Assessing Performance in Human-AI Teams:

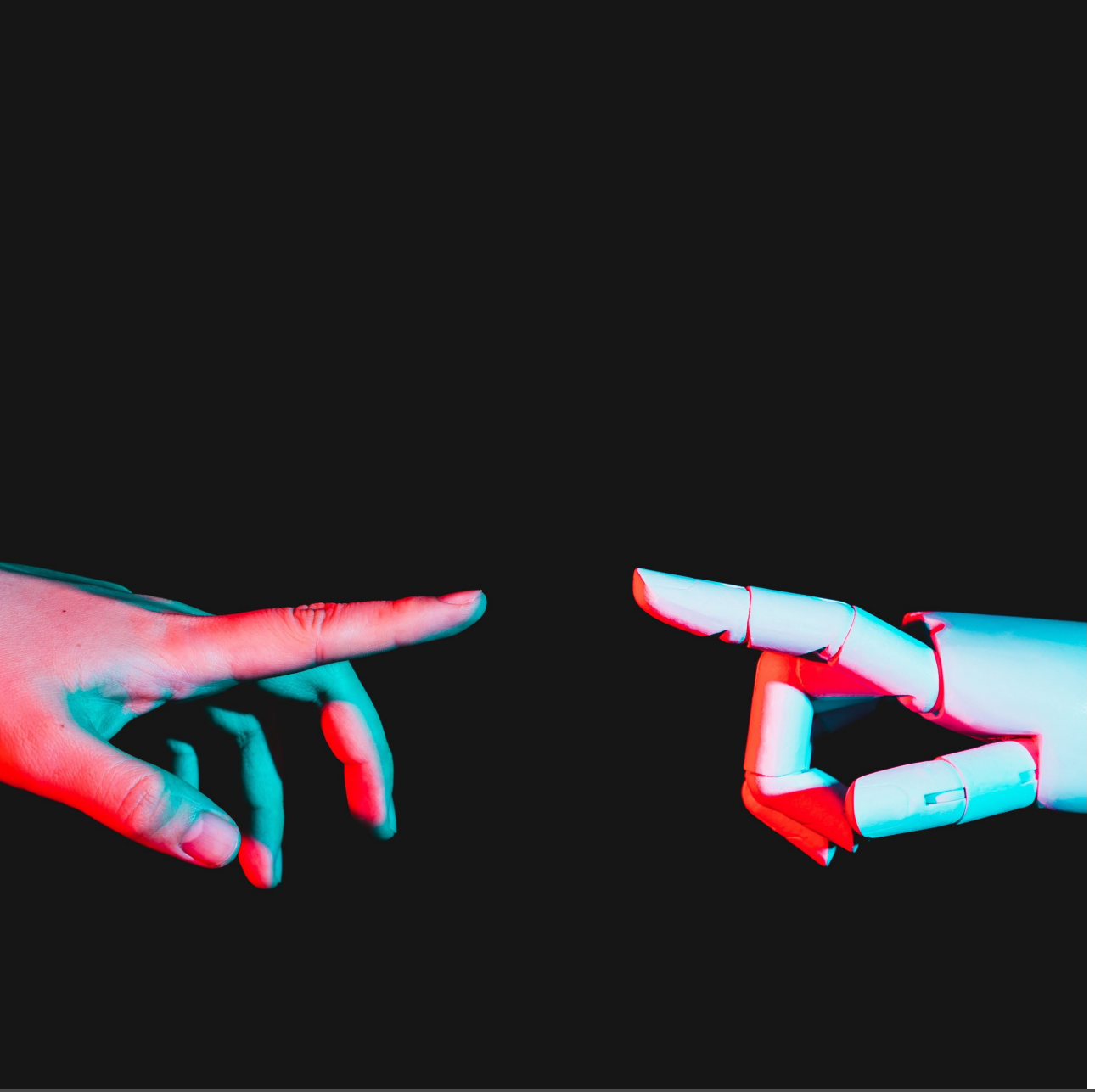
Empirical Evaluation and Moderators for Optimizing Outcomes

Beau G. Schelble

BOHSI Webinar: Performance Optimization in Human-AI Teams
September 9th, 2025

THE UNIVERSITY OF TENNESSEE KNOXVILLE

THE ARCS LAB



Human-Machine Teaming

Definition

- Autonomous Technology Collaborating with Humans as Teammates to **Undertake Taskwork and Teamwork Functions**

- A human-AI team is 2 or more members working **interdependently** toward a common goal
- At least one autonomous agent…
- …where the autonomous agent(s) occupy unique roles on the team and possess **significant agency**.

THE ARCS LAB bschelbl@utk.edu

Human-AI Team Performance

What is Human-AI Team Effectiveness?

- Individual Objective Outcomes
  - e.g., Taskwork Accuracy, Efficiency
- Team Objective Outcomes
  - e.g., Team Goal Completion Accuracy, Efficiency, Compliance
- Affective Outcomes
  - e.g., Satisfaction with Team, Satisfaction with AI Teammate

THE ARCS LAB bschelbl@utk.edu

Human-AI Team Performance

Measuring Objective Outcomes

- Function of Individual and Team-Level **Goal Priorities and Execution**
- Context Dependent
- e.g.
  - Time to Goal Completion
  - Accuracy
  - Resources Expended
  - Compliance with Standards
- AI Teammate's Actions **Should not Dominate** Measures of Human-AI Team Performance

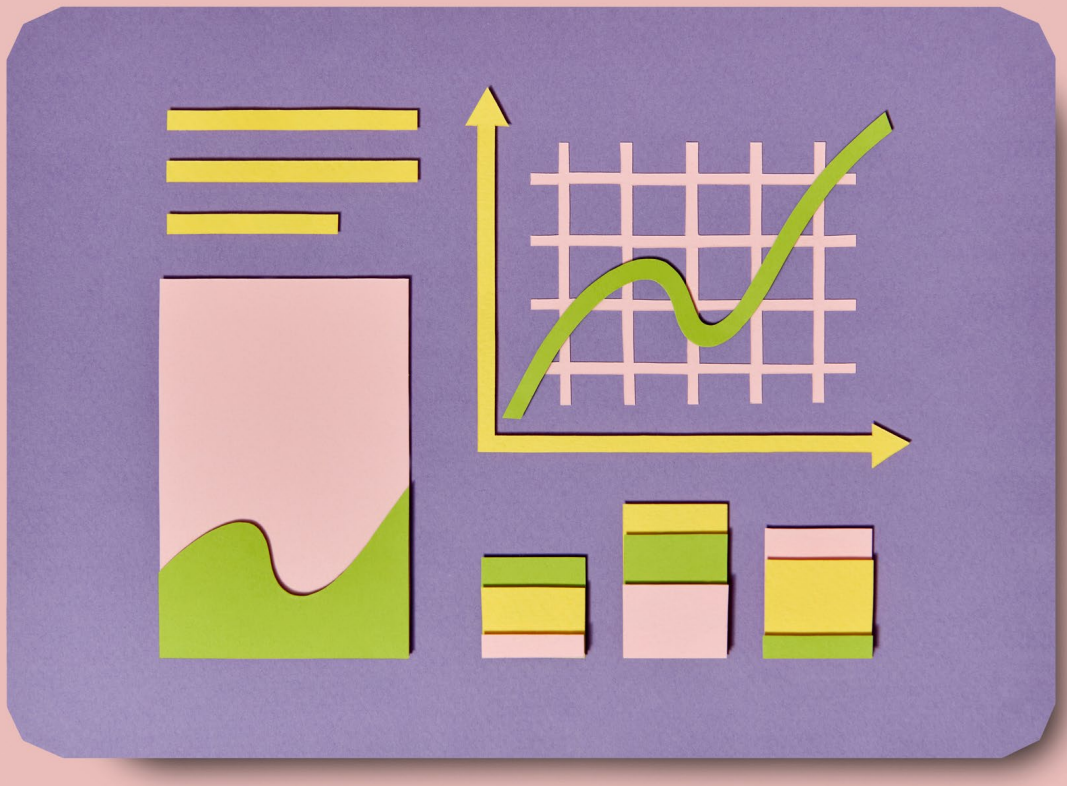
THE ARCS LAB bschelbl@utk.edu

Human-AI Team Performance

Measuring Objective Outcomes

- Weighting Depends on the Context
  - Maximize Accuracy when Errors are Costly
  - Maximizing Compliance/Safety Procedures when Risk is High
  - Maximizing Time-Based Efficiency when Speed is Critical
  - Balance Weights when Trade-Offs are Acceptable
- Performance Composites for AI Teammate Training will Require **Rewards for Teamwork**

THE ARCS LAB bschelbl@utk.edu



Composite Score Example

$$TPS_i = 1000 - \left(\sum Team\ Penalty_j\right)$$

- Missed Waypoint Penalty (Weight = 318.63)
- Missed Photos Penalty (Weight = 314.96)
- Alarm Penalty (Weight = 393.22)
- Warning Penalty (Weight = 112.02)

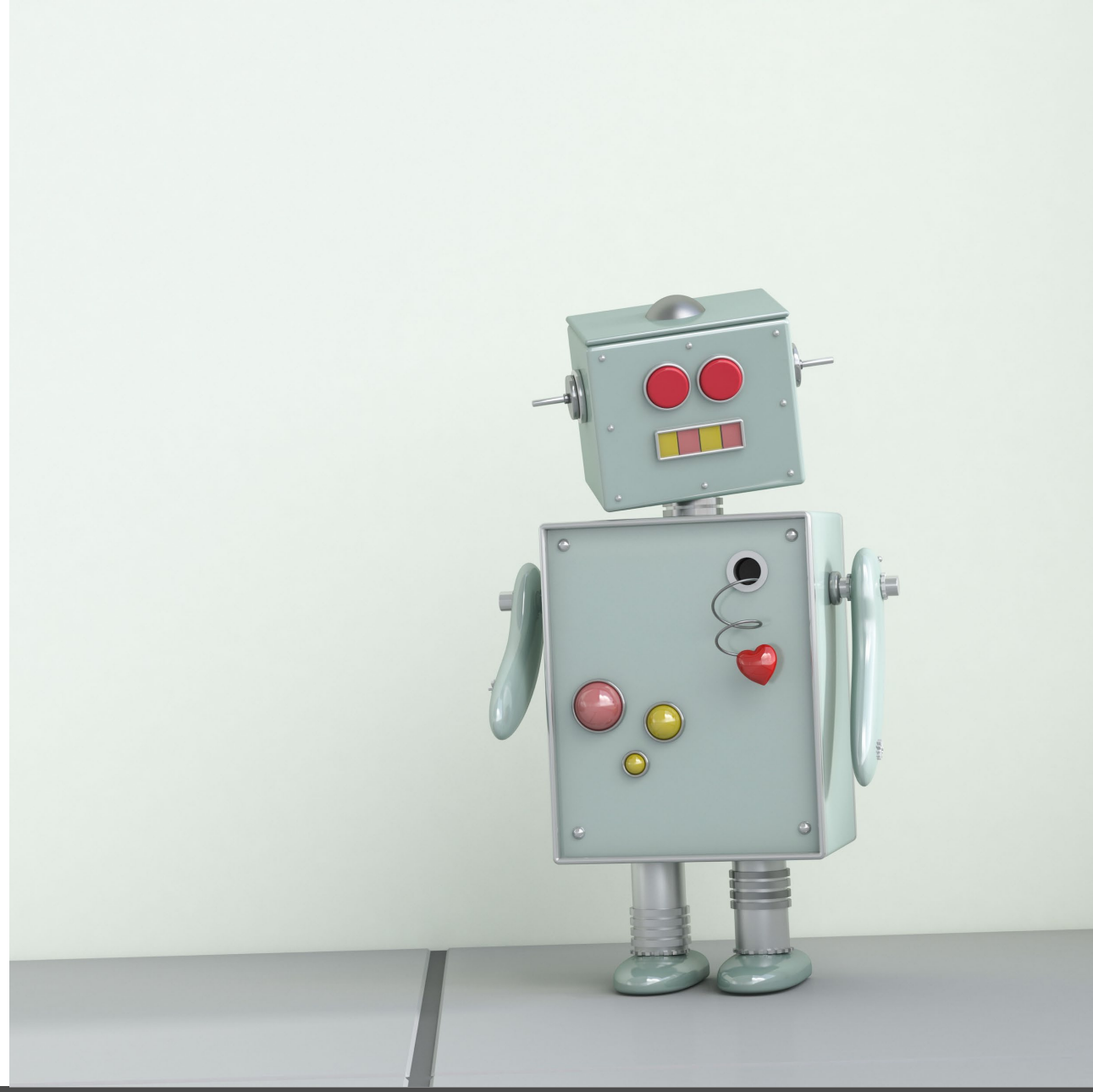
Weights Reflect **Priorities** (Warnings versus Alarms)

Cooke, N. J., Gorman, J., Pedersen, H., Winner, J., Duran, J., Taylor, A., ... & Rowe, L. (2007). *Acquisition and Retention of Team Coordination in Command-and-Control* (No. AFRLHEAZTR20070041).

Human-AI Team Performance

Measuring Affective Outcomes

- How did Team Members Feel About Their Experience?
- Was their Experience with the AI Positive?
- Typically Measured Using Surveys
  - **Impact** Future Human-AI Team Outcomes (Schelble et al., 2020; Hafizoğlu & Sen, 2018a; 2018b)

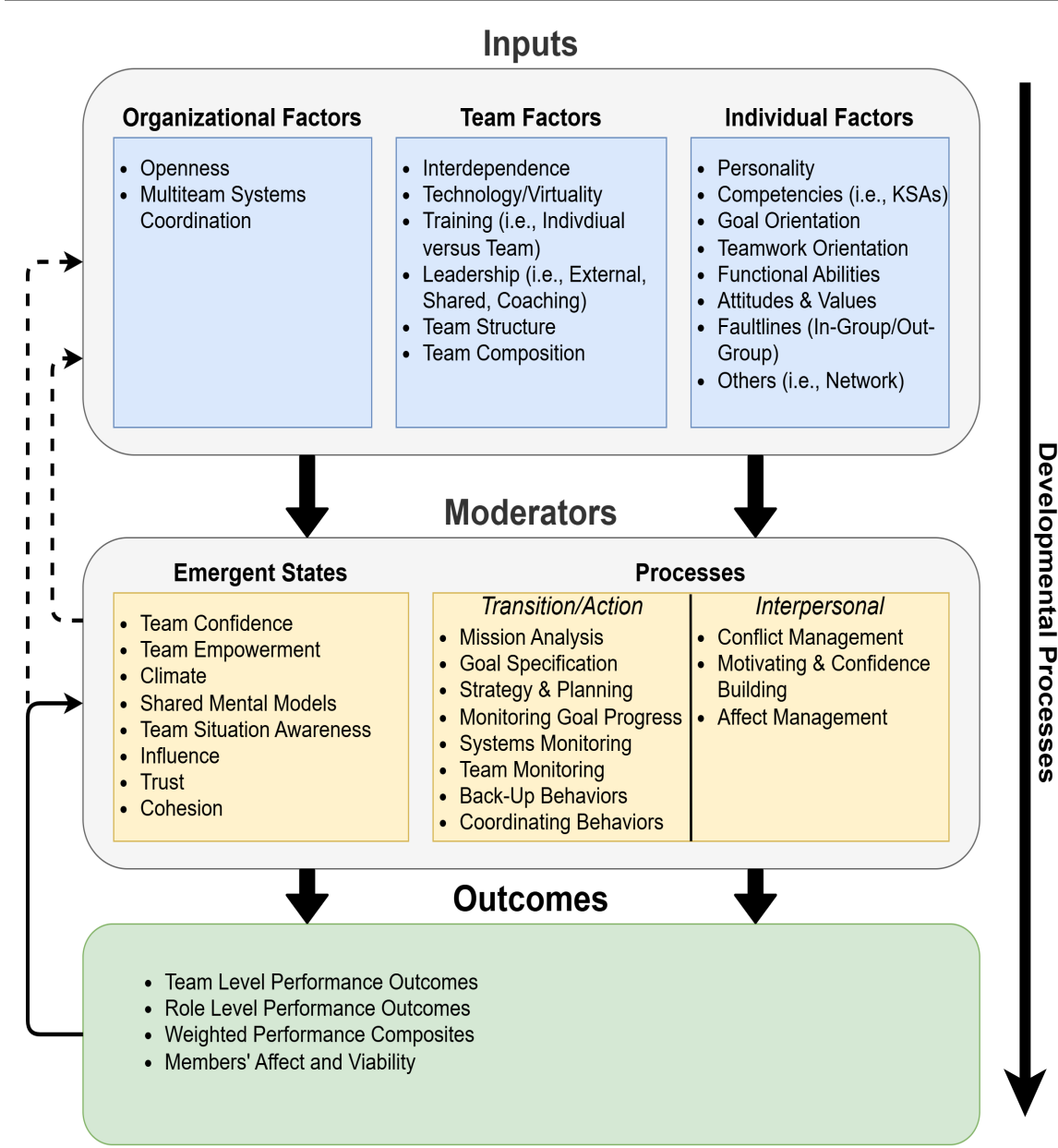


THE ARCS LAB bschelbl@utk.edu

Performance Antecedents

How Do Human-AI Teams Get to Outcomes?

- AI Excels in Rote Taskwork (Schelble et al., 2022a)
- However, Teamwork is a Struggle for AI (McNeese et al., 2018)
- These Teamwork and Taskwork Moderators **Explain the Variance in Outcomes**
- E.g.
  - Trust (Emergent)
  - Team Situation Awareness (Emergent)
  - Monitoring (Process)
  - Coordination (Process)



bschelbl@utk.edu

Performance Antecedents

Optimizing Human-AI Team Performance

Inputs

- Individual Differences (Jorge et al., 2023)
- Team Composition (Schelble et al., 2022a)

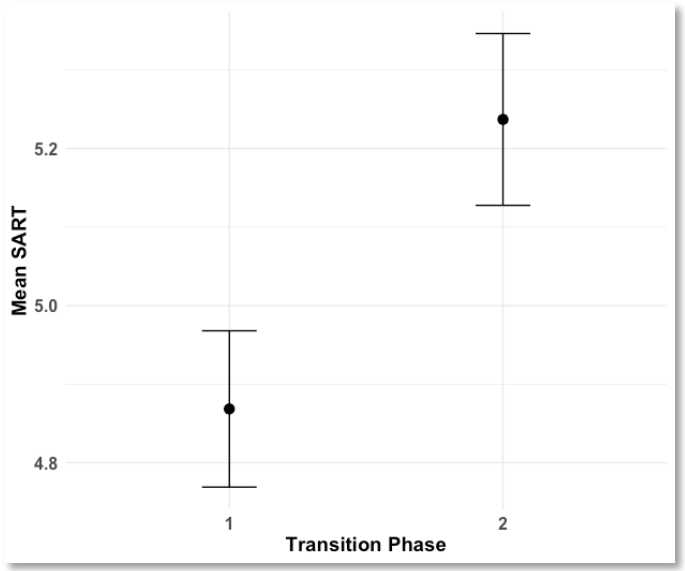
Emergent States

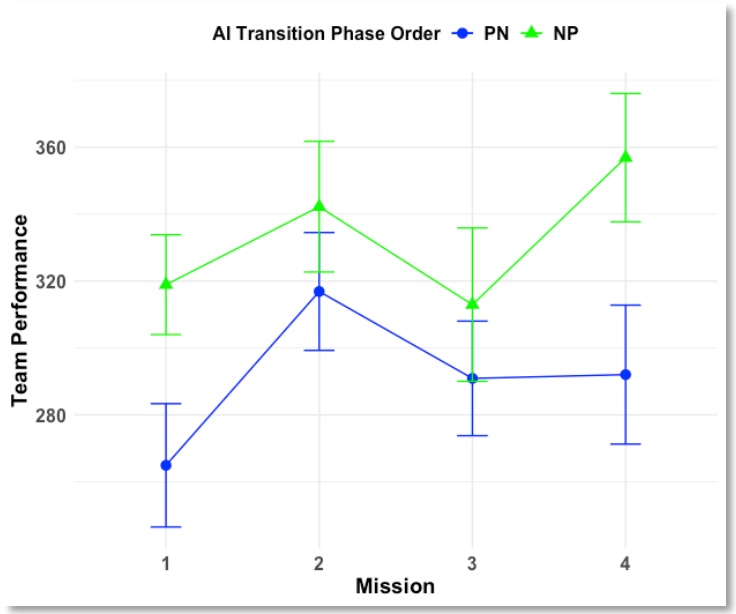
- Shared Mental Models (Schelble et al., 2022b)
- Trust (Schaefer et al., 2017)
- Situation Awareness (Endsley, 2023)

Processes

- Communication (McNeese & Cooke, 2016)
- Adaptation (Hauptman et al., 2023)

THE ARCS LAB bschelbl@utk.edu





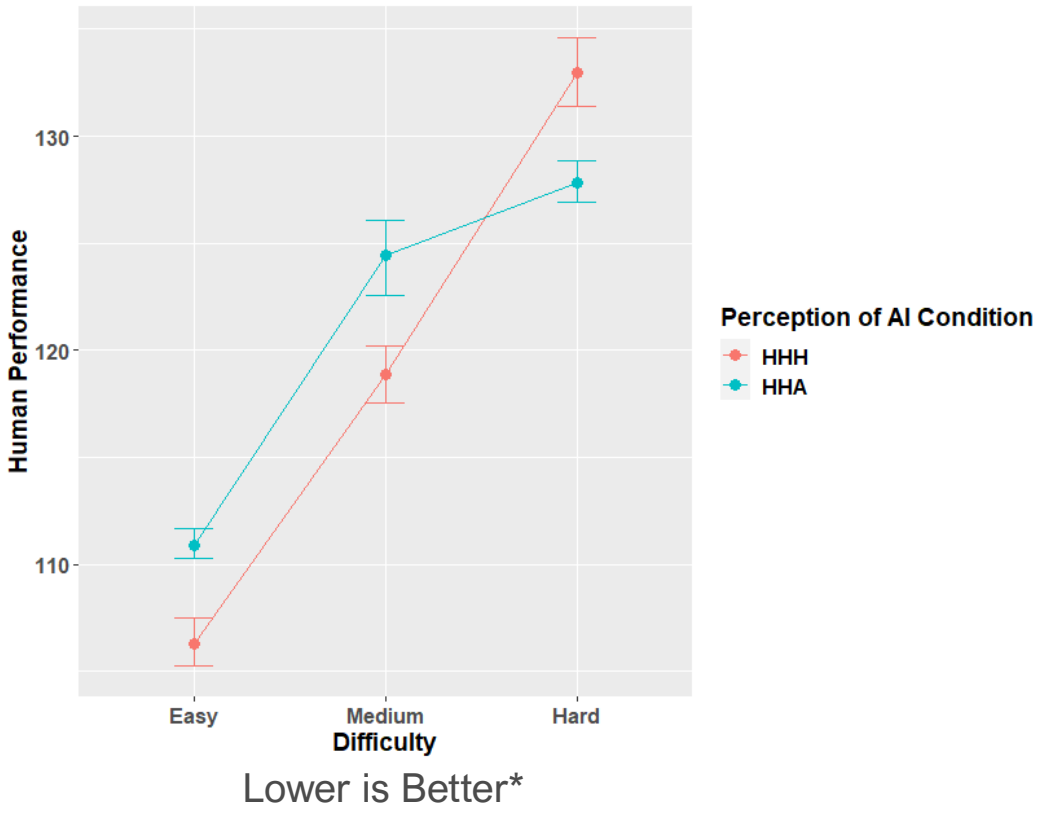
Performance Antecedents Example

When Should AI Join Strategy Discussions?

- Effects **Varied by Timing** in the Team Life Cycle
- Stronger Benefits **Later in the Life Cycle**
- Direct Effects on…
  - Situational Awareness
  - Objective Performance

Schelble, B. G., Mallick, R., Hauptman, A., & McNeese, N. (2025). Should AI Teammates Give All the Answers? Examining the Role of Different AI Information-Sharing Techniques on Team Cognition in Human-AI Teams. *International Journal of Human–Computer Interaction*, 1-26.

Performance Antecedents Example



Lower is Better*

Perceived Teammate Artificiality Affects Performance

- Performance **Improved in Higher-Difficulty** Phases
- Affective States **Influenced Outcomes** in Human-AI Teams

Schelble, B. G., Flathmann, C., McNeese, N. J., O'Neill, T., Pak, R., & Namara, M. (2023). Investigating the Effects of Perceived Teammate Artificiality on Human Performance and Cognition. *International Journal of Human–Computer Interaction*, *39*(13), 2686-2701.

Takeaway #1

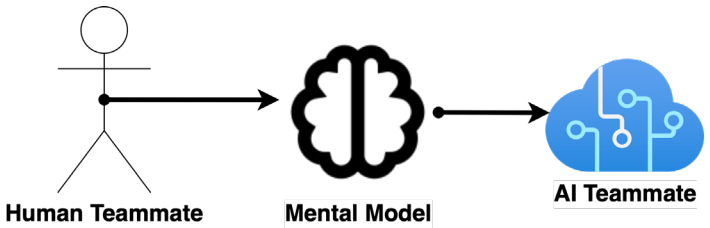
Human-AI Team Performance is Multifaceted

- Objective Performance
  - Measured as a Function of Accuracy, Efficiency, and Compliance
  - Weighting Based Upon Context
- Affective Outcomes
  - Survey Measures for Affective Outcomes
  - May also Utilize Expert or Supervisor Evaluation of Team Outputs

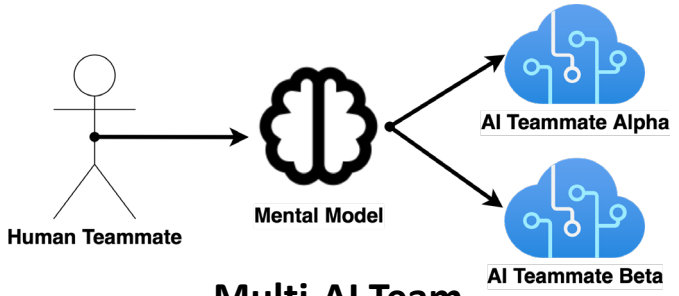
Takeaway #2

Composition Increases Performance Variance

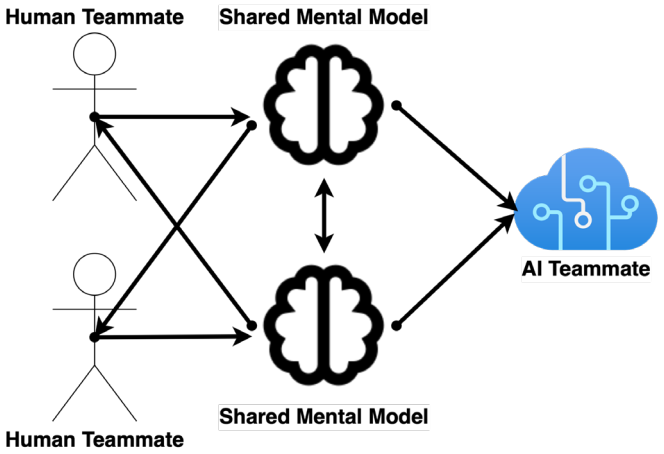
- Expanding Human-AI Teams **Increases the Variance** in Performance Outcomes
- These Complexities from Inputs and Moderators **Have Always Existed** and Involve Several Constructs
- Increased Complexity Highlights the Need to **Ensure AI is Working Towards the Shared Goal**



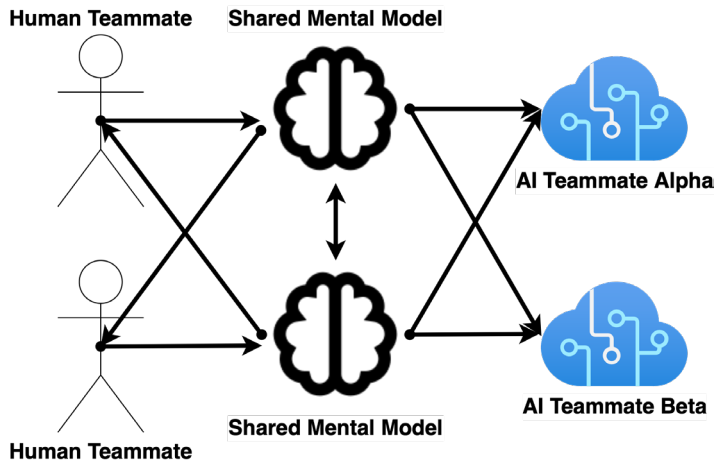
Dyad



Multi-AI Team



Multi-Human Team



Multi-Human Multi-AI Team

bschelbl@utk.edu

Takeaway #3

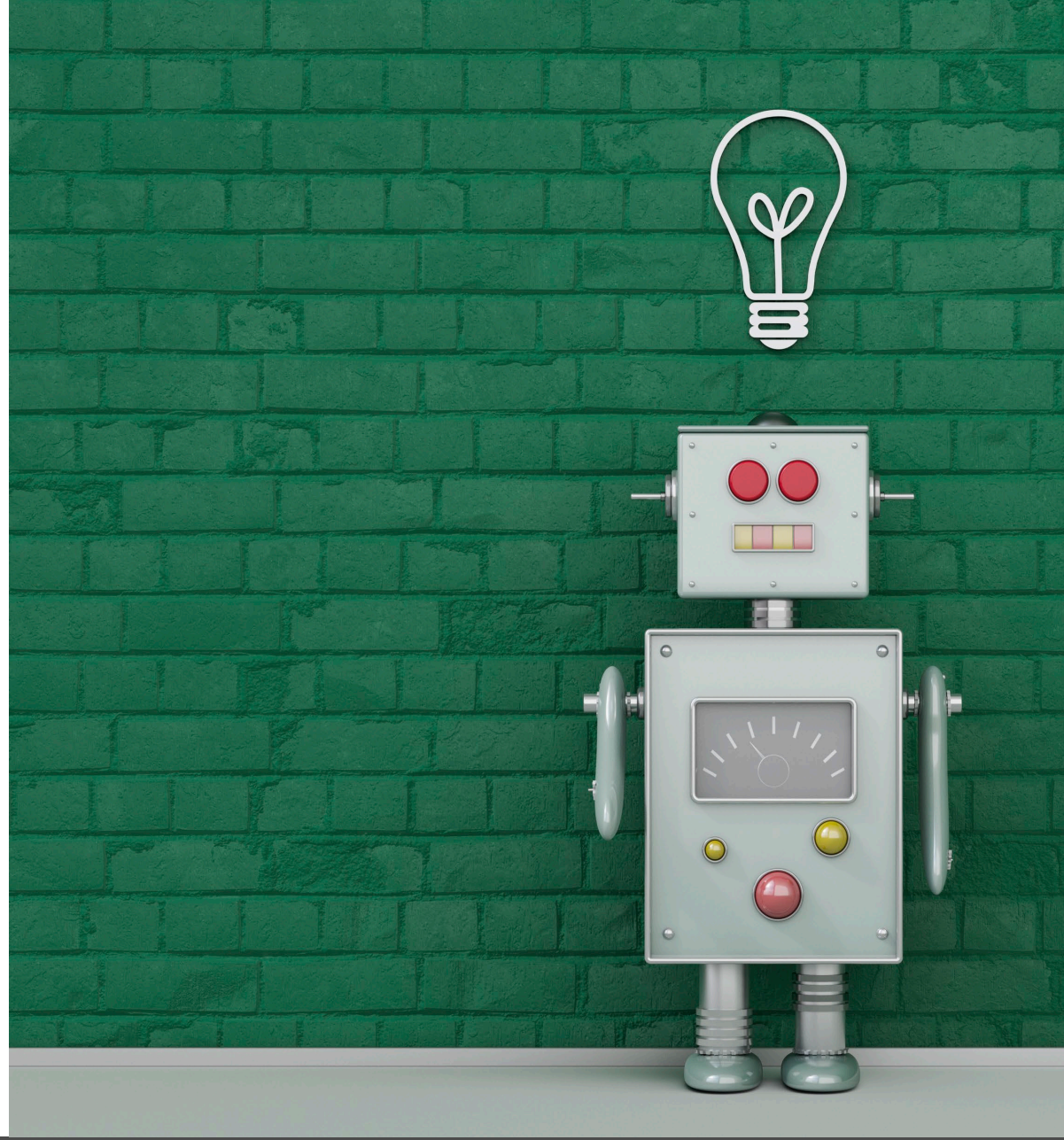
AI Teammates Must Support Teamwork Behaviors to Scale

- Optimal AI Teamwork Behaviors **will Likely Differ** from Human Teammates
- Communication and Coordination **Remain Fundamental to Team Performance**
- AI can Support the Creation of **Novel Emergent States and Team Processes** to Drive Improvements to Team Outcomes
- Leverage AI to **Manage Complexity in Moderators** to Reduce Performance Variance

Takeaway #4

Human-AI Complementarity

- Design Team Roles to **Individual Strengths**
  - Human Judgement in Uncertainty
  - AI Speed and Consistency
- **Adapt Decision-Making Authority** Using Predictors of *Individual* Performance
- Human-AI Team Performance is Wholly Reliant on **Bidirectional Performance Improvement** Between Human and AI Teammates



Conclusion

Align Performance Metrics **with Human *and* AI Decision Making**

Equip AI to Support Teamwork Behaviors to Scale Human-AI Teams

Emphasize Antecedents to Optimize those Performance Metrics