Rethinking accuracy in theory of mind measures: influence of individual-related and task-related factors in adults

Kit Ling Yeung (kxy090@student.bham.ac.uk), Ian Apperly and Rory Devine
School of Psychology, University of Birmingham

Introduction
- Mindreading/ Theory of mind (ToM): attribution of mental states to others
- Most existing measures have pre-determined correct answers
  - Interpreting ambiguous social stimuli: lacking a definitive ground truth
  - Current studies: Crowdsourcing → Alignment rather than accuracy
- Study 1: Would different groups of neurotypical individuals endorse different but equally legitimate interpretations of the same social stimuli?
- Study 2: Would task format (open-ended vs. forced choice) influence interpretations of the same social stimuli?

Study 1
- Group 1) Younger adults (18-26 years, M\text{age} = 23.25; n=83);
  Group 2) Older adults (53-60 year, M\text{age} = 57.14; n=84)
- “What do you think the circled person is thinking/feeling?”
  - Six pictorial stimuli (items)
  - Open-ended;
  - Multiple interpretations allowed, ranked by perceived possibility
  - First-ranked interpretation coded and analysed
- Coding scheme generated by inductive content analysis → Text to category
  - Multiple coding allowed for each text response (e.g. confused/unsure + angry/irritated)

<table>
<thead>
<tr>
<th>Age</th>
<th>Scoring</th>
<th>Young</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>Same-group condition</td>
<td>Crossed-group condition</td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>Crossed-group condition</td>
<td>Same-group condition</td>
<td></td>
</tr>
</tbody>
</table>

Group alignment score on an item
- Proportion of group members endorsing the same category (adjusted in same-group condition) averaged across endorsed categories within participant

Results
- Permutation test (left)
  - Overall alignment score (across groups; red line on the right)
  - Exceeded 95% percentile of alignment calculated from simulated random responses over 1000 iterations (blue line on the left)
  - Participants did agree on certain interpretations (categories)
  - But 2-5 popular interpretations were found for each item

Paired t-tests (right)
- Significant effect of scoring condition within either of the two groups will indicate presence of differences in how the two groups’ interpretations
- Significant scoring condition effect in the younger group (same-group > crossed-group alignment; Cohen’s d=.330)

Study 2
- n = 44 (18-25 years, M\text{age} = 22.41)
- Forced-choice version of the task in study 1
- “What do you think the circled person is thinking/feeling?”
  - Forced-choice options: Verbatim responses from the top 4 popular categories for each item in study 1
  - Different processes: “Generation” vs. “recognition”

Format alignment score
- Proportion of participants endorsing the same option/category as the most likely interpretation
- Same-format scoring (FC):
  - Based on the current sample (adjusted)
- Crossed-format scoring (OPEN):
  - Based on younger adult group in study 1

Results
- Paired t-test
  - Compared between format scoring
  - Significant format effect (same-format > crossed-format alignment)

Conclusion
- Individual (age) and task-related (format) factors influenced interpretations of mental states in ambiguous social scenarios
- “Decoding” of mental states from observable features is not a sufficient explanation for mindreading success
- The ability to interpret stimuli in a contextually sensitive manner may be at least as important
- Advanced mindreading assessments should be contextualized
- Challenges “measurement invariance”
- Researchers should be cautious about the applicability of tasks with various formats and across different populations