How We Make Judgments about Human and Automated Teammates

Lisa Vangsness, PhD
The University of Alabama in Huntsville

Monitoring During Human-Automation Teaming

- Automated parking assist requires drivers to monitor an automated system as it performs a task.

Monitoring During Human-Automation Teaming

- "… technology has now come to the rescue. Automatic self-parking systems are getting ever-more-prevalent."
(J.D. Power; Nerad, 2021)

Monitoring During Human-Automation Teaming

- "[U]nforseen circumstances can impair Autopark's ability to park Model 3… stay prepared to immediately take control."
(Tesla Model 3 Manual, 2022)

Monitoring During Human-Automation Teaming

- Human-Automation Teams perform high-stakes tasks…

  - reconnaissance

  - explosive ordinance disposal

  - threat identification

Theories of Metacognition

- Mental models in HAT are debated

  - **<u>Media Equation Hypothesis</u>** – mental models are identical for humans and automation
    (Nass & Lee, 2001; Reeves & Nass, 1996; Nass & Moon, 2000)

  - **<u>Unique Agent Hypothesis</u>** – mental models are different for humans and automation
    (Hoff & Bashir, 2015; Lee & See, 2004; Cohen et al., 1998)
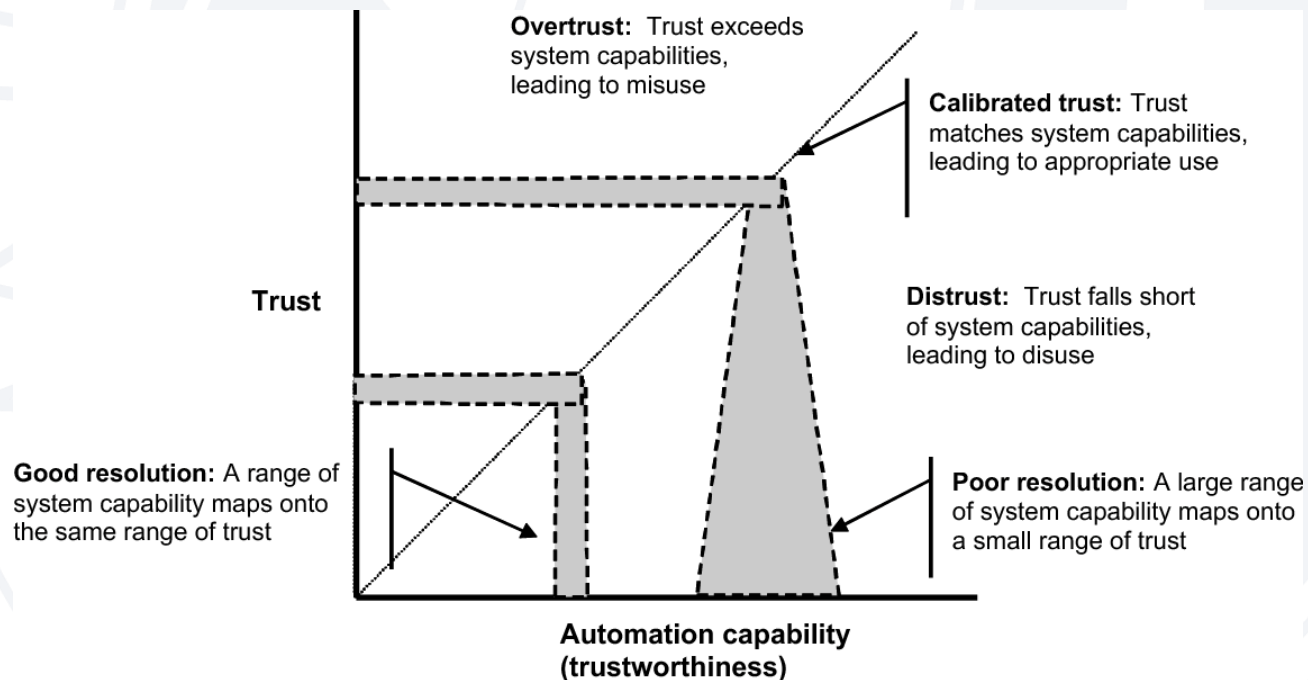
Monitoring During Human-Automation Teaming

How do people make judgments about human and automated teammates?

Are these judgments the same or different?

Theories of Metacognition

- Cue use provides a framework by which to understand calibration (Lee & See, 2004)

Overtrust: Trust exceeds system capabilities, leading to misuse

Calibrated trust: Trust matches system capabilities, leading to appropriate use

Distrust: Trust falls short of system capabilities, leading to disuse

Good resolution: A range of system capability maps onto the same range of trust

Poor resolution: A large range of system capability maps onto a small range of trust

Trust

Automation capability (trustworthiness)

Theories of Metacognition

• Studying metacognition allows us to ask deeper, more nuanced questions…
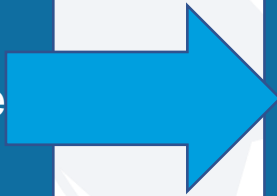
BEFORE

NOW

The mental models may be different…
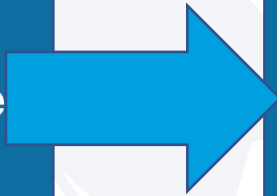
If the mental models are different, then why?

Theories of Metacognition

- and to identify the mechanisms that underlie poor calibration…

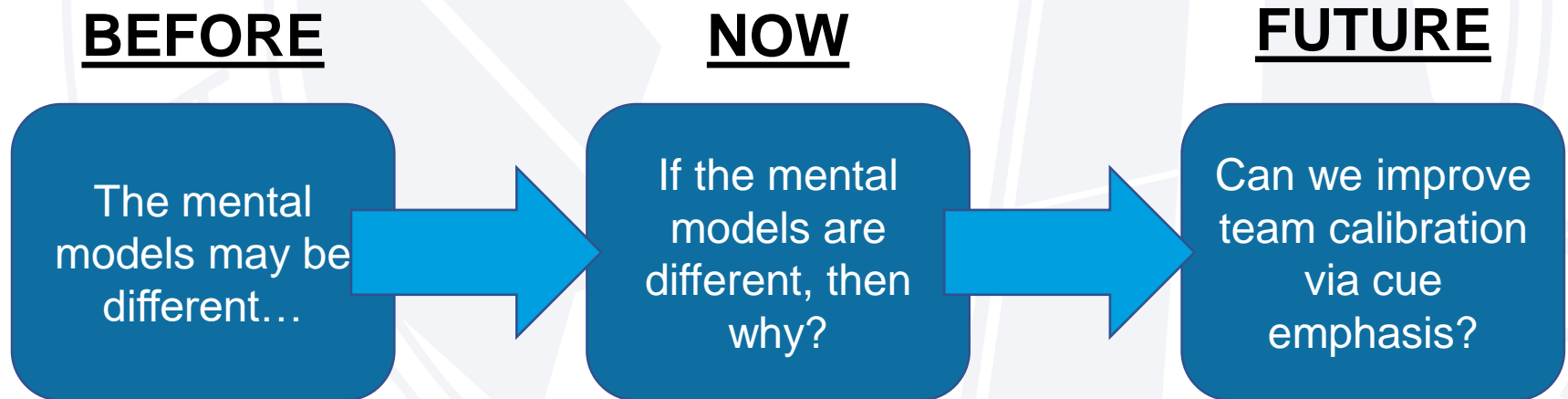**BEFORE**

NOW

The mental models may be different…

→

If the mental models are different, then why?

Theories of Metacognition

- So that we can ultimately improve the performance of agile HAT teams.

BEFORE

NOW

FUTURE

The mental models may be different…

If the mental models are different, then why?

Can we improve team calibration via cue emphasis?

Theories of Metacognition

- **Metacognition** – thinking about thinking (Flavel, 1979)

- **Theory of mind** – how we represent and making judgments about others who may be very different from ourselves in their knowledge and skill (Röska-Hardy, 2009)

Theories of Metacognition

- The judgments we make while watching others are derived from…

    - **<u>Simulation Theory</u>** – our imagined performance
      (Gordon, 1986; Heal, 1996; Gallese & Goldman, 1998)

    - **<u>Theory Theory</u>** – the cues that arise as we observe
      (Gopnik & Wellman, 1992; Gopnik & Meltzoff, 1997; Carruthers, 2009)

Monitoring During Human-Automation Teaming

- Lens model framework maps cue use (Brunswik, 1943)
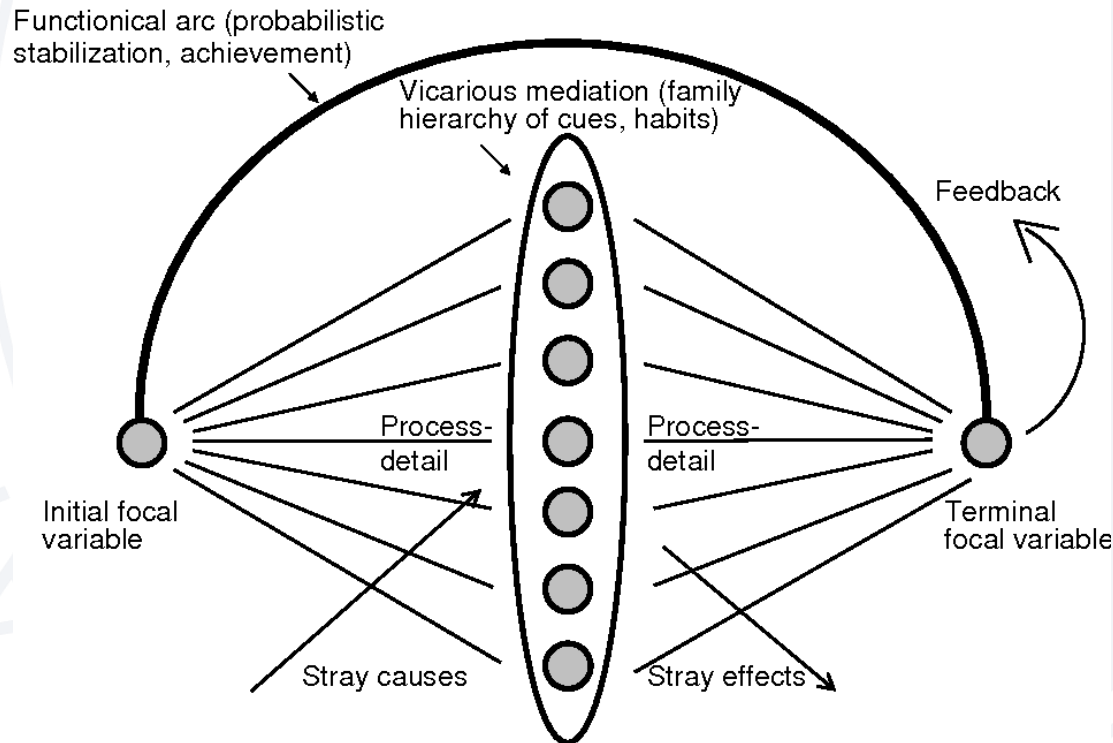
Overlap in Theories

- These theories can be disambiguated by studying peoples' cue use during monitoring

(Vangsness & Young, 2019; Vangsness & Young, 2021)

Overlap in Theories

- These theories can be disambiguated by studying peoples' cue use during monitoring

Simulation Theory
our imagined performance

Media Equation
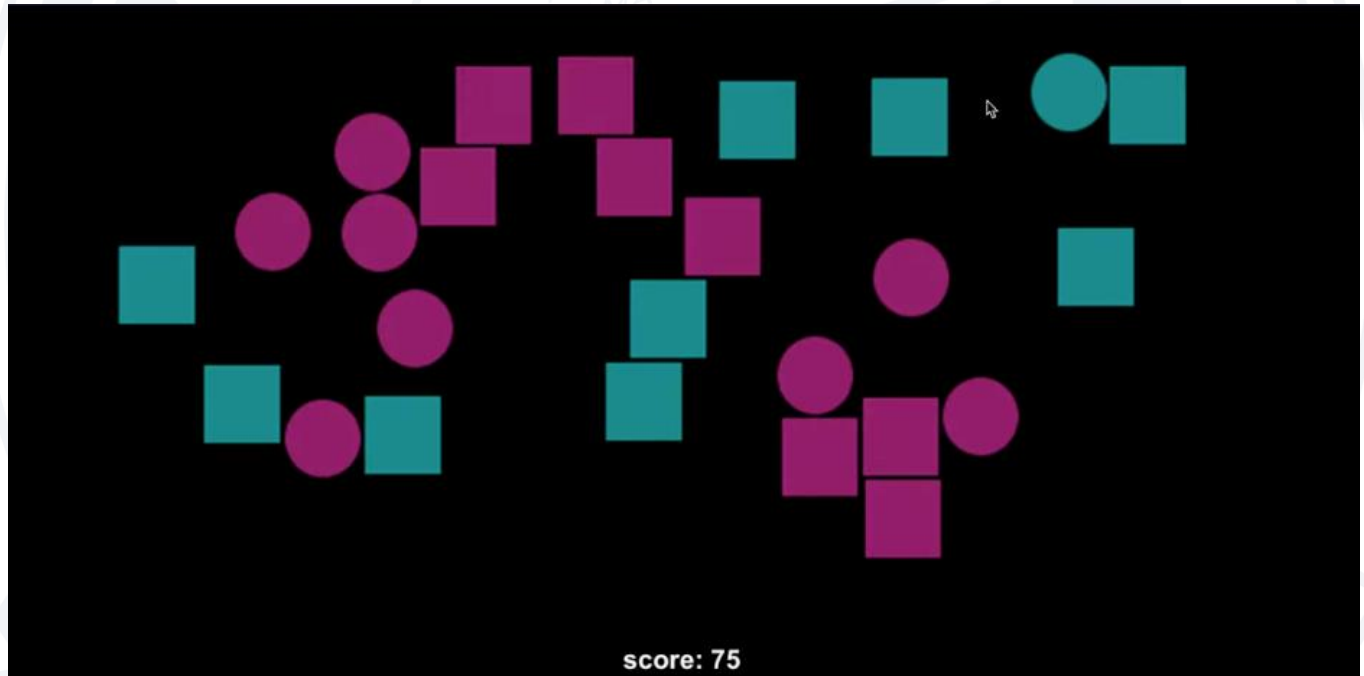judgments about automation mirror judgments about humans

Theory Theory
cues that arise as we observe

Unique Agent
judgments about automation differ from judgments about humans

Method

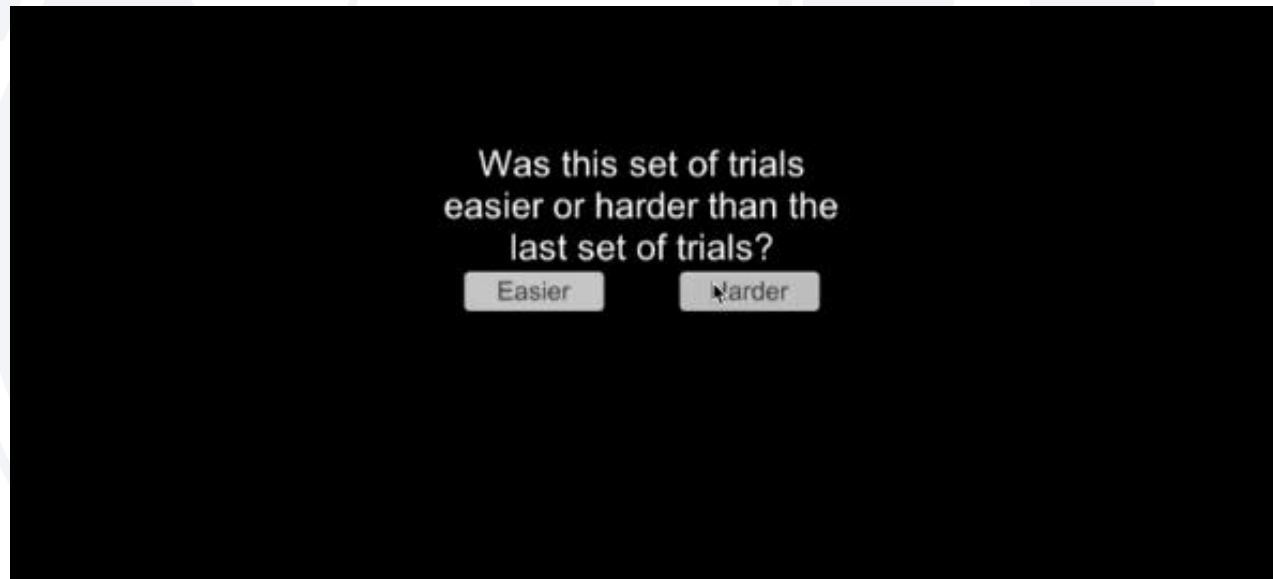- Three studies involving a **visual search task**

Scan for video!

score: 75

(Vangsness & Driggs, 2022; Vangsness & Young, under review; Driggs & Vangsness, in prep)

Method

- After each trial, participants made an **easier/harder JOD**

Was this set of trials
easier or harder than the
last set of trials?

Easier   Harder

(Vangsness & Driggs, 2022; Vangsness & Young,
under review; Driggs & Vangsness, in prep)

Scan for video!

Method

- **<u>Task difficulty</u>** was manipulated **<u>within-subject</u>** for generalizability:

dimension	range	fixed value
clicks	1 – 6 clicks	3 clicks
feedback	2 – 45 points	25 points
set size	2 – 45 items	25 items
timing	1.04 – 4.46 s	2.10s

- blocked by difficulty dimension and counterbalanced across students

(Vangsness & Driggs, 2022; Vangsness & Young, under review; Driggs & Vangsness, in prep)
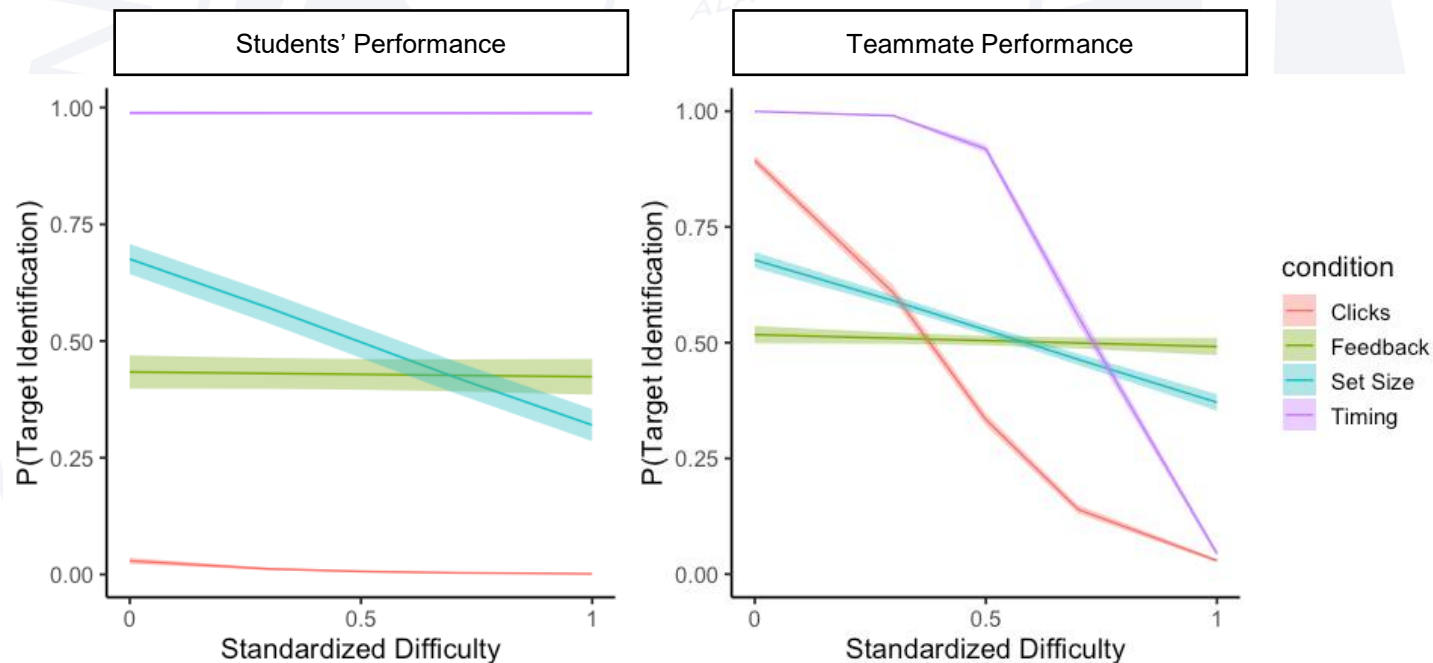
Method

- **Role** was manipulated **between-subject**:
  - **Observe first** – watched a teammate for the first half of each difficulty block; performed during the second half
  - **Perform first** – performed during the first half of each difficulty block; watched a teammate for the second half
  - **Interleaved** – alternated between watching and performing every 5 trials

(Vangsness & Driggs, 2022; Vangsness & Young, under review; Driggs & Vangsness, in prep)

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Method

- The **teammate**'s **reliability** was based on participants' performance during a pilot study

(Vangsness & Driggs, 2022; Vangsness & Young, under review; Driggs & Vangsness, in prep)

Method

- Trust in Automation Questionnaire
  (Jian et al., 2000)

- Propensity to Trust Machines Questionnaire
  (Merritt et al., 2013)

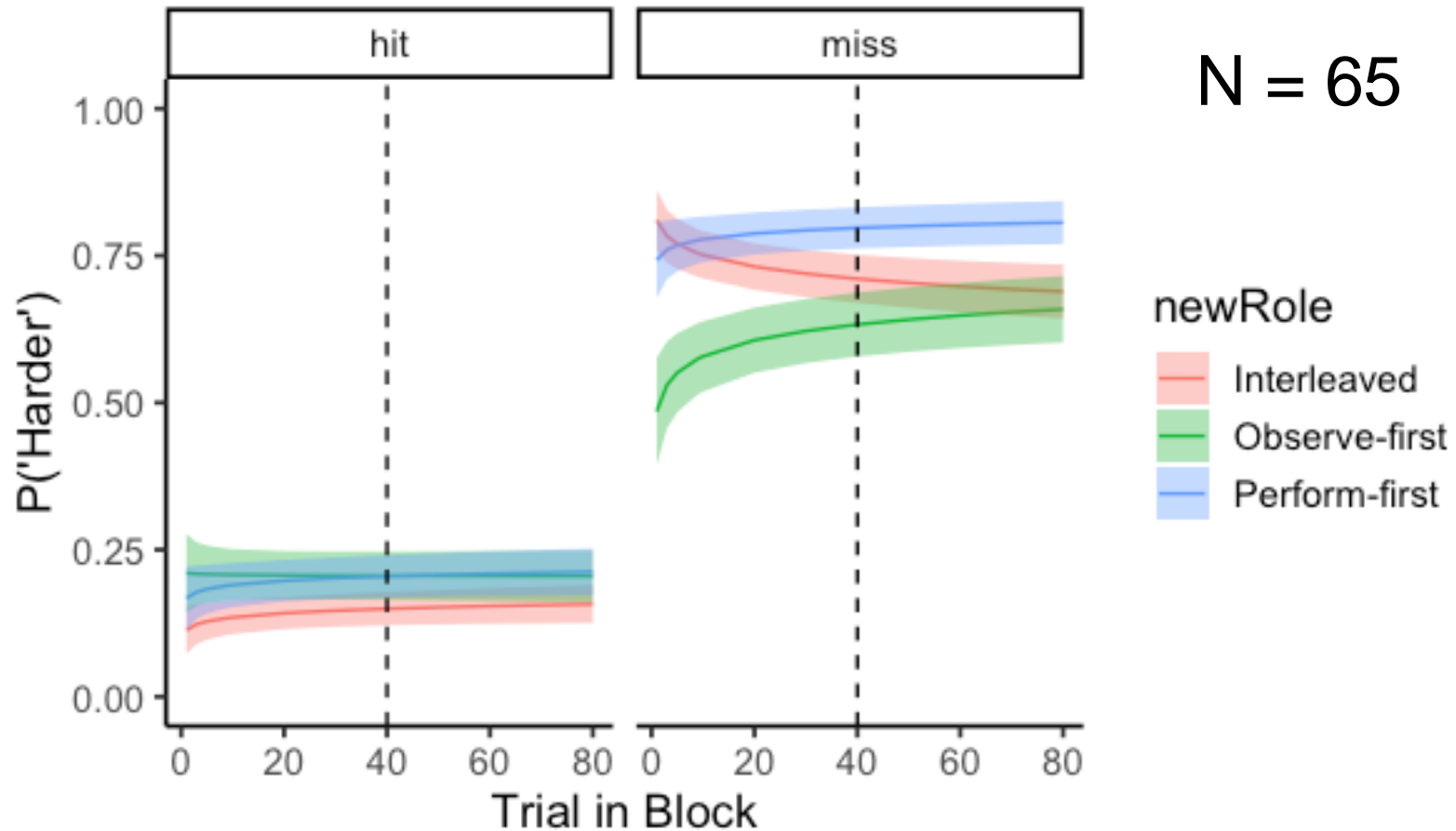(Driggs & Vangsness, 2022)

#ICHMS2022

Method

- Standardized **difficulty** was a **central cue** to difficulty

- **Performance** (successful/unsuccessful identification) was a **peripheral cue** to difficulty
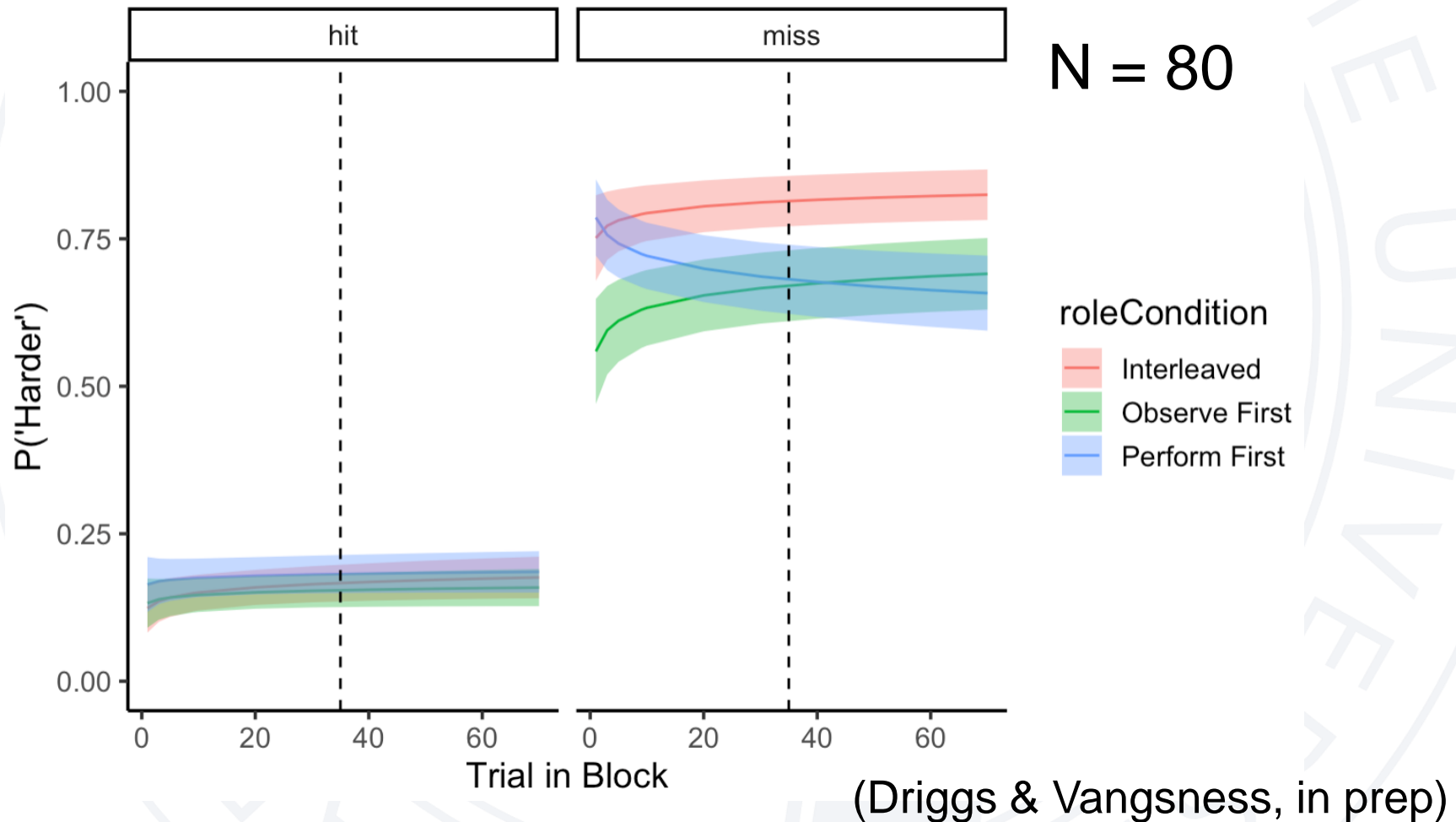
(Vangsness & Driggs, 2022; Vangsness & Young, under review; Driggs & Vangsness, in prep)

Results – Human Teammate

N = 65

(Vangsness & Young, under review)

Results – Automated Teammate

N = 80

(Driggs & Vangsness, in prep)

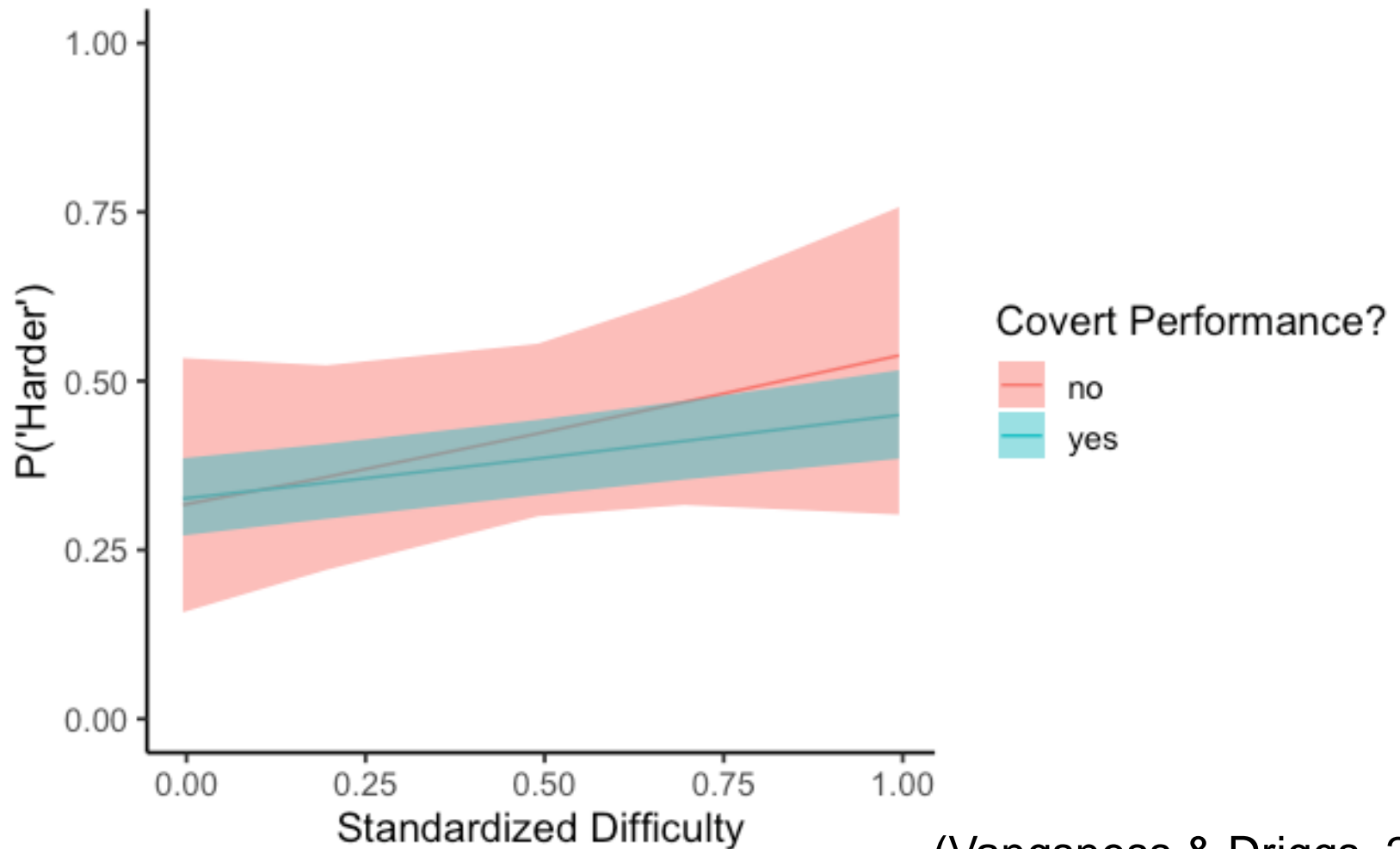THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Method

- **<u>Covert performance</u>** – whether people clicked on the target during watch trials
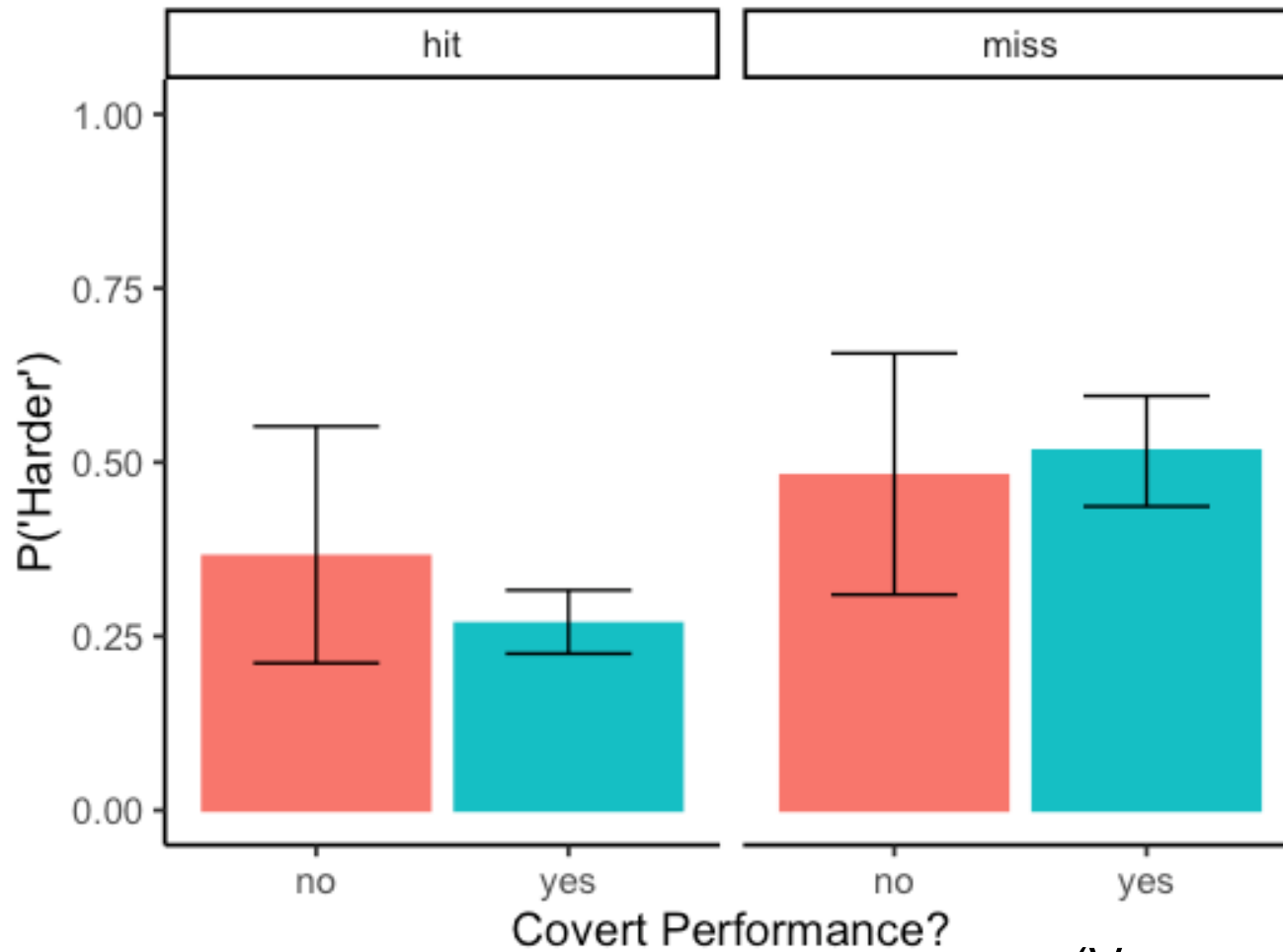
(Vangsness & Driggs, 2022)

Results

(Vangsness & Driggs, 2022)

Results

(Vangsness & Driggs, 2022)

Discussion

- People adhere to **theory theory** when **watching automated systems** (Driggs & Vangsness, in prep)

- People adhere to **simulation theory** when **watching other people** (Vangsness & Young, under review)

- Cue use is not affected by covert performance.

(Vangsness & Driggs, 2022; Vangsness & Young, under review; Driggs & Vangsness, in prep)

Overlap in Theories

- These theories can be disambiguated by studying peoples' cue use during monitoring

Simulation Theory
our imagined performance

Media Equation
judgments about automation mirror judgments about humans

Theory Theory
cues that arise as we observe

Unique Agent
judgments about automation differ from judgments about humans

Discussion

- These theories can be disambiguated by studying peoples' cue use during monitoring

Simulation Theory
our imagined performance

Media Equation
judgments about automation mirror judgments about humans

Theory Theory
cues that arise as we observe

Unique Agent
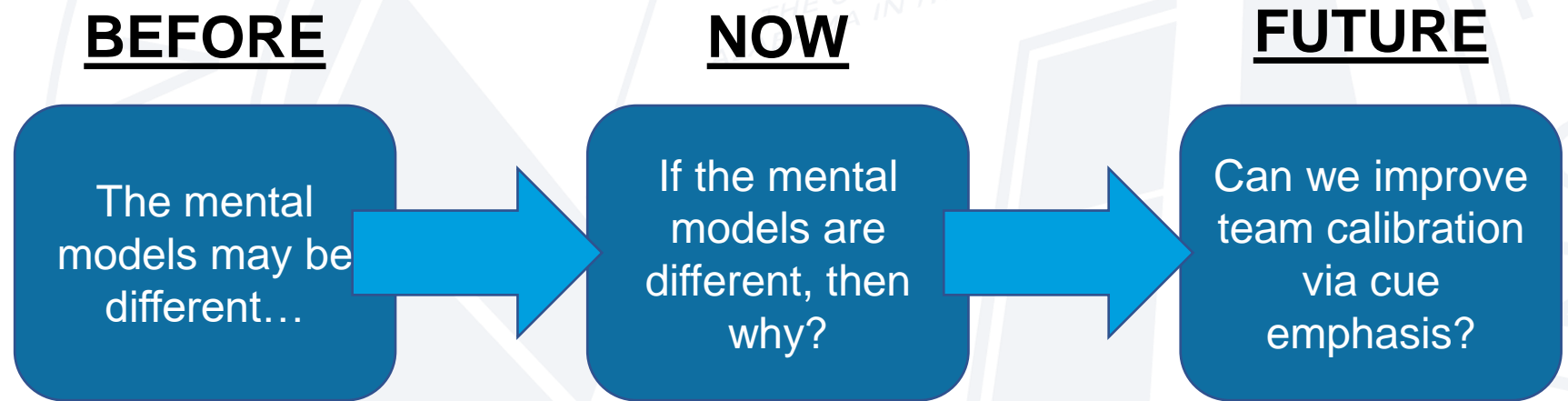judgments about automation differ from judgments about humans

Discussion

- Perhaps **<u>different mental models are employed</u>** when evaluating human vs. automated teammates
(e.g., Tenhundfeld et al., 2021; Tenhundfeld & Witt, 2020)

Discussion

- Extension of findings to other types of
  - automated systems
  - judgments

Future Directions

BEFORE

The mental models may be different…

NOW

If the mental models are different, then why?

FUTURE

Can we improve team calibration via cue emphasis?

Future Directions

BEFORE

NOW

FUTURE

The mental models are different…

→

… because people use different cues.

→

Can we improve team calibration via cue emphasis?

Future Directions

- Capture of **trial-by-trial, dynamic** trust
(Driggs; Vangsness)

- Signal Detection Analysis (SDT) of **trust calibration under different goal frameworks**
(Vangsness)

- Establishing **"ground truth" of trust** calibration
(Sutton; Vangsness)

Future Directions

- Assessing **the incentive structure of** different **explainability aids**
(Chesser; Vangsness)

- Mapping **latent patterns of trust** in text/speech
(Koehl; Vangsness)

- Measuring the **influence of usability on trust**
(Monroe; Vangsness)

- Measuring the **influence of automation on perceptions of risk**

(Wilson; Niven; Koehl; Hendrix; Vangsness)

Acknowledgements

Jade
Driggs, MS

Stephen
Monroe, MA

Rachel
Sutton, MA

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Acknowledgements

Amber
Chesser, BA

Derek
Koehl, MA

Questions?