Reference game results and analysis

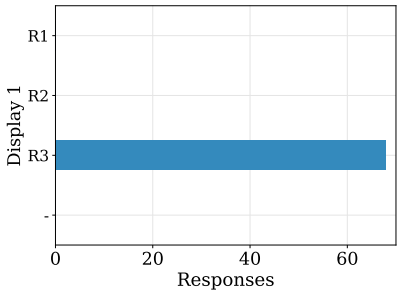
Chris Potts, Ling 130a/230a: Introduction to semantics and pragmatics, Winter 2024

Feb 22

This handout reports on the reference games experiment we did in class on February 13.

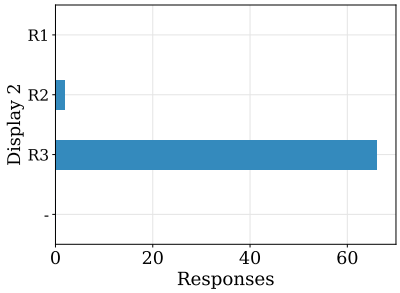
1 Results ($N = 68$)

1



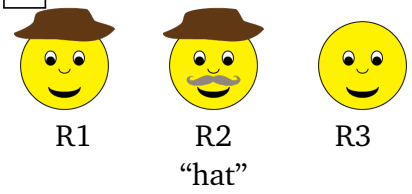
Purely truth conditional; expecting 'R3'.

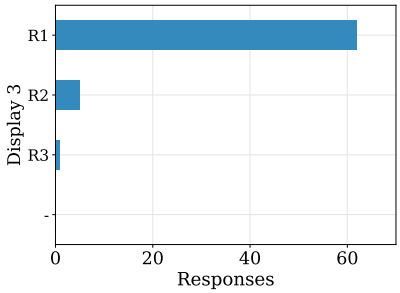
2



Purely truth-conditional; expecting 'R3'.

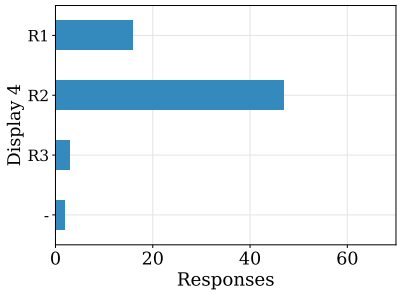
3





Expecting 'R1' because 'R2' could be 'mustache'.

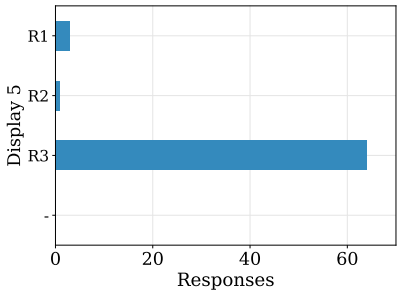
4



Impossible; maybe expecting 'R2' since others have named properties.

5

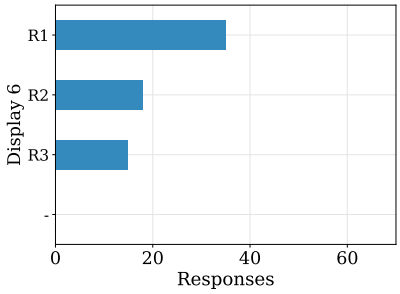
R1 R2 “glasses” R3



Expecting ‘R3’ because ‘R1’ could be ‘hat’

6

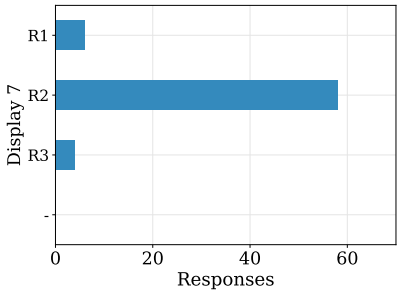
R1 R2 “glasses” R3



Unavoidable ambiguity; expecting ‘R1’ or ‘R3’

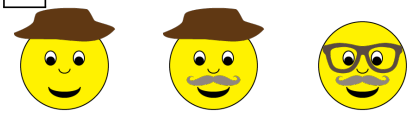
7

R1 R2 “mustache” R3

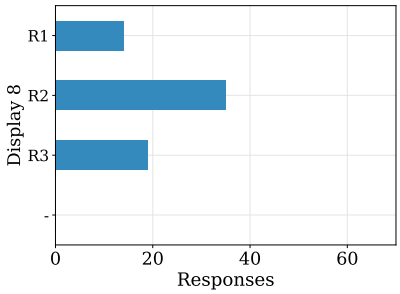


Expecting ‘R2’ because R3 could be ‘glasses’.

8



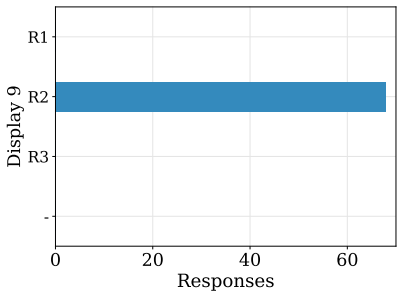
R1 R2 “mustache” R3



Very complex; in theory, expecting ‘R2’ because R1 is ‘hat’ and R3 is ‘glasses’.

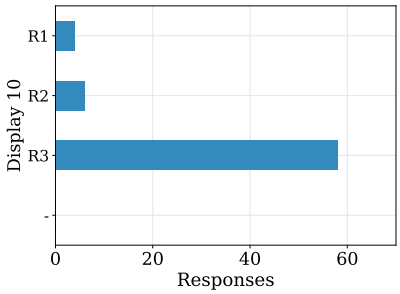
9

R1 R2 “glasses” R3



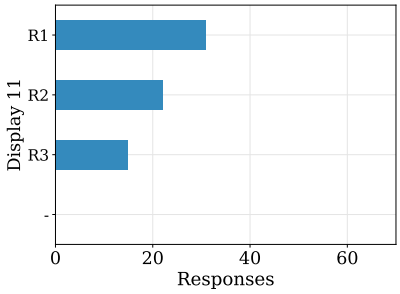
Purely truth conditional; expecting ‘R2’.

10



Expecting ‘R3’; prep for next item.

11



Very complex; in theory, expecting R1 because R3 is ‘mustache’, which makes R2 ‘hat’.

2 RSA analysis

2.1 Method

The RSA model is run on the individual scenarios, and we record the probabilistic predictions made by the literal listener and the pragmatic listener. For each of these agents, we concatenate all of the predictions into one long vector and compare them with the results from the experiment, arranged in the same way, as one long vector. This allows a correlation analysis.

2.2 Findings

We report on a pragmatic listener where the speaker agent has $\alpha = 4$. This high value does well with our *highly pragmatic* response data.

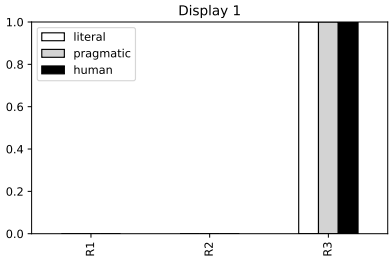
- Literal listener: correlation of 0.71 ($p < 0.0001$).
- Pragmatic listener: correlation of 0.87 ($p < 0.0001$).

Overall, the correlation is high for the pragmatic listener, and so it looks like the RSA model is a solid model of the data. However, there are certainly some unexplained aspects of the experimental data.

2.3 Individual cases

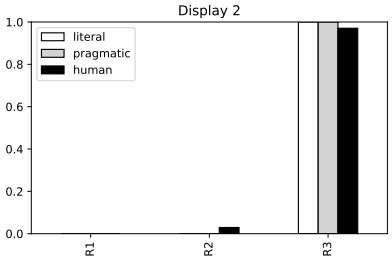
2.3.1 Purely truth conditional

Here, the literal and pragmatic listener predictions essentially completely align with the experimental responses:



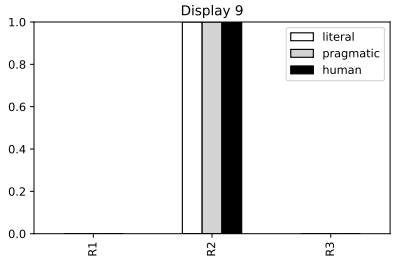
"hat"

Note: purely truth-conditional



"mustache"

Note: purely truth-conditional

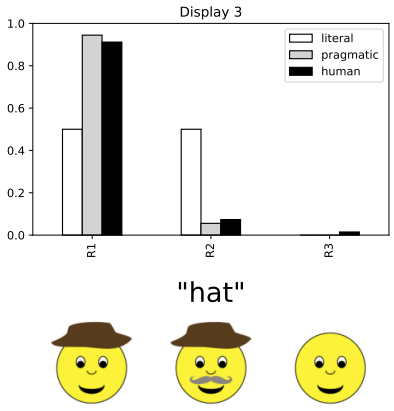


"glasses"

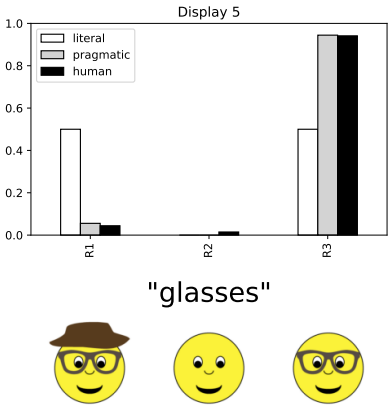
Note: purely truth-conditional

2.3.2 Standard implicature

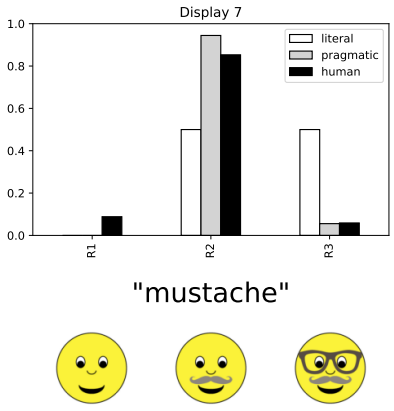
For these cases, the literal listener is very different from the experimental responses, as expected. By contrast, the pragmatic listener closely aligns with the responses overall.



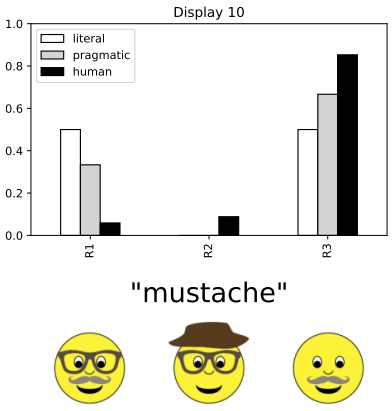
Note: basic scalar



Note: basic scalar



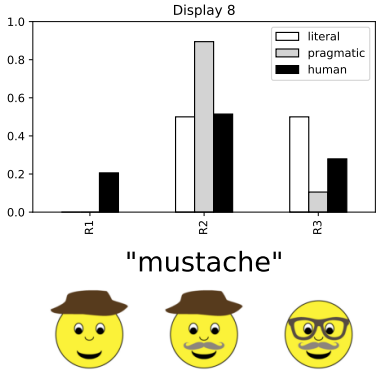
Note: basic scalar



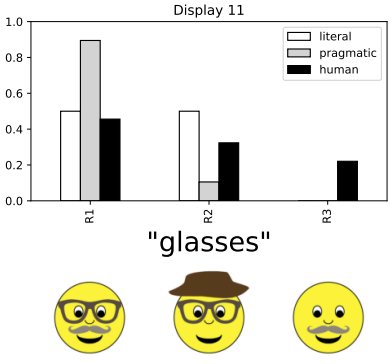
Note: basic scalar

2.3.3 Complex implicature

For these cases, the pragmatic listener appears to be “super-human”; the human responses seem split between pragmatic and more literal interpretations.



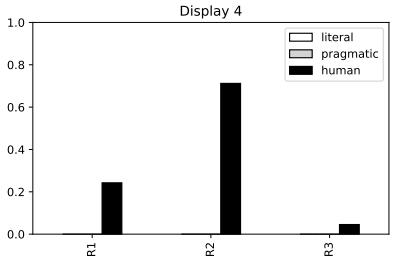
Note: complex scalar; expecting R2 because R1 is 'hat' and R3 is'glasses'



Note: complex; expecting R1 because R3 is 'mustache', creating scalar inference for R1 and R2

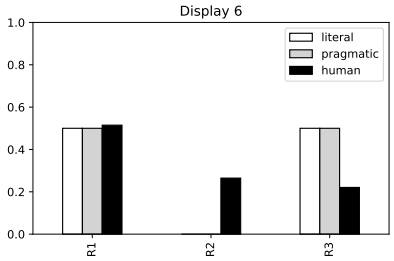
2.3.4 Non-literal signaling

For these cases, the literal and pragmatic models really have no hope of aligning with the human data. For the left case, we expect random responses from humans, whereas the actual data are pretty systematic. For the right case, we expect people to be split between R1 and R3, due to the unavoidable ambiguity, whereas we see evidence of very different strategies being employed.



"mustache"

Note: impossible



"glasses"

Note: unavoidable ambiguity