Quantifying Calibration: Bridging Trust and Reliance in Automation Across Cultural Values and Dispositional Factors

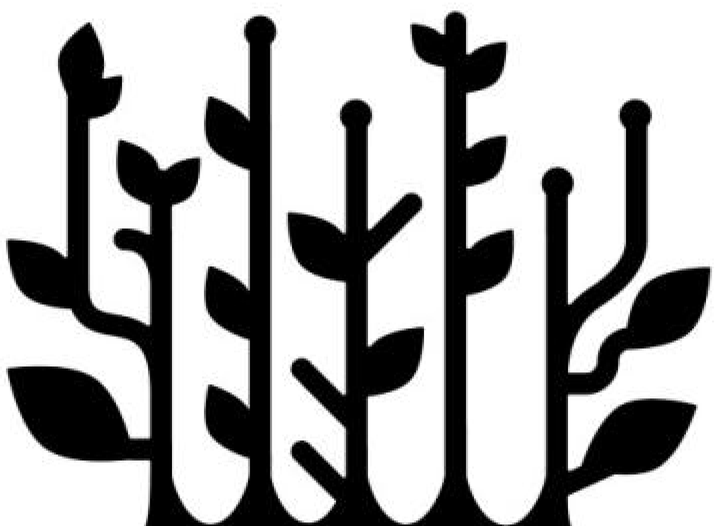
Evelyn S. Goroza

in partial fulfillment of the requirements for the degree of
Master of Science in Human Factors Engineering

Tufts University

August 2025

Adviser: Dave B Miller

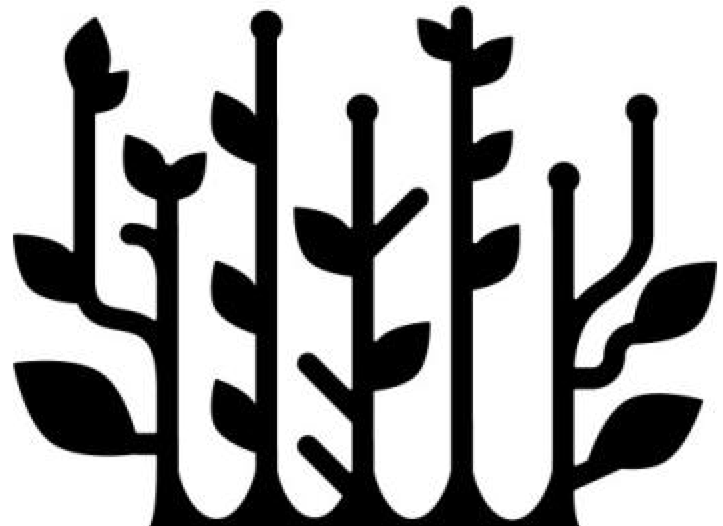


Acknowledgements

First, a **massive thank you** to everyone who stood by me during this wildly ambitious thesis adventure.
Otto and I couldn't have done it without you!

This work is dedicated to:

- **My parents, my family**, and **friends**
- **Dave Miller,** my advisor
- **Dan Hannon** and **Holly Taylor**, my additional committee members
- **Gavin McCarthy-Bui** and **Anne Zhao** – for bringing the Unity game to life
- **Ryan Veiga** – for guiding me through the statistical jungle and R coding
- **Elsa Ostenson, Parisa Arastu, Malu Sajith, Jacob Ratzliff, & Harlan Knightly** – for your help with the usability side of Calibratio
- **Every brilliant author and researcher** cited in my References – for being the kind of crazy I aspire to someday reach.

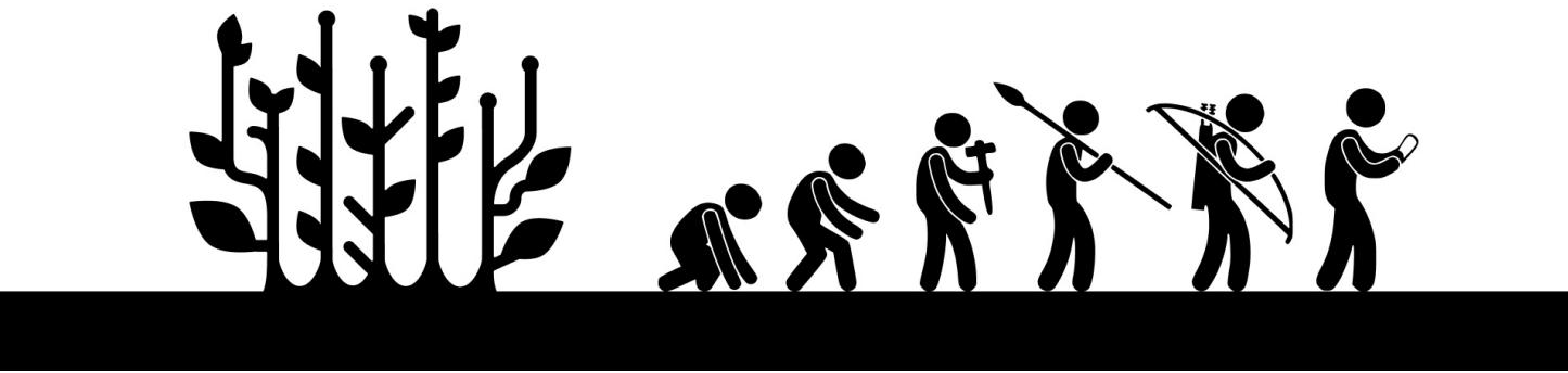


"Many animals make and use tools, but humans are distinctive in the complexity, diversity, sophistication, omnipresence, and obligatory nature of our reliance on tools.

Simply put, other animals use tools but only humans have technology. Indeed, humans inhabit a uniquely ***technological niche*** *that we ourselves have constructed"*

(Mangalam et al., 2022, p. 13).

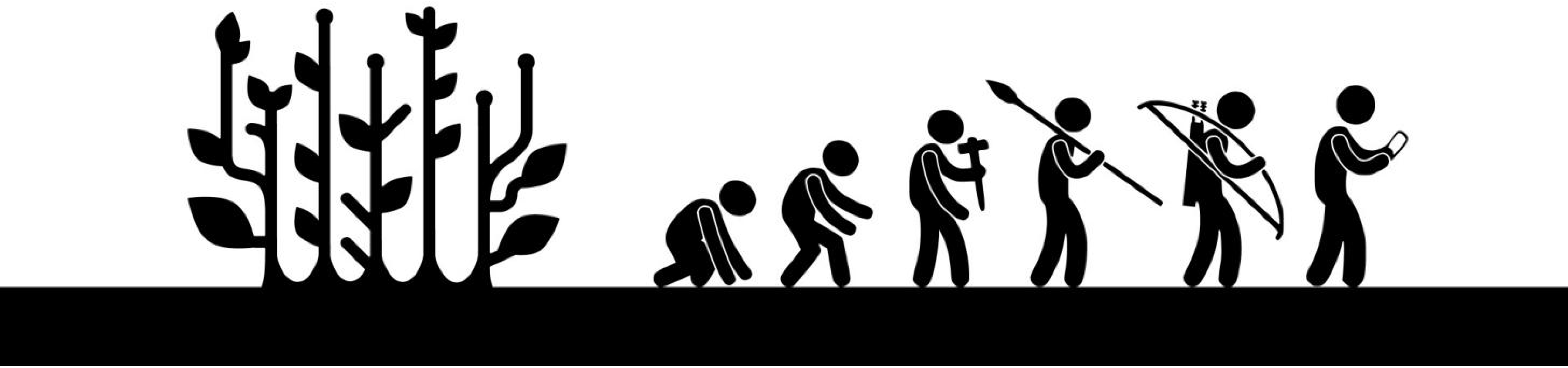
Tools can **extend** human capability.



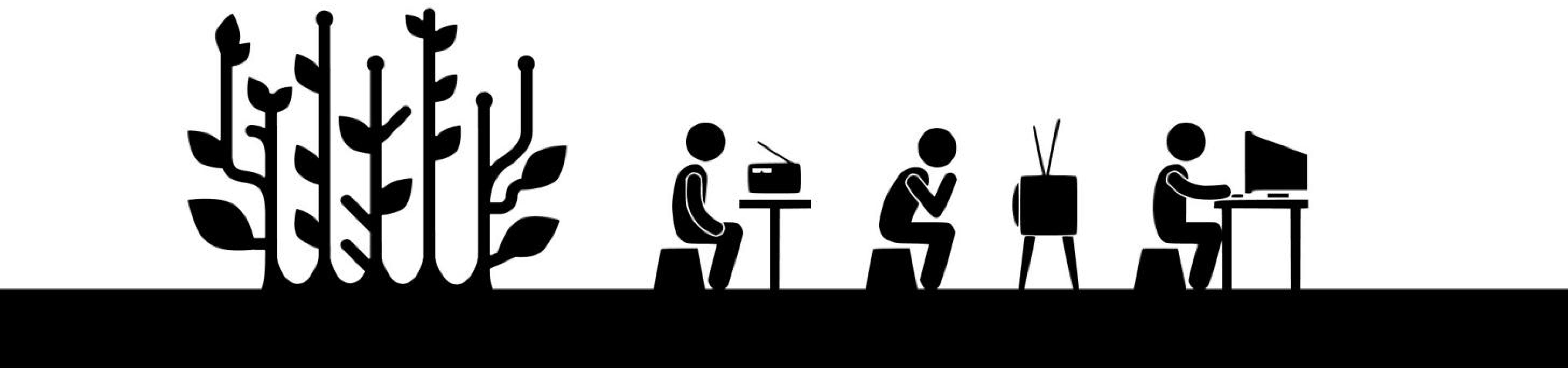
Humans inhabit a self-constructed **Technological Niche** (Mangalam et al., 2022).

The tools we make evolved from a manipulation of our **environment**.

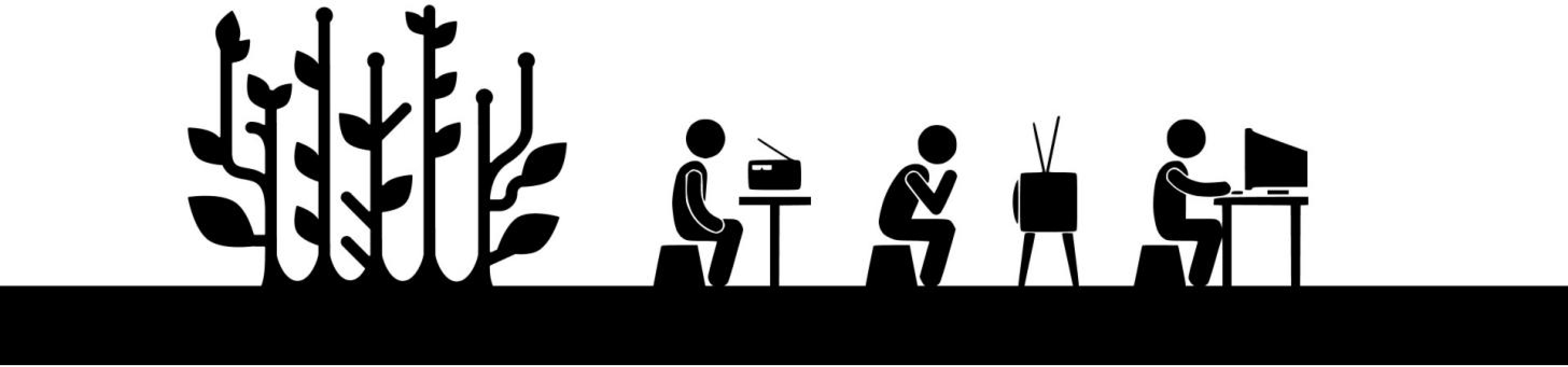
Physical Tools emerge **directly** from the environment (Osiurak, 2018).

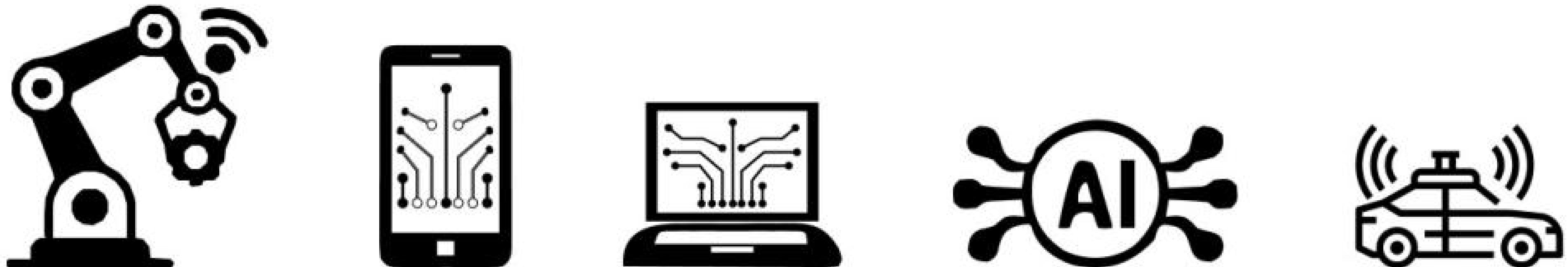


Sophisticated Tools enable us to solve more **complex** problems (Osiurak, 2018).

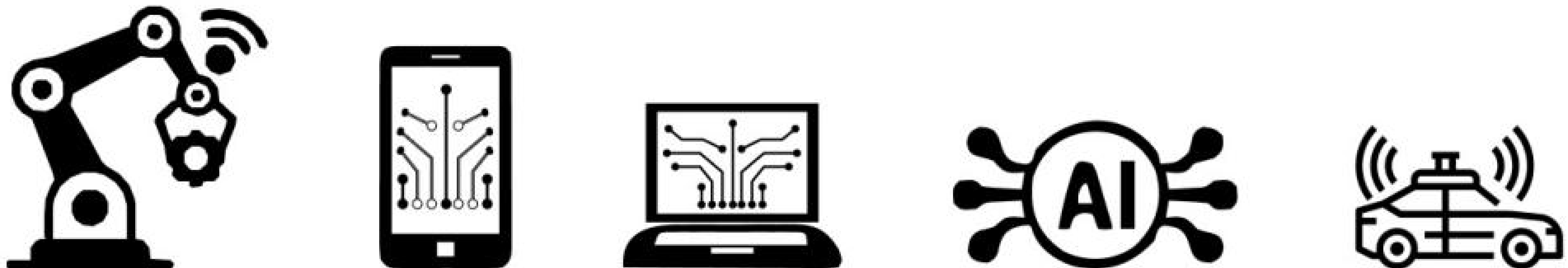


Sophisticated tools known as **automation** can replace manual processes.





Automation:
a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator” (Parasuraman, 2000).



Automation can be applied to four broad classes of functions:

1) information acquisition
2) information analysis
3) decision and action selection
4) action implementation

(Parasuraman et al., 2000, p. 286).

Automatica, Vol. 19, No. 6, pp. 775–779, 1983
Printed in Great Britain.

0005–1098/83 $3.00 + 0.00
Pergamon Press Ltd.

Brief Paper

Ironies of Automation*

LISANNE BAINBRIDGE†

Key Words—Control engineering computer applications; man–machine systems; on-line operation; process control; system failure and recovery.

Abstract—This paper discusses the ways in which automation of industrial processes may expand rather than eliminate problems with the human operator. Some comments will be made on methods of alleviating these problems within the 'classic' approach of leaving the operator with responsibility for abnormal conditions, and on the potential for continued use of the human operator for on-line decision-making within human–computer collaboration.

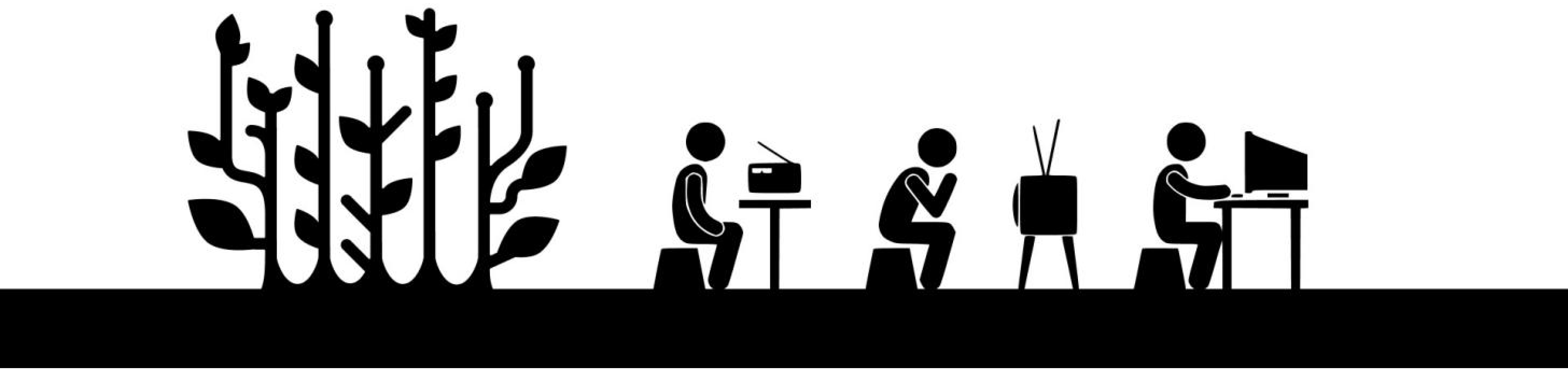
Irony: combination of circumstances, the result of which is the direct opposite of what might be expected.
Paradox: seemingly absurd though perhaps really well-founded statement.

THE classic aim of automation is to replace human manual control, planning and problem solving by automatic devices and computers. However, as Bibby and colleagues (1975) point out: "even highly automated systems, such as electric power networks, need human beings for supervision, adjustment, maintenance,

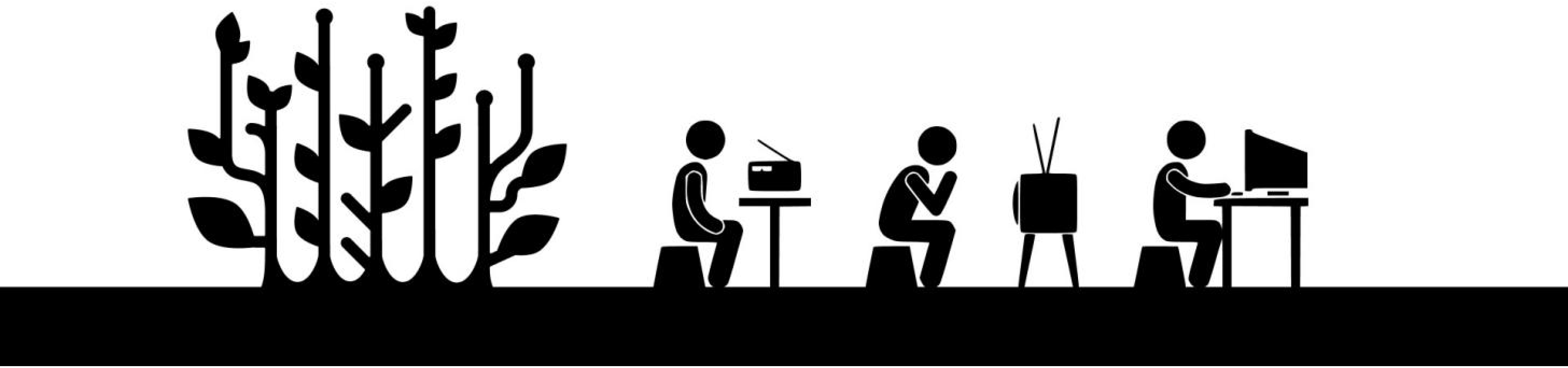
designer errors can be a major source of operating problems. Unfortunately people who have collected data on this are reluctant to publish them, as the actual figures are difficult to interpret. (Some types of error may be reported more readily than others, and there may be disagreement about their origin.) The second irony is that the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate. It is this approach which causes the problems to be discussed here, as it means that the operator can be left with an arbitrary collection of tasks, and little thought may have been given to providing support for them.

1.1. *Tasks after automation.* There are two general categories of task left for an operator in an automated system. He may be expected to monitor that the automatic system is operating correctly, and if it is not he may be expected to call a more experienced operator or to take-over himself. We will discuss the ironies of manual take-over first, as the points made also have implications for monitoring. To take over and stabilize the

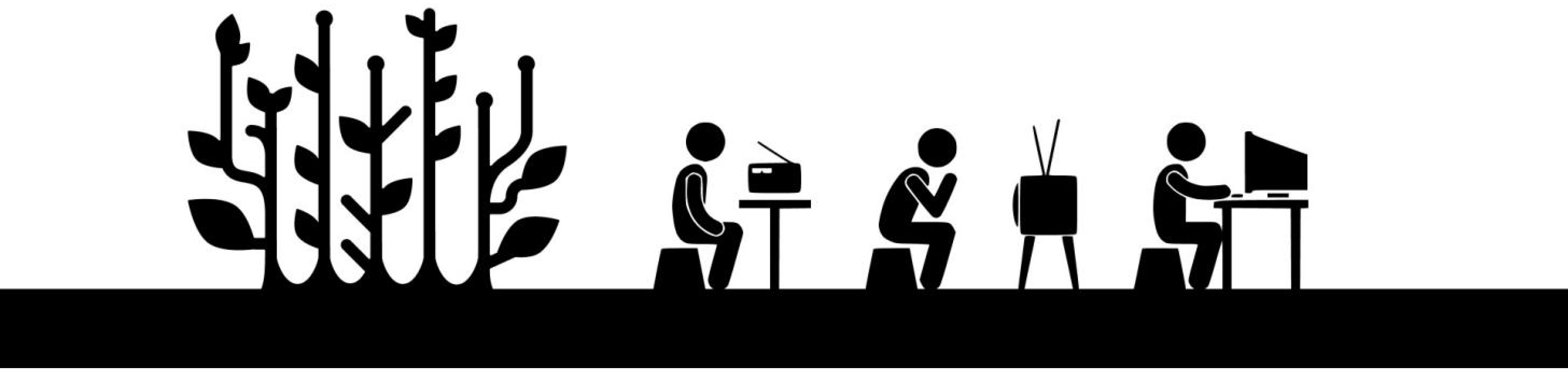
Well-designed automation facilitates **appropriate reliance**.



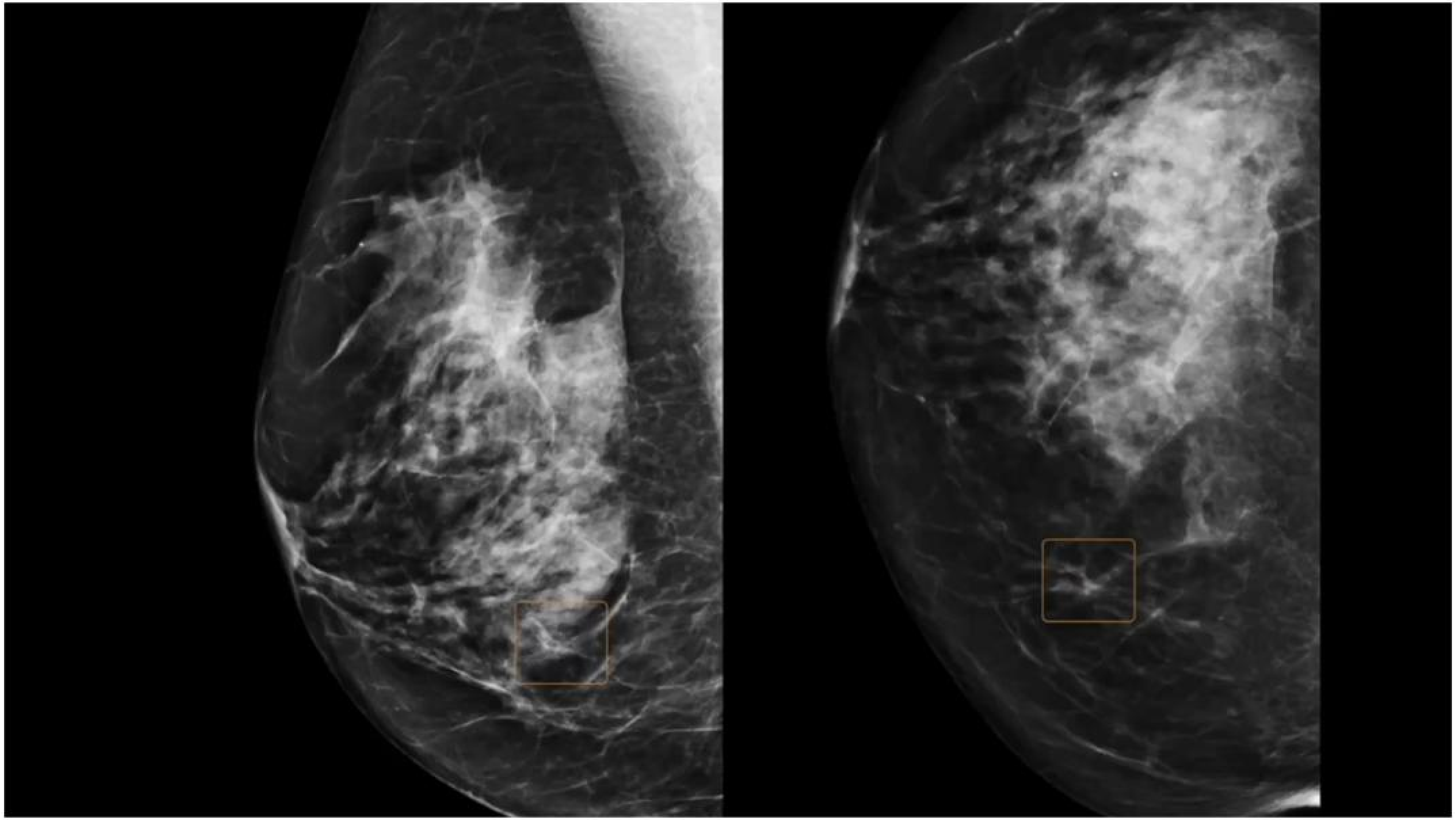
Reliance on automation, or **automation use**, is the behavior describing the voluntary engagement of an automated system.



Appropriate reliance is characterized by **calibrated use,** in which the operator's reliance on a system aligns with the **capability** of the system (Parasuraman & Riley, 1997; Lee & See, 2004).



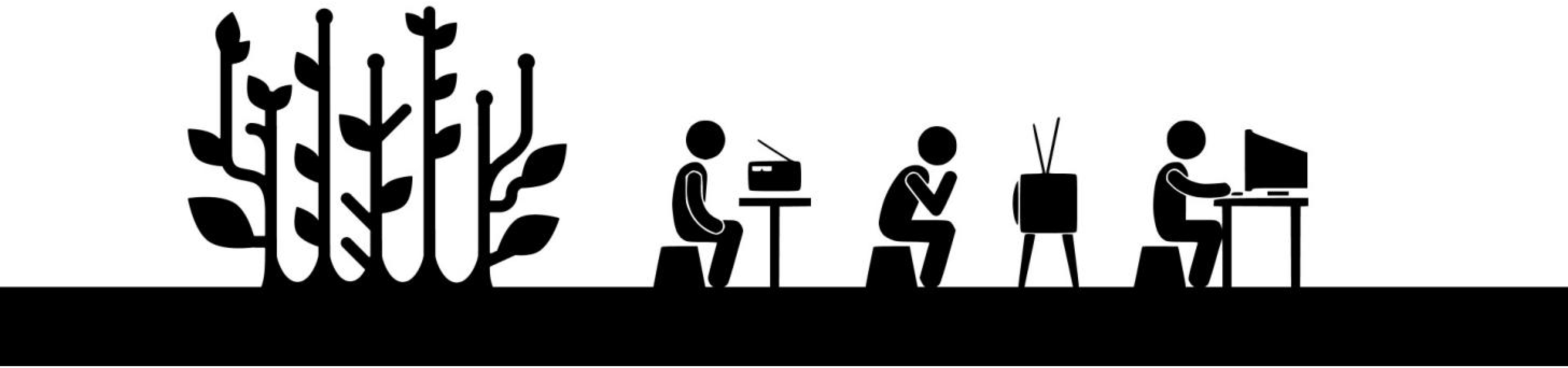
Example: Mammography Decision Support System



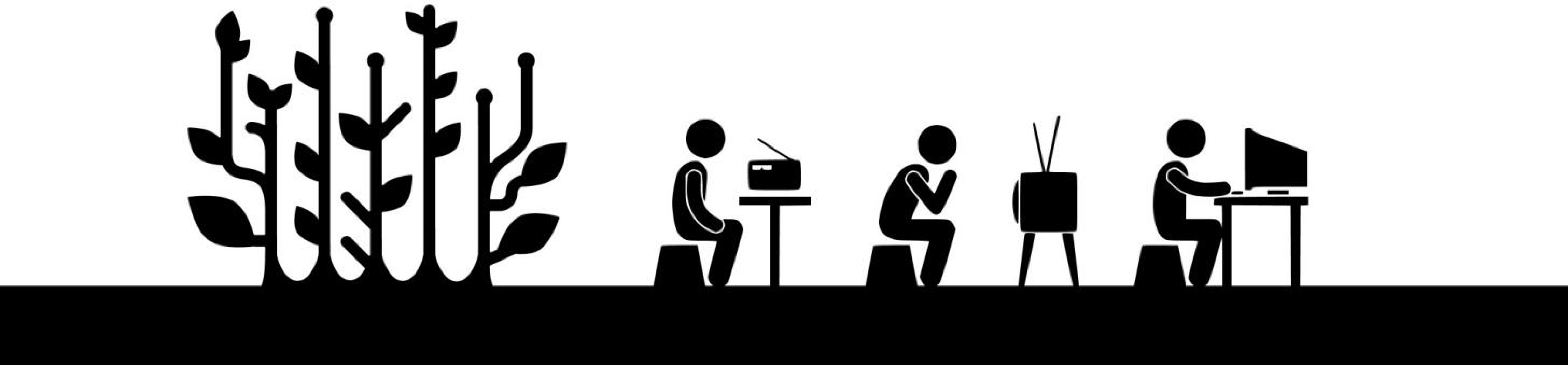
(Northwestern University, 2020)

https://www.youtube.com/watch?v=Mur70YjInmI&list=PPSV

Well-designed automation facilitates **appropriate reliance**.



Reliance is largely dependent on **trust**.



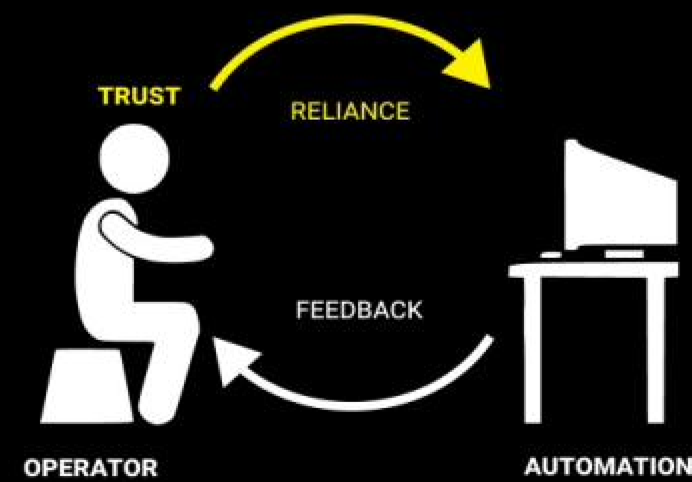
Trust in Automation

Trust is the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability (Mayer et al., 1995).

Background

Trust

Human-Machine System



Background

Lee & See 2004

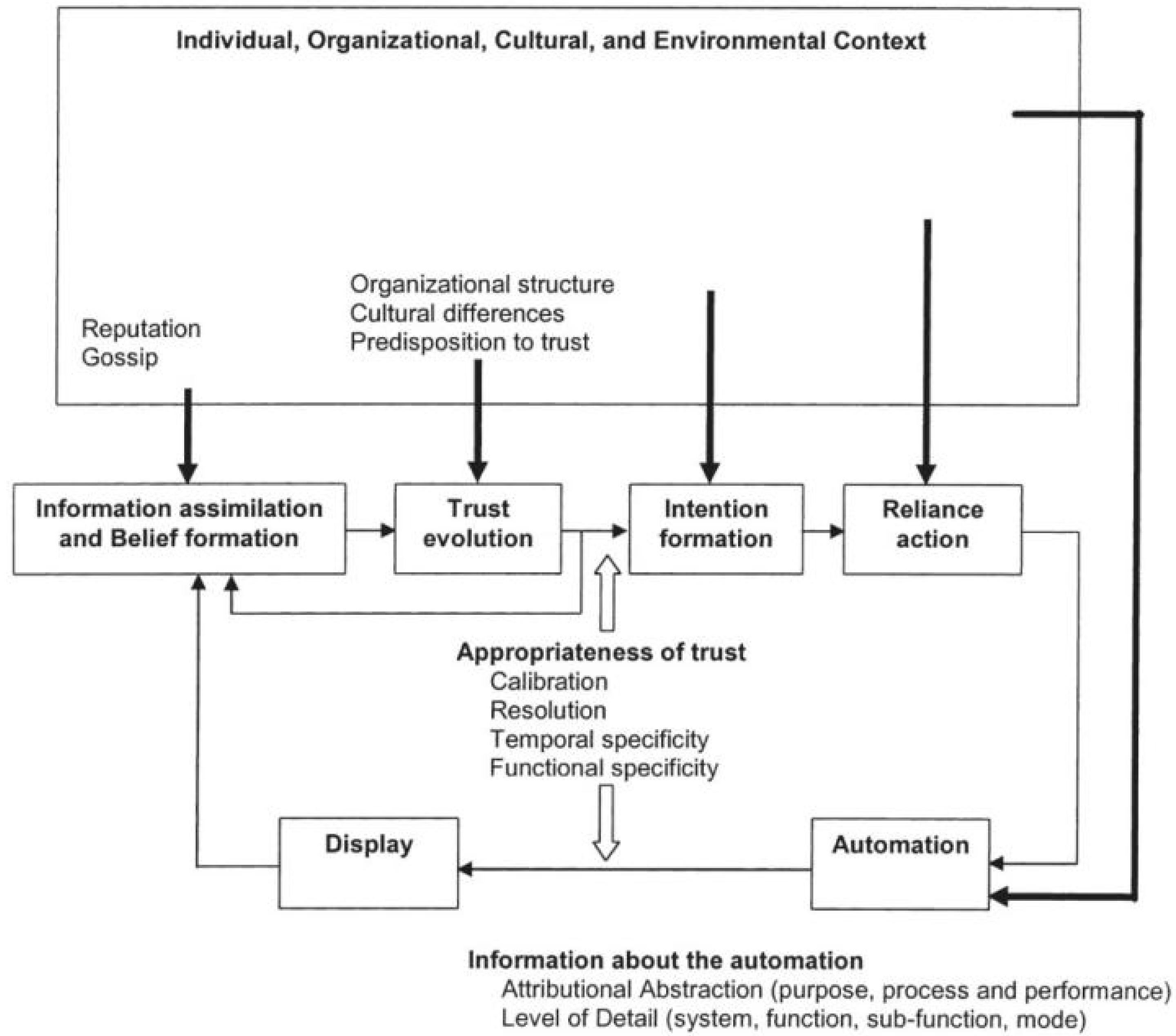
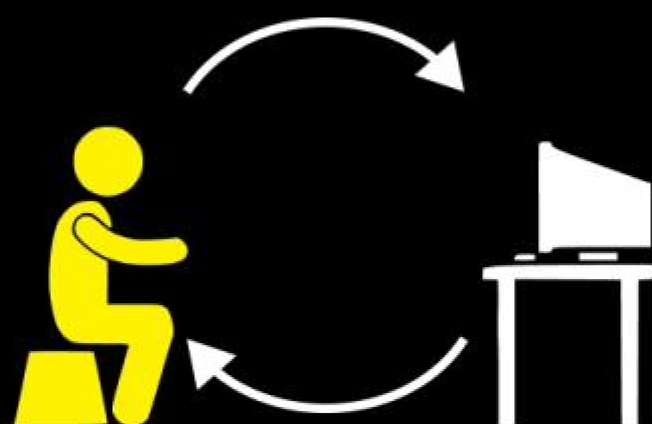


Figure 1. The interaction of context, agent characteristics, and cognitive properties with the appropriateness of trust.

Human-Machine System



Lee & See 2004

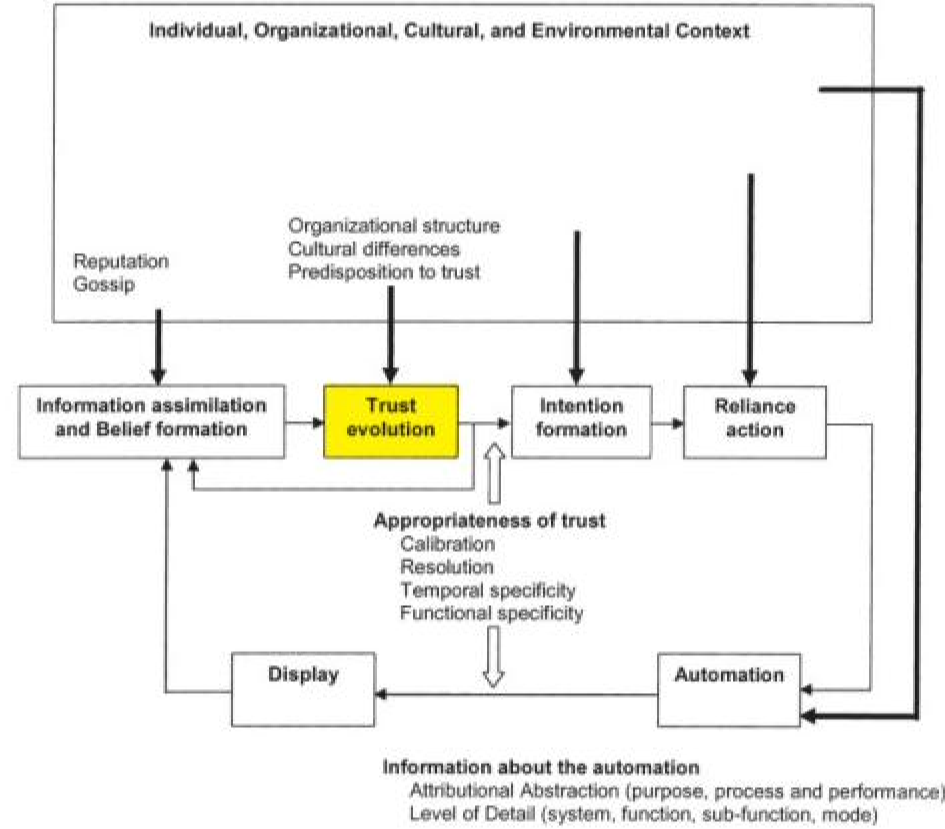


Figure 1. The interaction of context, agent characteristics, and cognitive properties with the appropriateness of trust.

Background

Hoff & Bashir 2015

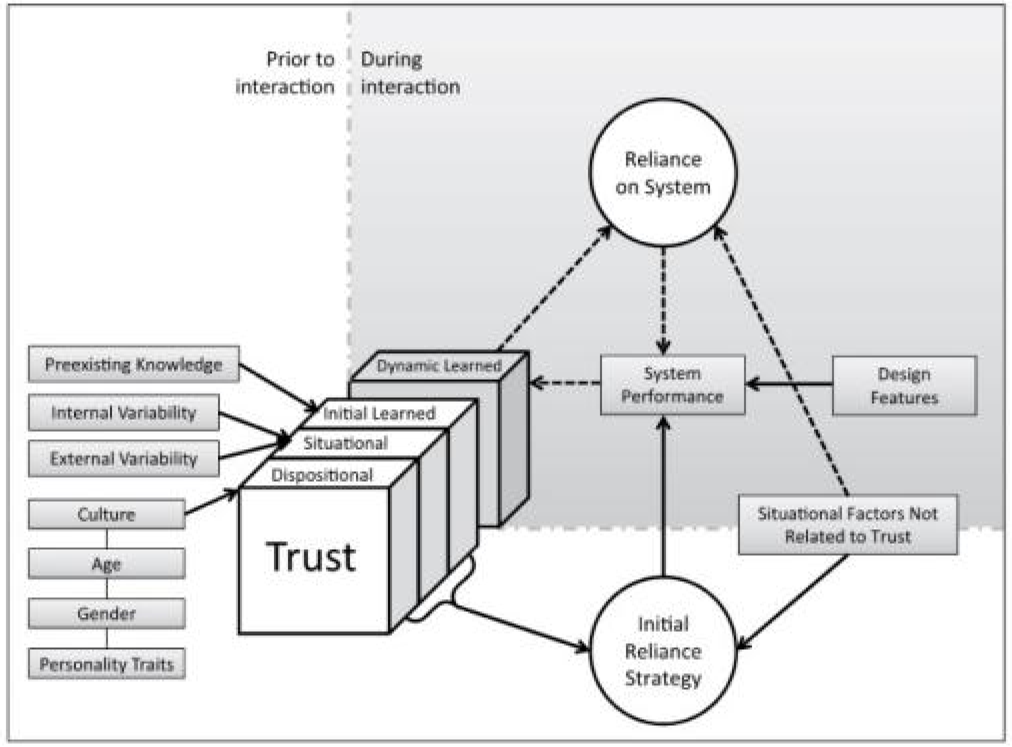
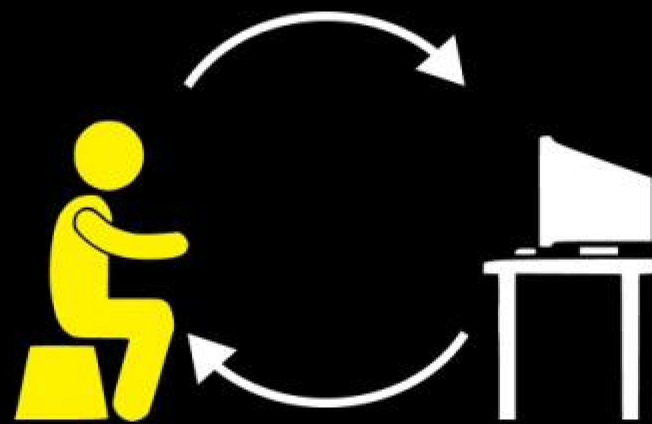


Figure 6. Full model of factors that influence trust in automation. The dotted arrows represent factors that can change within the course of a single interaction.

Human-Machine System



Background

Hoff & Bashir 2015

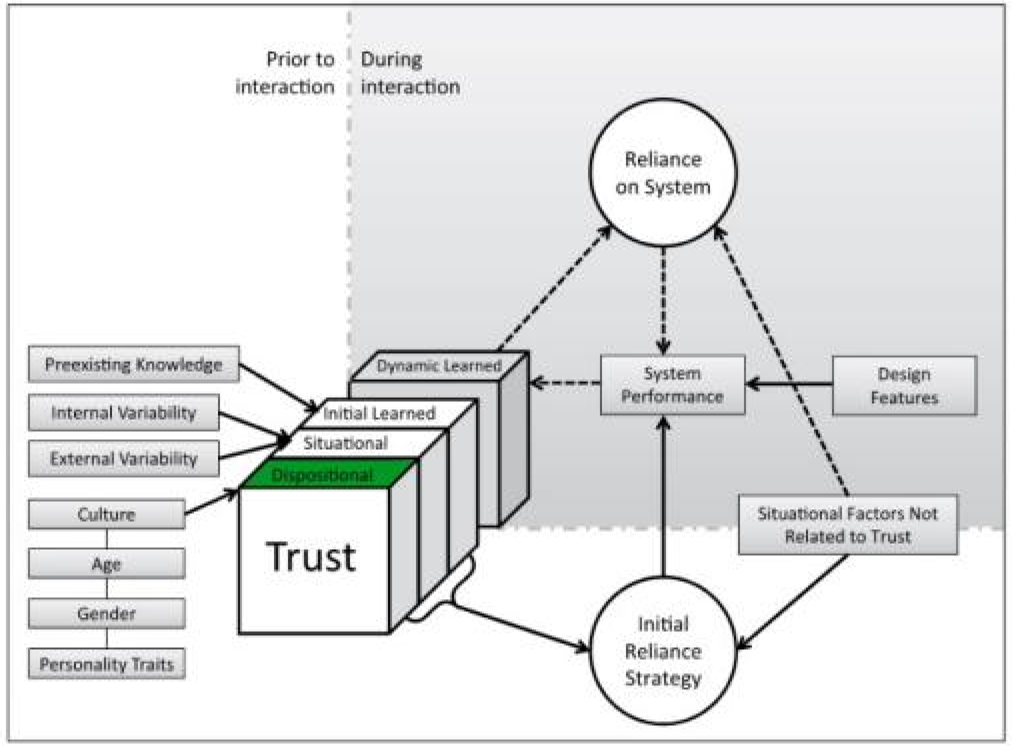
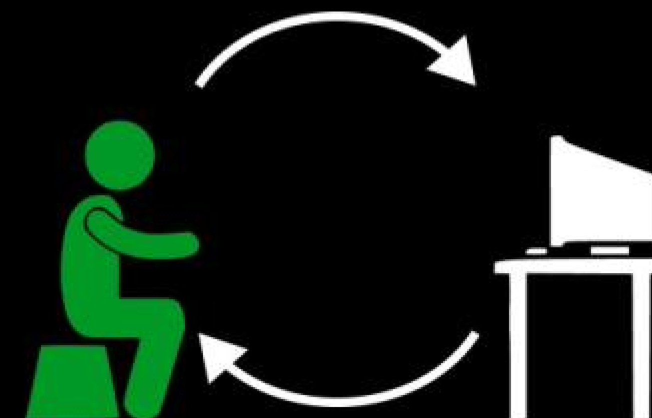


Figure 6. Full model of factors that influence trust in automation. The dotted arrows represent factors that can change within the course of a single interaction.

Human-Machine System



Background

Hoff & Bashir 2015

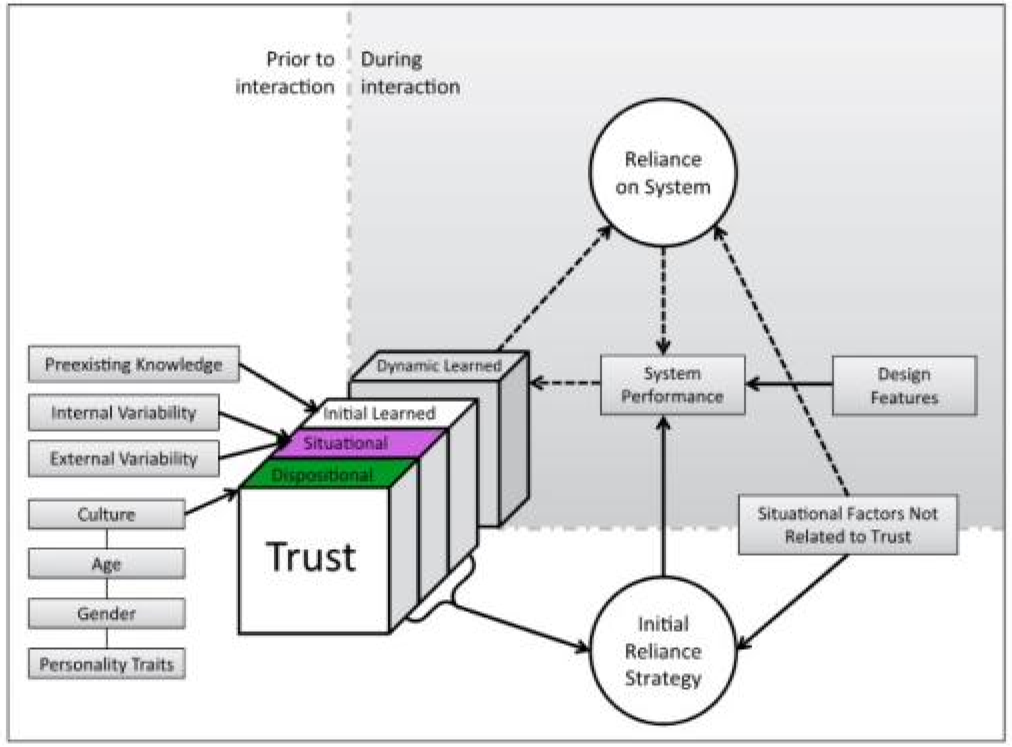
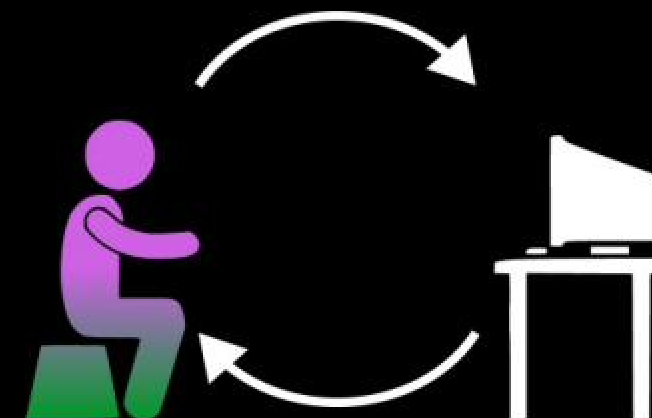


Figure 6. Full model of factors that influence trust in automation. The dotted arrows represent factors that can change within the course of a single interaction.

Human-Machine System



Background

Hoff & Bashir 2015

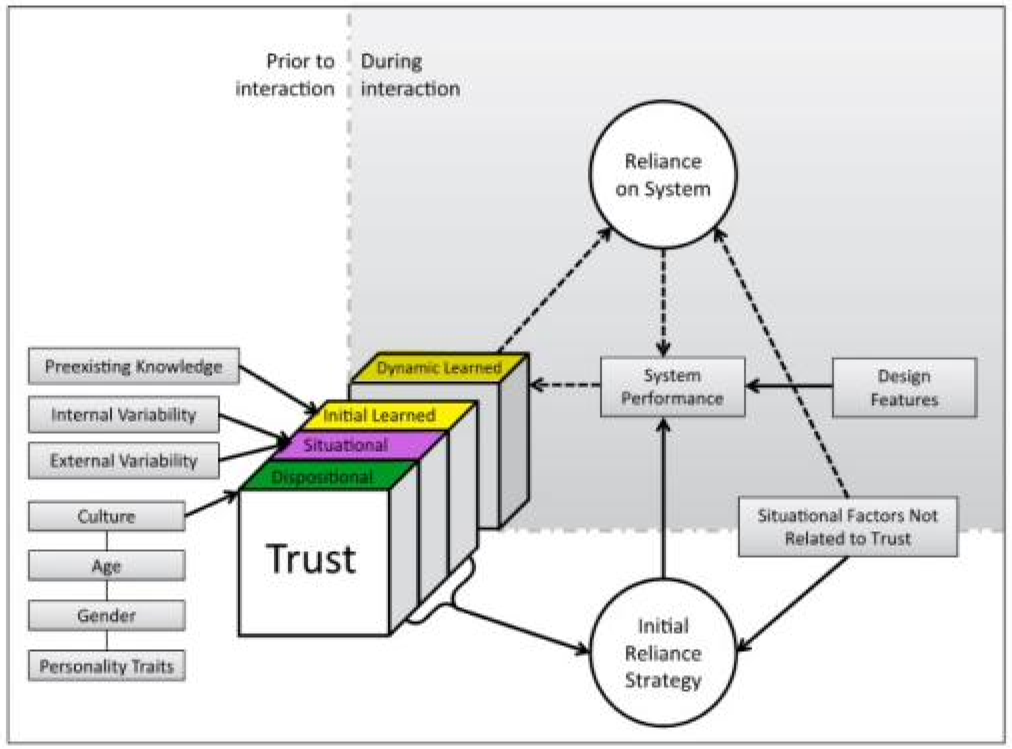


Figure 6. Full model of factors that influence trust in automation. The dotted arrows represent factors that can change within the course of a single interaction.

Human-Machine System

Background

Hoff & Bashir 2015

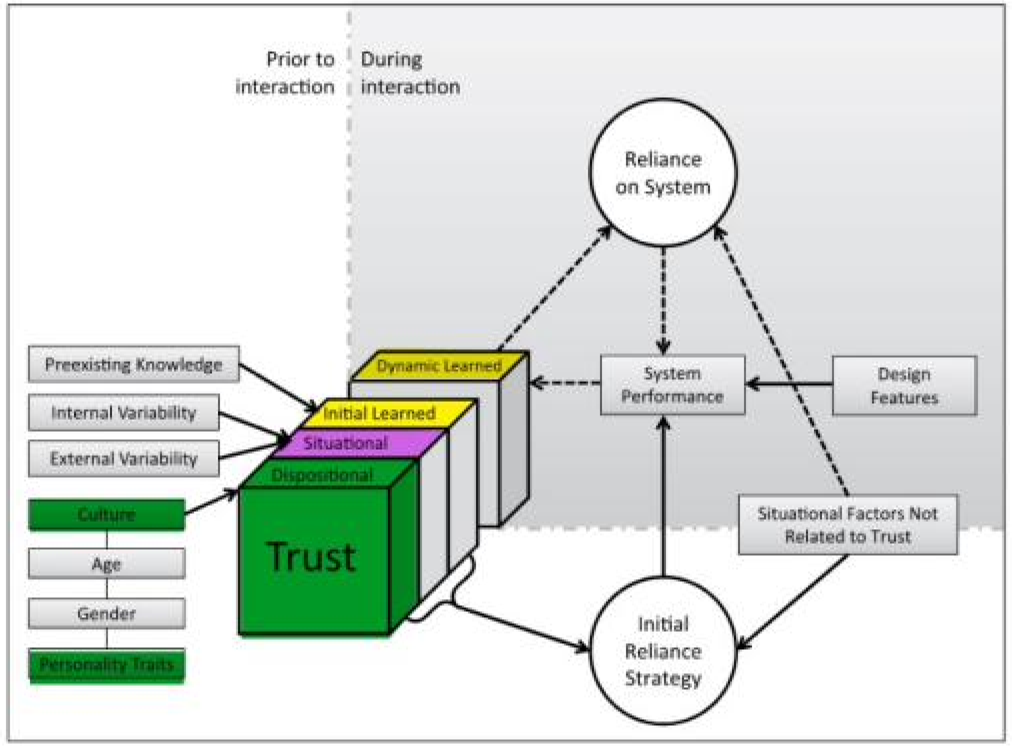


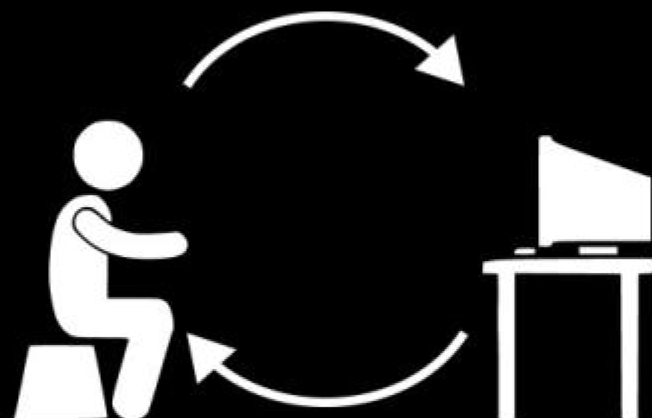
Figure 6. Full model of factors that influence trust in automation. The dotted arrows represent factors that can change within the course of a single interaction.

Human-Machine System

Background

Razin & Feigh, 2024

Human-Machine System



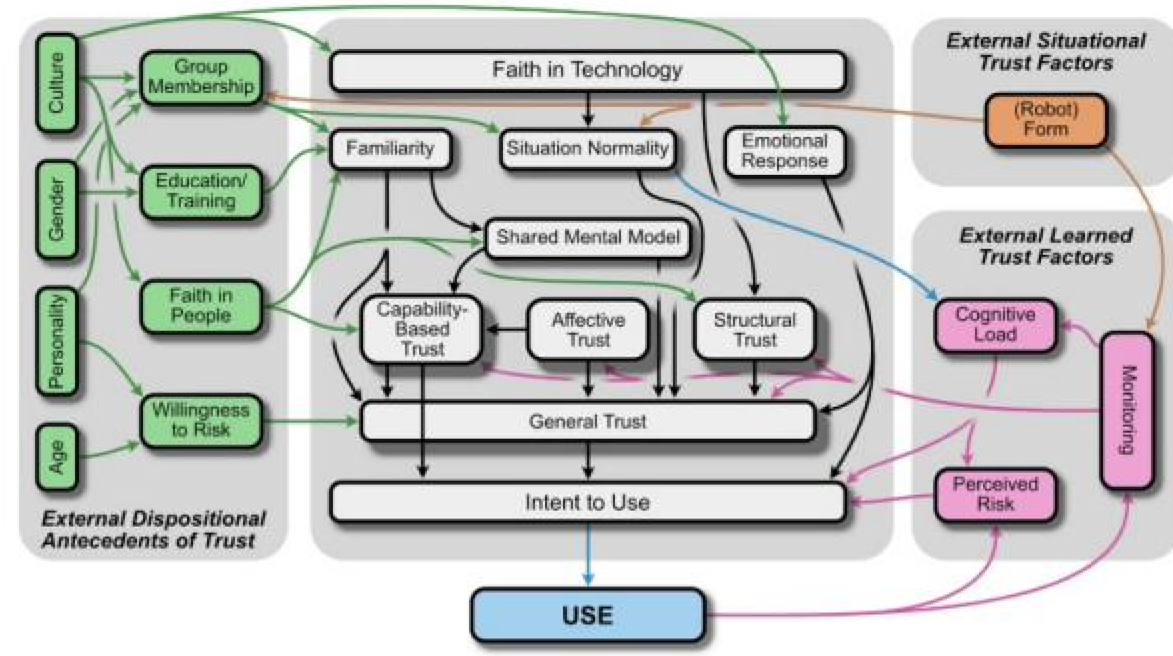
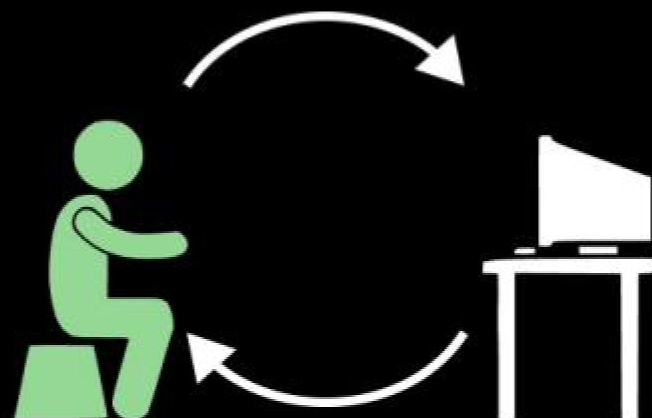


Fig. 5. The emergent trust model in relation to significant external factors.

Background

Razin & Feigh, 2024

Human-Machine System



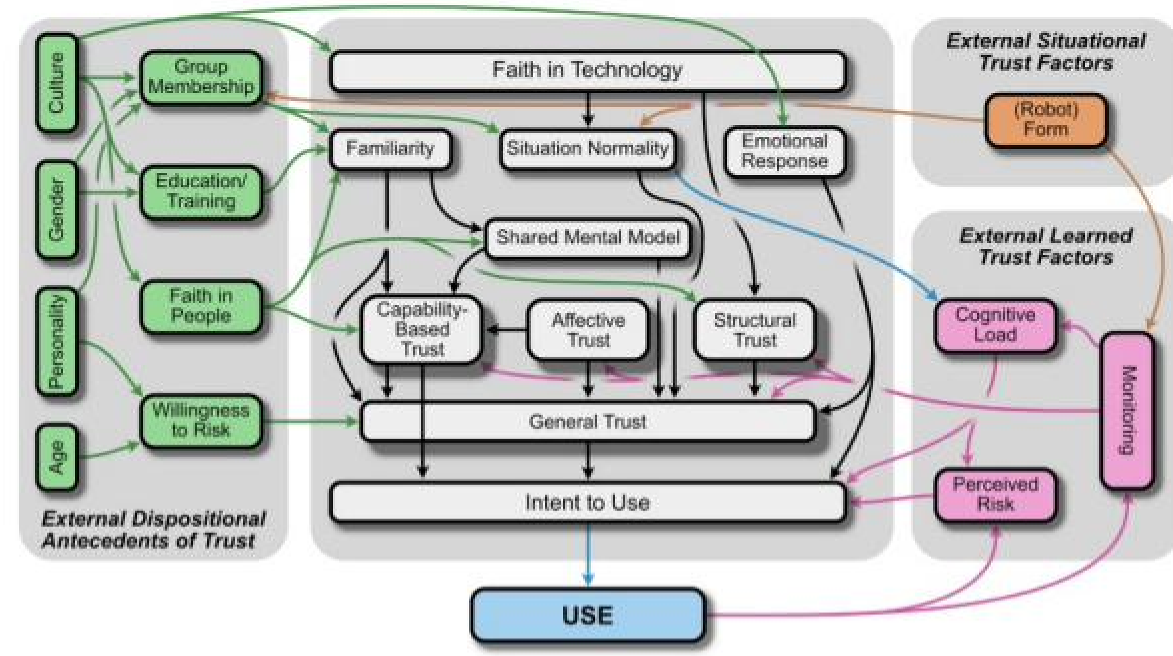
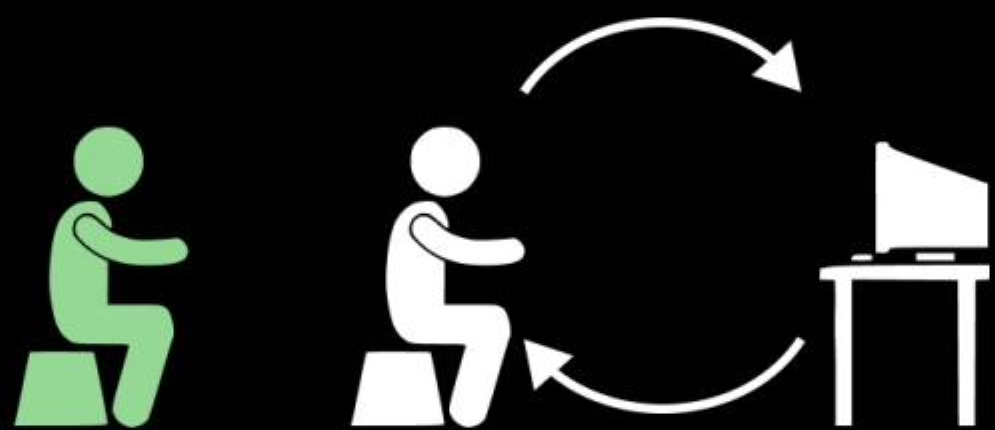


Fig. 5. The emergent trust model in relation to significant external factors.

Background

Razin & Feigh, 2024

Human-Machine System



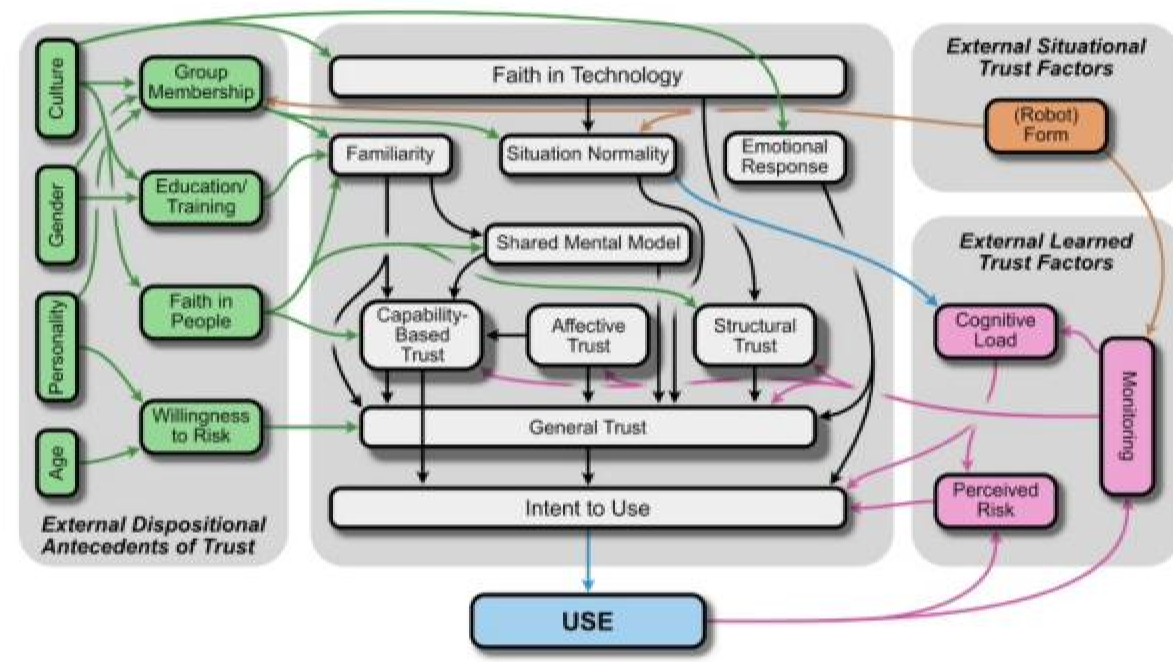
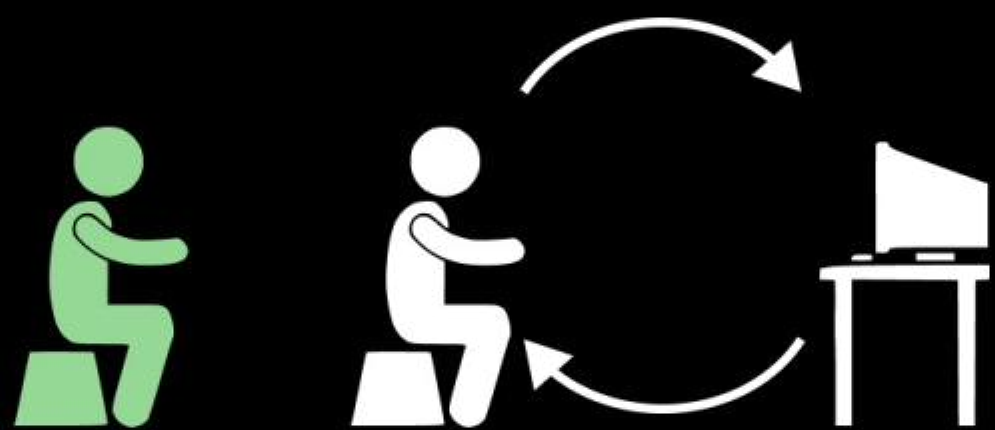


Fig. 5. The emergent trust model in relation to significant external factors.

Background

Razin & Feigh, 2024

Human-Machine System



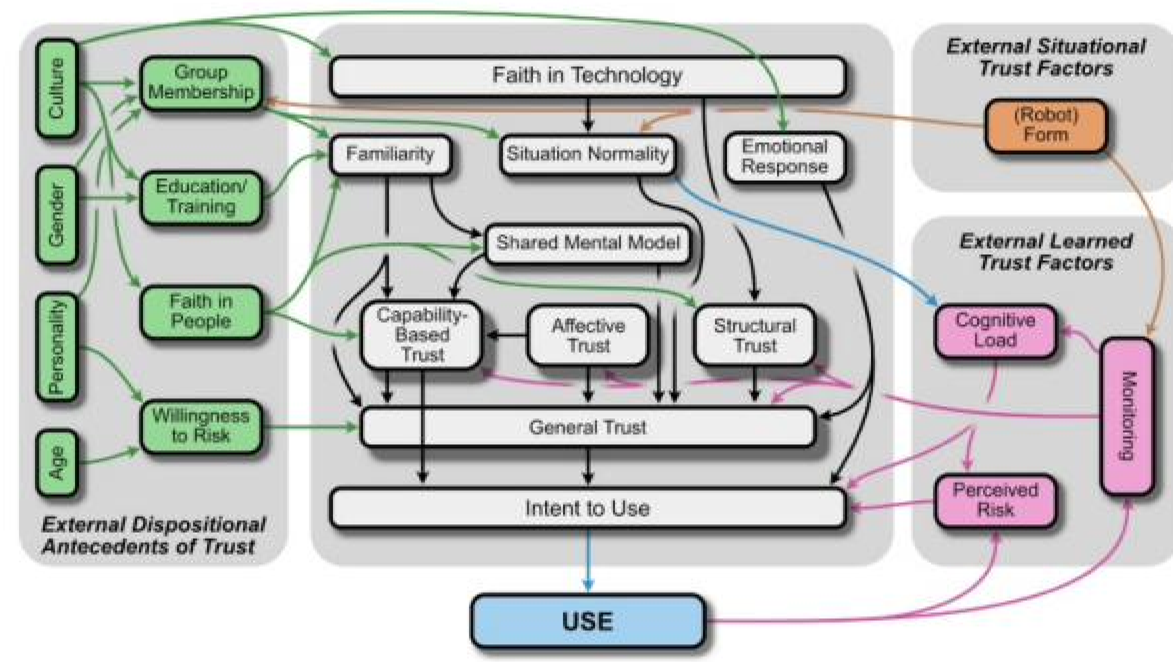


Fig. 5. The emergent trust model in relation to significant external factors.

Background

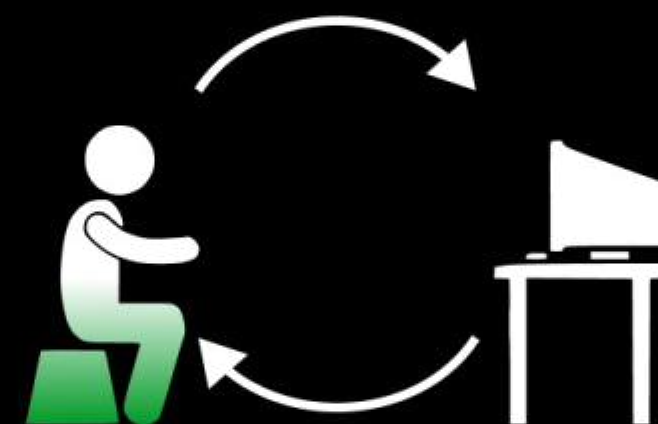
Research Gap: Dispositional Factors

Dispositional Trust: CVs & Personality

operator traits

personality

cultural values



"Who"

Background

Lee & See 2004

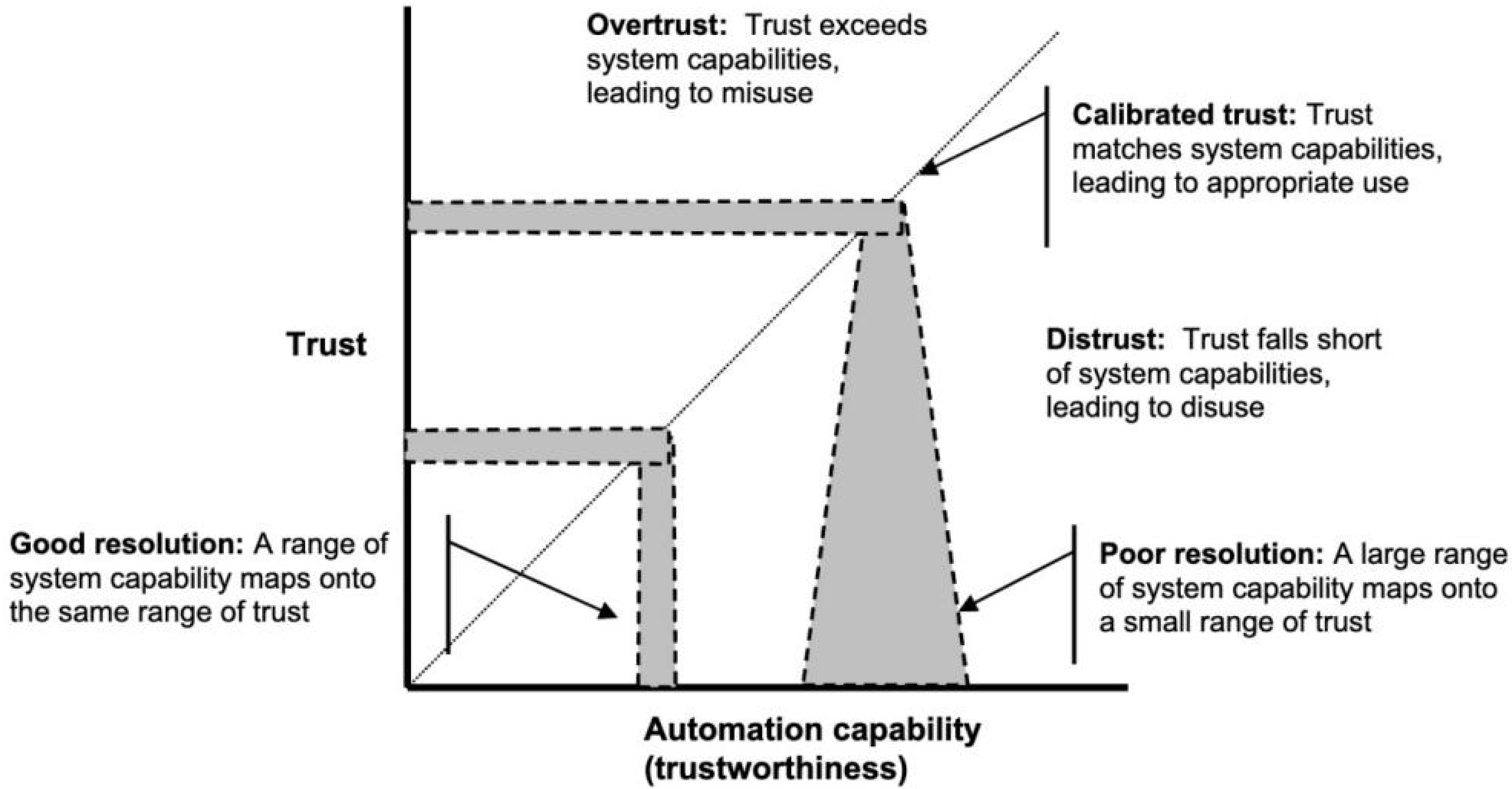


Figure 2. The relationship among calibration, resolution, and automation capability in defining appropriate trust in automation. Overtrust may lead to misuse and distrust may lead to disuse.

Background

Parasuraman and Riley, 1997

Ratio Quantity of Calibrated Use
(Interpreted)

$$R:C = \begin{cases} \text{misuse, if} > 1 \\ \text{disuse, if} < 1 \\ \text{calibrated use, if} = 1 \end{cases}$$

Background

Research Gap: Quantifying Calibrated Use

Ratio Quantity of Calibrated Use
(Interpreted)

$$R:C = \begin{cases} \text{misuse, if} > 1 \\ \text{disuse, if} < 1 \\ \text{calibrated use, if} = 1 \end{cases}$$

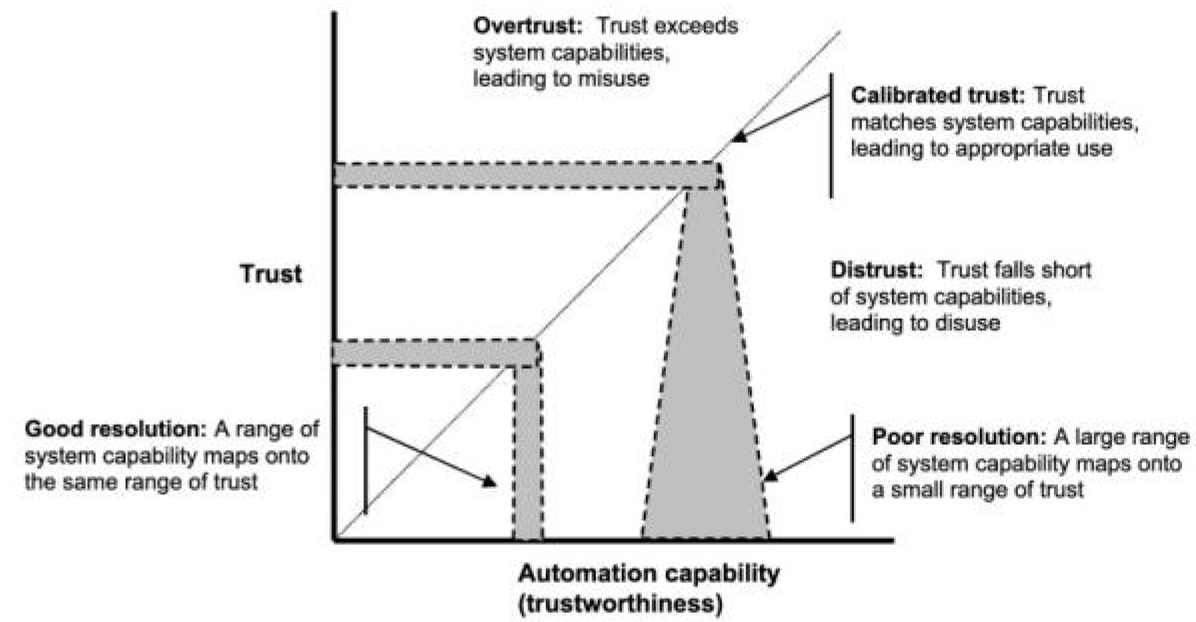


Figure 2. The relationship among calibration, resolution, and automation capability in defining appropriate trust in automation. Overtrust may lead to misuse and distrust may lead to disuse.

Method

Research Questions

1. How do dispositional traits and cultural values influence trust and reliance on an adaptable, automated system?

Operator Traits

Personality

Cultural Values

2. Can we quantify the alignment of an operator's trust and reliance to an adaptable system's actual capability?

Ratio Quantity of Reliance to Capability

$$R:C = \begin{cases} \text{misuse, if } >1 \\ \text{disuse, if } <1 \\ \text{calibrated use, if } = 1 \end{cases}$$

Method
Study Design

1. Operator Trait Measures + Baseline Trust (Qualtrics)
2. Trust and Reliance Metrics (Unity-based delegation task)

PREDICTOR	CONSTRUCT	SCALES
Personality Traits	Measures of Dispositional Trust Propensities	Propensity to Trust (Frazier, 2013) Faith in General Technology (McKnight, 2011) General Propensity to Trust Machines (McKnight, 2011)
Cultural Values	Hofstede's Cultural Dimensions - Individual Level	CVscale (Yoo et al., 2011)
Baseline Trust in Otto	Baseline Trust	Adapted from Functionality and Reliability Subscales of General Trusting Stance (McKnight, 2011)

Method

Study Design

Trait Measures: Examples

Please rate your level of agreement with each of the following statements.

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
People in higher positions should not ask the opinions of people in lower positions too frequently.	○	○	○	○	○
People in higher positions should avoid social interaction with people in lower positions.	○	○	○	○	○
People in lower positions should not disagree with decisions by people in higher positions.	○	○	○	○	○

Please rate your level of agreement with each of the following statements.

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
It is important to closely follow instructions and procedures.	○	○	○	○	○
Rules and regulations are important because they inform me of what is expected of me.	○	○	○	○	○
Instructions for operations are important.	○	○	○	○	○

Please rate your level of agreement with each of the following statements.

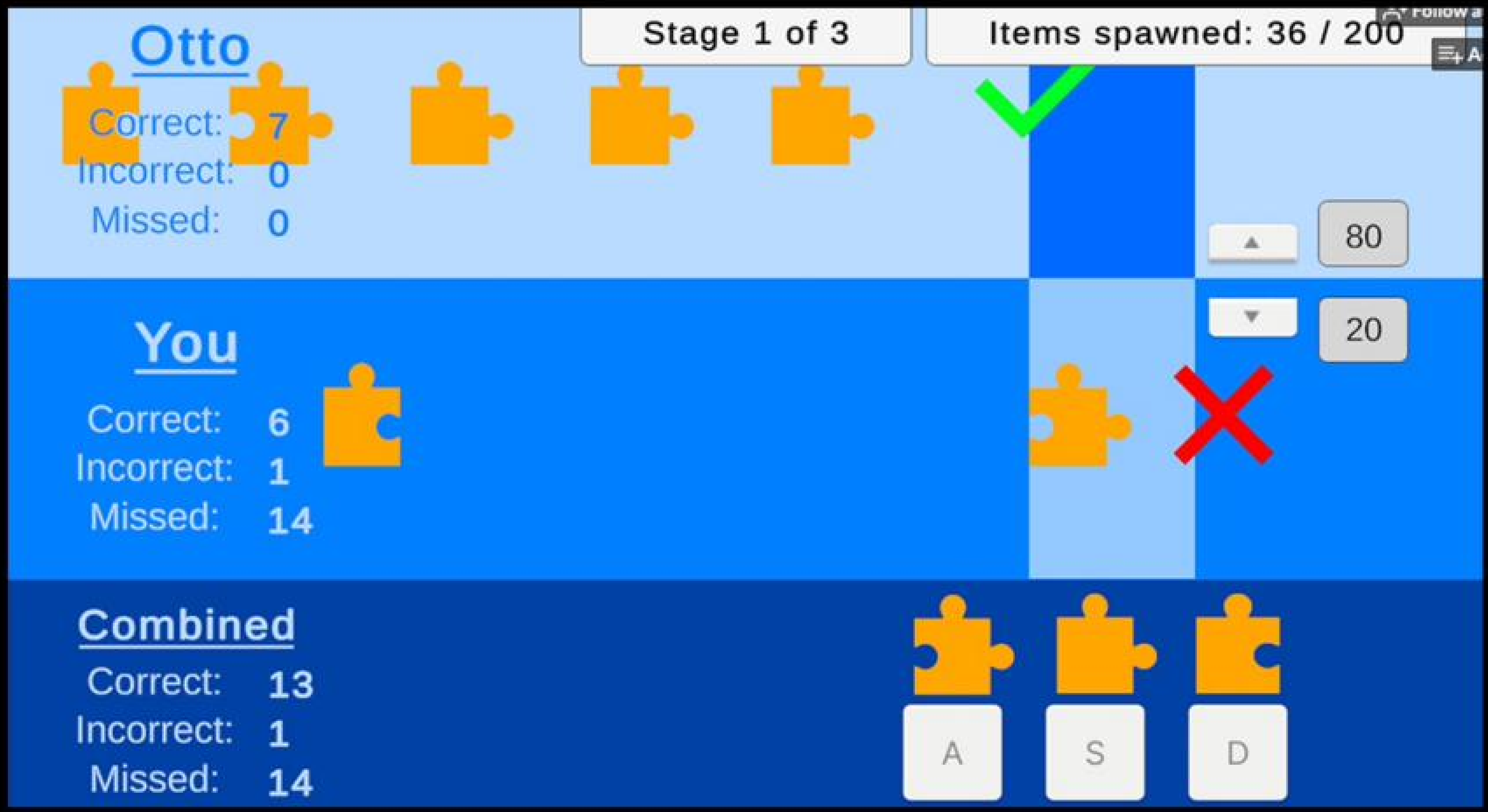
	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
It is more important for men to have a professional career than it is for women.	○	○	○	○	○
Men usually solve problems with logical analysis; women usually solve problems with intuition.	○	○	○	○	○
Solving difficult problems usually requires an active, forcible approach, which is typical of men.	○	○	○	○	○

Method

Trust and Reliance Metrics

Calibratio: A Collaborative Sorting Task with Otto the Adaptable System

Baseline followed by
3 Gameplay blocks of system capability (40, 60, 80%) counterbalanced across participants

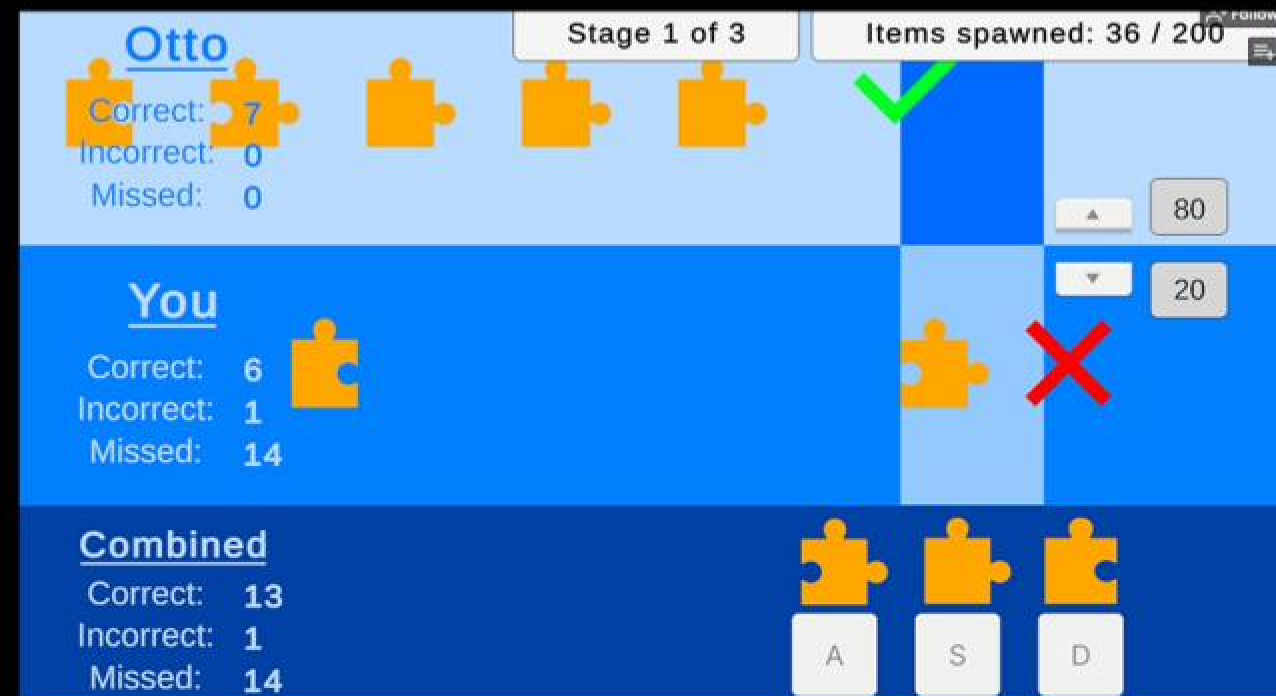


Method

Trust and Reliance Metrics

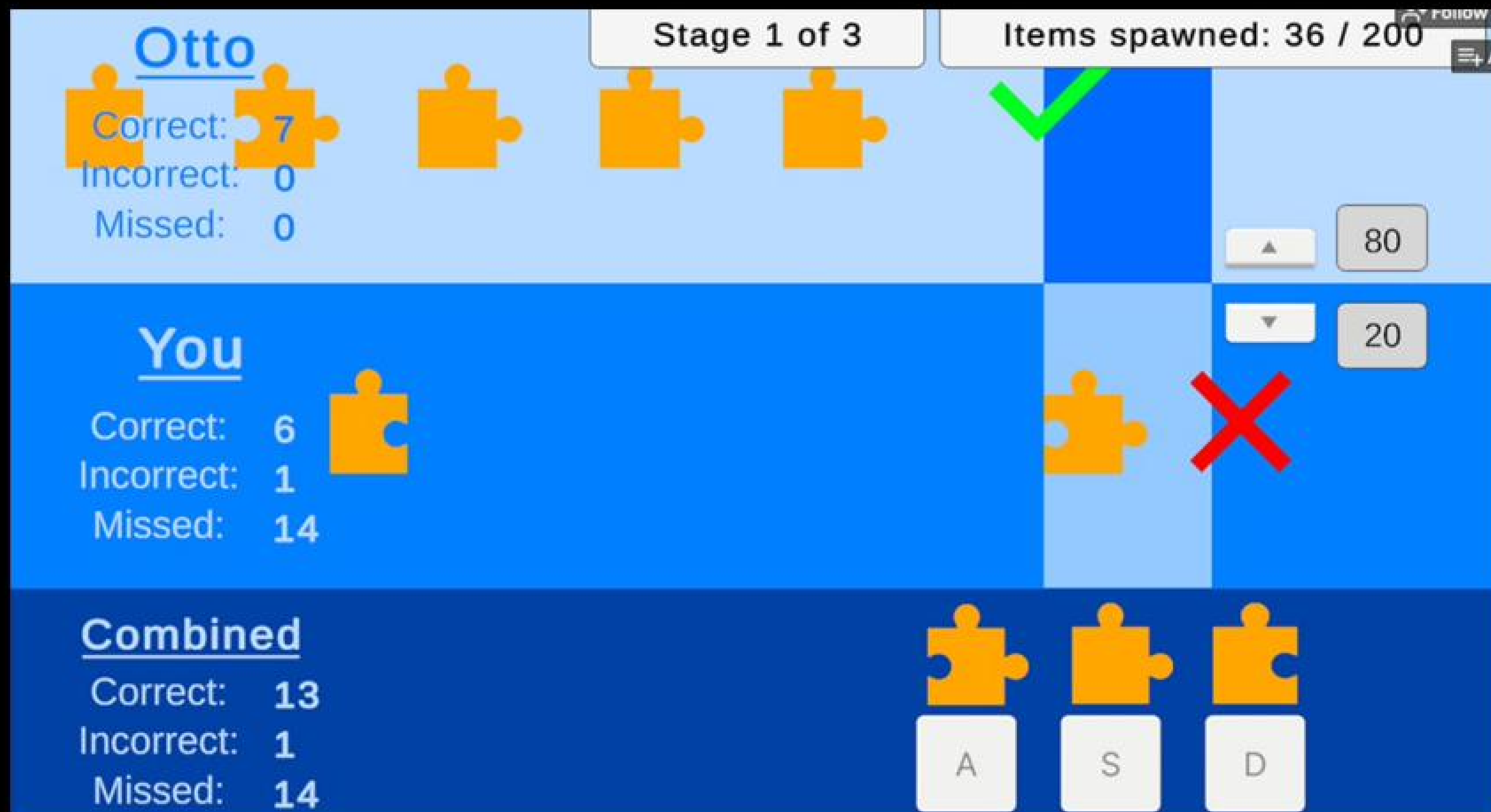
Calibratio: Metrics

1. In-Task Trust
2. Self-Assessment of Performance
3. Continuous Reliance on Otto (0-100, every 0.5s)



Method
Trust and Reliance Metrics

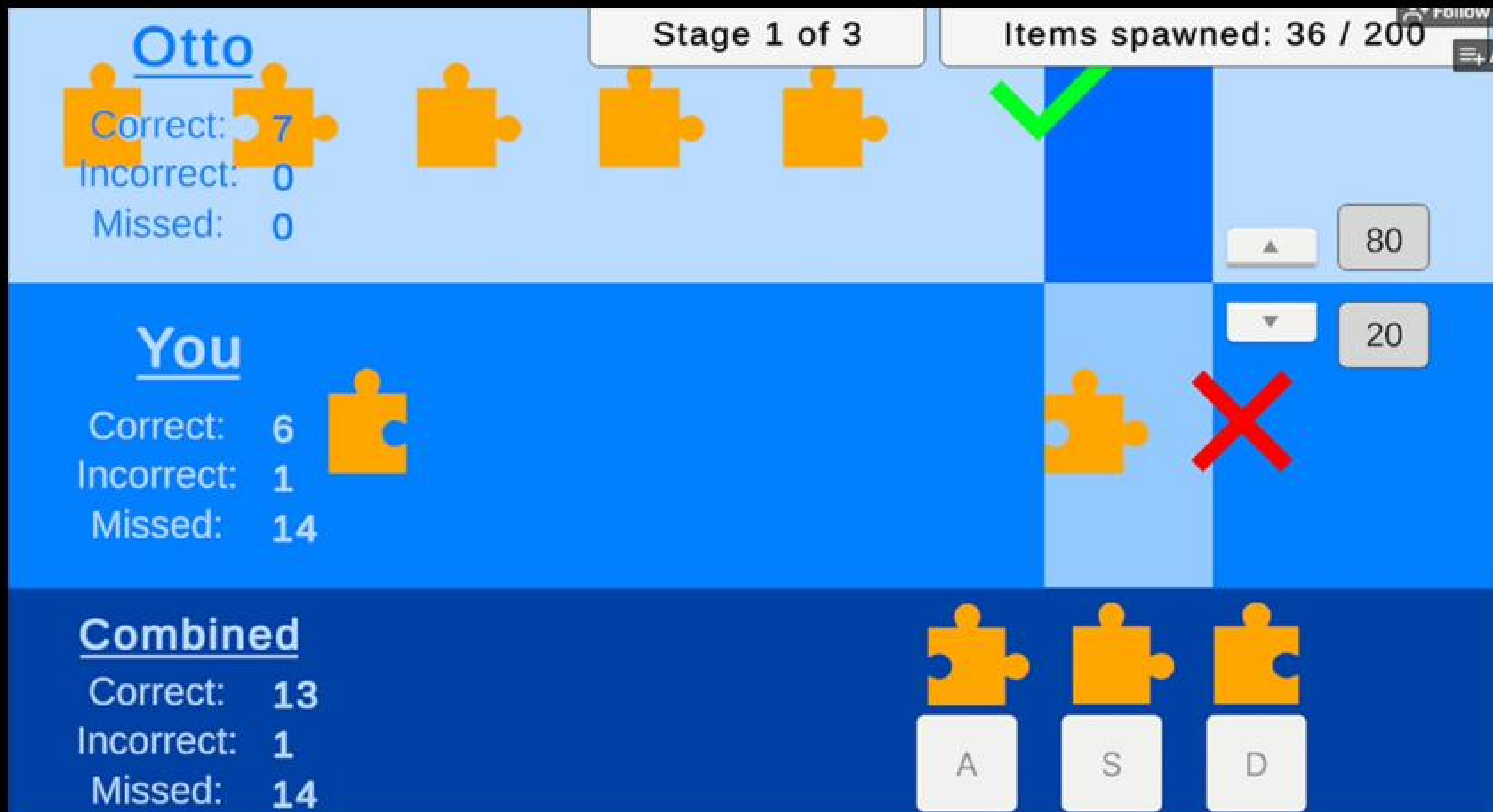
Calibratio: Gameplay



Method

Trust and Reliance Metrics

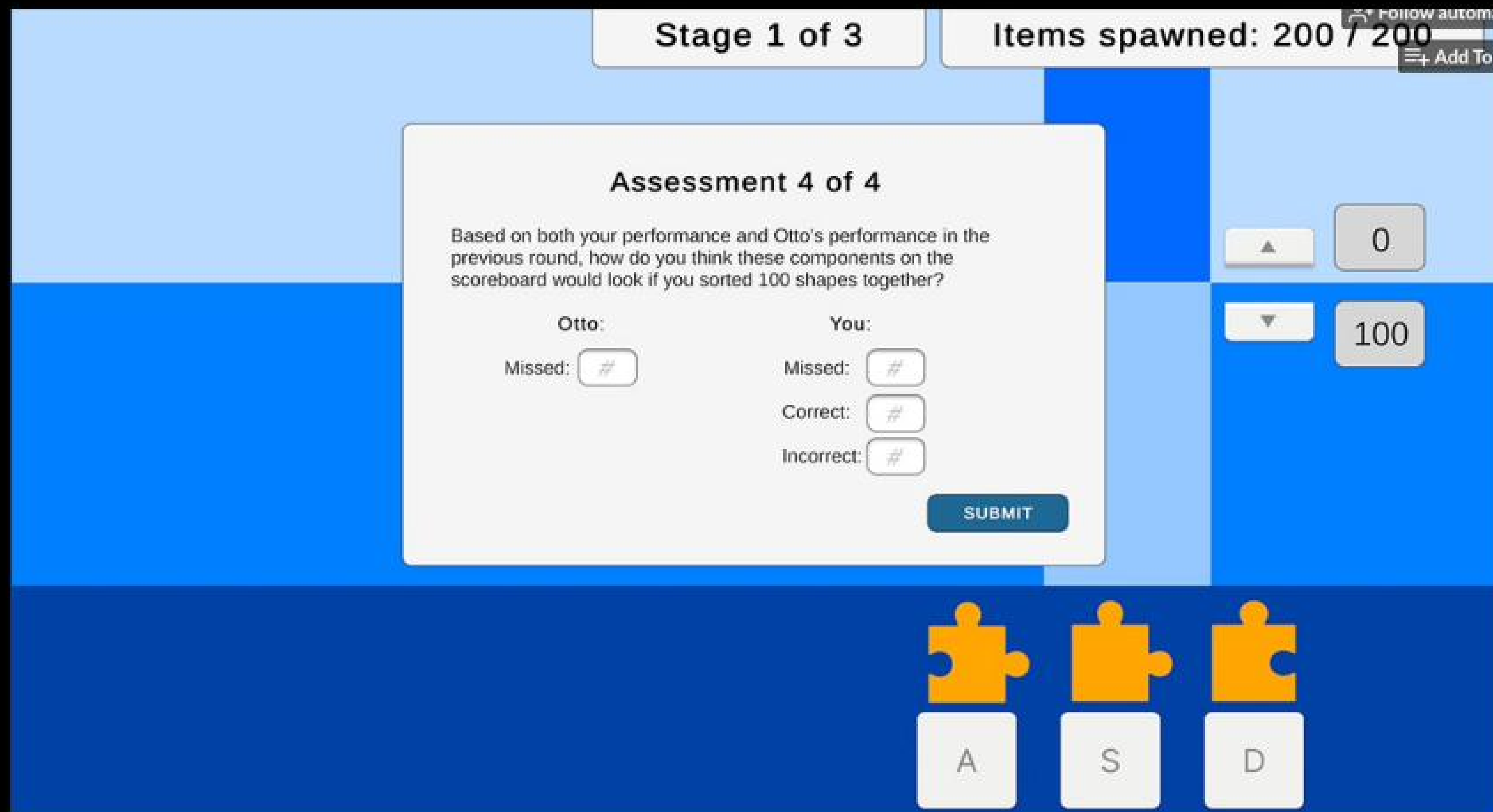
Calibratio: Gameplay



Method

Trust and Reliance Metrics

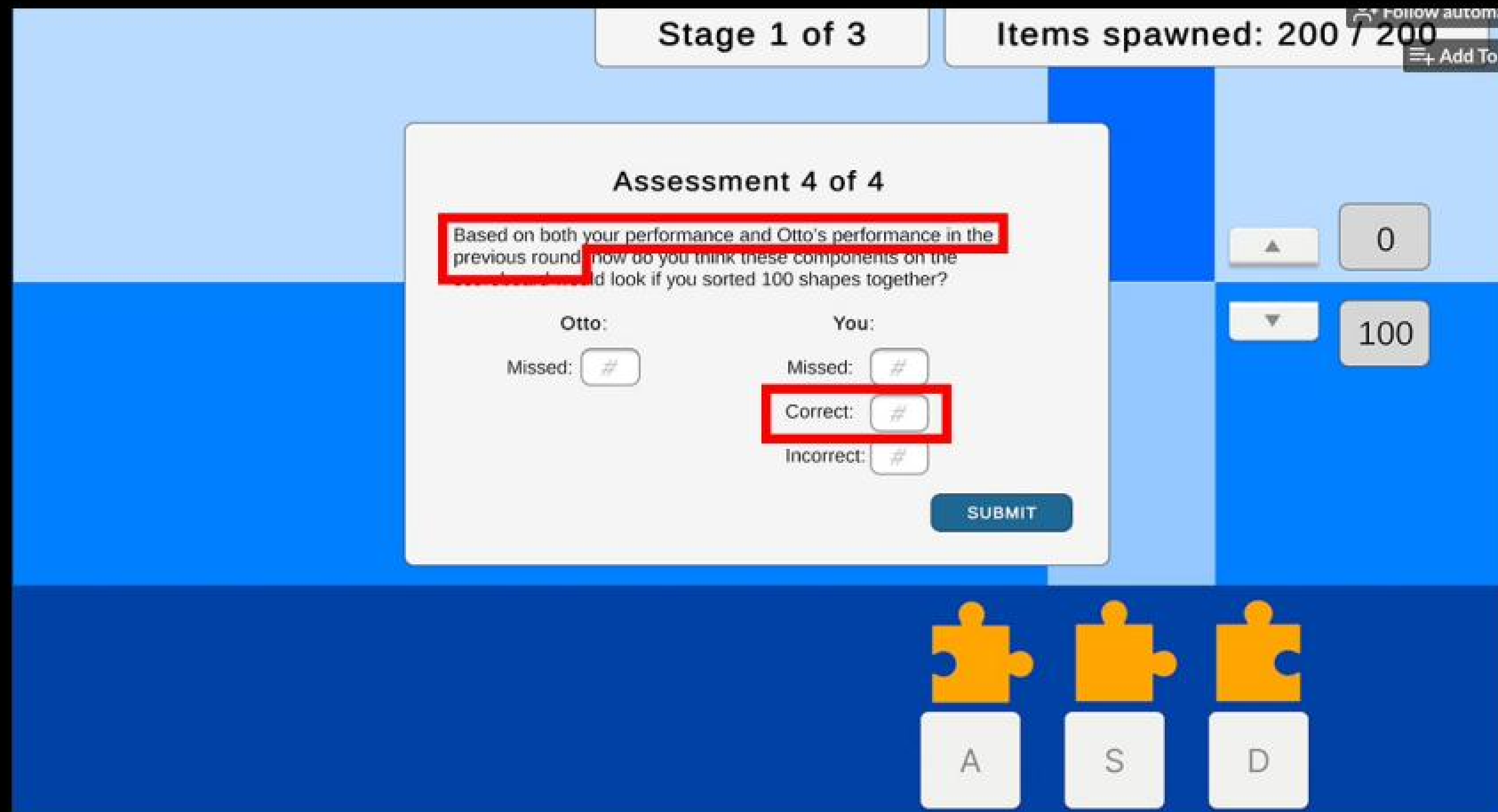
Calibratio: Self-Assessment of Performance



Method

Trust and Reliance Metrics

Calibratio: Self-Assessment of Performance



Analysis

Overview

1. Structural Equation Modeling (SEM) - Path Model

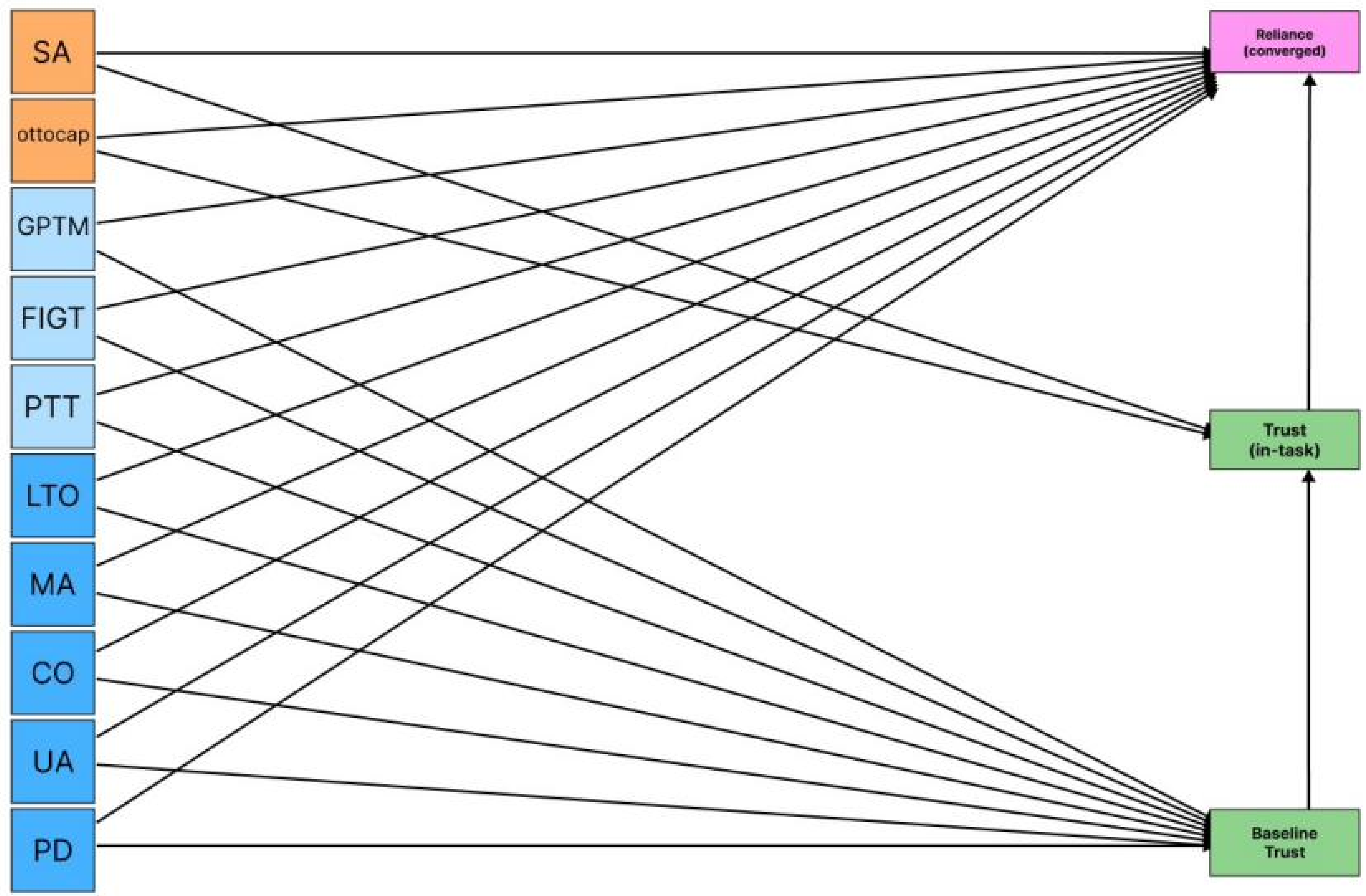
2. Exploratory Analysis (%TIR)

a priori power analysis

G*Power small-to-med effect

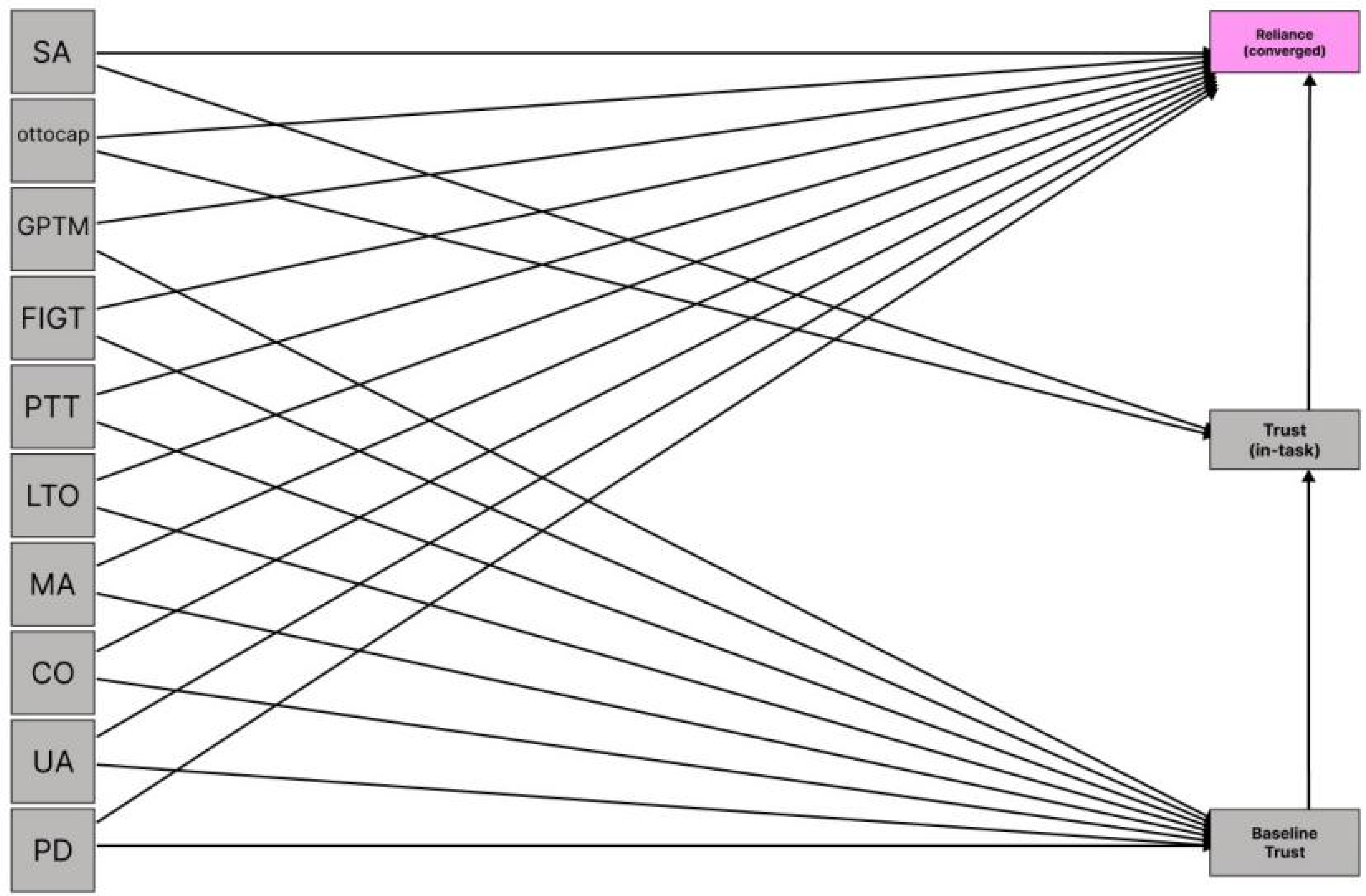
Analysis

Path Model



Analysis

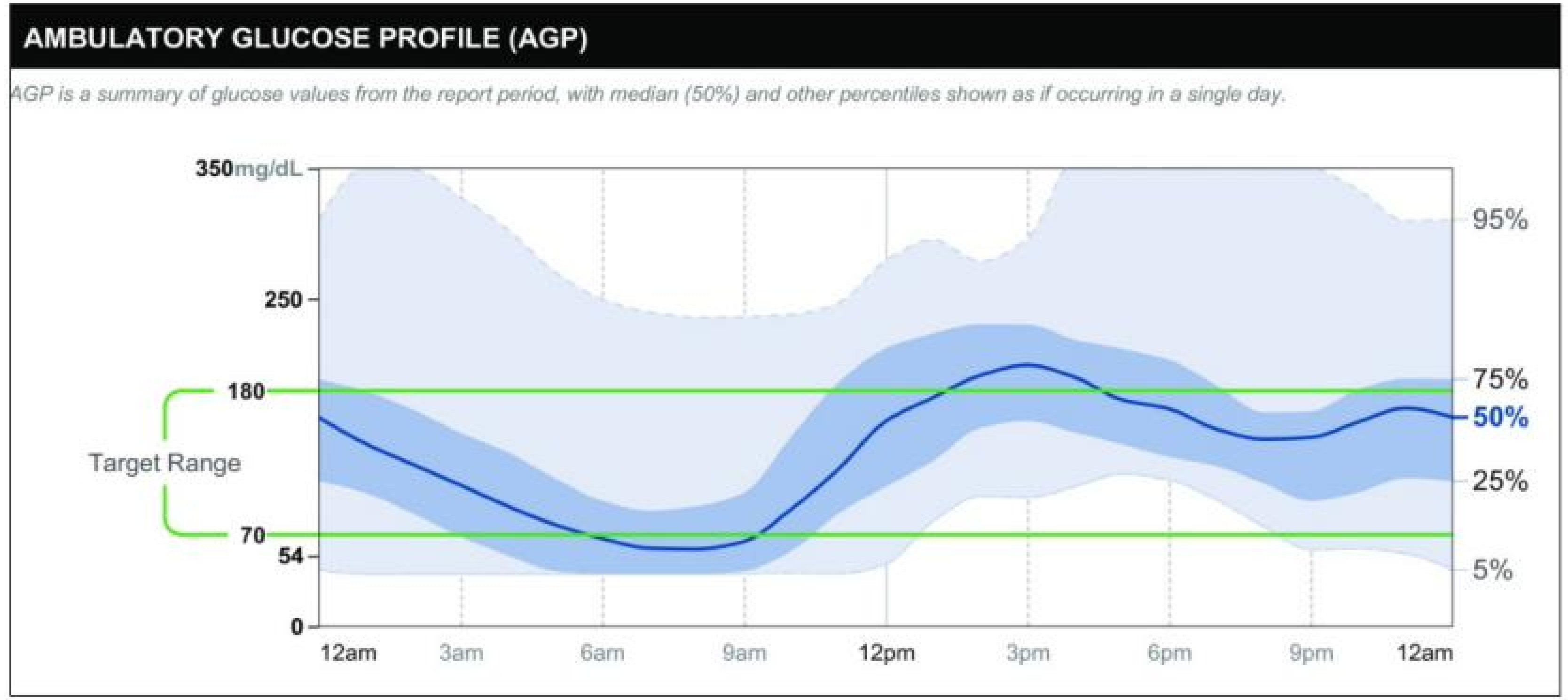
Path Model

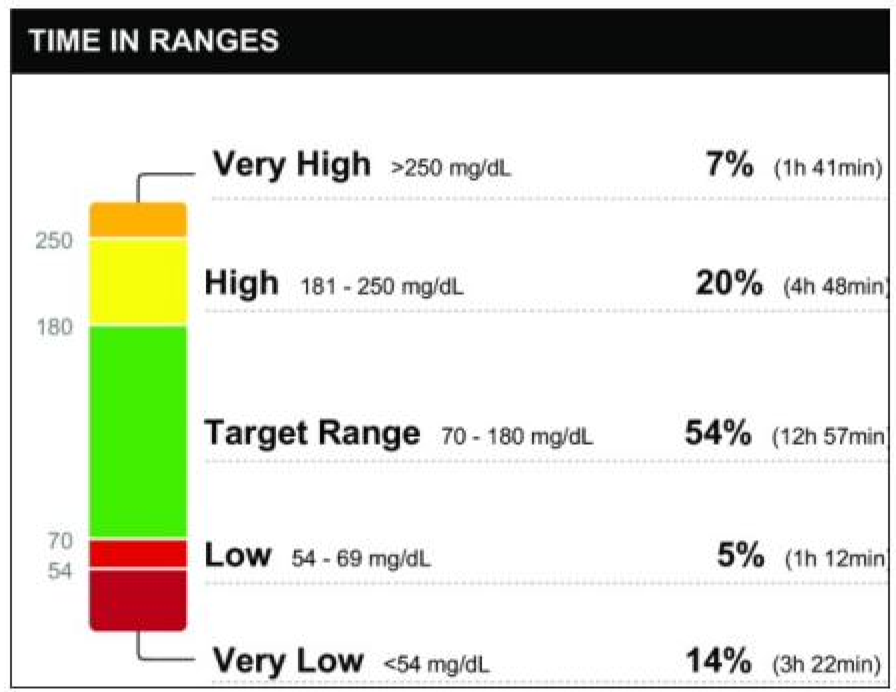


Pilot Trials:
Median value for
last 30s of game

Analysis

%TIR





Wright et al., 2020

Analysis

%TIR

Exploratory Analysis (UIR)

$$U = \frac{\sum_i \sum_j (C_{ij} \pm 10)}{\sum_i \sum_j T_{ij}}$$

$U_{ij} = \textit{proportion of pieces allocated to Otto within} \pm 10\% \textit{ of } C_{ij}$

T_{ij} = total pieces of stage j for user i

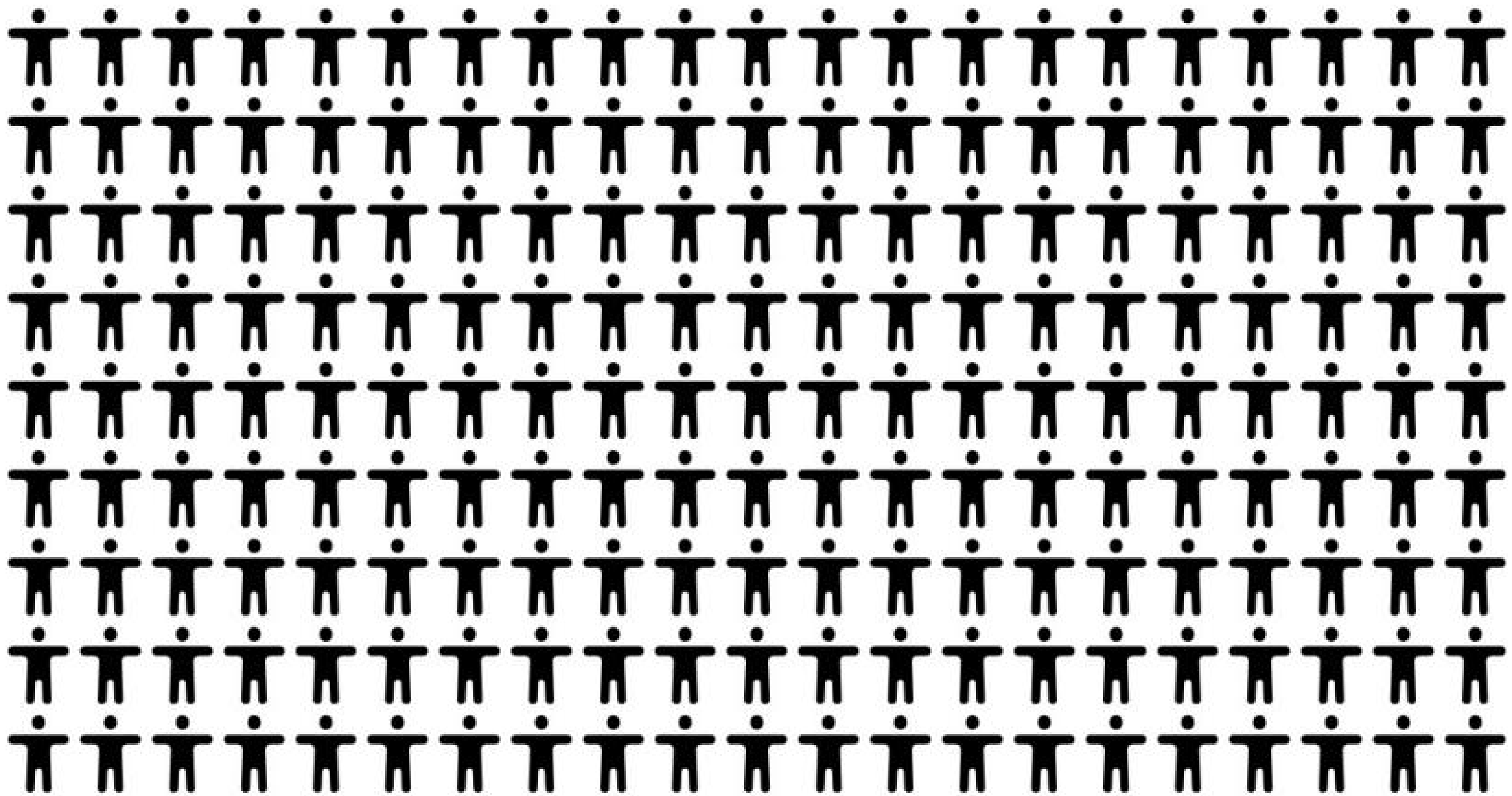
C_{ij} = Otto's capability for user i at stage j

Results

Overview

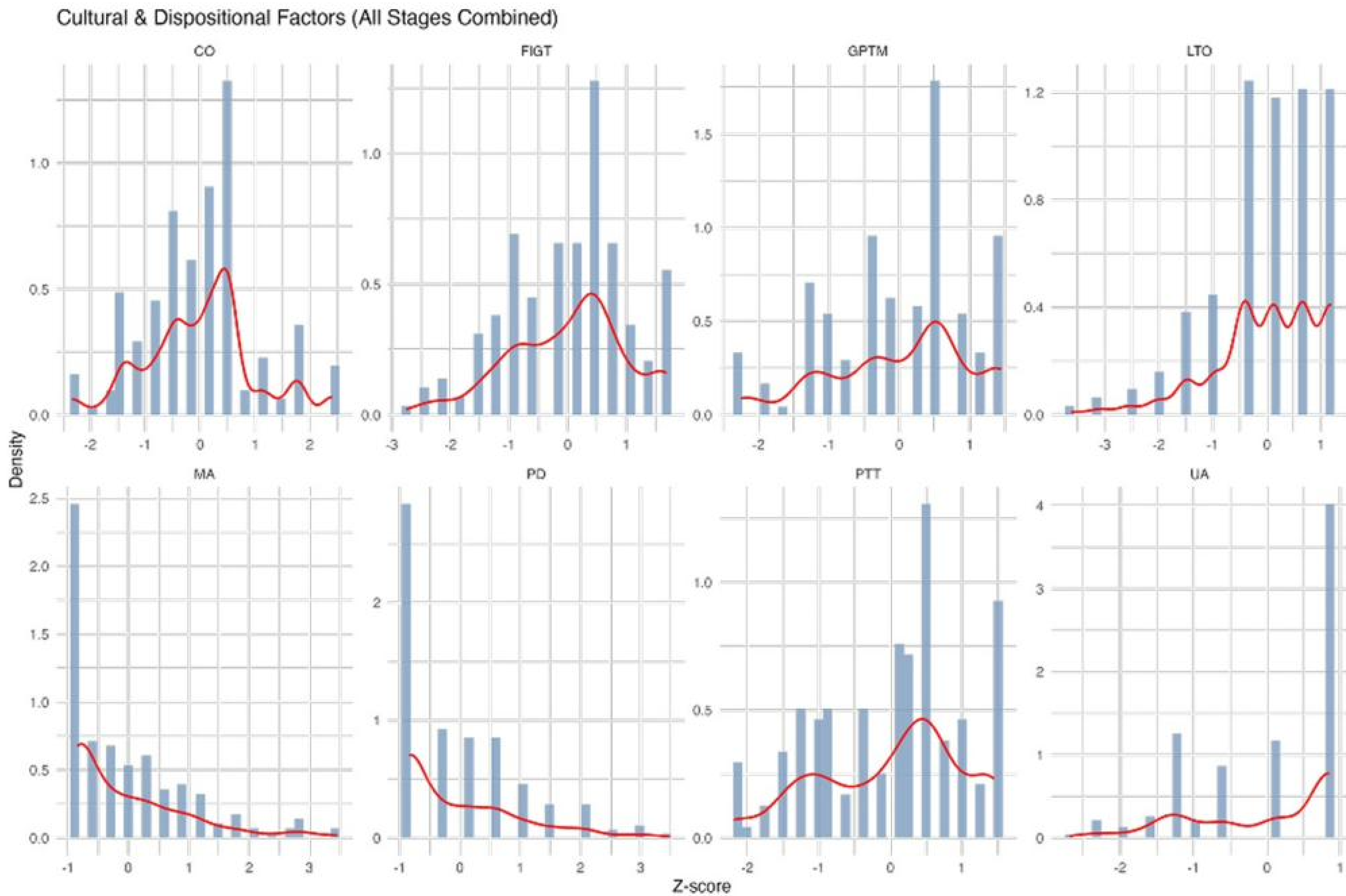
1. Structural Equation Modeling (SEM)

2. Exploratory Analysis (%TIR) - Descriptive Statistics



N = **189** N observations = **567**

Sample was not representative of all cultural values and traits - skewed distributions



Analysis

SEM Estimators

R (lavaan 0.6-19)

Global Fit Indices

- Robust Maximum Likelihood (MLR)

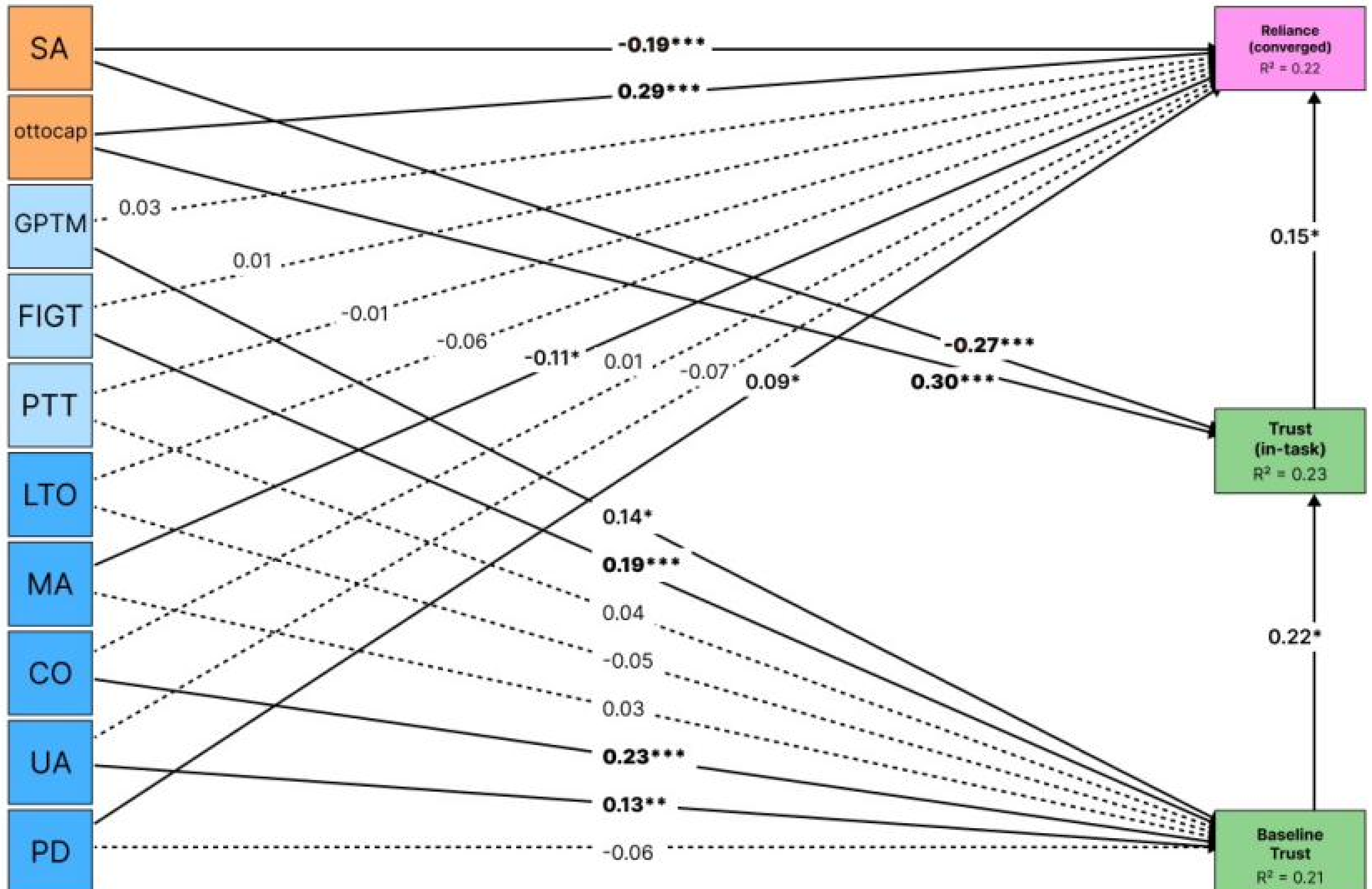
Parameter Inference

- Paths + Indirect Effects
- ML Bootstrap 95% CIs (1000 resamples)
- Skewed distributions

Model	χ^2 (p)	df	RMSEA	RMSEA 90% CI	SRMR	CFI	TLI	Overall
SEM	16.19 (p = .134)	11	0.029	0.000–0.057	0.017	0.985	0.954	Good
Interpretation			Good		Good	Good	Good	Good

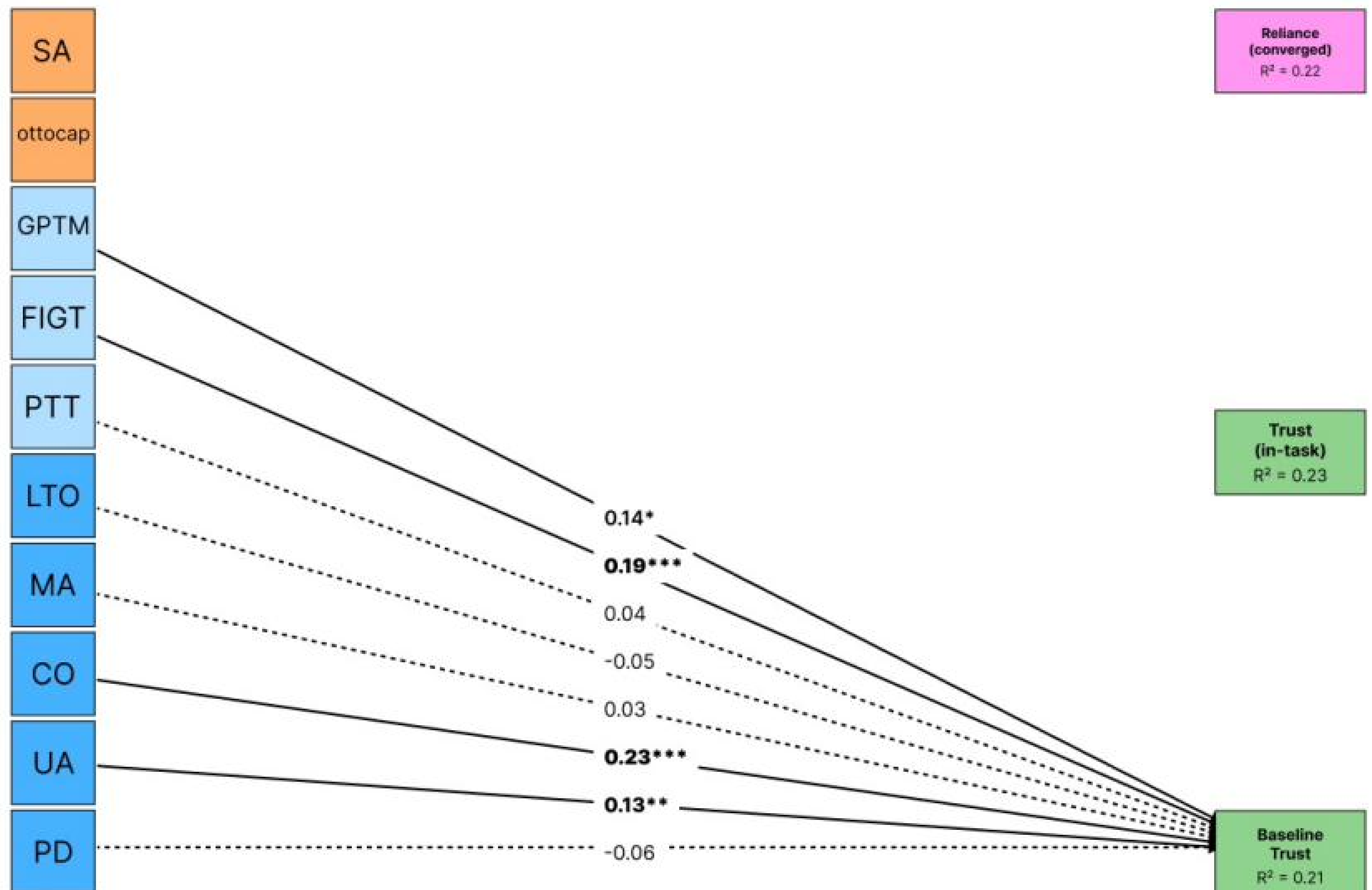
Analysis

Structural Equation Modeling

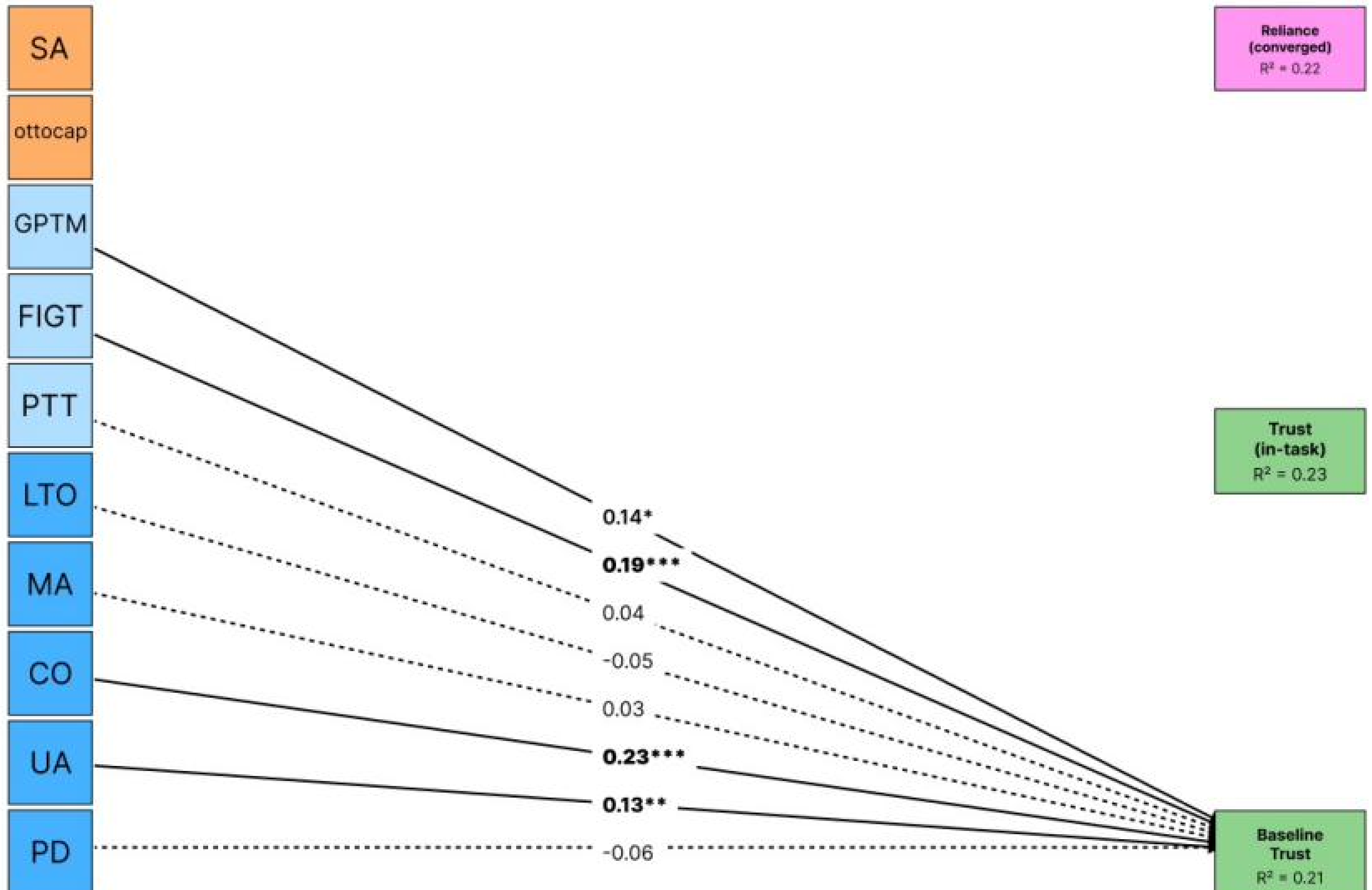


Path 1: Predictors of Baseline Trust

$$T_{comp_0} = \beta^1 \cdot PD + \beta^2 \cdot UA + \beta^3 \cdot CO + \beta^4 \cdot MA + \beta^5 \cdot LTO + \beta^6 \cdot PTT + \beta^7 \cdot FIGT + \beta^8 \cdot GPTM + \varepsilon^1$$

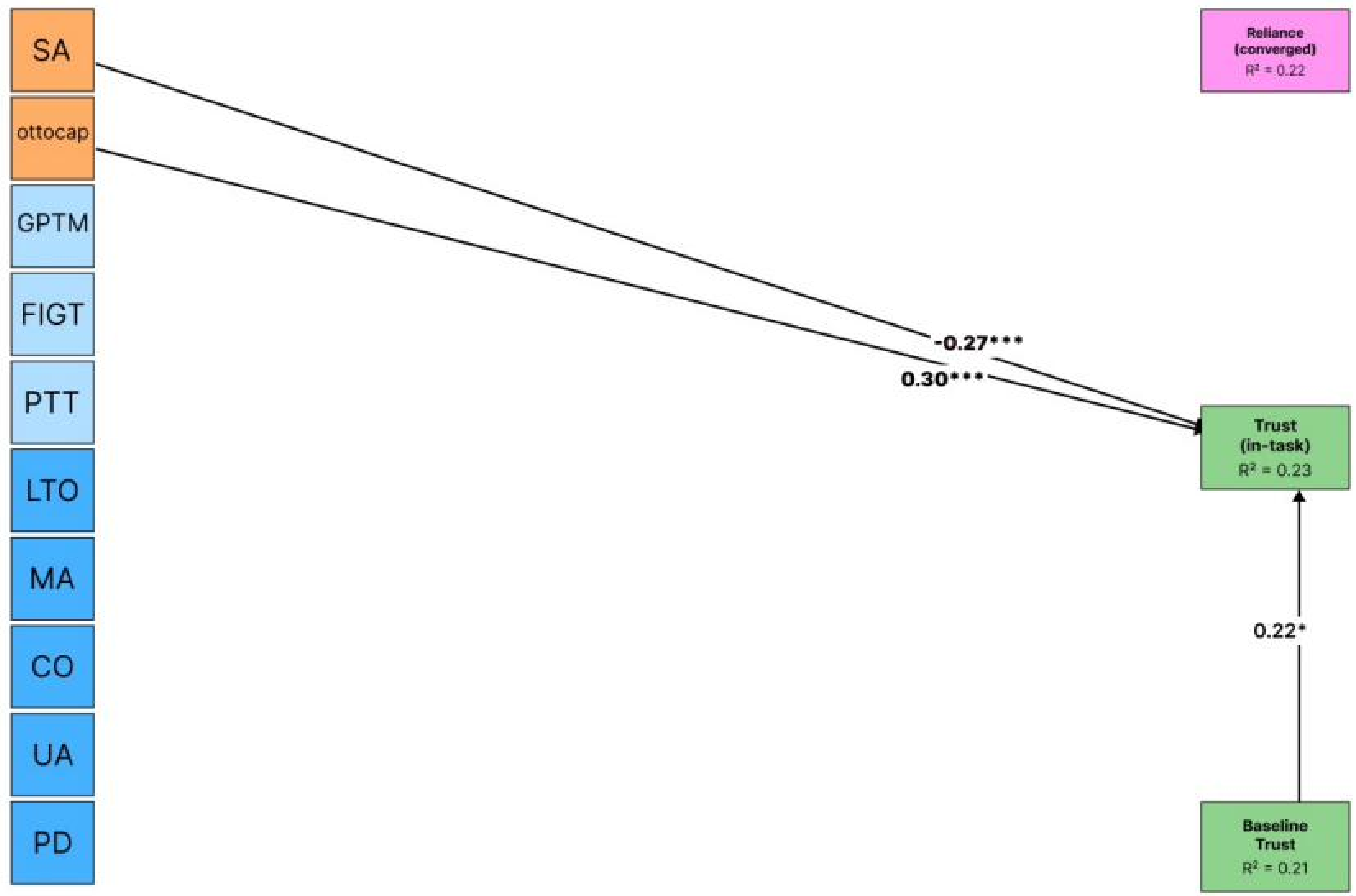


Baseline Trust is shaped by Faith in General Tech, Collectivism, General Propensity to Trust Machines, Uncertainty Avoidance

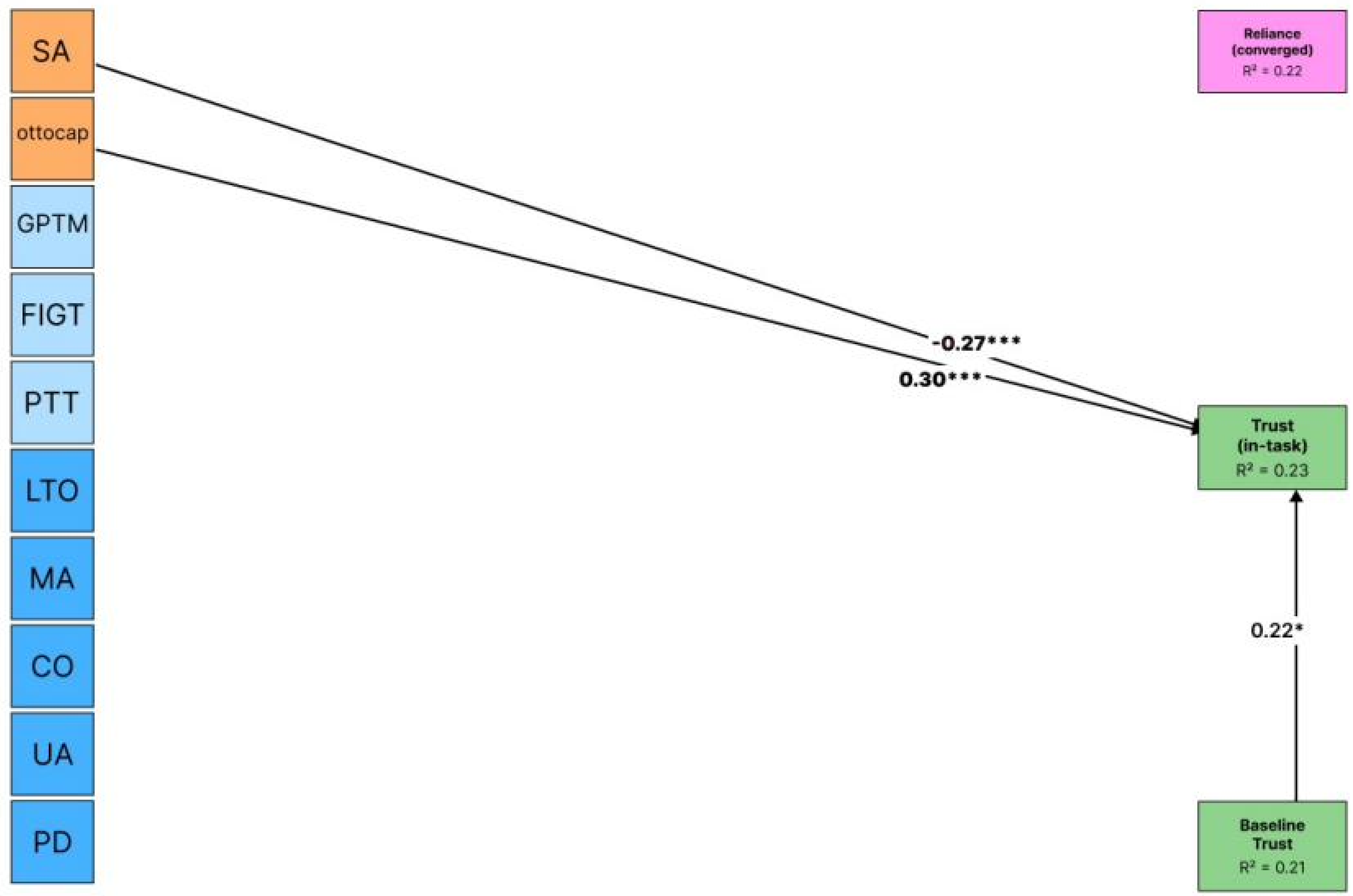


Path 2: Predictors of In-Task Trust

$$T_{comp} = \gamma^1 \cdot T_{Comp0} + \gamma^2 \cdot cap + \gamma^3 \cdot SA + \varepsilon^2$$

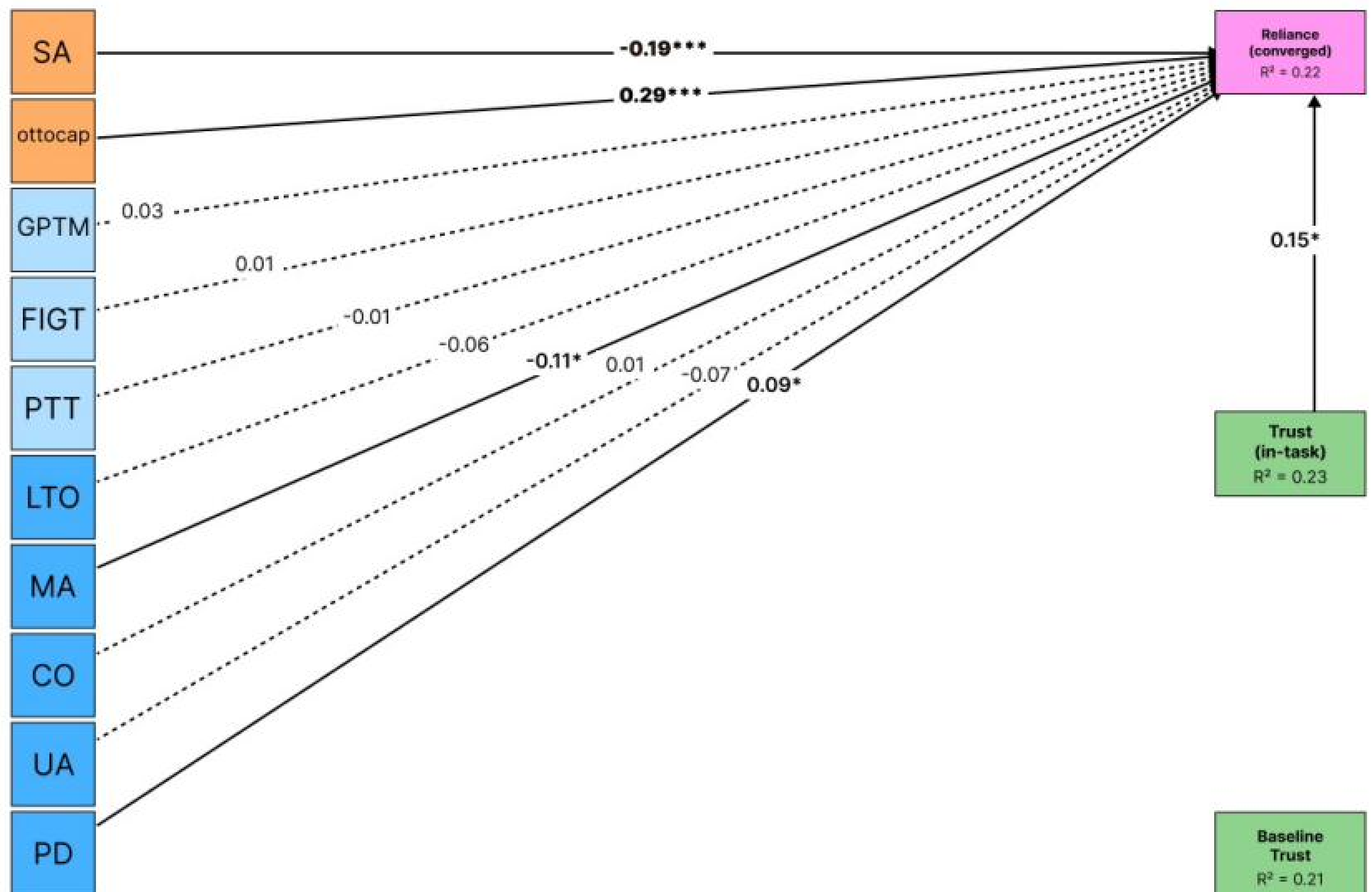


In-Task Trust is shaped by Self-Assessed Performance, capability

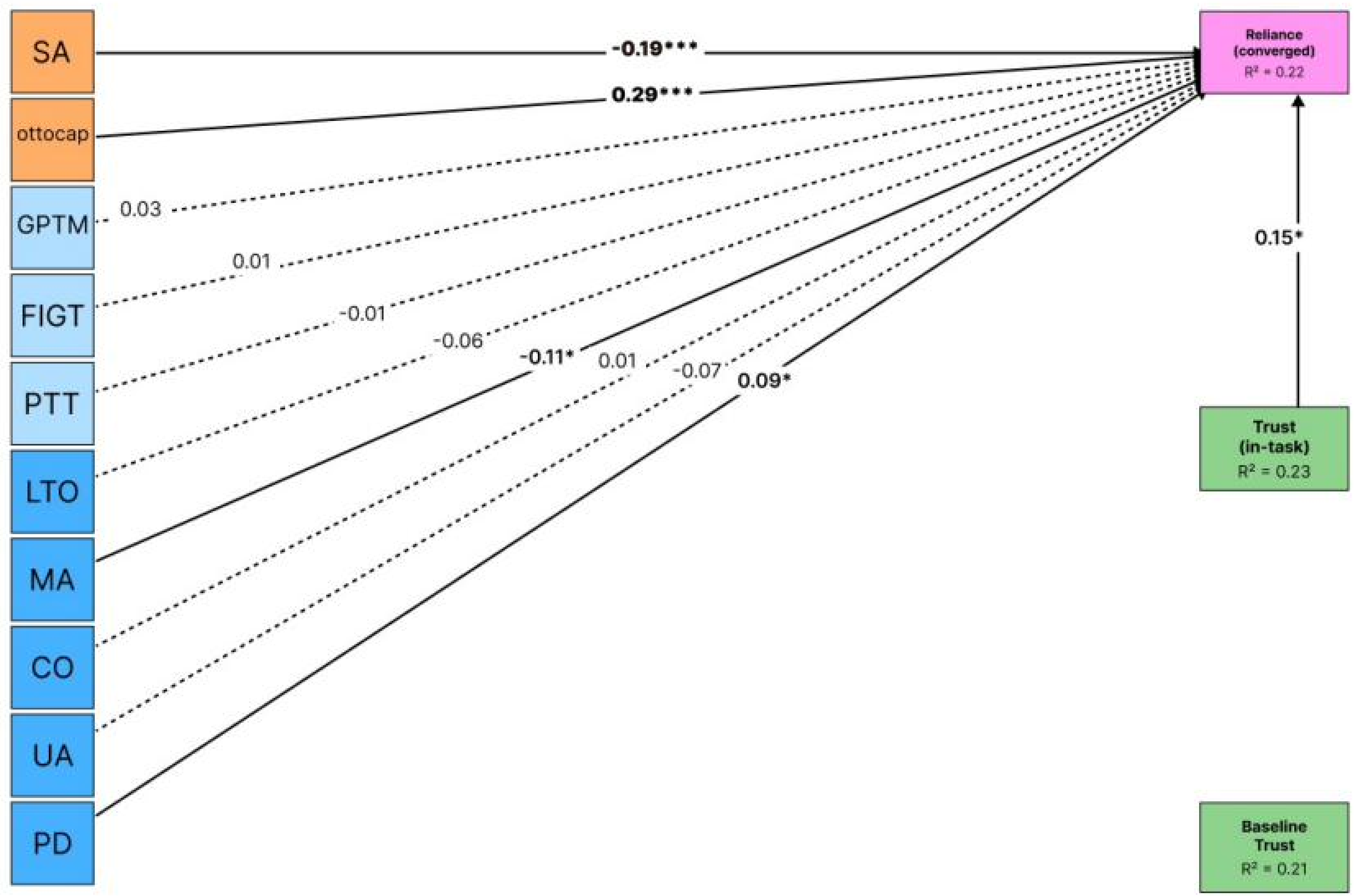


Path 3: Predictors of Converged Reliance

$$R_{cnv} = \delta^1 \cdot T_{Comp} + \delta^2 \cdot cap + \delta^3 \cdot SA + \delta^4 \cdot PD + \delta^5 \cdot UA + \delta^6 \cdot CO + \delta^7 \cdot MA + \delta^8 \cdot \dot{L}TO + \delta^9 \cdot PTT + \delta^{10} \cdot FIGT + \delta^{11} \cdot GPTM + \varepsilon^3$$



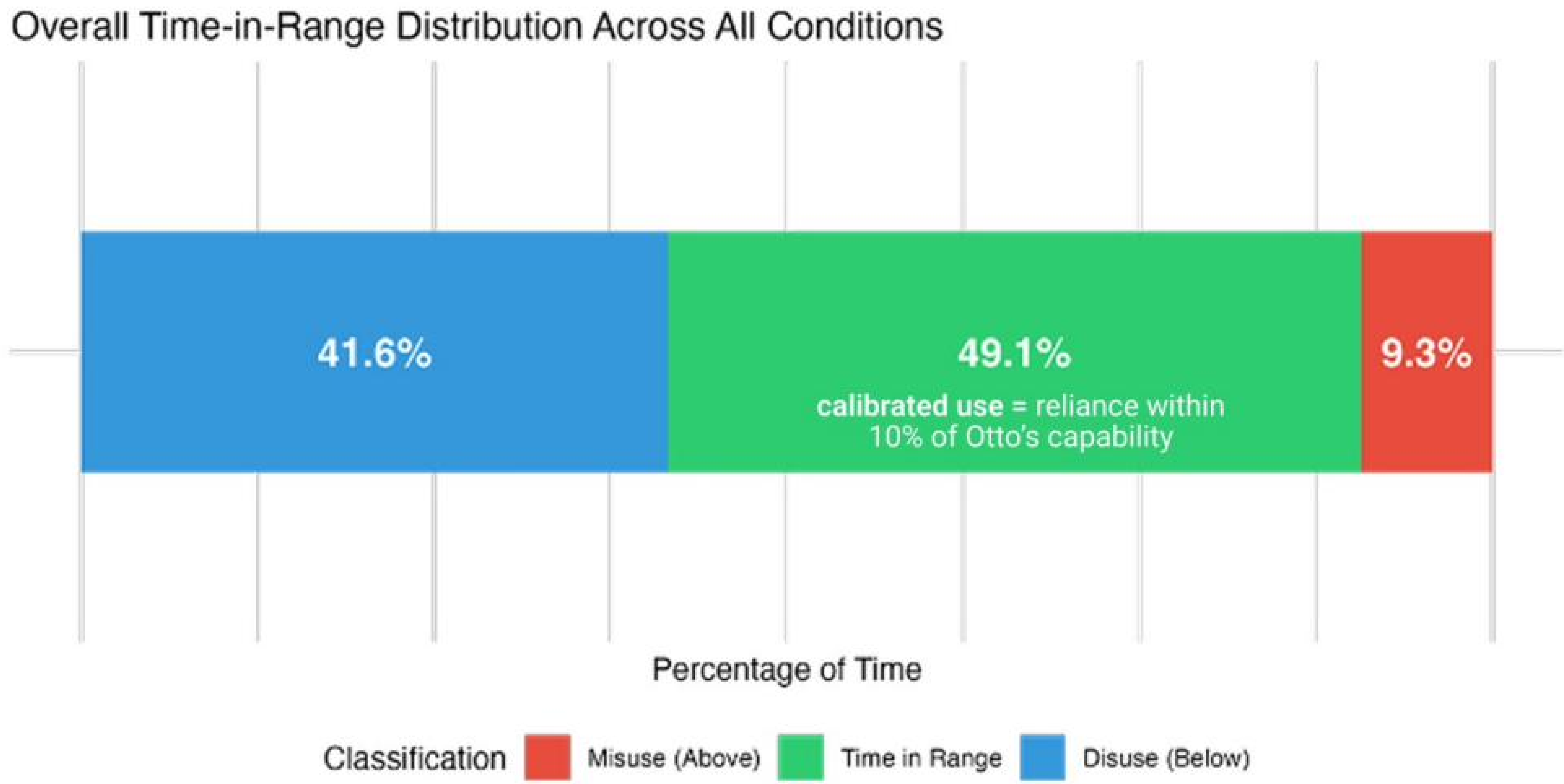
Converged Reliance is shaped by Self-Assessed Performance, capability, Masculinity, Power Distance



Analysis

%TIR

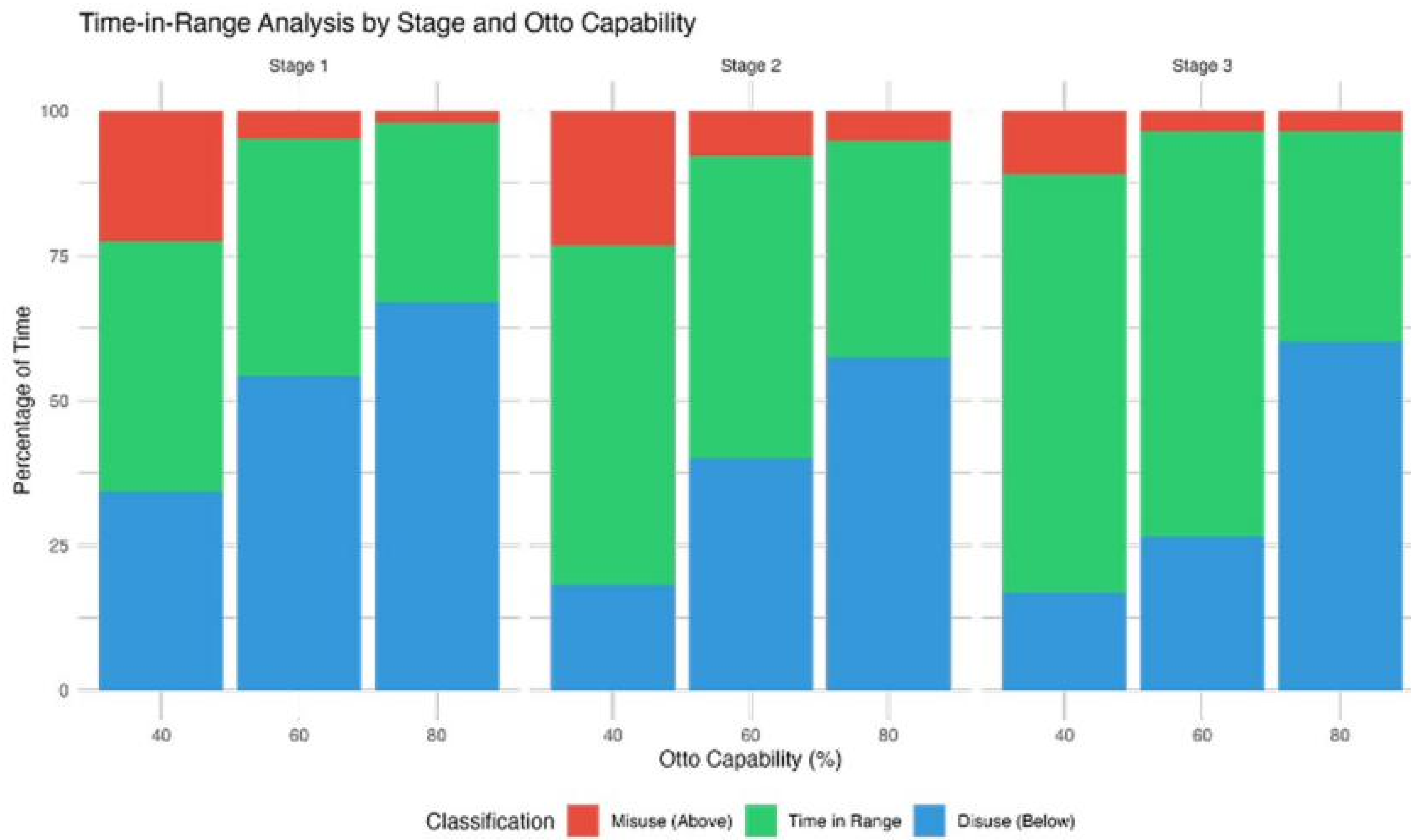
Frequency of Calibrated Trust > Misuse > Disuse (collapsed)



Analysis

%TIR

Frequency of Calibrated Trust > Misuse > Disuse (by Stage, capability)



Discussion

SEM Findings

- **Baseline Trust** is shaped by Faith in General Tech, Collectivism, General Propensity to Trust Machines, Uncertainty Avoidance

- **In-Task Trust** is shaped by Self-Assessed Performance, capability

- **Converged Reliance** is shaped by Self-Assessed Performance, capability, Masculinity, Power Distance

Discussion

%TIR Findings

- **Calibrated Trust was the most frequent case,** occurring about half of the time across the entire game & across all participants
- **Disuse was the second most frequent**, occurring about two-fifths of the time across the entire game & across all participants
- **Misuse was the second most frequent**, occurring about one-tenth of the time across the entire game & across all participants

Conclusions

SEM Findings

- **Baseline Trust** in a novel adaptable system depends on pre-conceived notions of technology and the operator's Uncertainty Avoidance
    - **Uncertainty Avoidance** = The extent to which people feel threatened by ambiguous situations and have created beliefs and institutions that try to avoid these.

Conclusions

SEM Findings

- **Trust during a task** depends on the operator's perception of their own performance, and the system's reliability

Conclusions

SEM Findings

- **Automation use** depends on the operator's perception of their own performance, the system's actual reliability, and the operator's Masculinity and Power Distance
  - **Masculinity** = a situation in which the dominant values in society are success, money, and things (vs. others and the quality of life) (Hofstede & Bond, 1984, pp. 3–4)
  - **Power Distance** = the extent to which the less powerful members of institutions and organizations accept that power is distributed unequally (Hofstede & Bond, 1984, pp. 3–4)

Conclusions

Overview

This study shows that trust, reliance, and calibration are shaped jointly by dispositional traits (including cultural values and dispositional propensities to trust), operator's perceived performance, and system capability.

I demonstrate the use of an empirical framework for measuring calibration of use in human–automation interaction.

Thank You!

Questions?

Tufts University
August 2025

Adviser: Dave B Miller