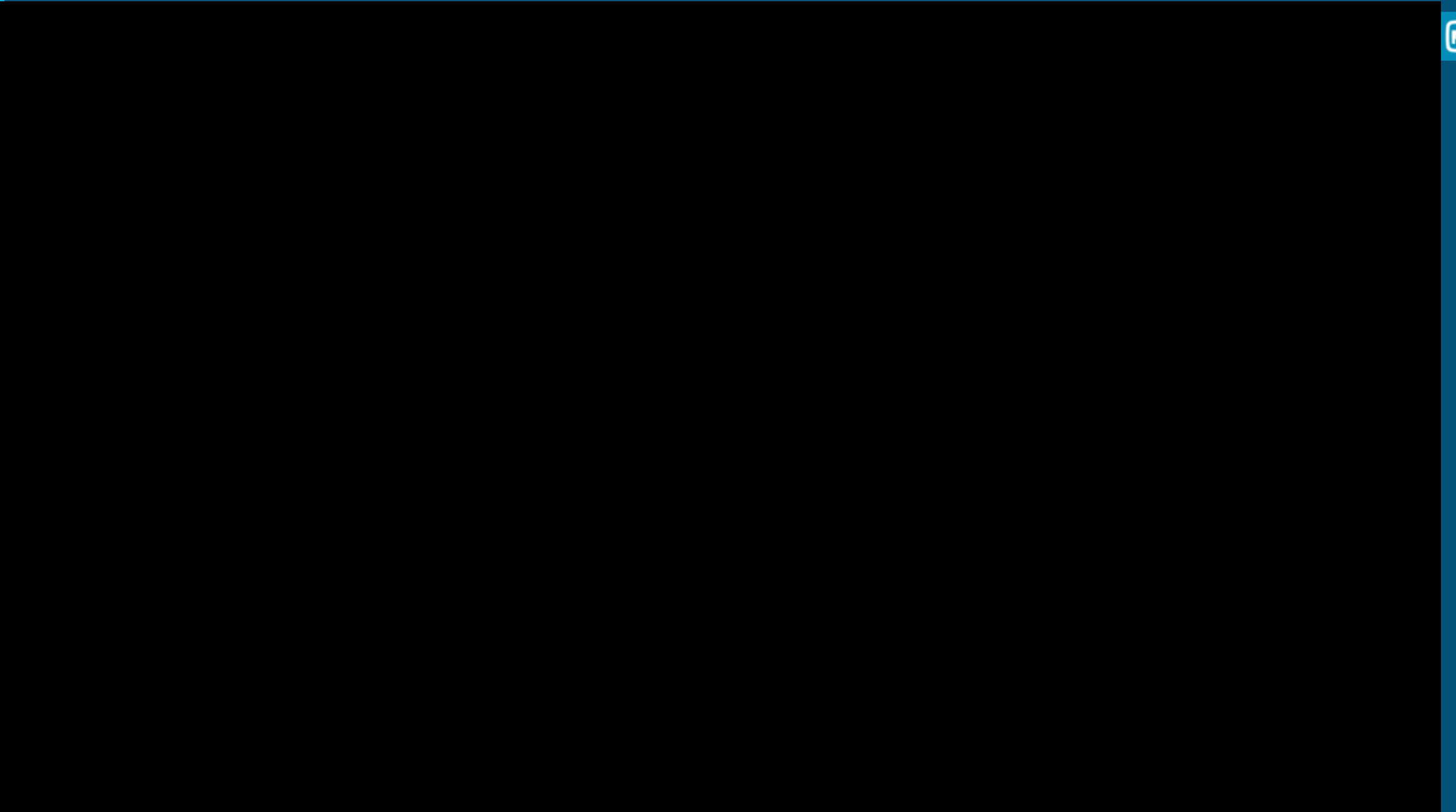
Live



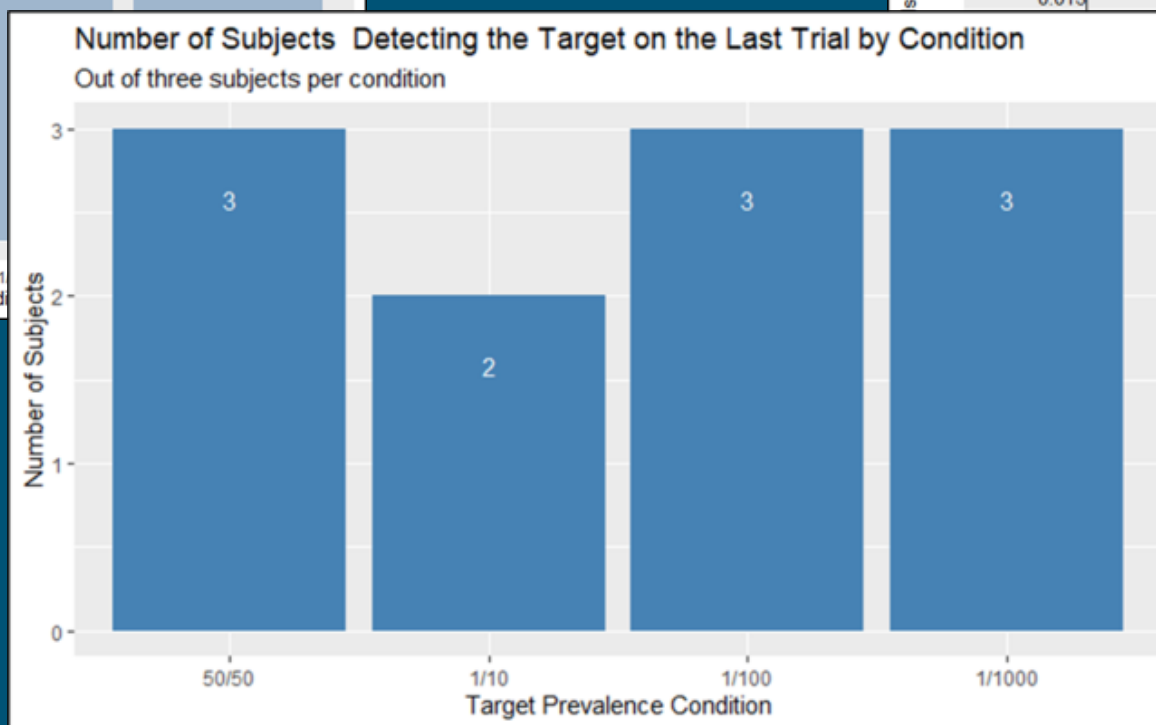
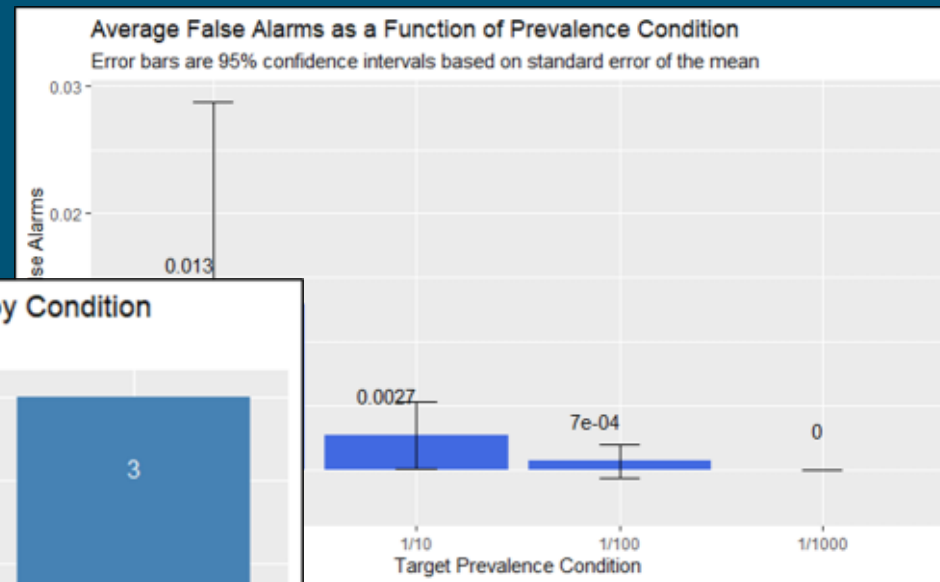
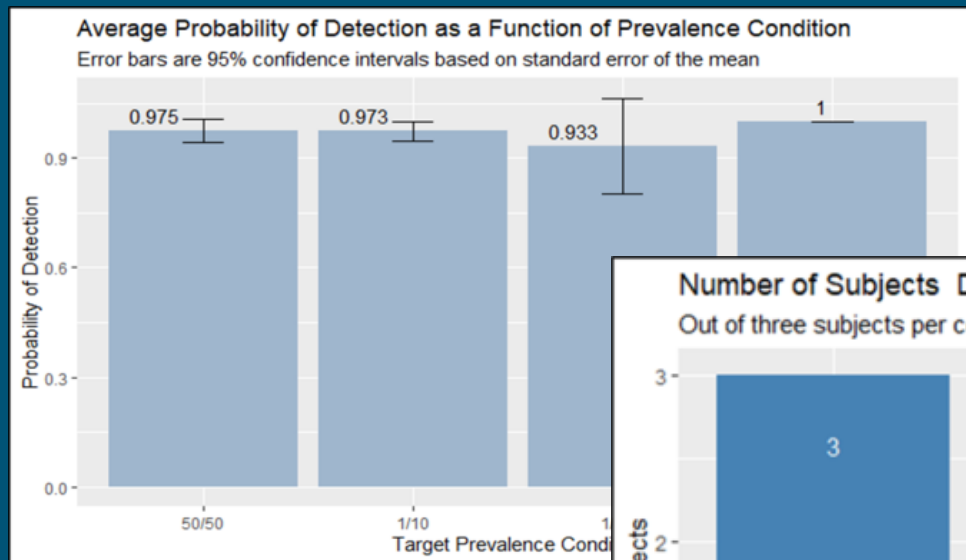
Replay

Alarm Time
-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5



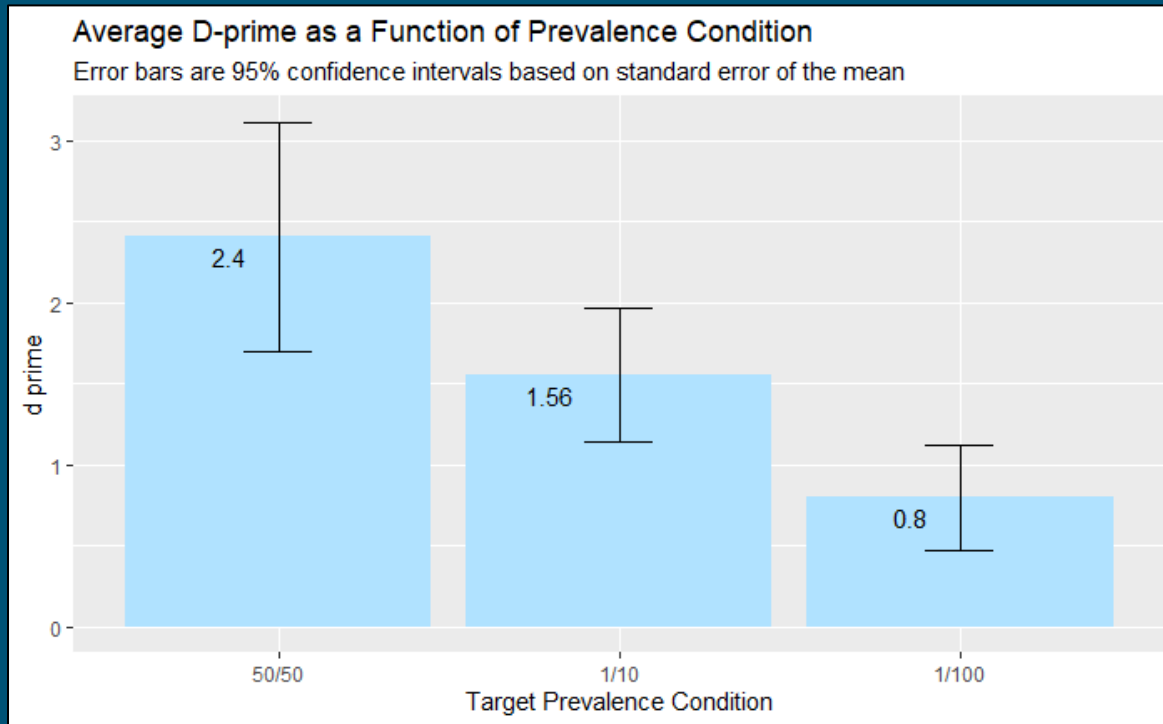


Results: Notional physical facility – 12 total subjects

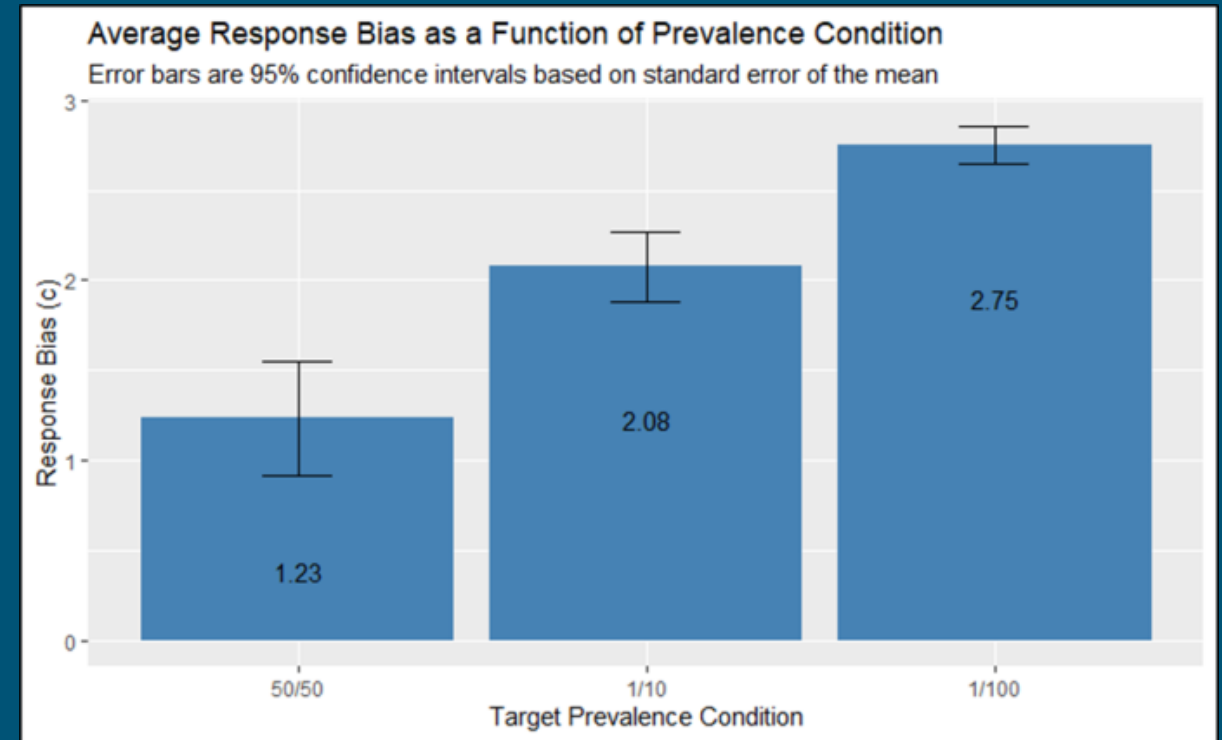


These results indicate our stimuli were too easy

Physical Facility Study: Response bias and d' as a function of target prevalence condition



Average d' as a function of threat prevalence condition.



Average response bias as a function of threat prevalence condition.

Despite stimulus problems, we are still seeing patterns we expect to see, especially in response bias

Conclusions

- Near-operational paradigms are useful, but difficult to construct well
 - Pilot testing is critical before engaging time-limited professionals
 - Experimental control is difficult to balance with operational realism
- Grounding near-operational experiments in basic research paradigms has both advantages and disadvantages.
- Despite shortcomings in our second experiment, we now have a platform for experimental investigations into the human element of physical security systems.



Thanks for your attention?
Questions?

Contact info:

Ann Speed – aespeed@sandia.gov

Jason Wheeler - jwwheel@sandia.gov