Can neural networks identify speaker commitment?

Marie-Catherine de Marneffe

The Ohio State University
Linguistics

Speaker commitment

🙋🏻‍♀️ People born before 1975 are now eligible for a vaccine.

🤷🏻‍♀️ People born before 1975 might now be eligible for a vaccine.

Inferences from speaker commitment are pervasive

(1) Did Sarah, <u>the CEO of Acme Inc</u>, call this morning?

(2) Did <u>Kim run</u> QUICKly?

(3) If anyone discovers that <u>the method is also wombat-proof</u>, they will want to scoop me.

Speaker commitment to content of clausal complements

At the heart of the universe there is cruelty. We are predators and are preyed upon, every living thing. Did you know that <u>wasps lay their eggs in ladybirds, piercing the weak spot in their armour</u>?

What is the source of speaker commitment to the content of clausal complements?

Factive [Karttunen 1971]

Does Judith know that Marie is in Brussels?

[i.a., Heim 1983, van der Sandt 1992, Geurts 1999, Potts 2005, Murray 2014]

What is the source of speaker commitment to the content of clausal complements?

Non-factive

Does Judith believe that Marie is in Brussels?

[i.a., Heim 1983, van der Sandt 1992, Geurts 1999, Potts 2005, Murray 2014]

Role of context in speaker commitment

If anyone discovers that <u>the method is also wombat-proof</u>, they will want to scoop me. 🙋

I haven't tried this with wombats though, and if anyone discovers that <u>the method is also wombat-proof</u>, I'd really like to know.

[Beaver 2010] 🤷

[i.a., Kadmon 2001, Simons 2001, Abrusán 2013, Simons et al. 2017]

CommitmentBank

[de Marneffe, Simons & Tonhauser SuB 2019]

CommitmentBank

Question
Negation
Conditional
Modal

Clause-embedding verb

Entailment-canceling operator

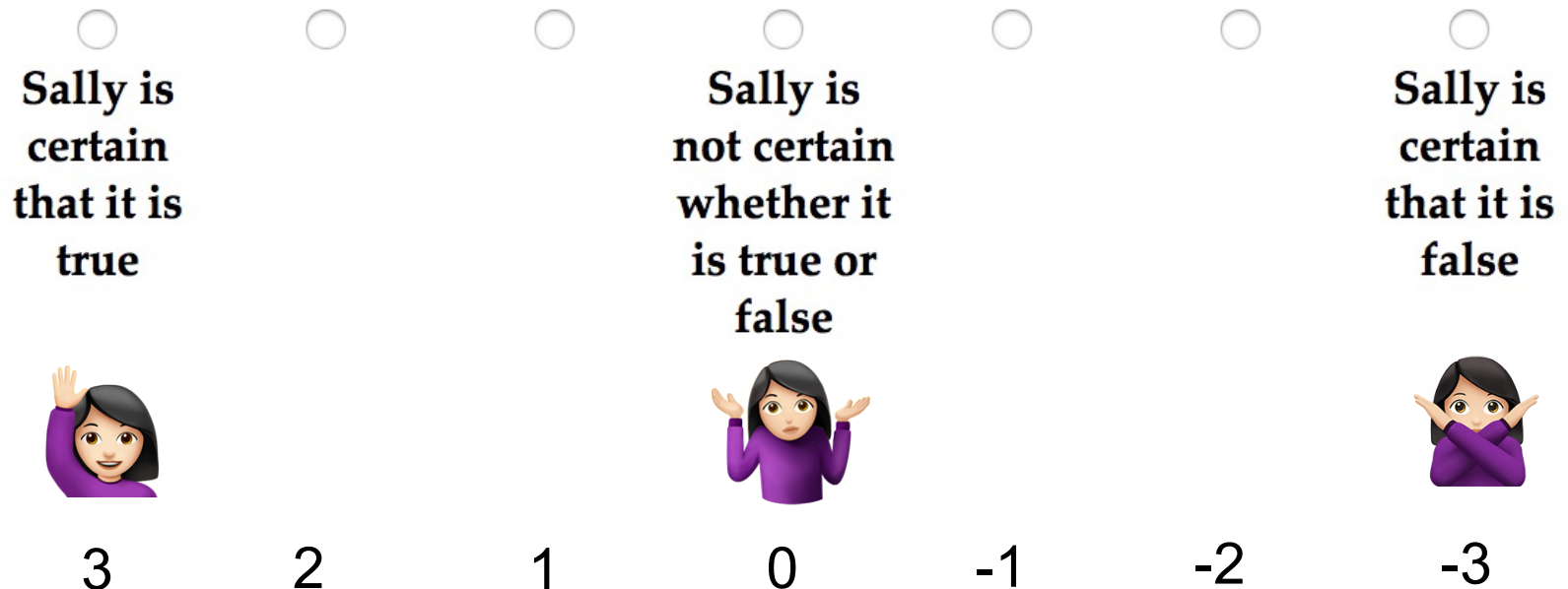Do you know that there is still a chance?

embedded clause

↑

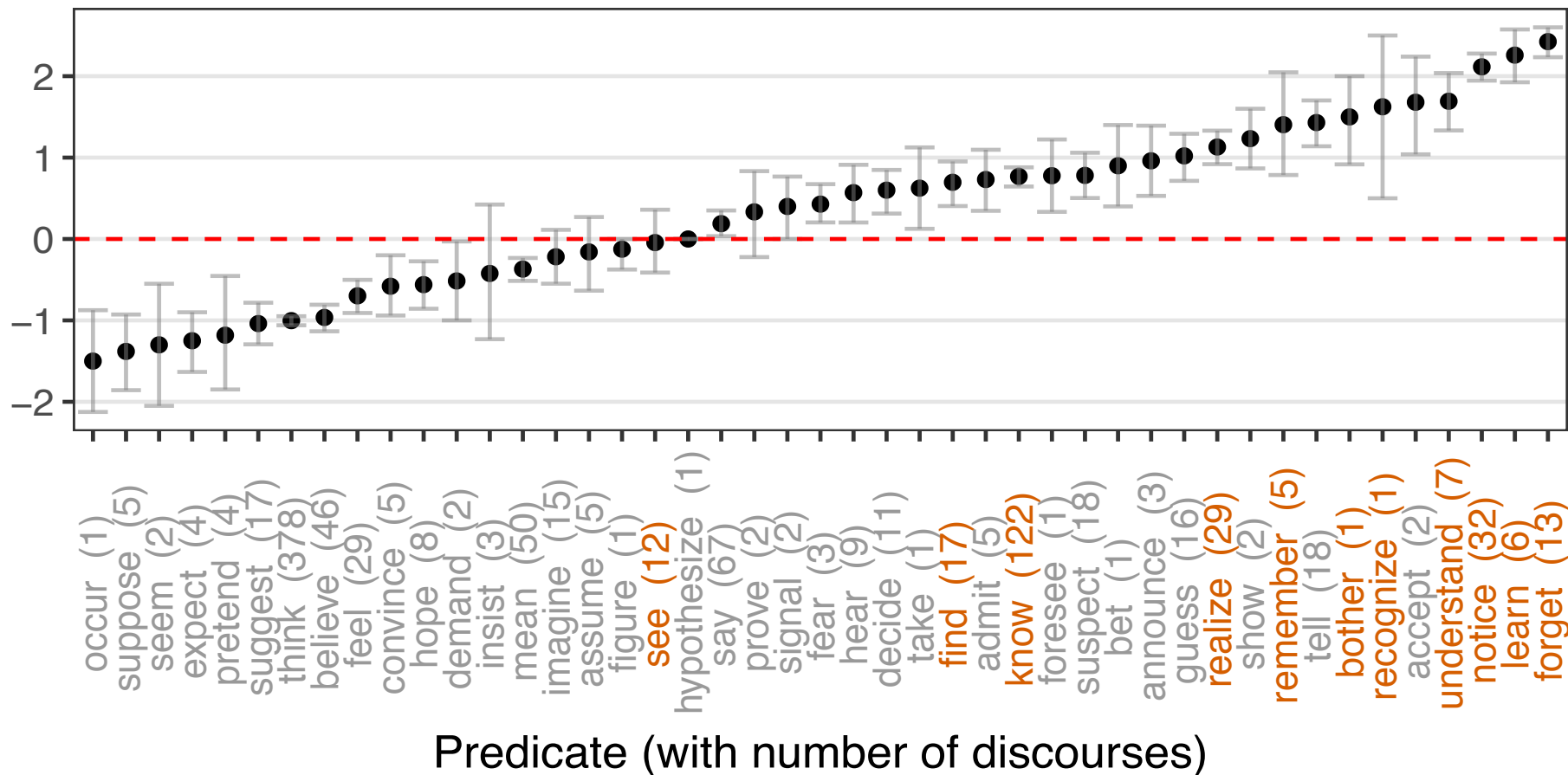commitment

Judgments obtained with MTurk

Sally: While the rest of the gang dived into the pub opposite to use the toilets, I called in at H. R. Higgins (Coffee-man) Ltd and bought six gift boxes of coffee (assorted) and two of tea (scented). That was my Christmas shopping sewn up. Who said it was stressful?

Tell us how certain Sally is that Christmas shopping was stressful.

Sally is certain that it is true			Sally is not certain whether it is true or false			Sally is certain that it is false
🙋🏻‍♀️			🤷🏻‍♀️			🙅🏻‍♀️
3	2	1	0	-1	-2	-3

Overall, gradience in speaker commitment

Mean speaker commitment

Predicate (with number of discourses)

occur (1)
suppose (5)
seem (2)
expect (4)
pretend (4)
suggest (17)
think (378)
believe (46)
feel (29)
convince (5)
hope (8)
demand (2)
insist (3)
mean (50)
imagine (15)
assume (5)
figure (1)
see (12)
hypothesize (1)
say (67)
prove (2)
signal (2)
fear (3)
hear (9)
decide (11)
take (1)
find (17)
admit (5)
know (122)
foresee (1)
suspect (18)
bet (1)
announce (3)
guess (16)
realize (29)
show (2)
remember (5)
tell (18)
bother (1)
recognize (1)
accept (2)
understand (7)
notice (32)
learn (6)
forget (13)

Context matters for speaker commitment

His hair was white, as my daughters reported when they went to view the body before it was given to the Odonata. Now he is known as The Man Who Changed the World, and there are statues to him everywhere.
No one remembers he had a younger brother. 🙋🏻‍♀️

Oh I did I did! I was lucky. I would have liked brothers and sisters but
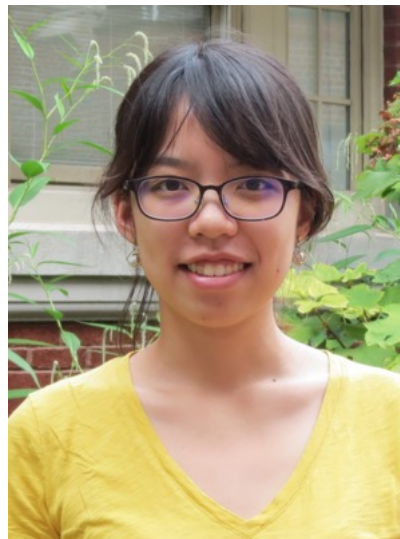**I don't remember that I was ever lonely.** 🤷🏻‍♀️

Machine Learning Confronts the Elephant in the Room | Quanta Magazine

quantamagazine.org

[Rosenfeld, Zemel & Tsotsos, arXiv 2018]

Evaluating state-of-the-art models of speaker commitment

[Jiang & de Marneffe, ACL 2019]

"Rule-based" TruthTeller

[Stanovsky et al. 2017]

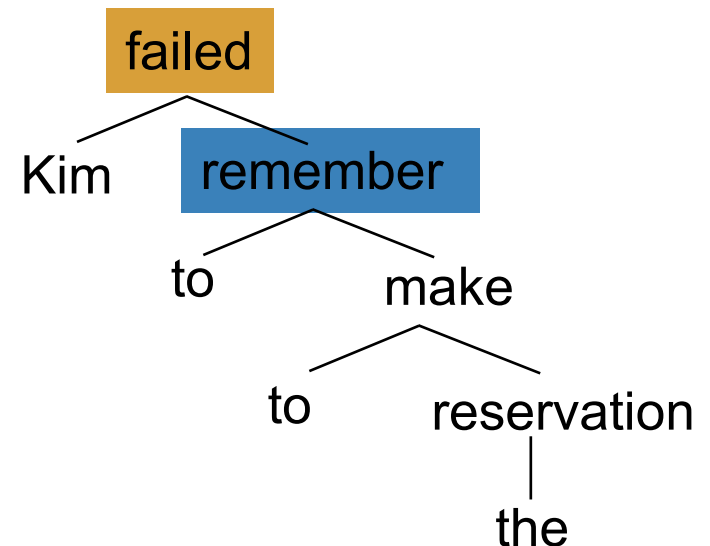Kim remembered to make the reservation.

Kim made the reservation. 🙋🏻‍♀️

[Nairn et al. 2006, Karttunen 2012]

"Rule-based" TruthTeller

[Stanovsky et al. 2017]

Kim failed to remember to make the reservation.

Kim made the reservation. 🙅🏻‍♀️

failed
Kim   remember
         to   make
              to   reservation
                         the

[Nairn et al. 2006, Karttunen 2012]
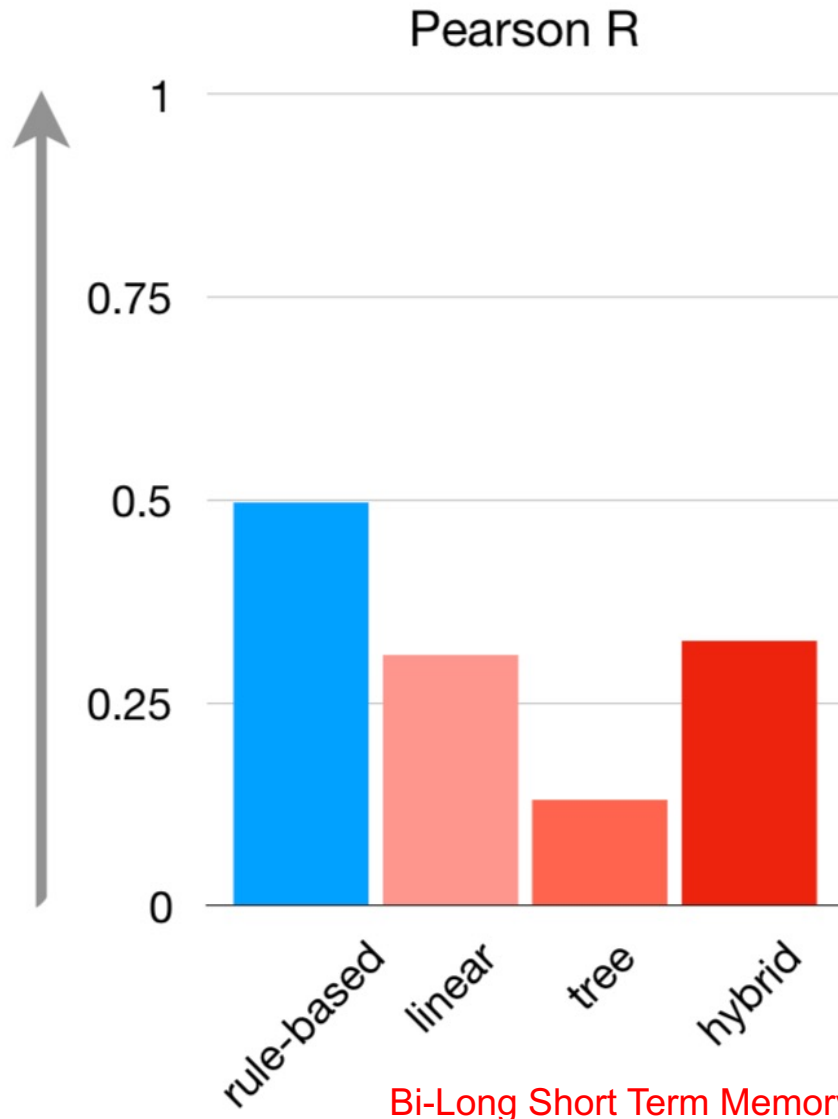
Bi-LSTM models

[Rudinger et al. 2018]

3 typologies:

- linear

- tree

- hybrid

4 speaker commitment datasets

[-3, 3] annotations

- FactBank

- UW

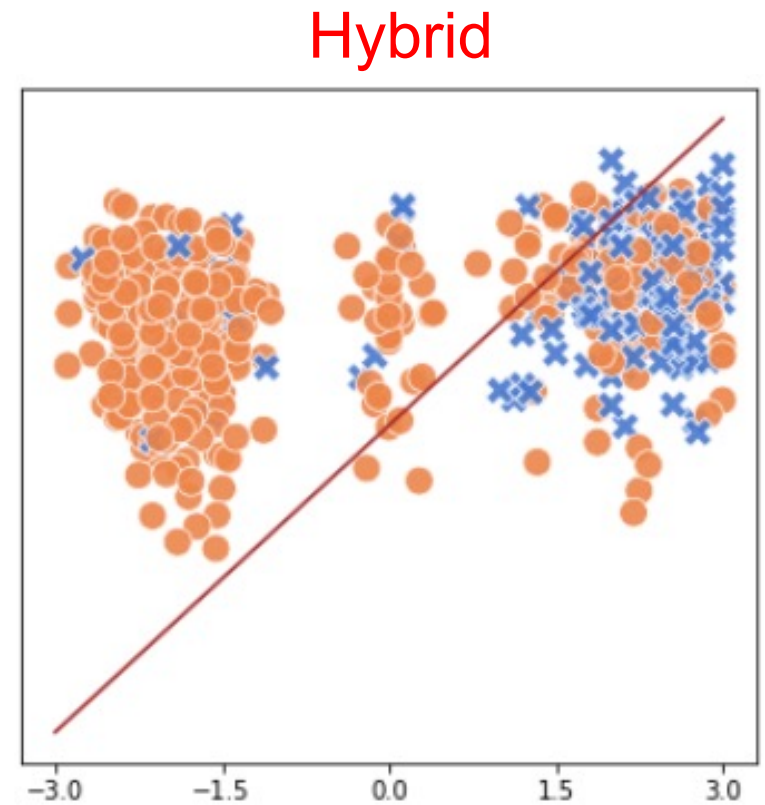- MEANTIME

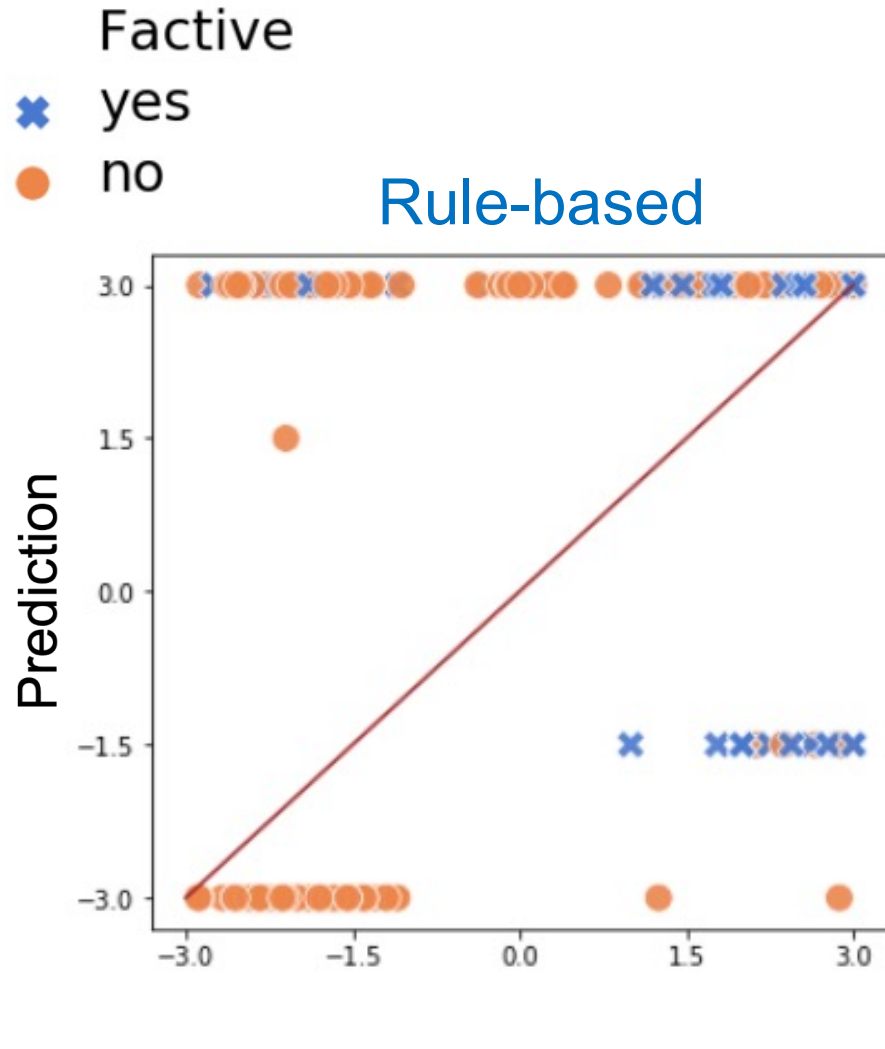- It Happened

Performance is not great overall

Pearson R

Bi-Long Short Term Memory

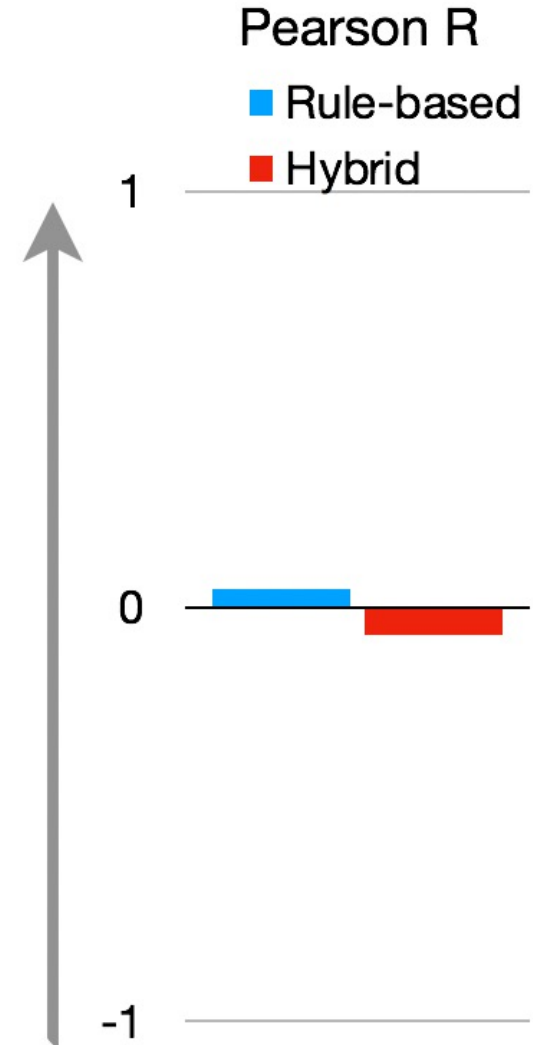[Stanovsky et al. 2017]   [Rudinger et al. 2018]
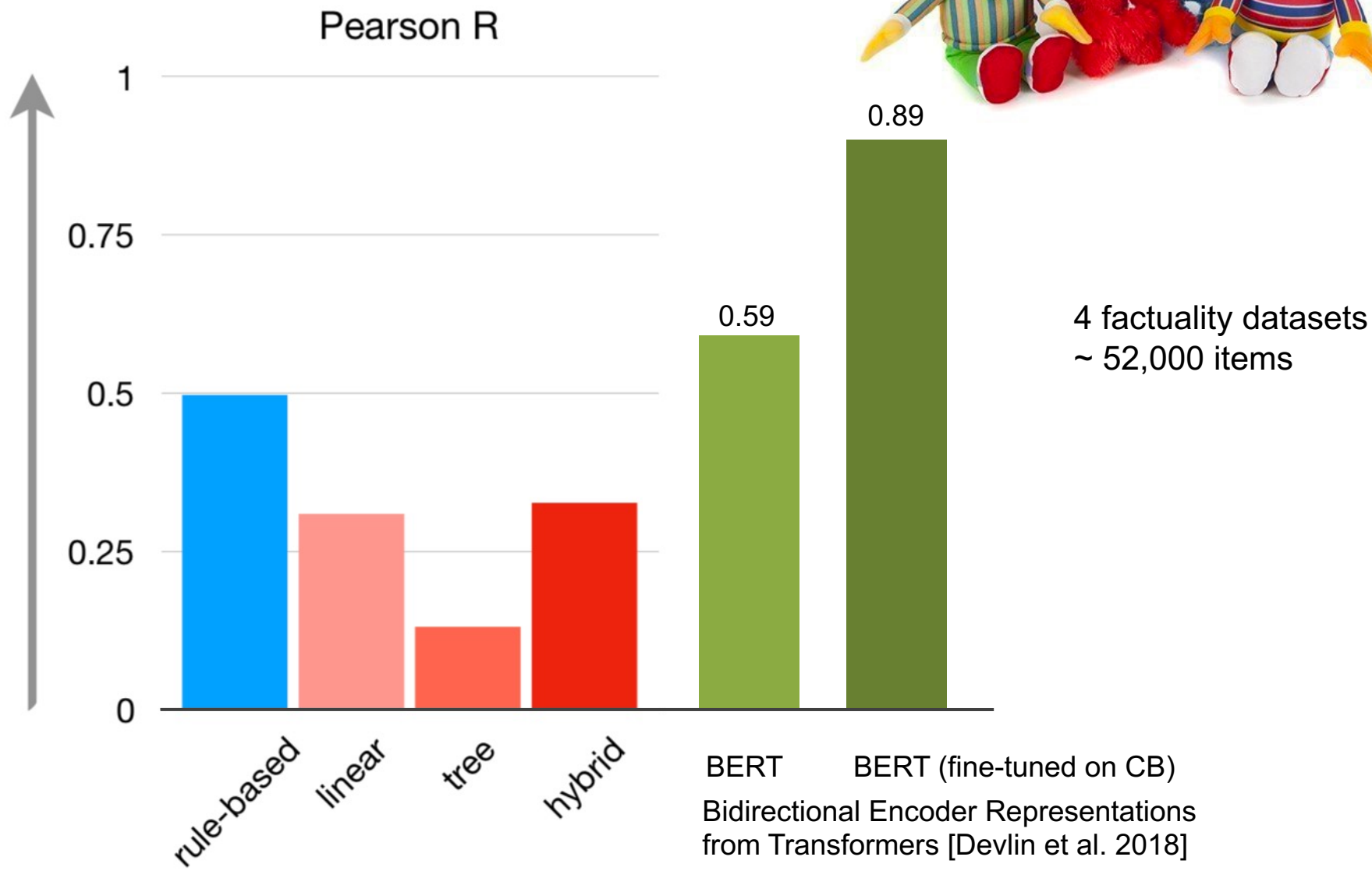
Models cannot deal with the gradience

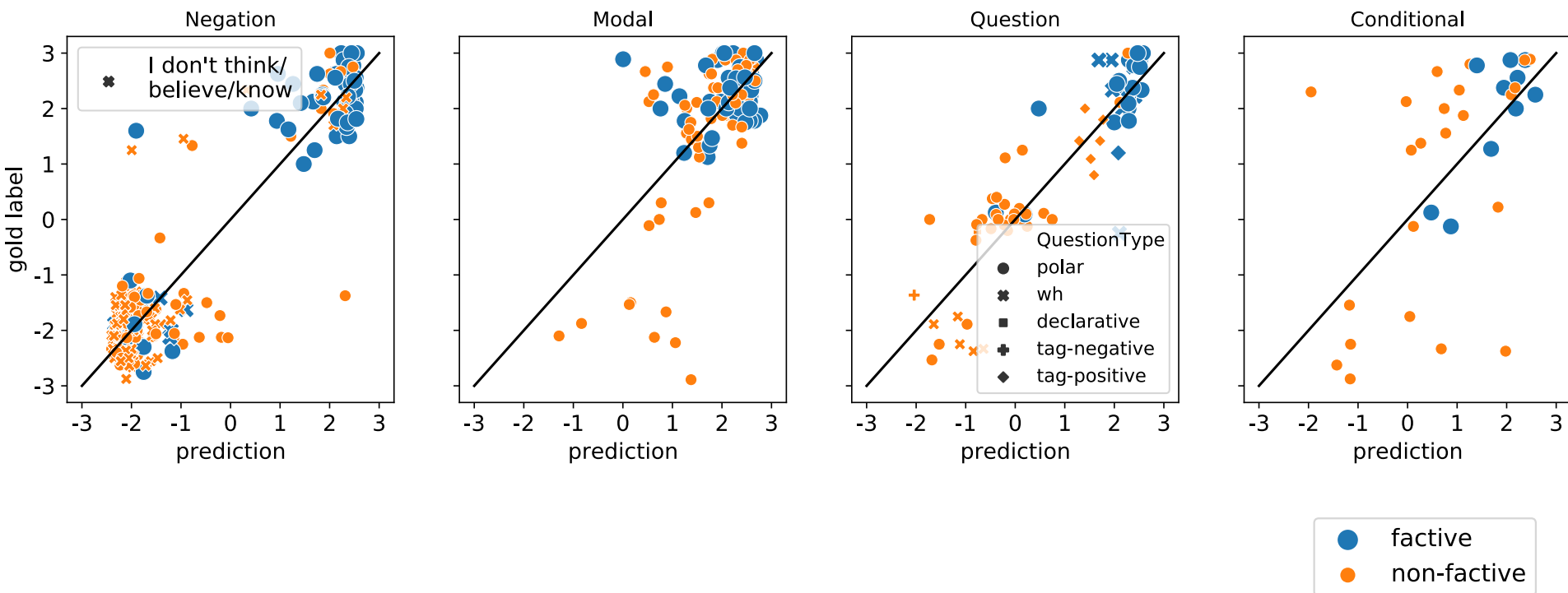# Neither models handle "neg-raising"

I don't think that's a great idea.

That's a great idea. 🙅🏻‍♀️

Pearson R
- 🟦 Rule-based
- 🟥 Hybrid

How does BERT fare?

Pearson R

4 factuality datasets
~ 52,000 items

rule-based   linear   tree   hybrid   BERT   BERT (fine-tuned on CB)

Bidirectional Encoder Representations
from Transformers [Devlin et al. 2018]

BERT learns patterns really well

BERT learns patterns really well

Negation · Modal · Question · Conditional

I don't think/believe/know

QuestionType
- ● polar
- ✕ wh
- ■ declarative
- ✚ tag-negative
- ◆ tag-positive

● factive
● non-factive
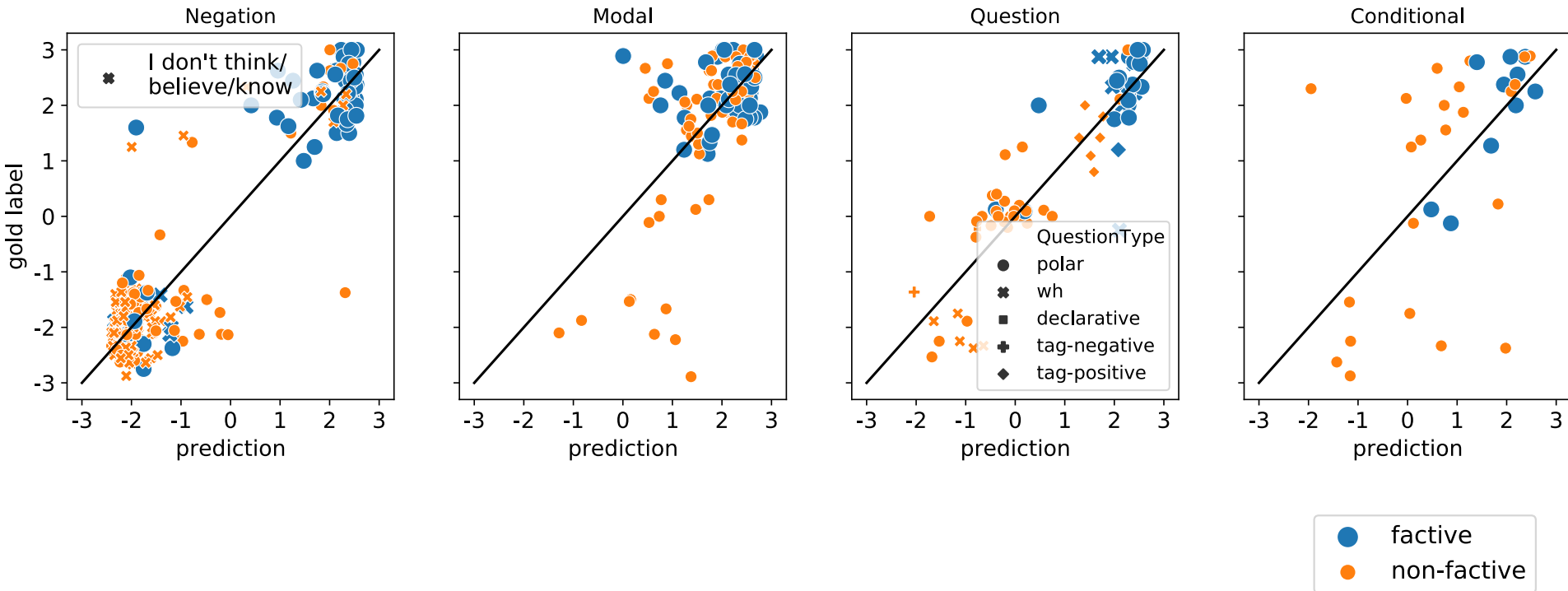
[…] I think it went to Lockheed, didn't it? 1.09 [1.52]

What goes wrong?

[…] He took the scuffed leather document case off the seat beside him and banged the door shut with the violence of someone who had not learned that car doors do not need the same sort of treatment as those of railway carriages.

Gold annotation: 2.63
Prediction: [0.96]

In a column lampooning Pat Buchanan, Royko did not write that Mexico was a useless country that should be invaded and turned over to Club Med.

Gold annotation: -3
Prediction: [-0.3]

BERT doesn't understand language

[…] And I think society for such a long time said, well, you know, you're married, now you need to have your family and I don't think it's been until recently that they had decided that two people was a family. 1.25 [-1.99]

B: All right, well. A: Um, short term, I don't think anything's going to be done about it or probably should be done about it. B: Right. Uh, are you saying you don't think anything should be done in the short term? 0 [-1.73]

I have myself devised many staff plans over the years and I do not believe I am being unduly boastful if I say that very few ever needed amendment. 2.3 [-1.94]

BERT doesn't understand language

[…] And I think society for such a long time said, well, you know, you're married, now you need to have your family and I don't think it's been until recently that they had decided that two people was a family. 1.25 [-1.99]

B: All right, well. A: Um, short term, I don't think anything's going to be done about it or probably should be done about it. B: Right. Uh, are you saying you don't think anything should be done in the short term? 0 [-1.73]

I have myself devised many staff plans over the years and I do not believe I am being unduly boastful if I say that very few ever needed amendment. 2.3 [-1.94]

Reminiscent of Chaves' work on filler-gap

[Chaves 2020]

"Are these models [LSTM RNN] actually learning filler-gap dependencies or are they simply learning surface-based contingencies that have little to do with the underlying syntactic and semantic mechanisms that cause island phenomena?"

*Which country weren't you born in _ ?

Which country haven't you visited _ yet?

*How fast didn't John drive _ ?

How fast is John required not to drive _ ?

Wrapping up

Linguistically-motivated features, conjoined with surface-level ones, are necessary to enable NLP applications

Speaker commitment to the content of complement clauses is a complex phenomenon that matters for correct information extraction/opinion mining, and we don't have yet a good handle on it

Thank you!
Merci !

Assistants:

Jemmin Chang, Willy Cheung, Tess Harty, Nanjiang Jiang, Taylor Mahler, Jack Parker, Abigail Salmon, Katherine Zhang