Trial by "Trial": An Item-Centric Investigation on the Director Task as an Individual Differences Measure of Theory of Mind

Han Hao[1], Manali Pathare[2], & Ester Navarro[3]

1. Tarleton State University 2. New Mexico State University 3. St.John's University

Introduction

The Director Task & Theory of Mind

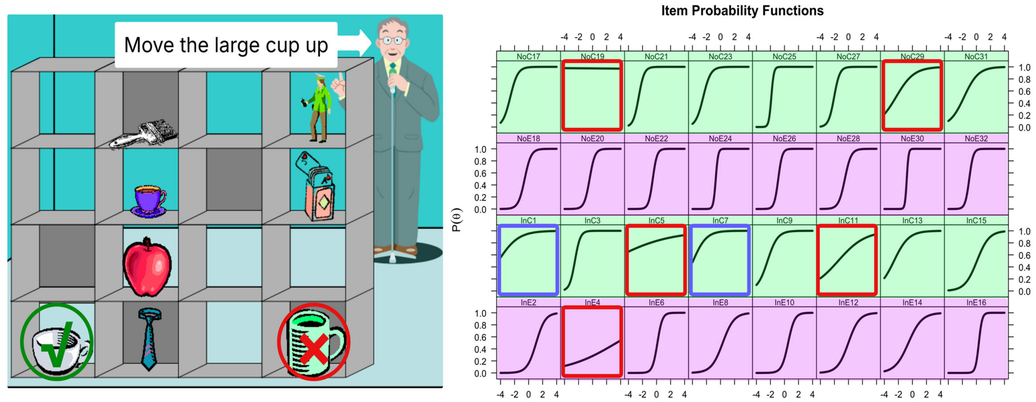
- **Theory of Mind (ToM)**: the ability to reason about others' perspectives and beliefs.
- The underlying mechanisms of ToM as a socio-cognitive construct remains unclear.
- Classic ToM task measures lack specificity and may fail to capture genuine perspective-taking (Quesque & Rossetti, 2020).
- **The Director Task (DT)**: Participants ignore their own perspective and select objects as seen by a "director" facing their opposite view.
- Task-level composites of DT are usually used as an indicator of cognitive individual differences in ToM while trial performance and properties across conditions remain underexplored.
- **Item Response Theory (IRT)** provides trial-level psychometric properties that may help clarify the relationship between experimental manipulations and trial responses.

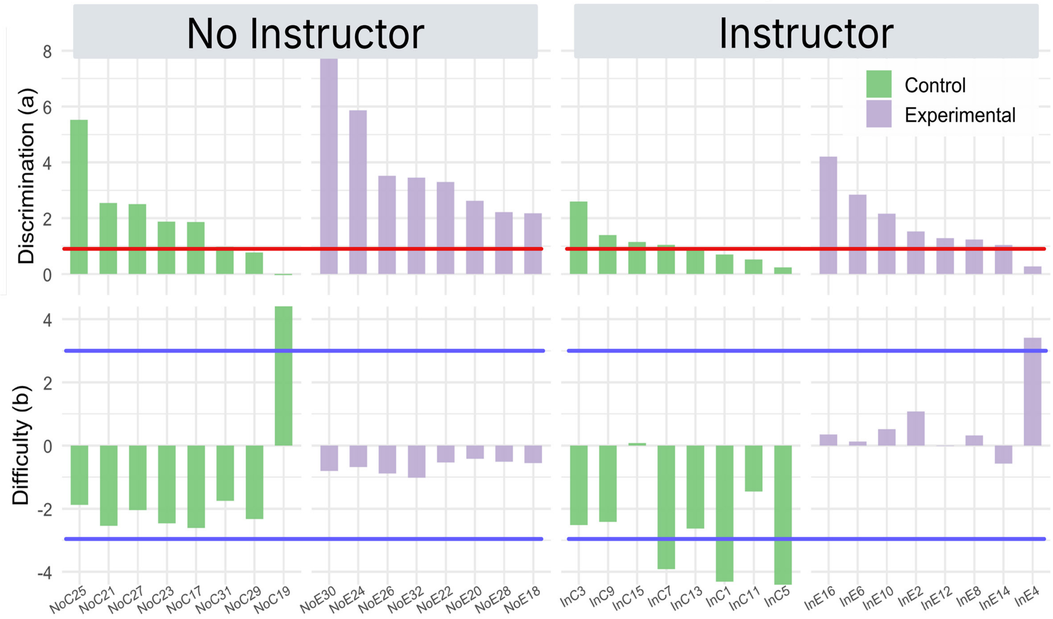
The Present Study

- We conducted an item-level analysis using a unidimentional 2-PL IRT model.
- Item parameters were compared in a 2×2 framework:
Instructor vs No-Instructor; Experimental vs Control

Aims: 1) identify poor items in the current DT task; 2) test how instructor presence and task demands affect item characteristics; and 3) provide insights to improve DT measurement and ToM theory.

Key Results





Methods & Results

Data & Statistical Procedures

- Total Sample: N = 158 (N_1 = 78; N_2 = 80)
- 2x2x8 = 32 Ins/NoIns & E/C trials
- **Uni-factor 2PL model** on dichotomous trial scores (all -1 rescored to 0)
- Item Discrimination (*a*) and Difficulty (*b*) for trials under different manipulations were compared **after** the estimation for condition-wise inference.

Summary of Results

Discrimination

- High *a* ≈ more reliable/sensitive to individual differences - Non-Instructor trials generally showed higher "*a*"s than Instructor trials
- Experimental trials generally have higher "*a*"s than Control trials
- The 3-5 trials with very small discrimination have low reliability and worse item quality as individual differences measure
- However, the trial(s) with extremely large discrimination may suggest potential separation rather than truly high reliability in this case

Difficulty

- Most control trials were overly easy therefore may not be informative on higher ability subjects (almost all *b*s ≈ -2)
- Experimental trials were slightly more difficult than Control trials (range around -1 to 1)
- Qualitatively, the presence of the virtual instructor generally introduced a slightly stronger experimental effect of "perspective taking"

Conclusion

- **The Director Task offers a promising paradigm but show uneven psychometric properties:**
- *No-Instructor items: Discriminate well at lower theta levels, i.e., can identify inhibition ability among individuals with below average inhibition.*
- *Instructor items: mixed item properties, i.e., generally more low discrimination trials and inconsistent difficulty.*
- **The virtual instructor increases ToM demands but reduces the accuracy and consistency of responses.**
- **The common practice of using composite score at the task level may be of potential issue when using DT as an individual differences measure of ToM.**