BEING TALL COMPARED TO BEING TALL AND BEING TALLER
THREE SENTENCES:

1. Blue is taller than Red
2. Blue is tall compared to Red
3. Blue is tall

TWO QUESTIONS:

‣ How (if at all) does 1 differ from 2?
‣ To what extent is that difference due to 2 being closer to 3 than 1 is?
Clustering to test influence of clusters on tallness judgments (Schmidt et al. 2009)

- Scaling included to increase salience of clusters—also affects absolute (but not relative) height differences between lines
TWO EXPERIMENTS:

- **CT1:** “The green line is tall.”

- **CT2:** *Comparative:* “The blue line is taller than the red line.” and *Positive:* “The blue line is tall compared to the red line.”
Participants and Stats

- Participants recruited on Amazon’s Mechanical Turk; estimated time 10-12 min.
  - CT1: n=20, $2.40 HIT
  - CT2: n=40, $2.00 HIT
- Stats included logistic (for judgment data) and linear (for RT data) mixed effects regressions with random intercepts for subjects.
- We report p-values based on model comparisons between a maximal model m and a model m' that subtracts a targeted variable. All analyses were conducted in R using RStudio and the lme4 package.
EXPERIMENT 1

- CT1 looks for signatures of tallness evaluations of individual lines, to leverage as a predictor for evaluations of “taller” vs. “tall compared to” sentences in CT2

“The green line is tall.”
EXPERIMENT 1

- Effect of the position of the green line (p < .0001)
- No effect of clustering (p = .75)
- An effect of scaling (p = .013)

- pictured: portion of “tall” judgments for each position in the distribution
EXPERIMENT 1

- Effect of the position of the green line ($p < .0001$)
EXPERIMENT 1

- Effect of scaling ($p = .013$)
EXPERIMENT 1

scaling*green.position (p = .17)
EXPERIMENT 1

- Takeaways:
  - Schmidt et al. (2009) found that a cluster-based model and a relative-height-by-range model both successfully captured their data.
  - Lack of clustering effect here is potentially surprising, given the success of cluster-based models in other work.
  - Scaling effect suggests magnitude of height differences is an important factor in tallness judgments.
  - However, scaling manipulates absolute, not relative, differences.
EXPERIMENT 2

CT2 looks for differences in the evaluation of *positive* and *comparative* statements of relative height, and compares them to CT1 “tallness values”.

*Positive*: The blue line is tall compared to the red line

*Comparative*: The blue line is taller than the red line
EXPERIMENT 2

‣ Potential views of the difference between “compared to” and “-er”:

‣ The **naive view** of “tall compared to” vs. “taller than”: the former involves computation of tallness *simpliciter*; the latter does not

‣ Prediction: “tallness values” from CT1 should influence “tall compared to” but not “taller than” judgments in CT2

‣ The **difference-based view**: “tall compared to” is sensitive to differences in height in a way parallel to tallness *simpliciter*, but does not involve calculation of tallness *simpliciter*

‣ Prediction: difference in height between blue and red will matter, but the tallness value of blue will not

‣ The **domain-restriction view**: ongoing work; feel free to ask
EXPERIMENT 2

First result:

- No effect of CT1 “tallness values” on CT2 responses
- Blue’s “tallness” didn’t impact judgments of either sentence type in cases where blue was taller than red
- Red’s “tallness” didn’t impact judgments in cases where blue was taller than red
- Red’s “tallness” didn’t impact judgments in cases where red was taller than blue
- Overall: no impact of CT1 values on CT2
EXPERIMENT 2

- We call a response “correct” if it accepts a sentence when $\text{height}(b) > \text{height}(r)$, and rejects it when $\text{height}(r) < \text{height}(b)$
- Responses were more likely to be “correct” as the positional difference between blue and red increased ($p = .0009$)
EXPERIMENT 2

- Participants were more likely to “correctly” reject sentences than to “correctly” accept them (p < .01), though rates of correctness are overall high.
EXPERIMENT 2

- This asymmetry is primarily driven by “tall compared to”, not “taller” (effect of sentence*winner on correctness, p < .0001)
EXPERIMENT 2

“Tall compared to” was more sensitive to the positional difference between red and blue than “taller than” \( (p = .018) \)
EXPERIMENT 2

- Takeaways:
  - No effect of CT1 values means that the naive view can be rejected
  - Participants more willing to reject “compared to” sentences than to accept them, compatible with difference-based view
  - Positional difference effect on “compared to” sentences, also compatible with difference-based view
APPENDIX 1: RT DATA FOR CT2

- Effect of scaling on RT ($p = .0007$)
- Not surprising: discriminations become finer as scaling gets flatter
APPENDIX 1: RT DATA FOR CT2

- Effect of position difference on RT (p < .0001)
- Not surprising: discriminations are finer when lines are adjacent
APPENDIX 1: RT DATA FOR CT2

- Effect of scaling * position on RT ($p = .07$)
- Not surprising: flatter distribution + closer lines ramifies discriminability
APPENDIX 2: ANALYSIS CHOICES FOR CT1

- Main slides talk about position of green line, not the absolute height of the green line
- Both analyses show no effect of clustering, an effect of scaling, and an effect of position/length
- Positional data is cleaner and more directly reflects design
APPENDIX 2: ANALYSIS CHOICES FOR CT1

- effect of absolute height is significant ($p < 0.0001$)
- clustering still isn’t ($p = 0.899$)
- scaling effect is even bigger ($p < 0.0001$)
- interaction between absolute height and scaling ($p < 0.0001$)
APPENDIX 2: ANALYSIS CHOICES FOR CT1

- graph of absolute difference effect bounces around a lot, because absolute height values are situated at different points on different distributions (which are subject to different scaling effects)
APPENDIX 3: ANALYSIS CHOICES FOR CT2

- Main slides talk about positional difference between red and blue lines, not absolute difference
- Both analyses show the same pattern of results
- Positional data is cleaner and more directly reflects design

- Main slides talk about “correctness” not “value” (true/false)
- For “value”, winner effects (is red or blue taller) drown out other effects; correctness preserves winner effects while revealing more information
Stats for position difference and absolute difference show the same effects.

<table>
<thead>
<tr>
<th></th>
<th>position (pos.diff)</th>
<th>length (abs.line.diff)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>value</td>
<td>correct</td>
</tr>
<tr>
<td>sentenceCode</td>
<td>p = 0.15</td>
<td>p = 0.22</td>
</tr>
<tr>
<td>winnerCode</td>
<td>*** p &lt; 0.0001</td>
<td>* p = 0.01</td>
</tr>
<tr>
<td>scalingCode</td>
<td>p = 0.81</td>
<td>p = 0.58</td>
</tr>
<tr>
<td>pos.diffCode</td>
<td>p = 0.25</td>
<td>** p = 0.009</td>
</tr>
<tr>
<td>sentenceCode:winnerCode</td>
<td>* p = 0.032</td>
<td>*** p &lt; 0.0001</td>
</tr>
<tr>
<td>sentenceCode:scalingCode</td>
<td>p = 0.47</td>
<td>p = 0.73</td>
</tr>
<tr>
<td>winnerCode:scalingCode</td>
<td>p = 0.74</td>
<td>p = 0.79</td>
</tr>
<tr>
<td>sentenceCode:pos.diffCode</td>
<td>p = 0.34</td>
<td>* p = 0.018</td>
</tr>
<tr>
<td>winnerCode:pos.diffCode</td>
<td>** p = 0.007</td>
<td>p = 0.28</td>
</tr>
<tr>
<td>scalingCode:pos.diffCode</td>
<td>p = 0.65</td>
<td>p = 0.48</td>
</tr>
<tr>
<td>sentenceCode:winnerCode:scalingCode</td>
<td>p = 0.85</td>
<td>p = 0.45</td>
</tr>
<tr>
<td>sentenceCode:winnerCode:pos.diffCode</td>
<td>* p = 0.014</td>
<td>p = 0.61</td>
</tr>
<tr>
<td>sentenceCode:scalingCode:pos.diffCode</td>
<td>p = 0.89</td>
<td>p = 0.23</td>
</tr>
<tr>
<td>winnerCode:scalingCode:pos.diffCode</td>
<td>p = 0.70</td>
<td>p = 0.75</td>
</tr>
</tbody>
</table>
Graphs for absolute differences come out messier, for the same reasons as in CT1

**Effect of position*winner on correctness**

\[ p = 0.009 \quad p = 0.0006 \]
APPENDIX 3: ANALYSIS CHOICES FOR CT2

- Graphs for absolute differences come out messier, for the same reasons as in CT1

  effect of position/length on RT

  \( p < .001 \)  \hspace{1cm}  \( p < .0001 \)
Appendix 3: Analysis Choices for CT2

- Graphs for absolute differences come out messier, for the same reasons as in CT1

  effect of position/length*scaling on RT

  \[ p = .074 \quad \text{and} \quad p < .0005 \]