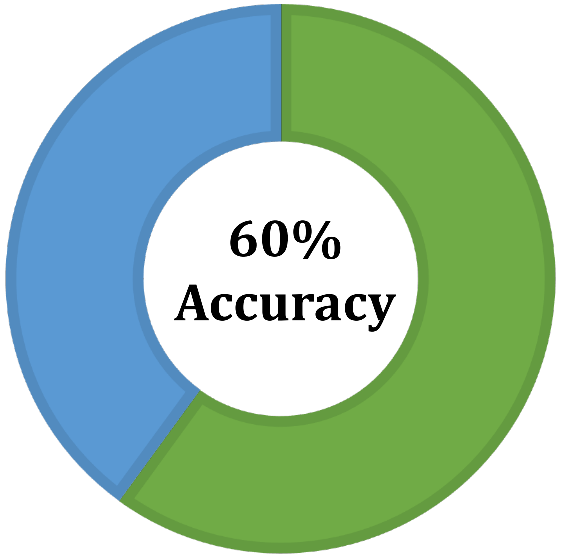
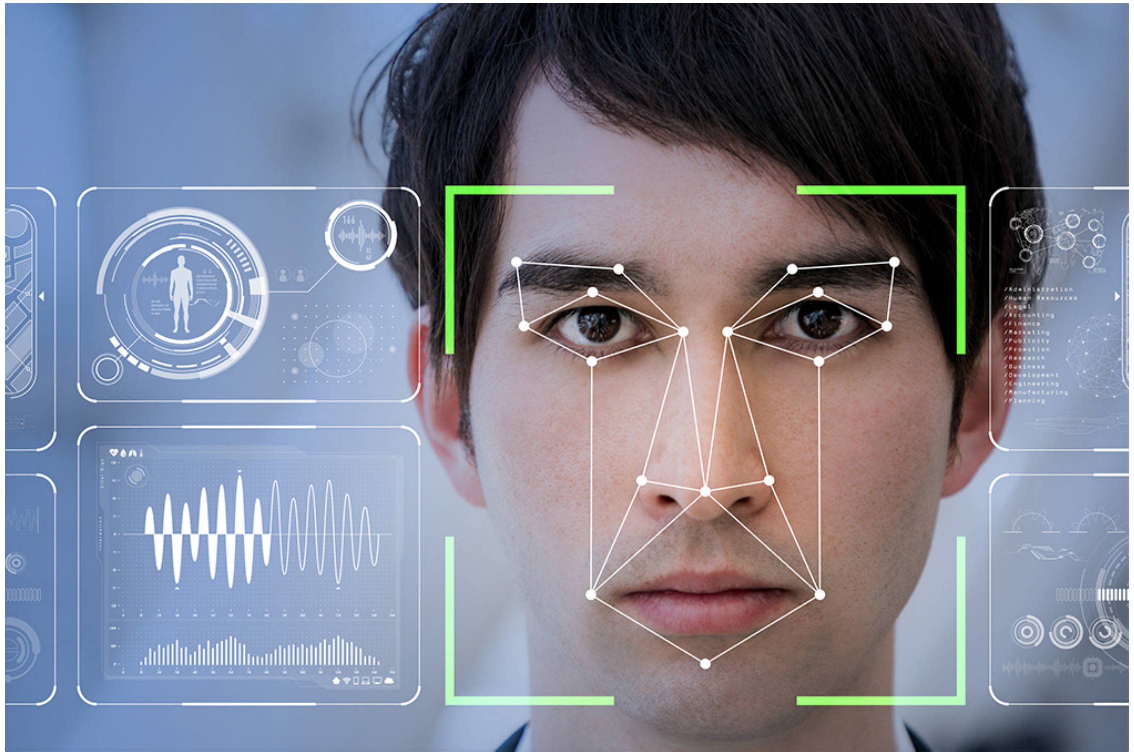
When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models



Amy Rechkemmer, Ming Yin
Purdue University

CHI 2022, New Orleans, Louisiana, April 30th - May 6th

Machine Learning is Everywhere…

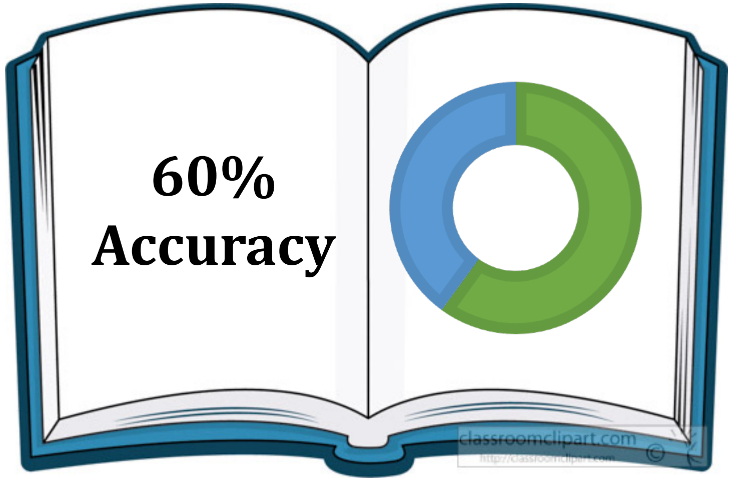


Critical Societal
Challenges

Everyday
Decision-Making

How do Performance Indicators Impact Trust?

Accuracy



Stated

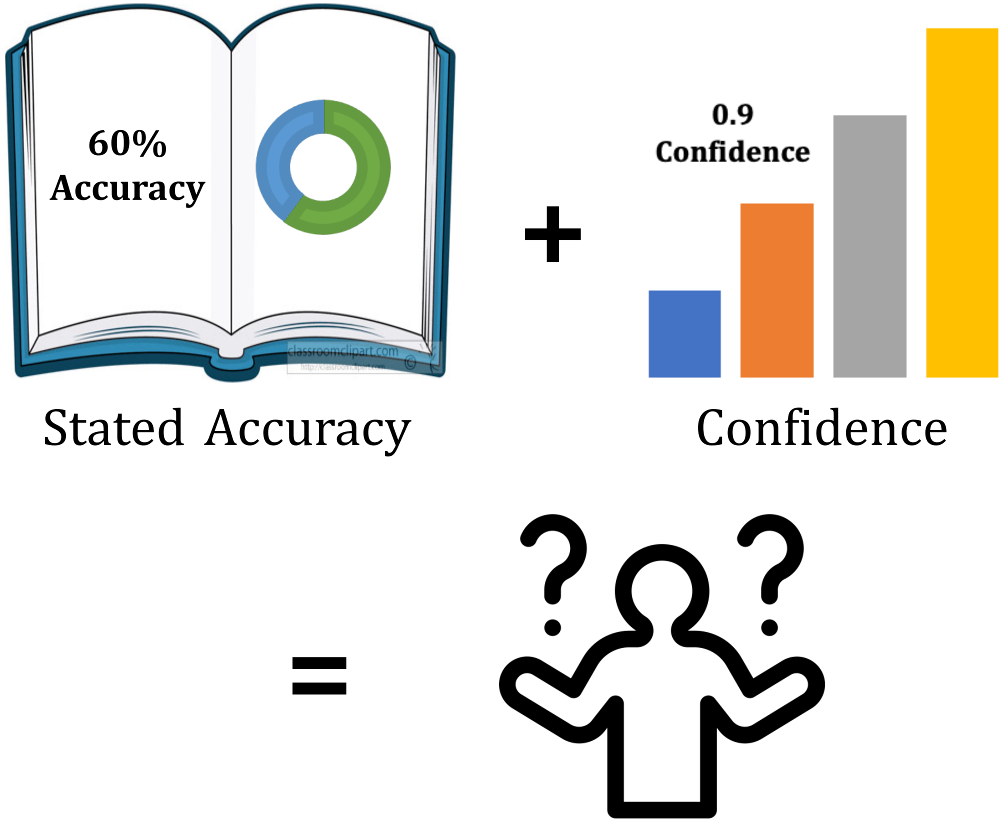
55%
Accuracy

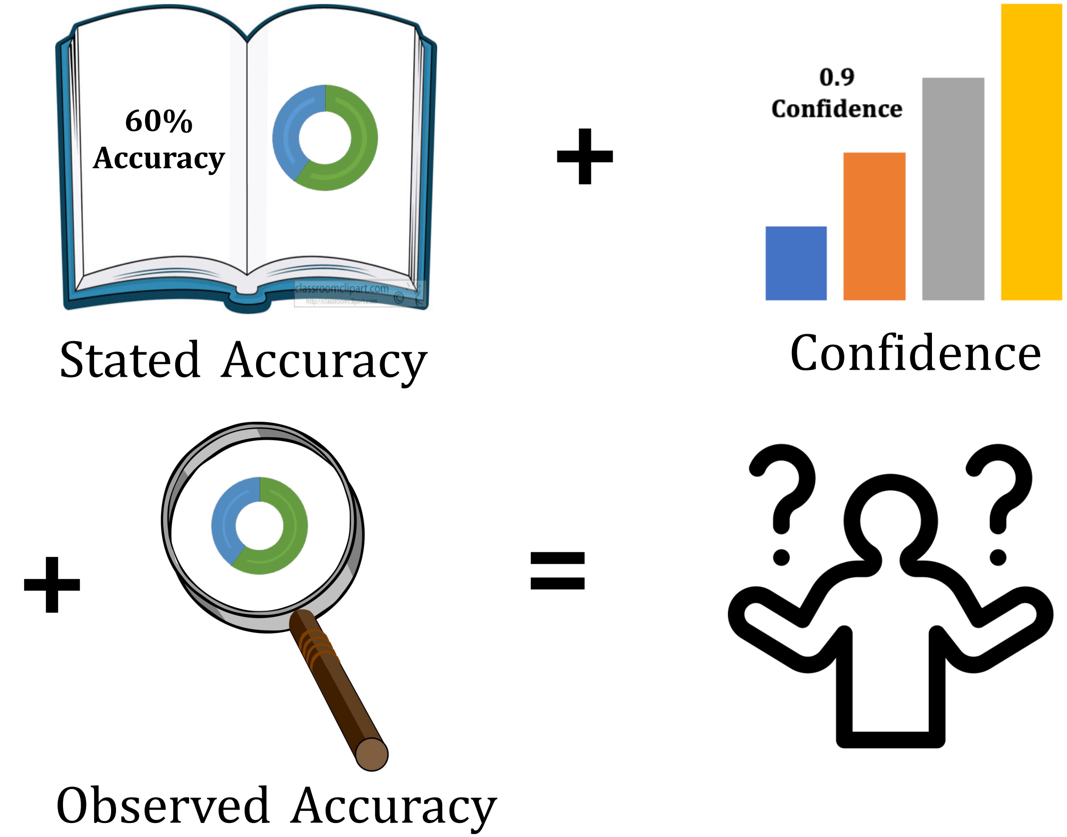
Observed

Confidence

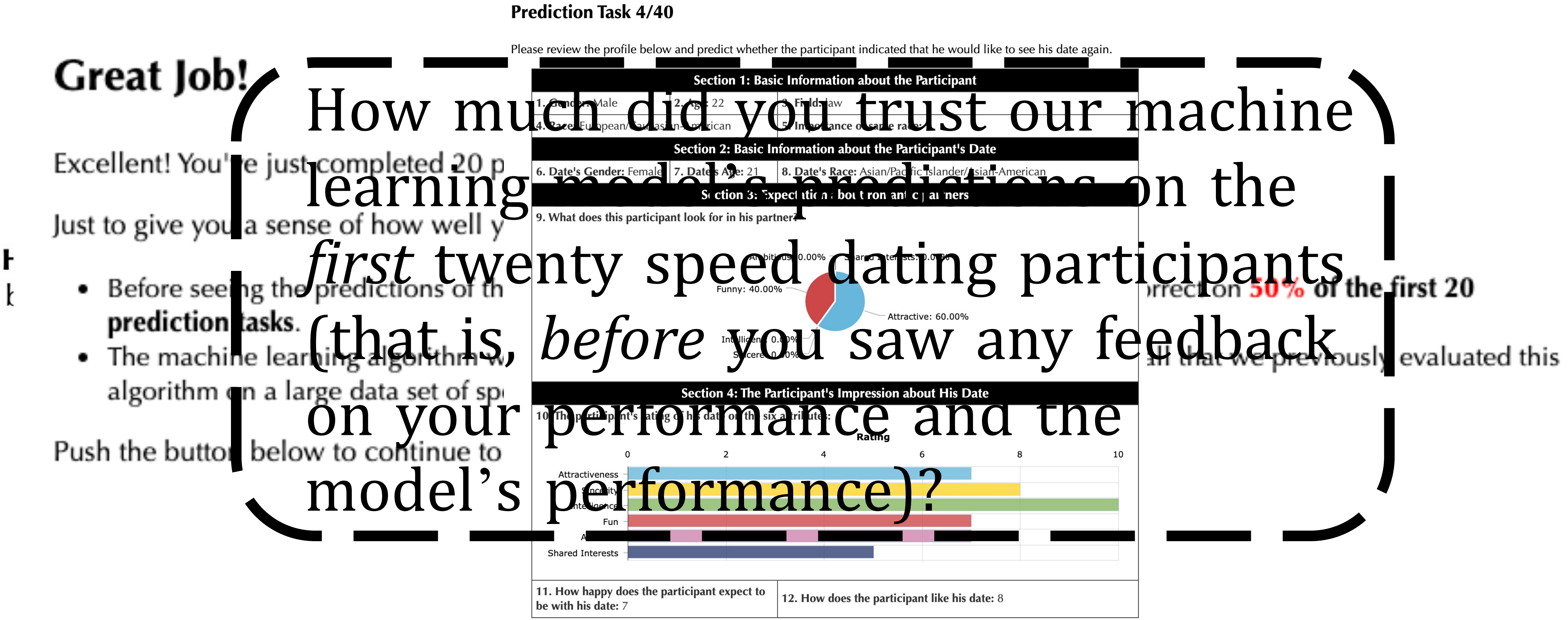
0.9
Confidence

RQ1: How is users’ trust impacted by **stated accuracy** and **confidence** *before* observing accuracy?





RQ2: How is users’ trust impacted by **stated accuracy**, **confidence**, and **observed accuracy** *after* observing accuracy?



Prediction Task 4/40

Please review the profile below and predict whether the participant indicated that he would like to see his date again.

Section 1: Basic Information about the Participant		
1. Gender: Male	2. Age: 22	3. Field: law
4. Race: European/Caucasian-American		5. Importance of same race:

Section 2: Basic Information about the Participant's Date		
6. Date's Gender: Female	7. Date's Age: 21	8. Date's Race: Asian/Pacific Islander/Asian-American

Section 3: Expectation about romantic partners

9. What does this participant look for in his partner?

Section 4: The Participant's Impression about His Date

10. The participant's rating of his date on the six attributes:

11. How happy does the participant expect to be with his date: 7	12. How does the participant like his date: 8

Great Job!

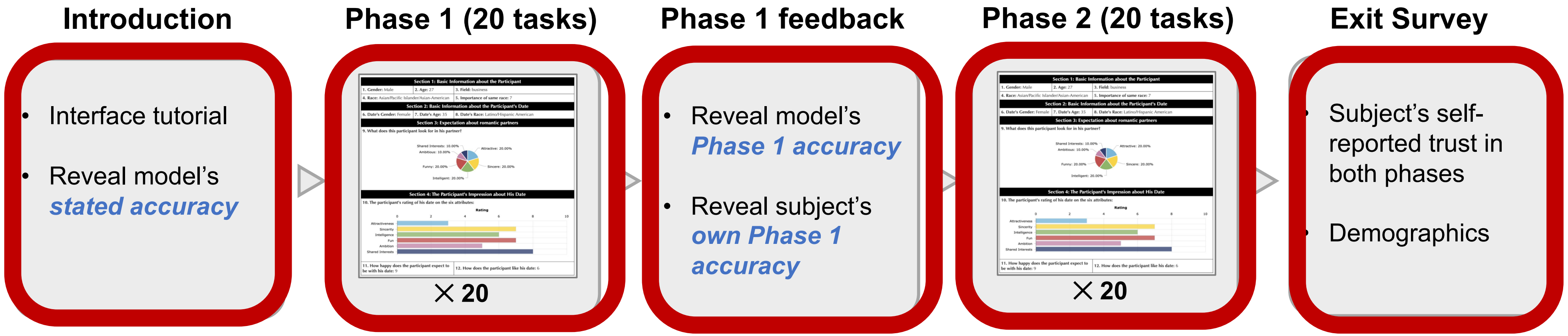
Excellent! You've just completed 20 p

Just to give you a sense of how well y

- Before seeing the predictions of th … rrection **50% of the first 20 prediction tasks.**
- The machine learning algorithm w … all that we previously evaluated this algorithm on a large data set of sp

Push the button below to continue to

How much did you trust our machine learning model's predictions on the *first* twenty speed dating participants (that is, *before* you saw any feedback on your performance and the model's performance)?



Introduction

- Interface tutorial
- Reveal model's ***stated accuracy***

Phase 1 (20 tasks)

× 20

Phase 1 feedback

- Reveal model's ***Phase 1 accuracy***
- Reveal subject's ***own Phase 1 accuracy***

Phase 2 (20 tasks)

× 20

Exit Survey

- Subject's self-reported trust in both phases
- Demographics

	Low	High
Confidence	0.5 – 0.8	0.8 - 1
Stated Accuracy	60%	90%
Observed Accuracy	55%	95%

Phase 1

55% Accuracy

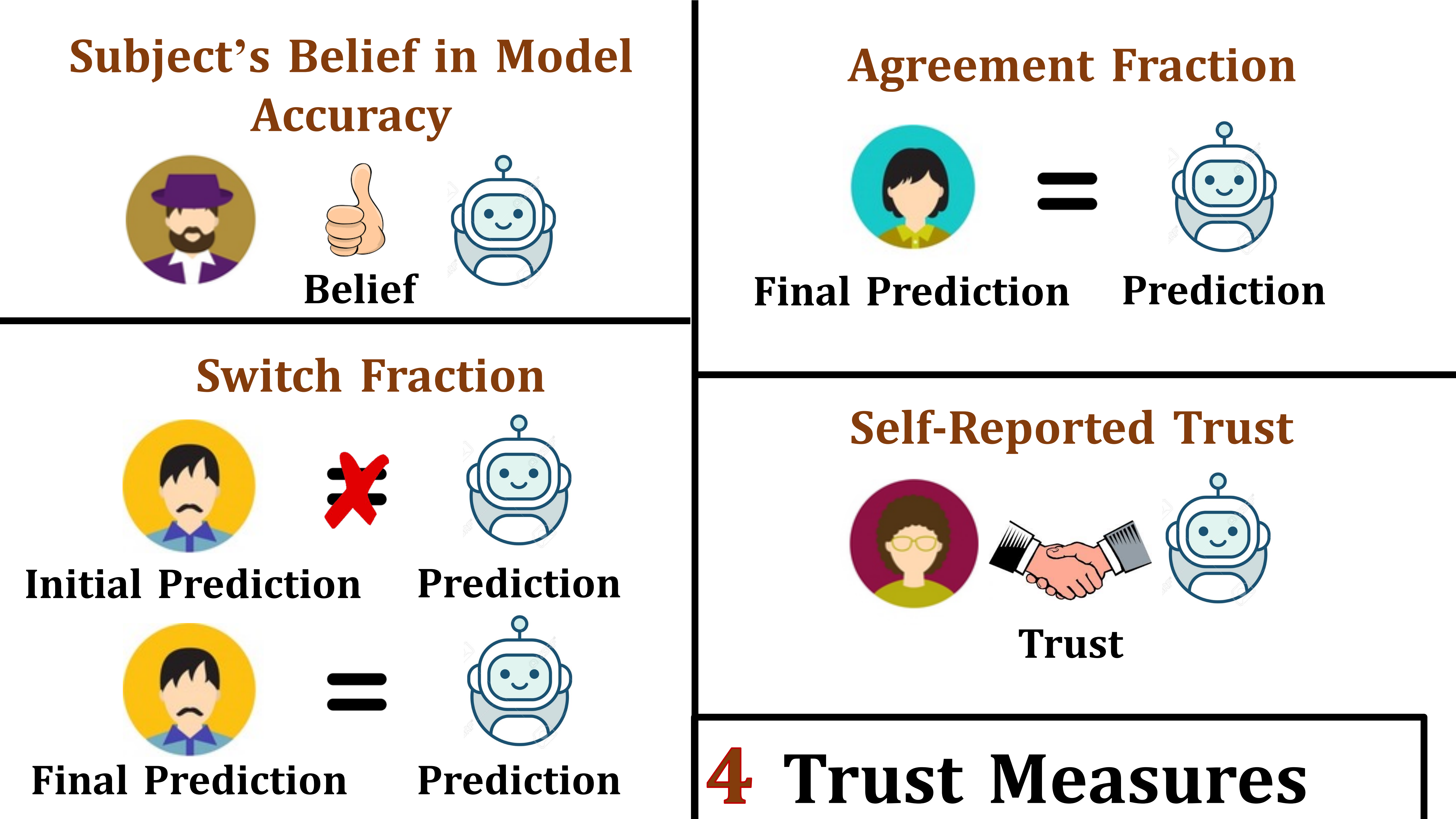
Same Predictions

95% Accuracy

Same Predictions

Phase 2

All Treatments See Same Predictions



Subject's Belief in Model Accuracy

Belief

Switch Fraction

Initial Prediction ≠ Prediction

Final Prediction = Prediction

Agreement Fraction

Final Prediction = Prediction

Self-Reported Trust

Trust

4 Trust Measures

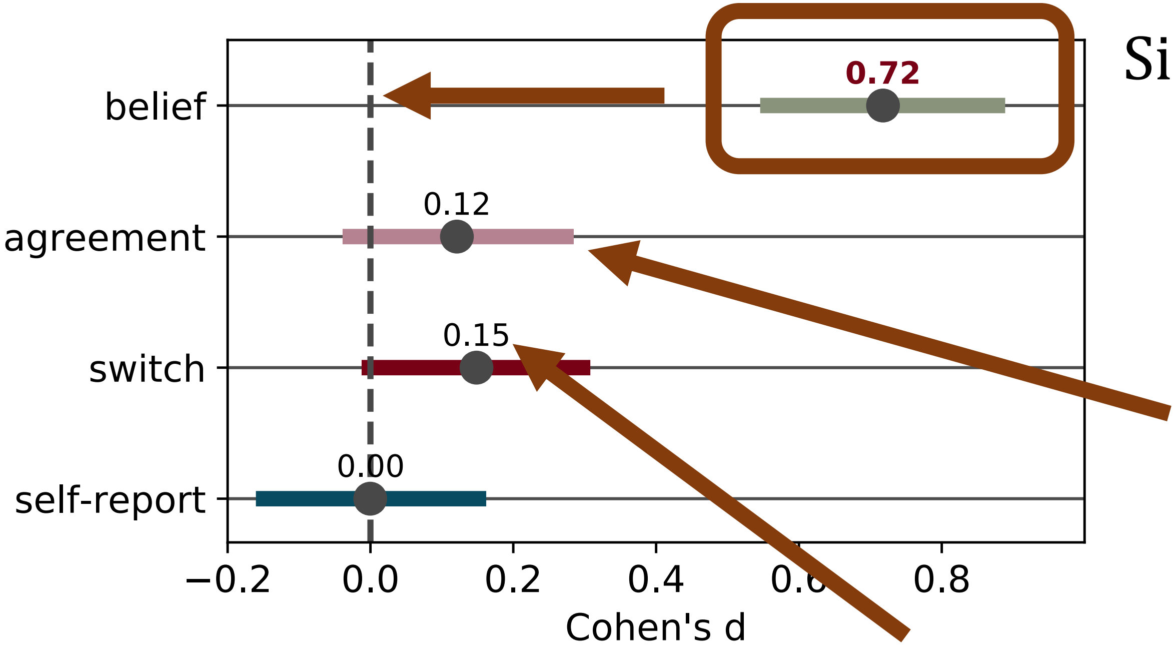
Analysis Method

Independent Variables

- Model Confidence
- Stated Accuracy

Dependent Variables

- Subject's Belief in Model Accuracy
- Agreement Fraction
- Switch Fraction
- Self-Reported Trust



Significant Result

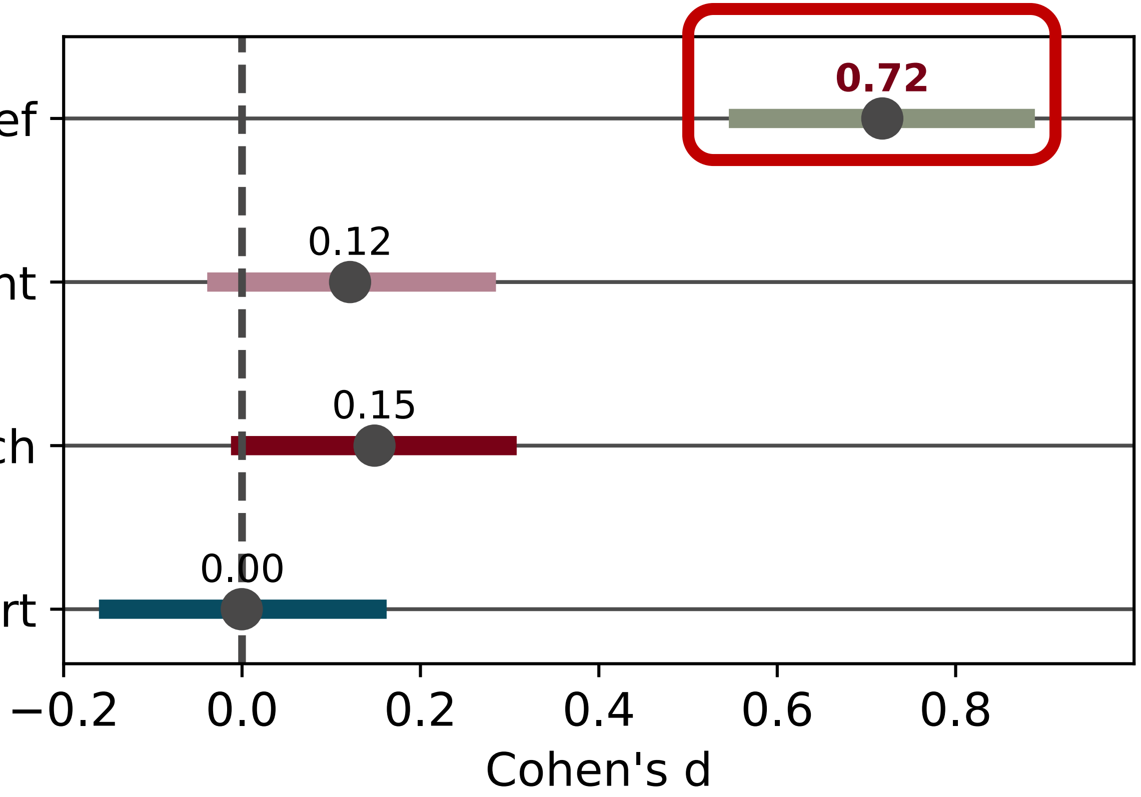
95% Bootstrap Confidence Interval

Confidence

Cohen's *d* Effect Size

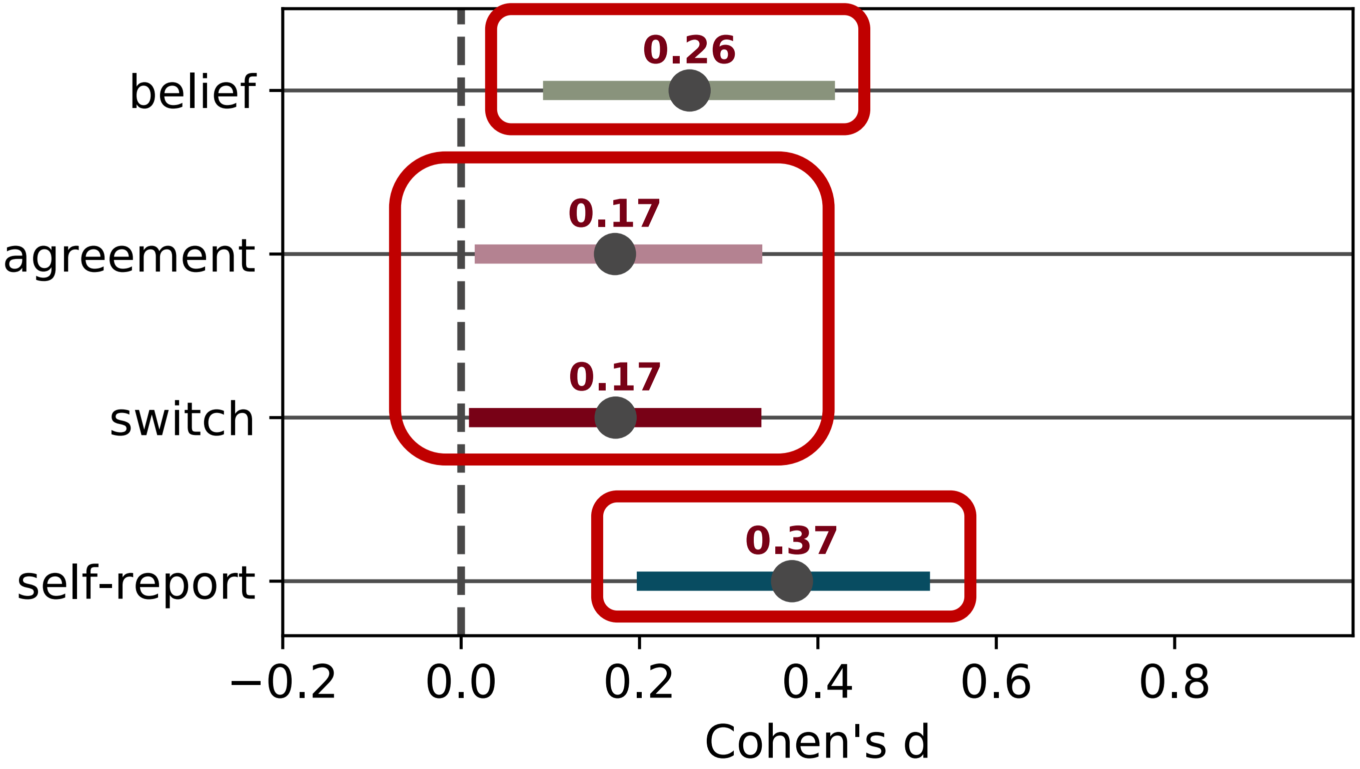
People **<span>believe</span>** **that a model with higher** **confidence** **is more accurate.**

RQ1: The Effect of Confidence in Phase 1



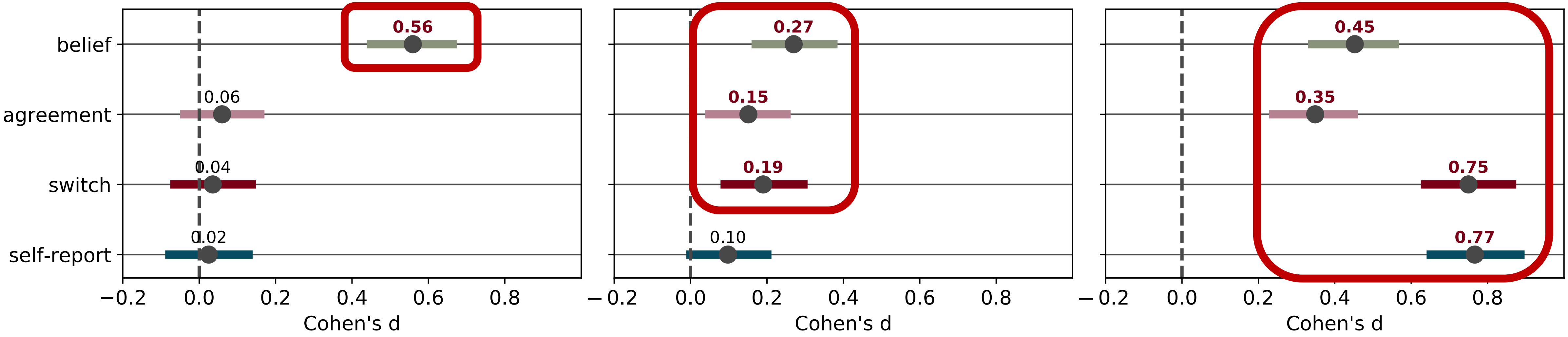
Before observing accuracy in practice, people believe that a model with higher confidence is more accurate.

RQ1: The Effect of Stated Accuracy in Phase 1



Before observing accuracy in practice, people trust a model with higher stated accuracy more and follow its predictions more often.

RQ2: Confidence, Stated Accuracy, and Observed Accuracy in Phase 2



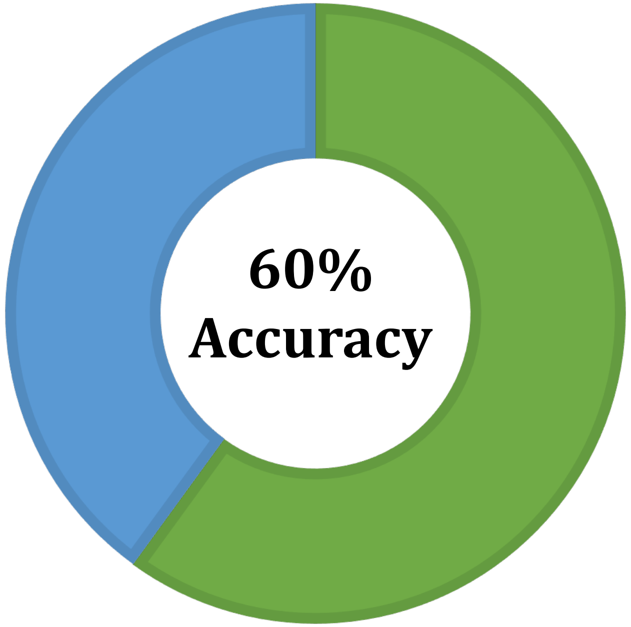
Confidence *Stated Accuracy* *Observed Accuracy*

After observing accuracy in practice, people believe that a model with higher confidence is more accurate, but trust a model with higher observed accuracy and follow its predictions more.

Conclusion and Implications

- Model confidence and accuracy play different roles in influencing trust.
- Confidence has a greater impact on belief in the model’s accuracy, while stated and observed accuracy influence people’s trust and willingness to follow the model.
- Shows importance of helping laypeople understand uncertainty of performance based on a small set of predictions and see the value of utilizing a calibrated confidence score

Thank You!



Contact: arechke@purdue.edu