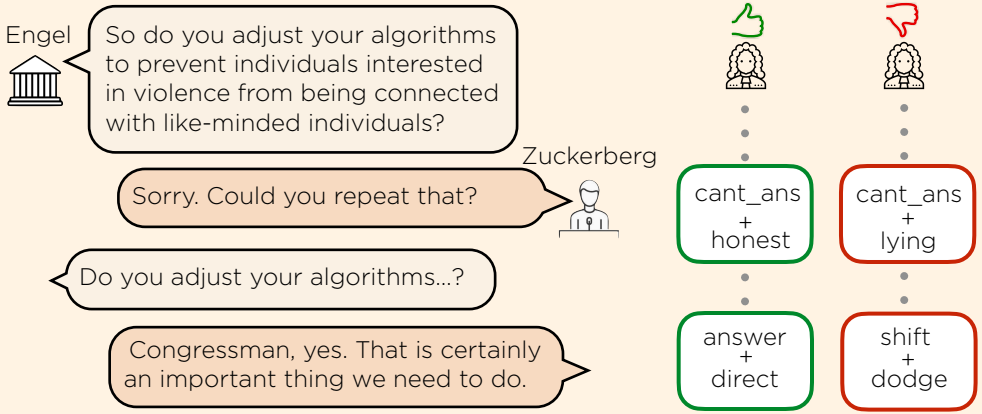
Did they answer? Subjective acts and intents in conversational discourse

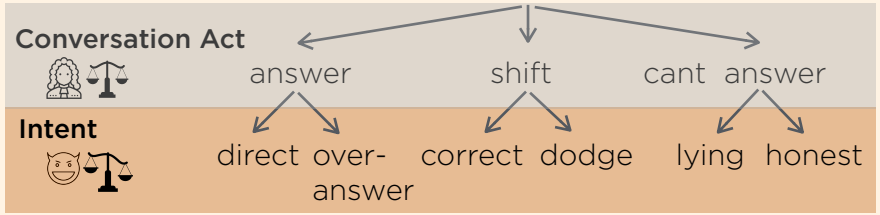
Elisa Ferracane, Greg Durrett, Jessy Li, Katrin Erk

The University of Texas at Austin

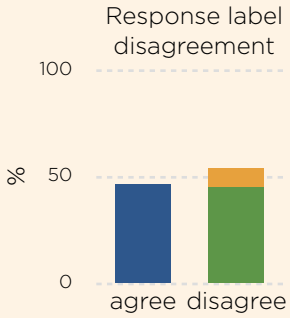
Dataset has *multiple* and *subjective* labels for interpreting each response:



Hierarchical taxonomy for response label:



Disagreements are frequent, with usually opposing intents:



IAA Krippendorff's α: 0.49

Conversation Act: 0.65

Intent: 0.38

Most frequent disagreements:

ans+direct vs. shift+dodge

shift+correct vs. shift+dodge

cant_ans+honest vs. cant_ans+lying

Taking into account annotator bias improves prediction of all labels:

Hierarchical: separate classifiers for CA and intent, with forced consistency during training

+Annotator: add annotator sentiment towards the witness