Harnessing the richness of the linguistic signal in predicting pragmatic inferences

Judith Degen

18.6.19

Interaction and the Evolution of Linguistic Complexity Workshop

Stanford University

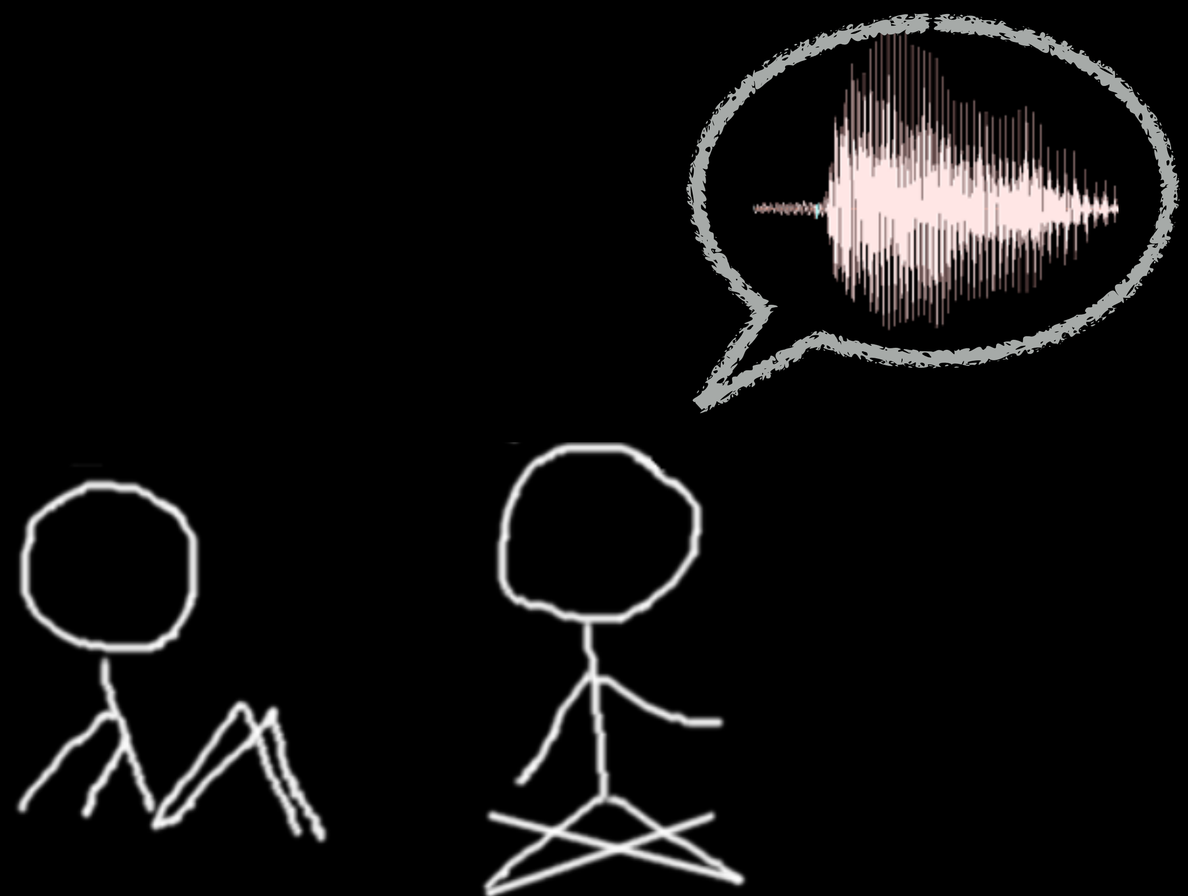
interActive Language Processing Lab Stanford (ALPS)

Harnessing the richness of the linguistic signal in predicting pragmatic inferences

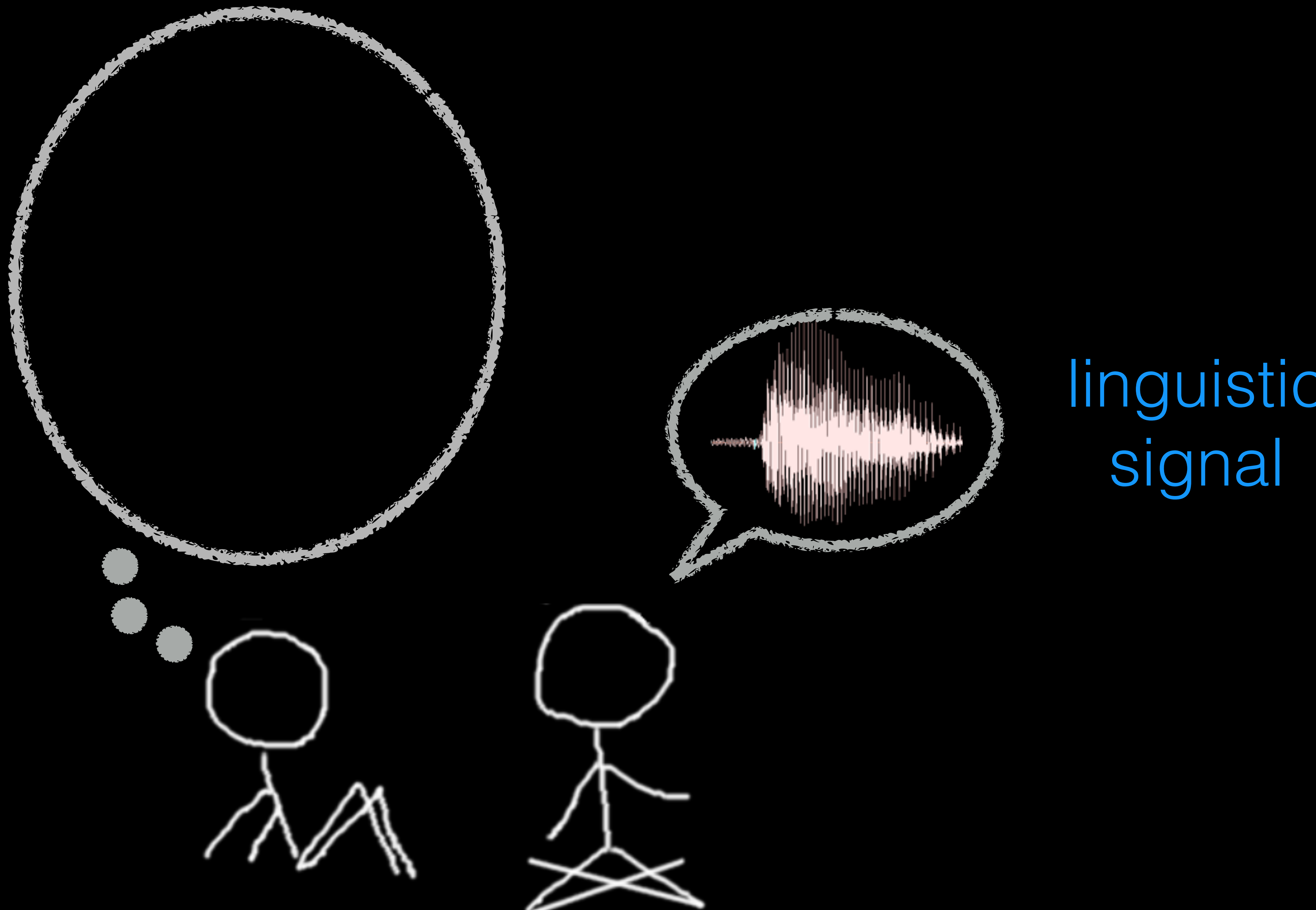
Judith Degen

18.6.19

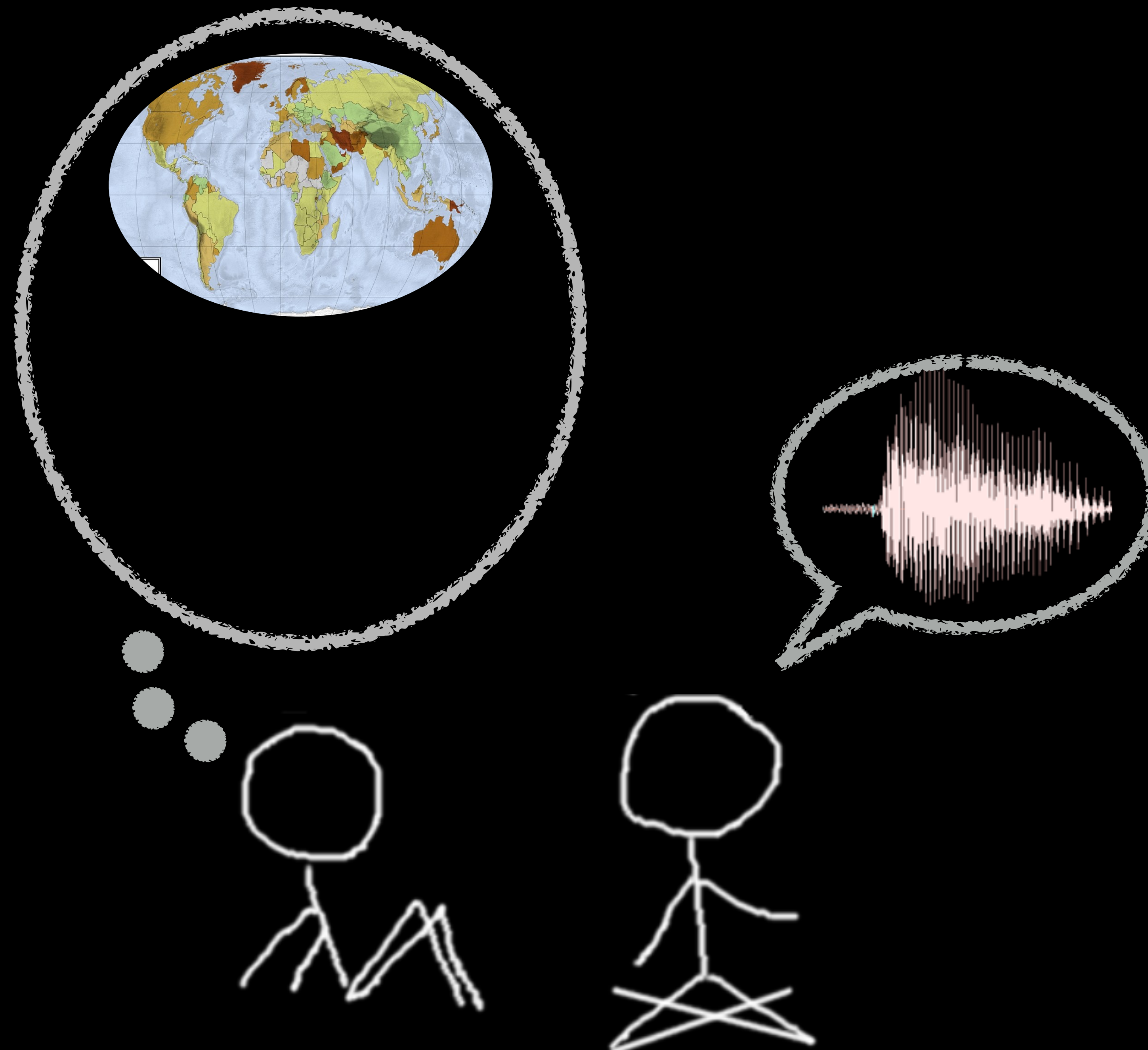
Interaction and the Evolution of Linguistic **Complexity** Workshop



linguistic
signal


linguistic
signal

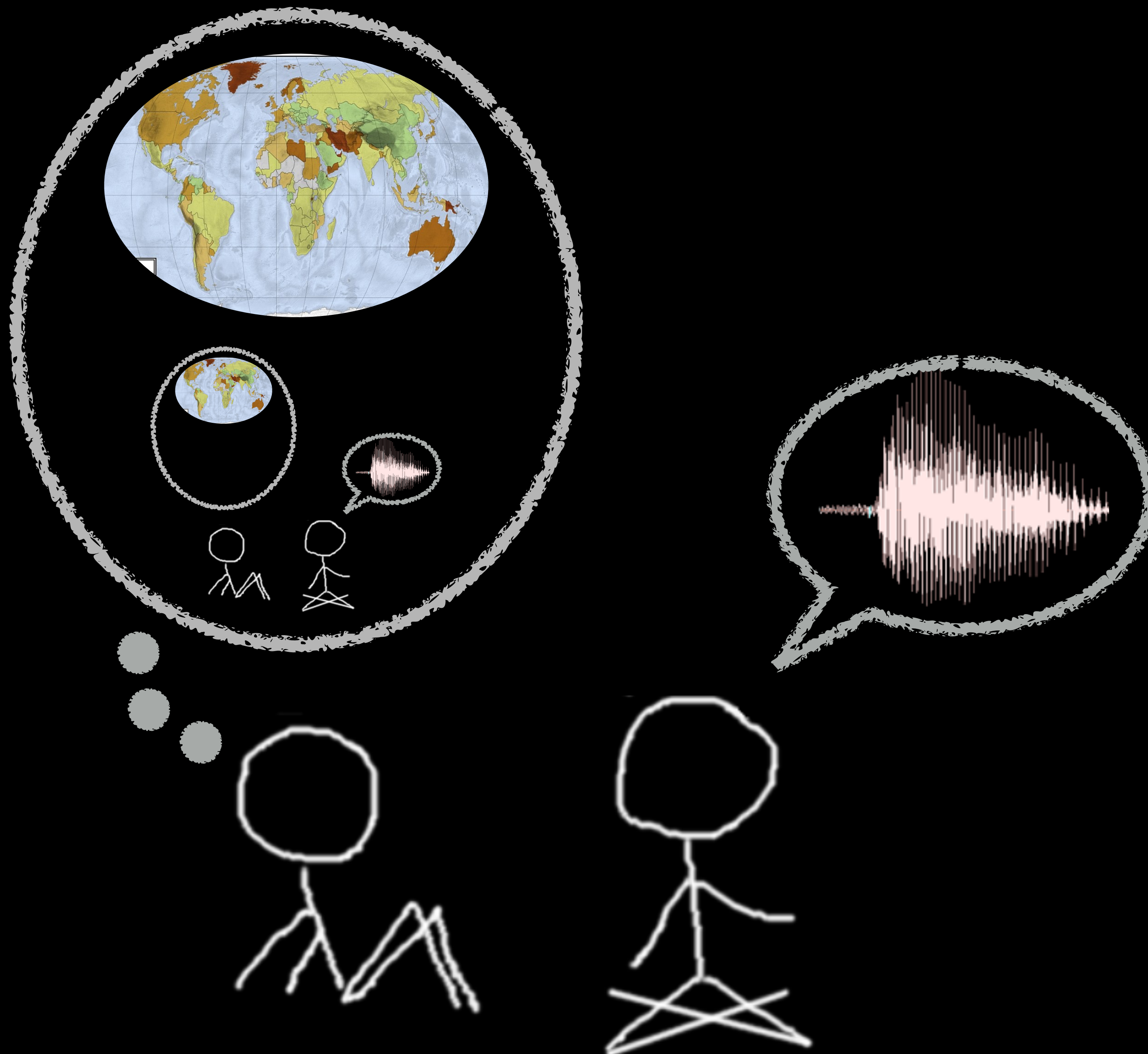
world
knowledge



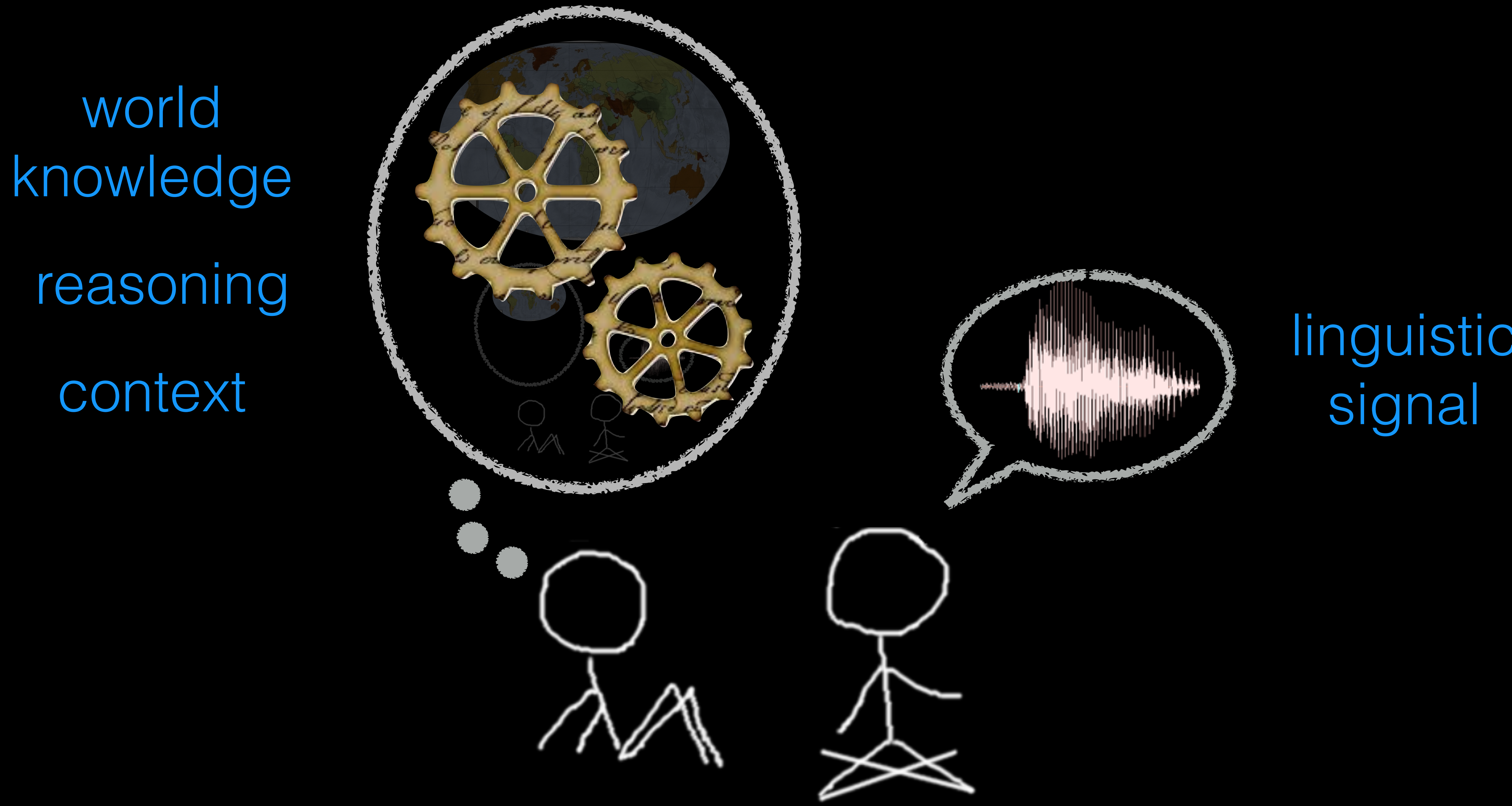
linguistic
signal

world knowledge

context



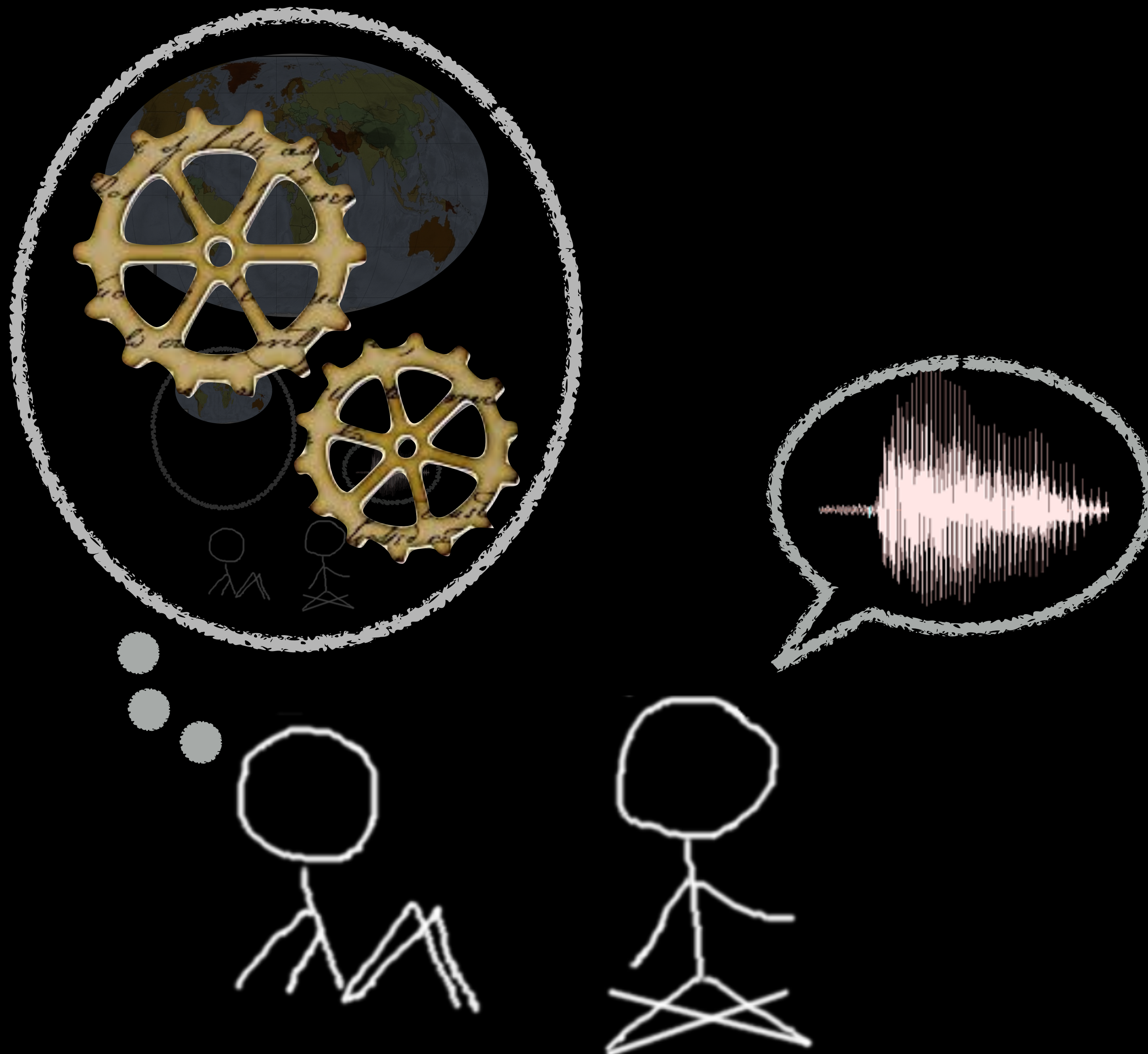
linguistic signal


world
knowledge
reasoning
context
linguistic
signal

world knowledge

reasoning

context



linguistic signal

PRAGMATICS

Scalar implicature

(1) John: Was the exam easy?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.

Scalar implicature

(1) John: Was the exam easy?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.

(2) John: Who came to the party?
Mary: Ann or Greg.
Inference: Either Ann or Greg came, but not both.

Scalar implicature

(1) John: Was the exam easy?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.

(2) John: Who came to the party?
Mary: Ann or Greg.
Inference: Either Ann or Greg came, but not both.

(3) John: How was your date?
Mary: It was OK.
Inference: The date was OK, but not great.

Why study scalar implicature?

(1) John: Was the exam easy?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.
Inference: The exam was not easy.

Why study scalar implicature?

(1) John: Was the exam easy?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.
Inference: The exam was not easy.

(1) John: Is the teacher doing a good job?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.
~~**Inference**: The exam was hard.~~
Inference: The teacher isn't doing a good job.

Why study scalar implicature?

(1) John: Was the exam easy?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.
Inference: The exam was not easy.

ROBUST

(1) John: Is the teacher doing a good job?
Mary: Some of the students failed.
Inference: Some, but not all of the students failed.
~~**Inference**: The exam was hard.~~
Inference: The teacher isn't doing a good job.

Accounts of scalar implicature

Accounts of scalar implicature

The default account

Levinson 2000

Basic assumptions:

- context is hard to integrate

Solution: two types of inferences

- fast, automatic, context-independent inferences
  Generalized Conversational Implicature
- slow, effortful, context-dependent inferences
  Particularized Conversational Implicature

Accounts of scalar implicature

The default account
Levinson 2000
Basic assumptions:
- context is hard to integrate

Solution: two types of inferences
- fast, automatic, context-independent inferences
  Generalized Conversational Implicature
- slow, effortful, context-dependent inferences
  Particularized Conversational Implicature

A contextualist account
Degen & Tanenhaus 2015
Basic assumptions:
- context is easy to integrate

Solution: efficient use of context
- listeners acquire a context-dependent speaker model: P(utterance | context, meaning)
- listeners use available contextual cues to infer speaker meaning: P(meaning | utterance, context)

Accounts of scalar implicature

The default account

Levinson 2000

Basic assumptions:

- context is hard to integrate

Solution: two types of inferences

- fast, automatic, context-independent inferences
  Generalized Conversational Implicature
- slow, effortful, context-dependent inferences
  Particularized Conversational Implicature

A contextualist account

Degen & Tanenhaus 2015

Basic assumptions:

- context is easy to integrate

Solution: efficient use of context

- listeners acquire a context-dependent speaker model: P(utterance | context, meaning)
- listeners use available contextual cues to infer speaker meaning: P(meaning | utterance, context)

Accounts of scalar implicature

??CONTEXT??

The default account

Levinson 2000

Basic assumptions:

- context is hard to integrate

Solution: two types of inferences

- fast, automatic, context-independent inferences
  Generalized Conversational Implicature
- slow, effortful, context-dependent inferences
  Particularized Conversational Implicature

A contextualist account

Degen & Tanenhaus 2015

Basic assumptions:

- context is easy to integrate

Solution: efficient use of context

- listeners acquire a context-dependent speaker model: P(utterance | context, meaning)
- listeners use available contextual cues to infer speaker meaning: P(meaning | utterance, context)

Sources of data in experimental pragmatics

- historically: introspective judgments
- judgment data from controlled experiments
- processing data from controlled experiments

Variability in scalar implicature

attributed to

- properties of the scale van Tiel et al 2016
- stress on cognitive system de Neys & Schaeken 2007
- idiosyncratic properties of participants
- context for a review: Degen & Tanenhaus 2019

What’s lacking

- a clear picture of the naturalistic contexts that speakers produce scalar expressions in
- a clear picture of whether listeners make use of the contextual information available to them in naturalistic contexts

Overview

1. A study combining corpus analysis & web-based experiments on “some”
2. Using distributed meaning representations to predict human inference ratings

There is much more variability in scalar inferences than commonly assumed — but it’s systematically context-dependent, and we can capture a lot of it by inspecting the naturalistic signal!

Case study: “some”

Scalar implicatures in the wild

Degen 2015

1. I like **some country music**.
2. It would certainly help them to appreciate **some of the things we have here**.
3. You sound like you have **some small ones** in the background.

Scalar implicatures in the wild

Degen 2015

1. I like **some country music**.

   Inference? I like some, but not all, country music

2. It would certainly help them to appreciate **some of the things we have here**.

   Inference? …to appreciate some, but not all…

3. You sound like you have **some small ones** in the background.

   Inference? … some, but not all small ones…

Combining corpora & the web

1. extracted all 1390 utterances containing *some* from the Switchboard corpus of spoken American English

2. collected inference strength ratings for each item on Mechanical Turk (10 judgments per item)

Speaker A: i mean, they just have beautiful, beautiful homes and they have everything. the kids only wear name brand things to school and it's one of these things,

Speaker B: oh me. well that makes it hard for you, doesn't it.

Speaker A: well it does, you know. it really does because i'm a single mom and i have a thirteen year old now and uh, you know, it does.

Speaker B: oh, me.

Speaker A: i mean, we do it to a point but uh, not to where she feels different ,

Speaker B: yeah.

Speaker A:
but some of them are very rich

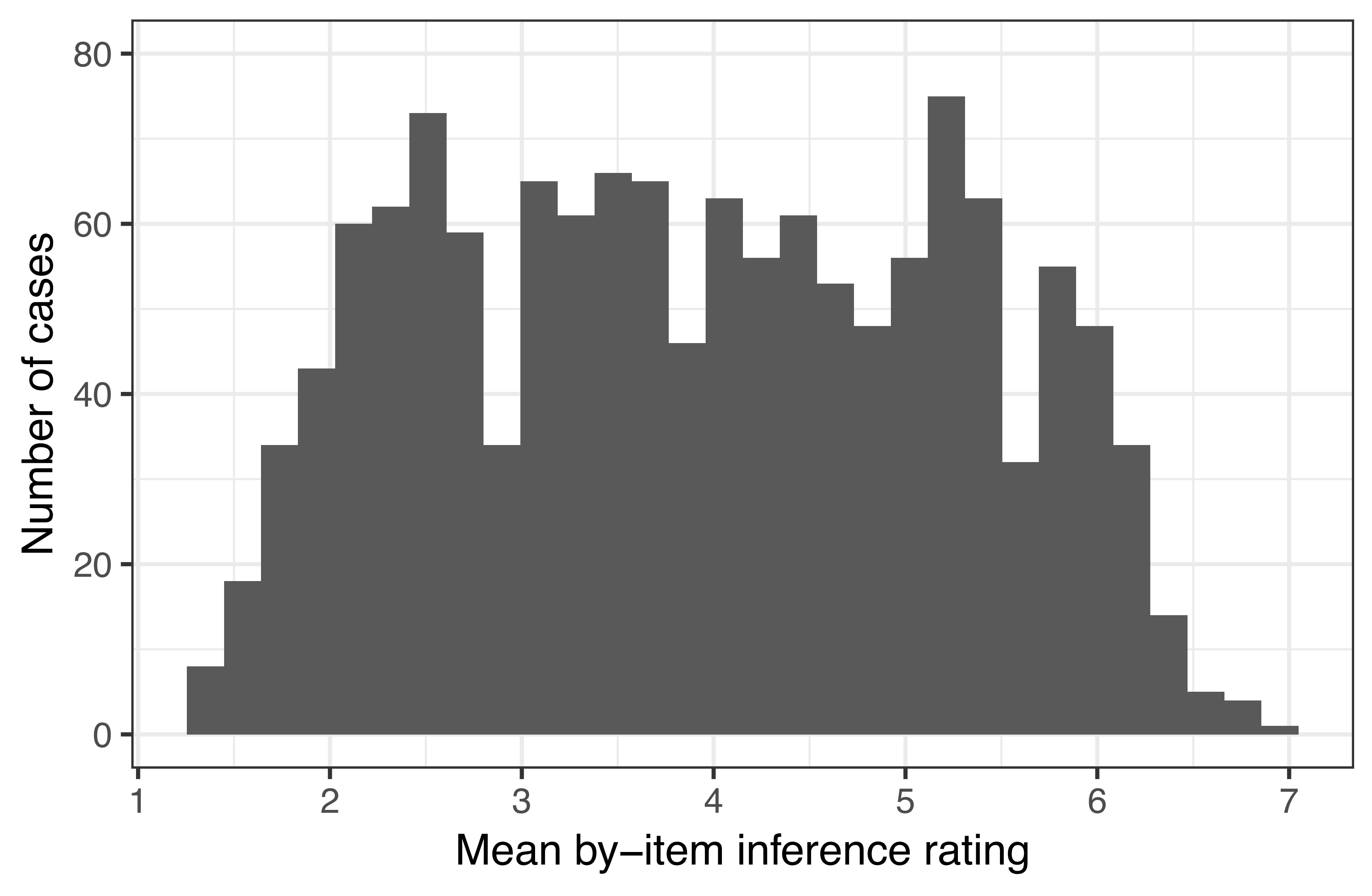
but **some, but not all** of them are very rich

How similar is the statement with 'some, but not all' (green) to the statement with 'some' (red)?

Very different meaning	1	2	3	4	5	6	7	Same meaning

Continue

Variability in inference strength



large amount of variability in inference strength

Just noise?

Qualitative investigation

1. I like **some country music**.

6.9

2. It would certainly help them to appreciate **some of the things we have here**.

4

3. You sound like you have **some small ones** in the background.

1.5

Qualitative investigation

1. I like **some country music**.

6.9 **Inference?** I like some, but not all, country music

2. It would certainly help them to appreciate **some of the things we have here**.

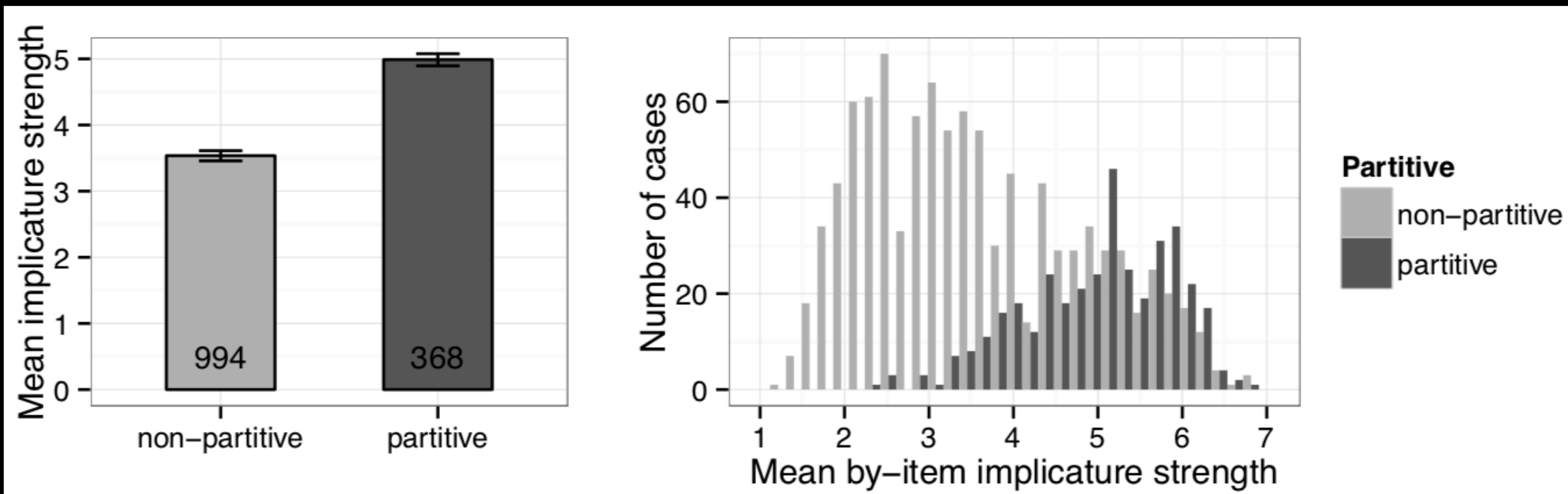
4 **Inference?** …to appreciate some, but not all…

3. You sound like you have **some small ones** in the background.

1.5 **Inference?** … some, but not all small ones…

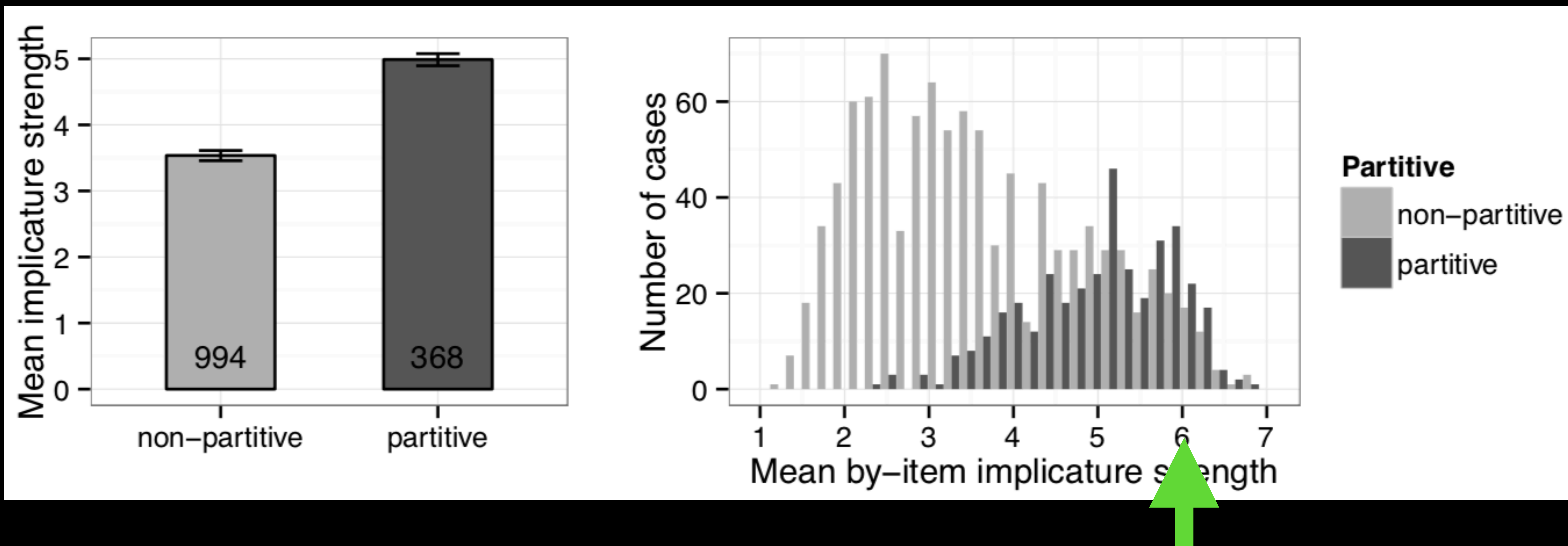
Stronger inferences…

…with **partitive** *some*-NPs.



Stronger inferences…

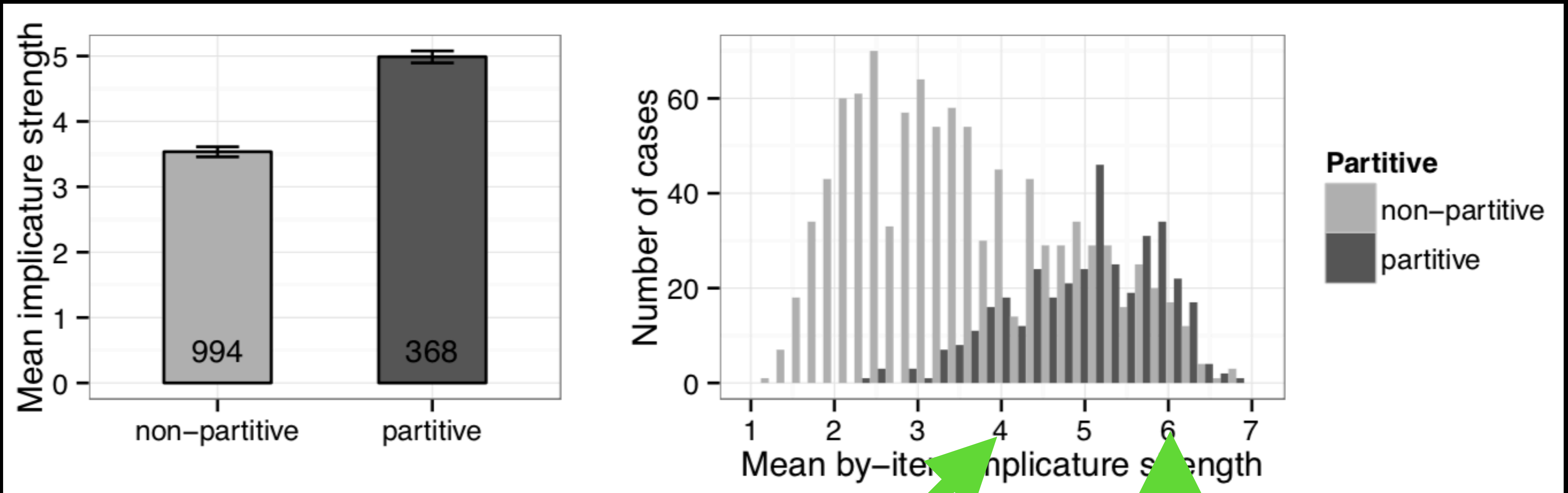
…with **partitive** *some*-NPs.



*I've seen **some of them** on repeats*

Stronger inferences…

…with **partitive** *some*-NPs.

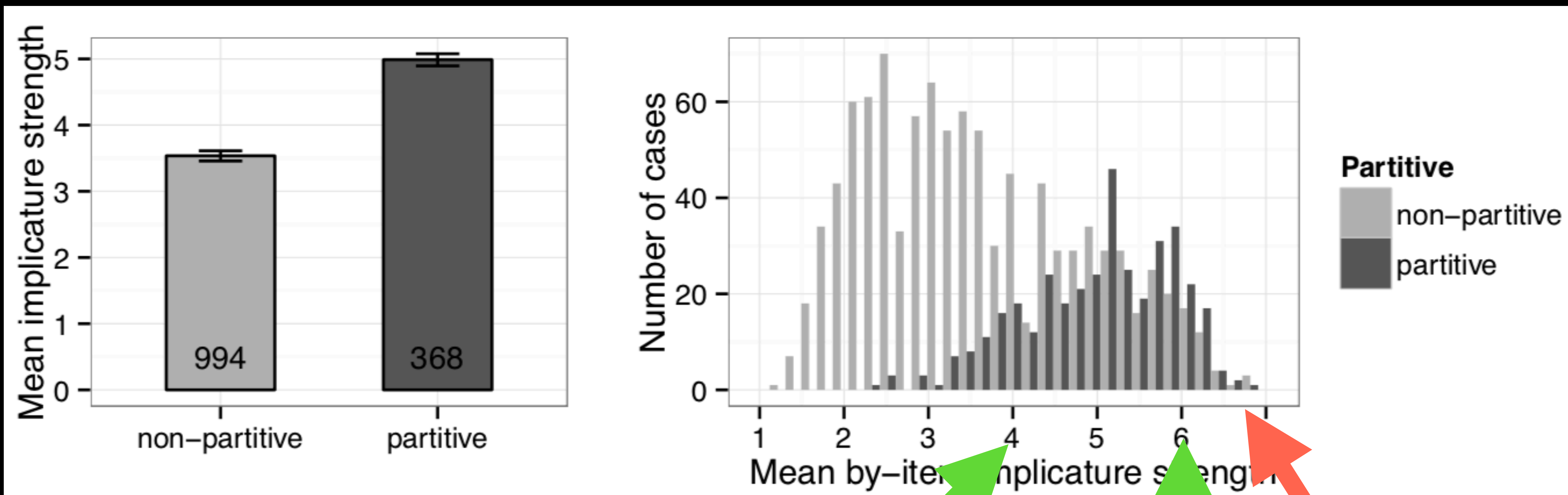


I've seen ***some of them*** *on repeats*

It would certainly help them to appreciate
some of the things we have here*.*

Stronger inferences…

…with **partitive** *some*-NPs.



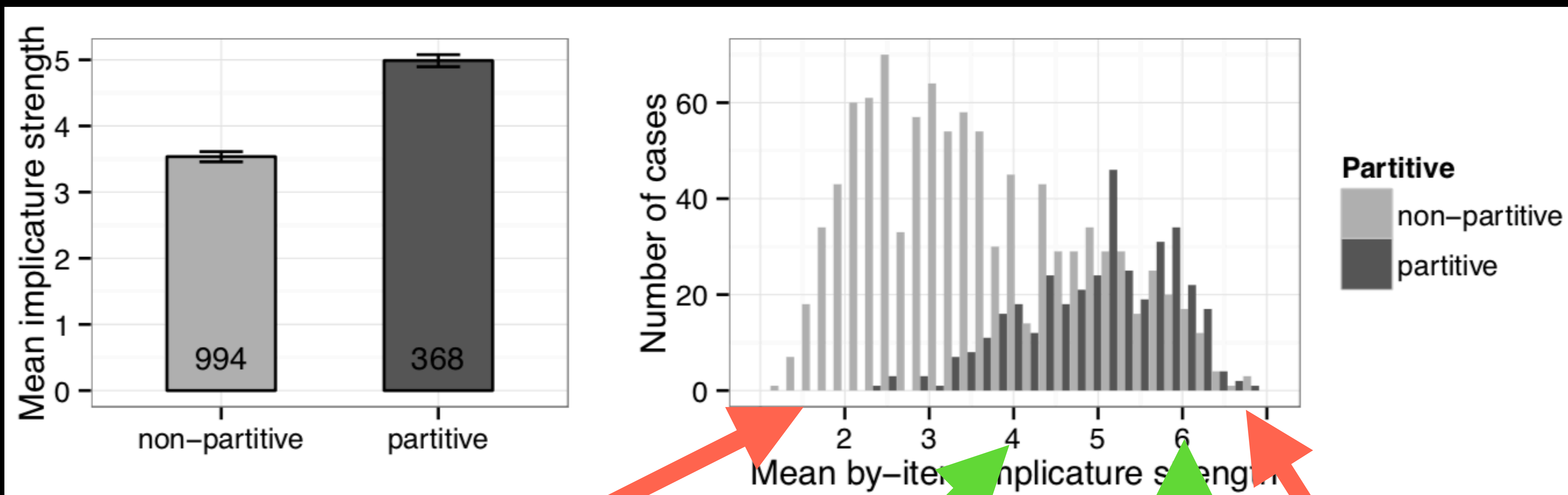
I've seen ***some of them*** *on repeats*

It would certainly help them to appreciate ***some of the things we have here****.*

I like ***some country music.***

Stronger inferences…

…with **partitive** *some*-NPs.



*I've seen **some of them** on repeats*

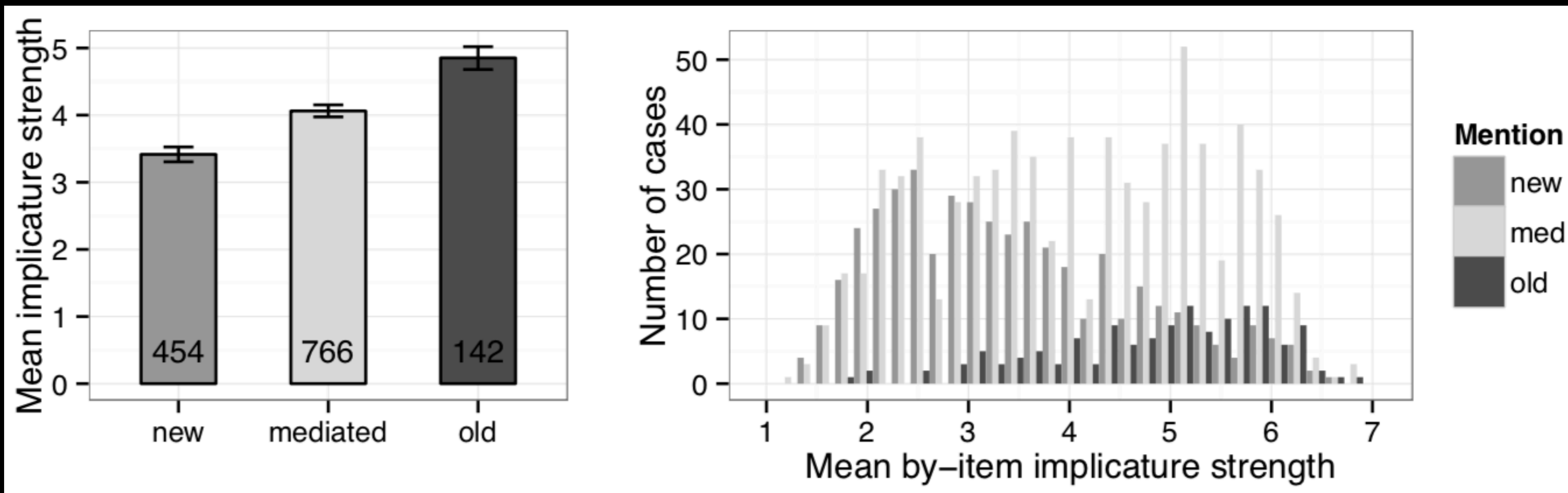
*It would certainly help them to appreciate **some of the things we have here**.*

*You sound like you have **some small ones** in the background.*

*I like **some country music.***

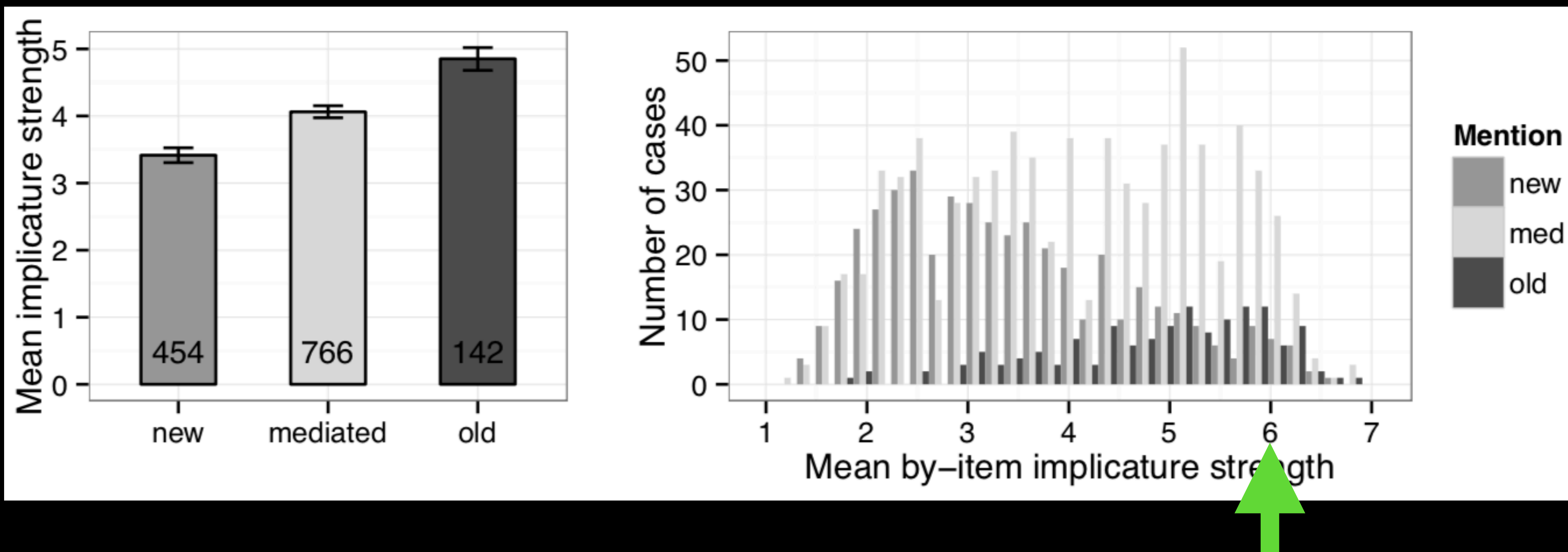
Stronger inferences…

…with **previously mentioned** NP referents.



Stronger inferences…

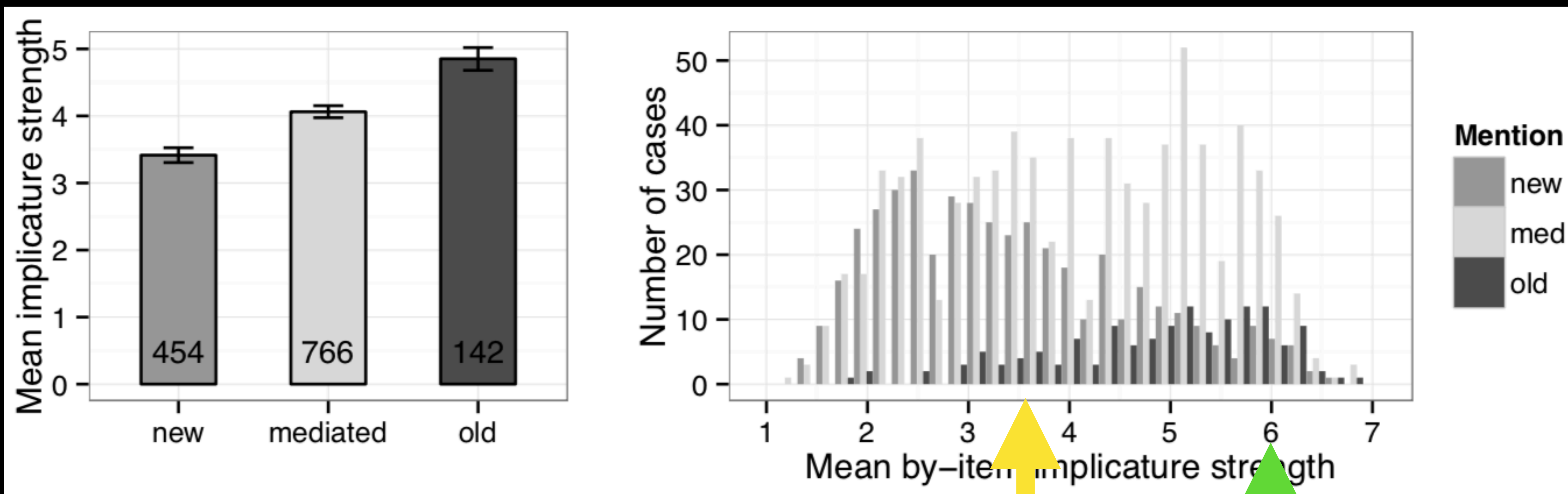
…with **previously mentioned** NP referents.



*I've seen **some of them** on repeats*

Stronger inferences…

…with **previously mentioned** NP referents.

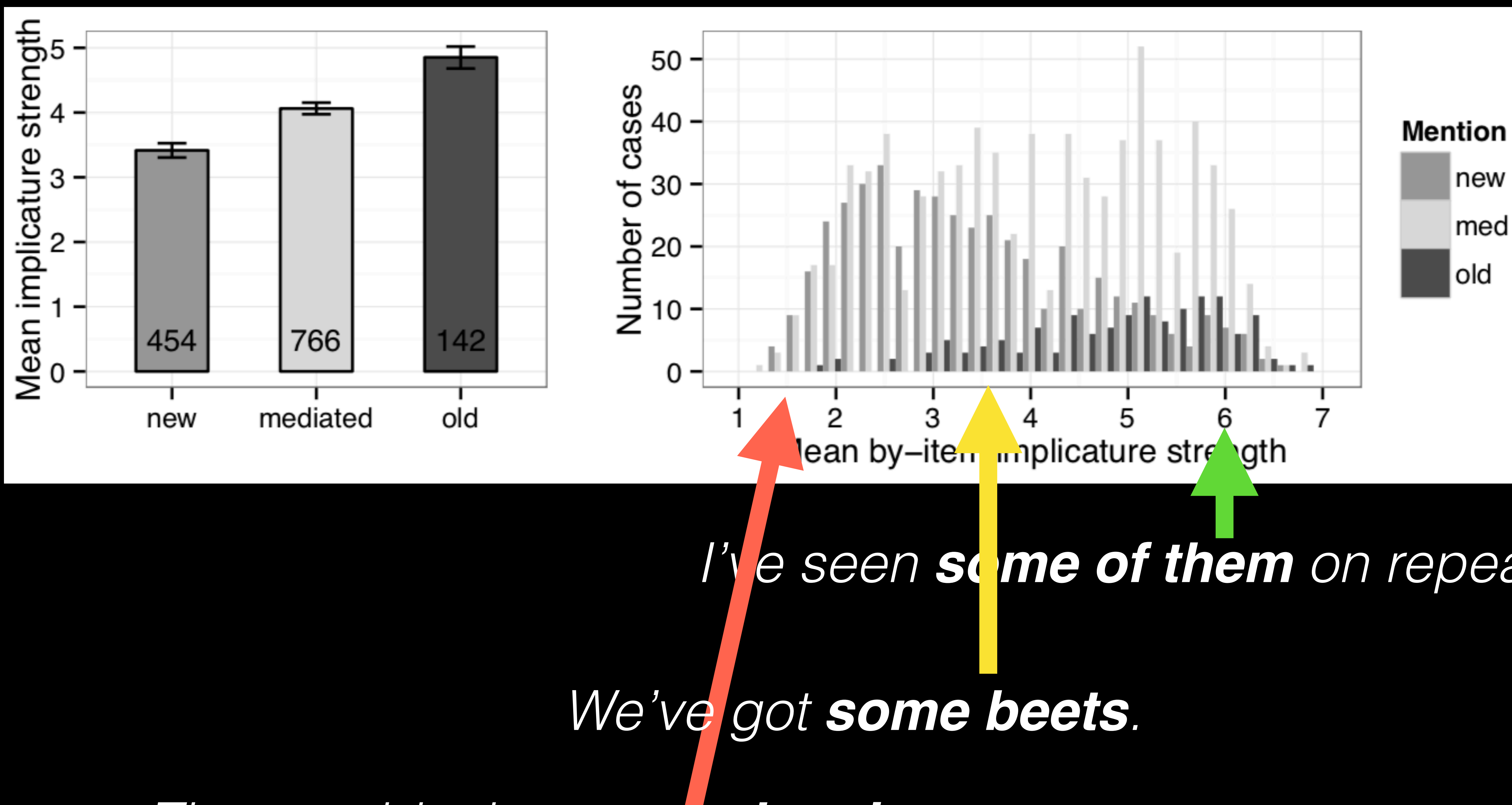


*I've seen **some of them** on repeats*

*We've got **some beets**.*

Stronger inferences…

…with **previously mentioned** NP referents.



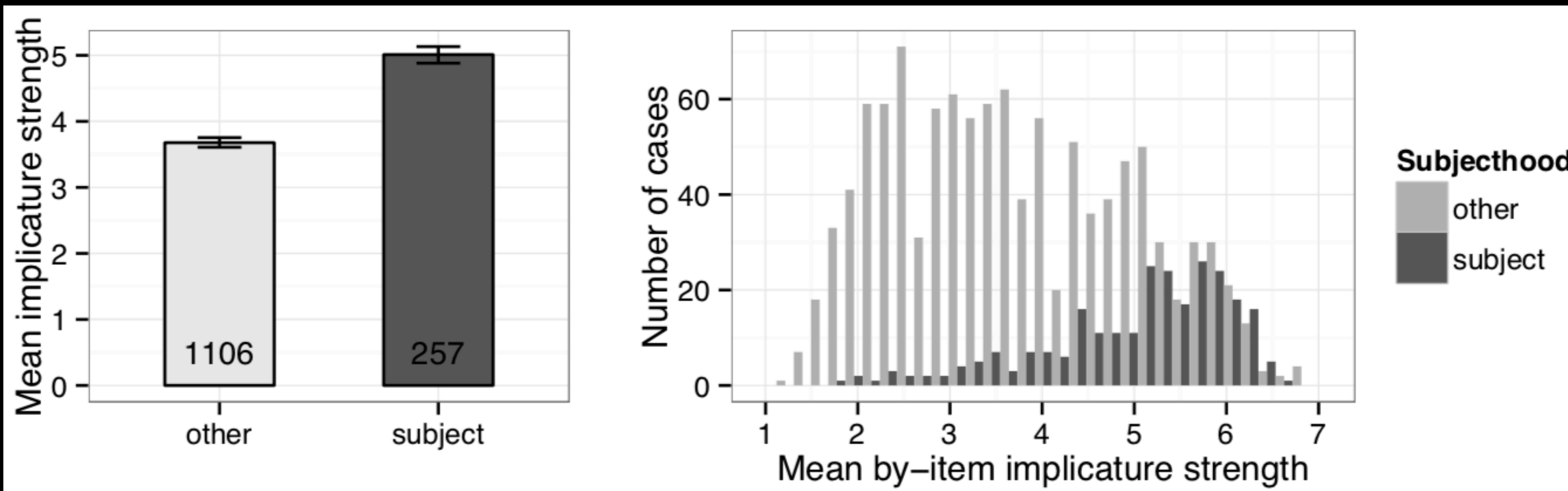
*I've seen **some of them** on repeats*

*We've got **some beets**.*

*That would take **some planning**.*

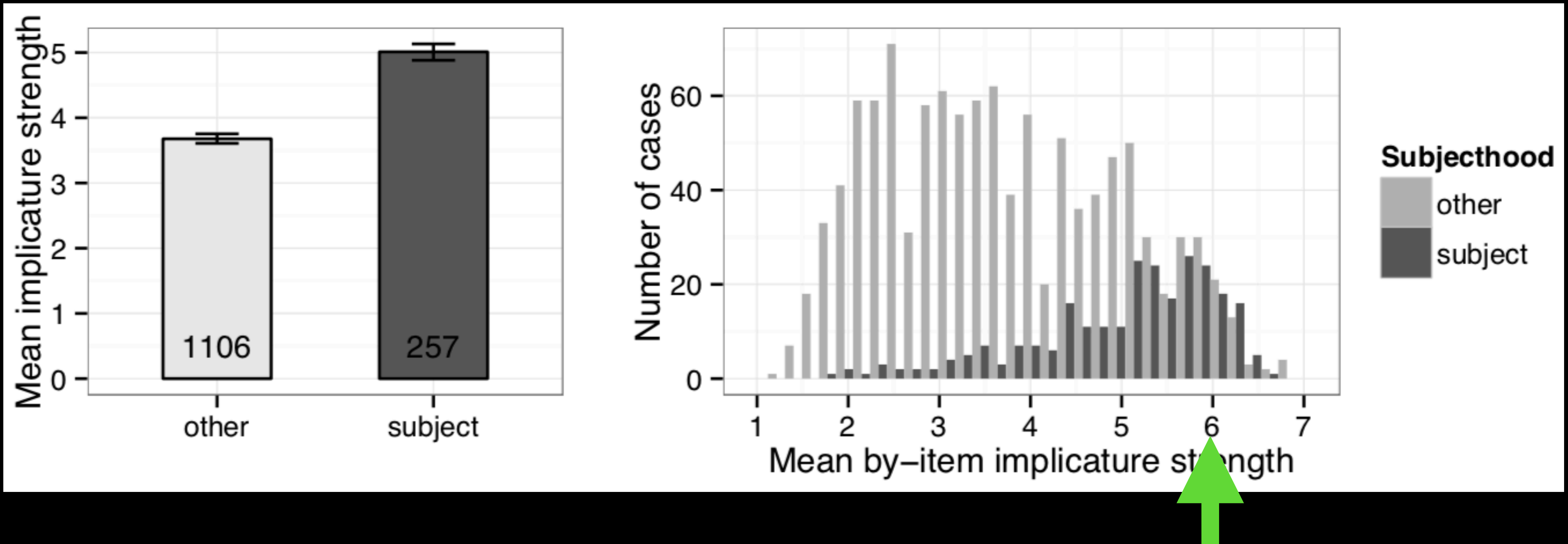
Stronger inferences…

…with *some*-NPs in **subject** position.



Stronger inferences…

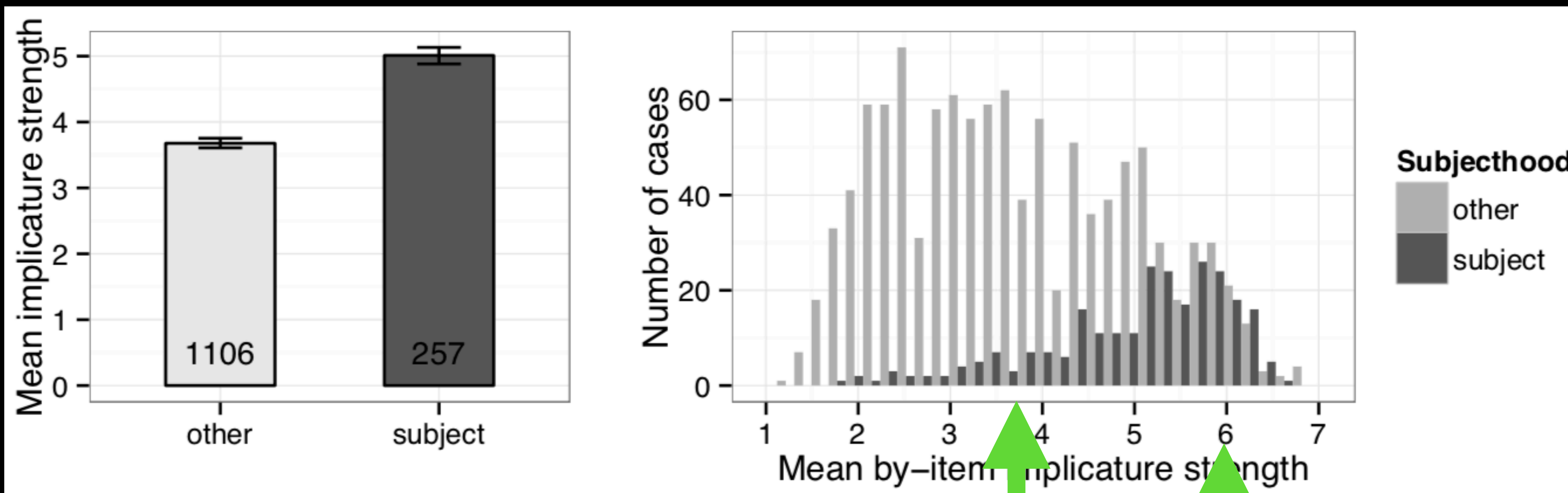
…with *some*-NPs in **subject** position.



***Some kids** are really having it.*

Stronger inferences…

…with *some*-NPs in **subject** position.

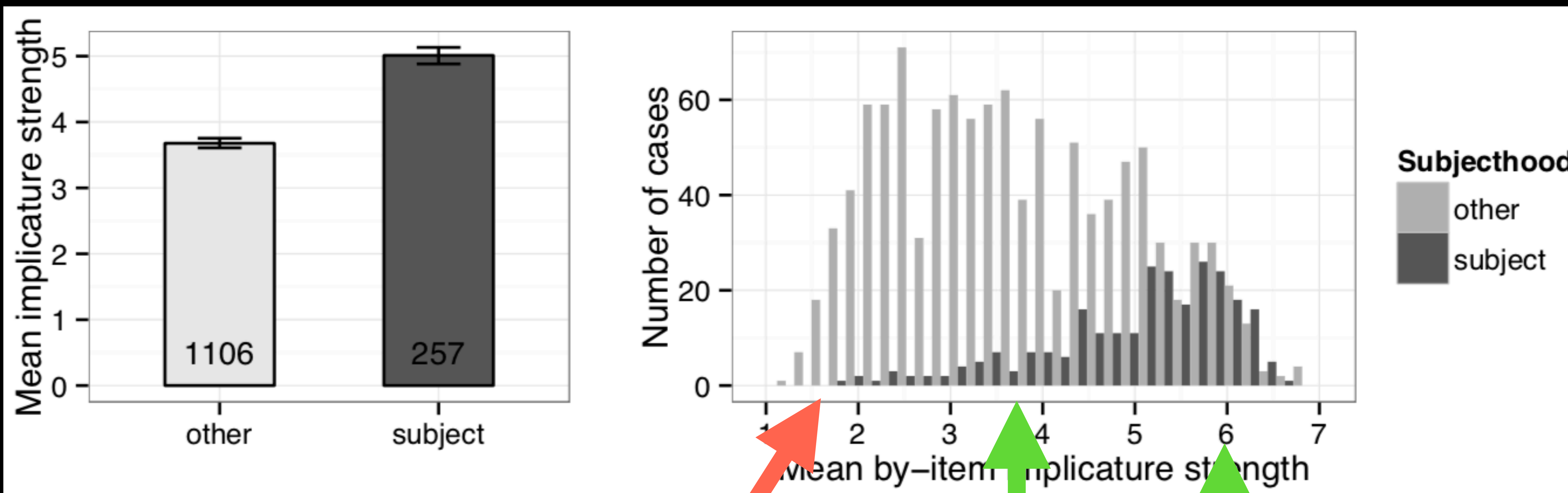


***Some kids** are really having it.*

*Occasionally, **some ice skating** will come on.*

Stronger inferences…

…with *some*-NPs in **subject** position.



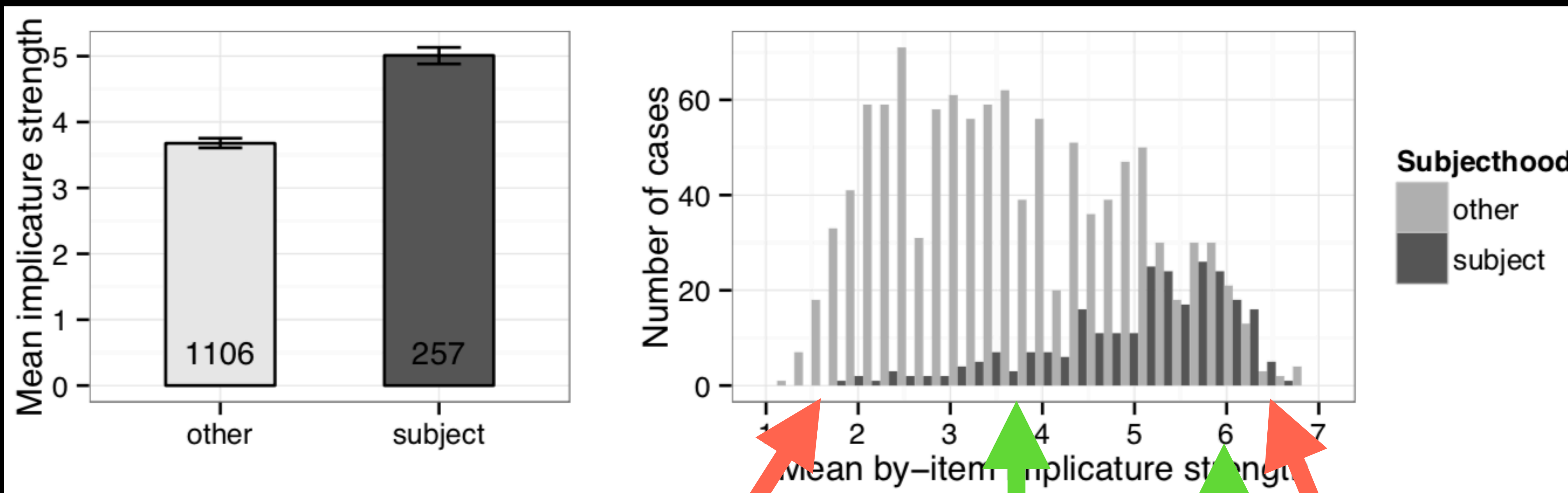
***Some kids** are really having it.*

*Occasionally, **some ice skating** will come on.*

*That would take **some planning**.*

Stronger inferences…

…with *some*-NPs in **subject** position.



***Some kids** are really having it.*

*Occasionally, **some ice skating** will come on.*

*That would take **some planning**.*

*I like **some country music.***

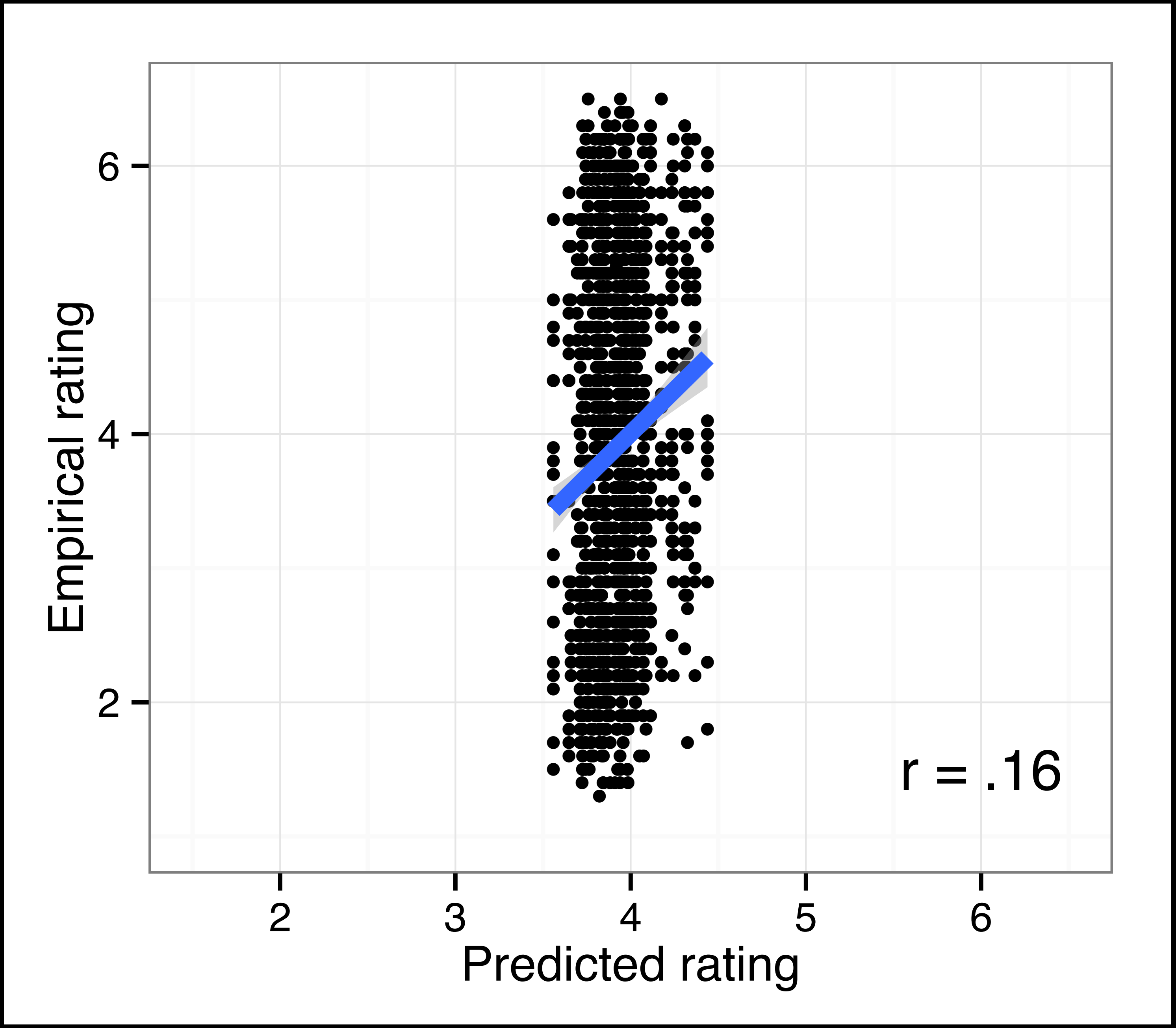
	Coef β	SE(β)	**t**	p
Intercept	4.01	0.06	68.7	**<.0001**
Partitive	0.91	0.09	9.6	**<.0001**
Strength	−0.50	0.05	−9.5	**<.0001**
Linguistic mention	0.31	0.07	4.4	**<.0001**
Subjecthood	0.41	0.10	4.2	**<.0001**
Modification	0.12	0.06	2.0	**<.05**
Sentence length	0.15	0.05	3.2	**<.01**
Partitive:Strength	0.39	0.10	4.1	**<.0001**
Linguistic mention:Subjecthood	0.17	0.21	0.8	<.44
Linguistic mention:Modification	0.34	0.13	2.6	**<.01**
Subjecthood:Modification	0.27	0.17	1.6	<.12
Linguistic mention:Subjecthood:Modification	0.61	0.42	1.4	<.16

Table 5 Model coefficients for the full model.

	Coef β	SE(β)	**t**	p
Intercept	4.01	0.06	68.7	**<.0001**
Partitive	0.91	0.09	9.6	**<.0001**
Strength	−0.50	0.05	−9.5	**<.0001**
Linguistic mention	0.31	0.07	4.4	**<.0001**
Subjecthood	0.41	0.10	4.2	**<.0001**
Modification	0.12	0.06	2.0	**<.05**
Sentence length	0.15	0.05	3.2	**<.01**
Partitive:Strength	0.39	0.10	4.1	**<.0001**
Linguistic mention:Subjecthood	0.17	0.21	0.8	<.44
Linguistic mention:Modification	0.34	0.13	2.6	**<.01**
Subjecthood:Modification	0.27	0.17	1.6	<.12
Linguistic mention:Subjecthood:Modification	0.61	0.42	1.4	<.16

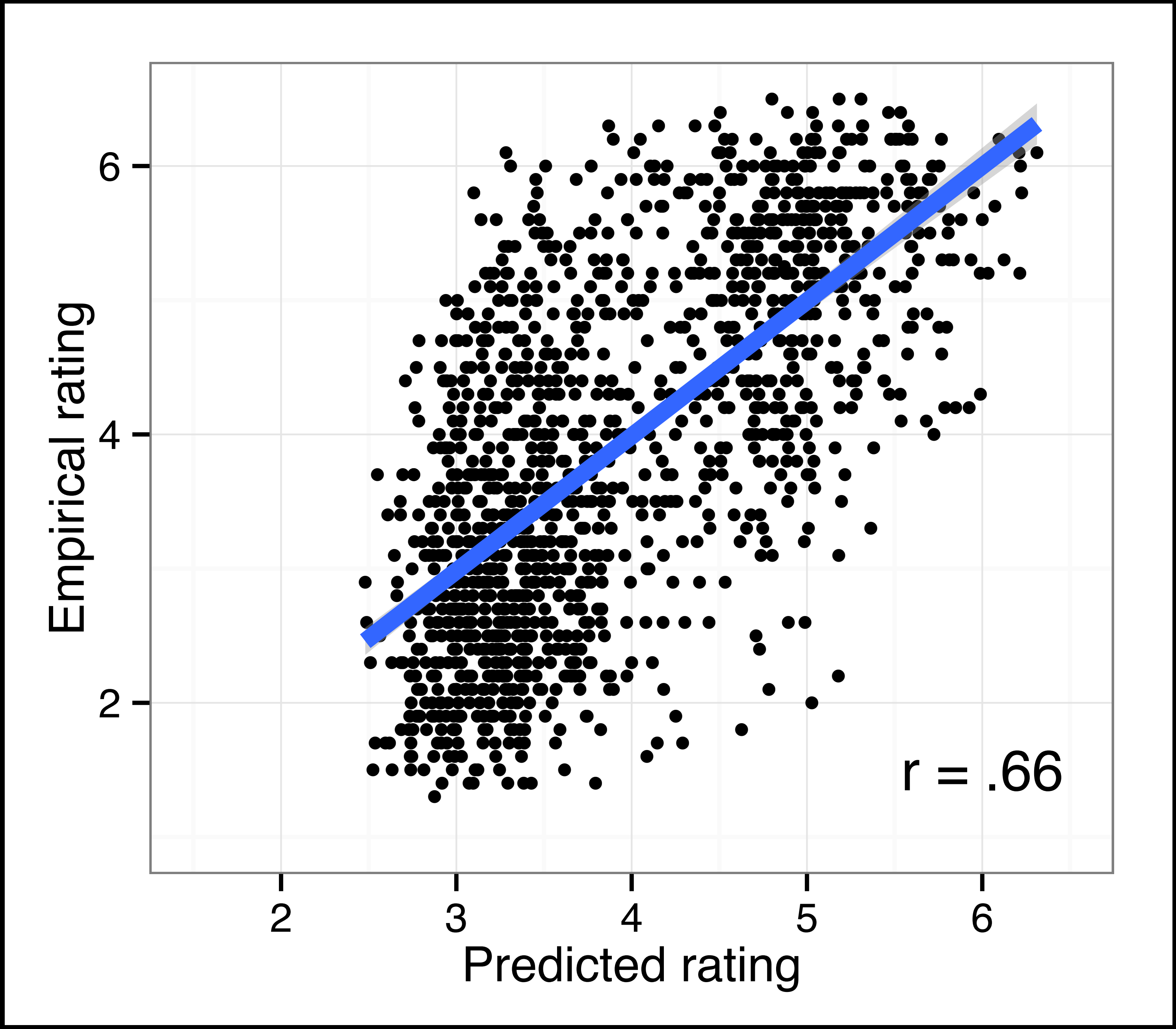
Table 5 Model coefficients for the full model.

Model fit



by-participant intercepts only

Model fit



after adding fixed effects of context

Just noise?

Just noise?

No. Variability in ratings is systematically predicted by syntactic, semantic, and pragmatic features of context.

But in some of these cases, “all” isn’t even an alternative!

You sound like you have **some small ones** in the background.

We've got **some beets**.

That would take **some planning**.

I like **some country music**.

I sold **some of them**.

I think **some parents** go a little bit overboard.

You sound like you have **all small ones** in the background.

We've got **all beets**.

That would take **all planning**.

I like **all country music**.

I sold **all of them**.

I think **all parents** go a little bit overboard.

1.5 You sound like you have **all** **small ones** in the background.

2.7 We've got **all** **beets**.

1.4 That would take **all** **planning**.

6.9 I like **all** **country music**.

6.8 I sold **all** **of them**.

6.4 I think **all** **parents** go a little bit overboard.

1.5 You sound like you have **all small ones** in the background.

2.7 We've got **all beets**.

1.4 That would take **all planning**.

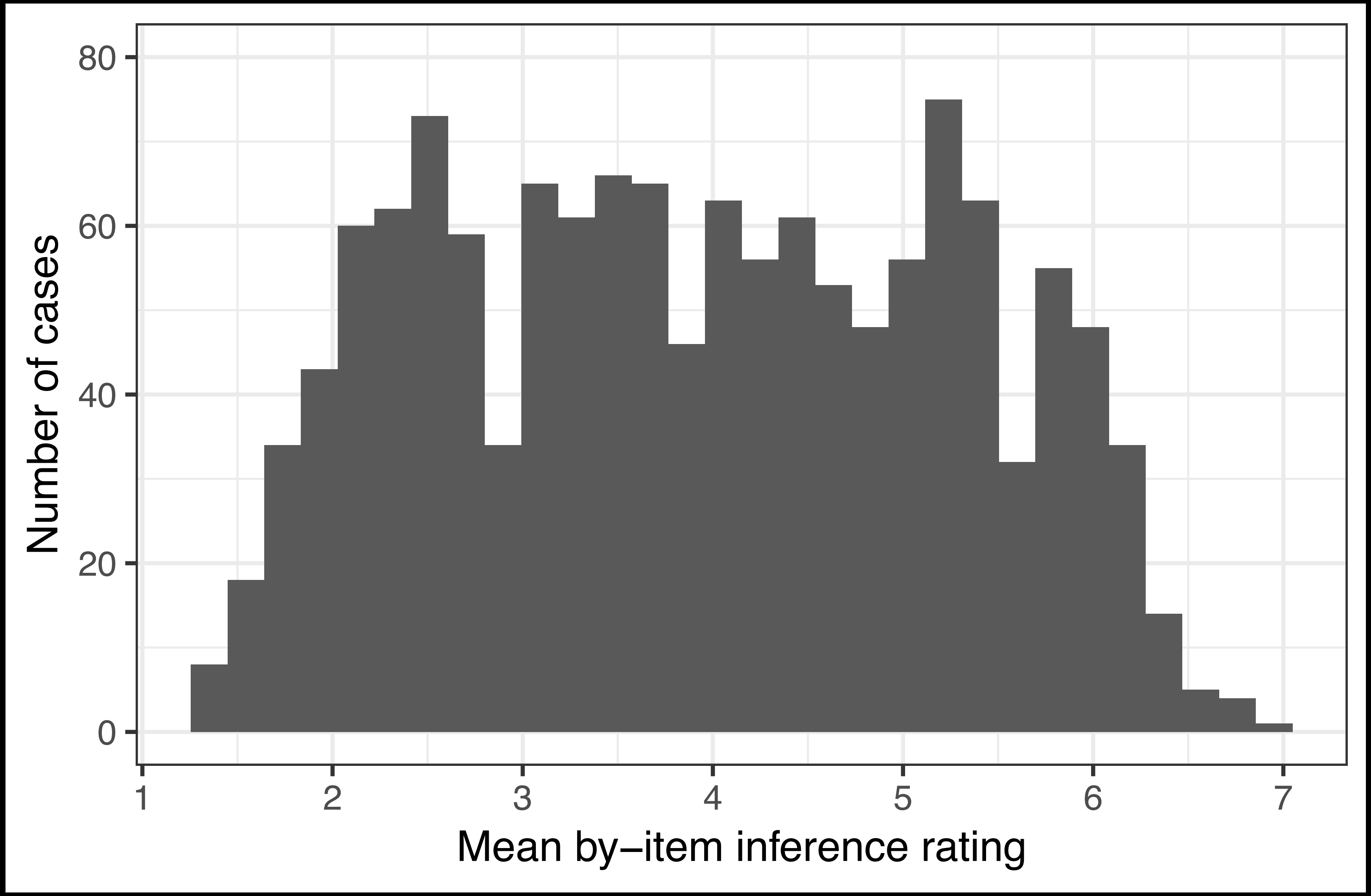
6.9 I like **all country music**.

6.8 I sold **all of them**.

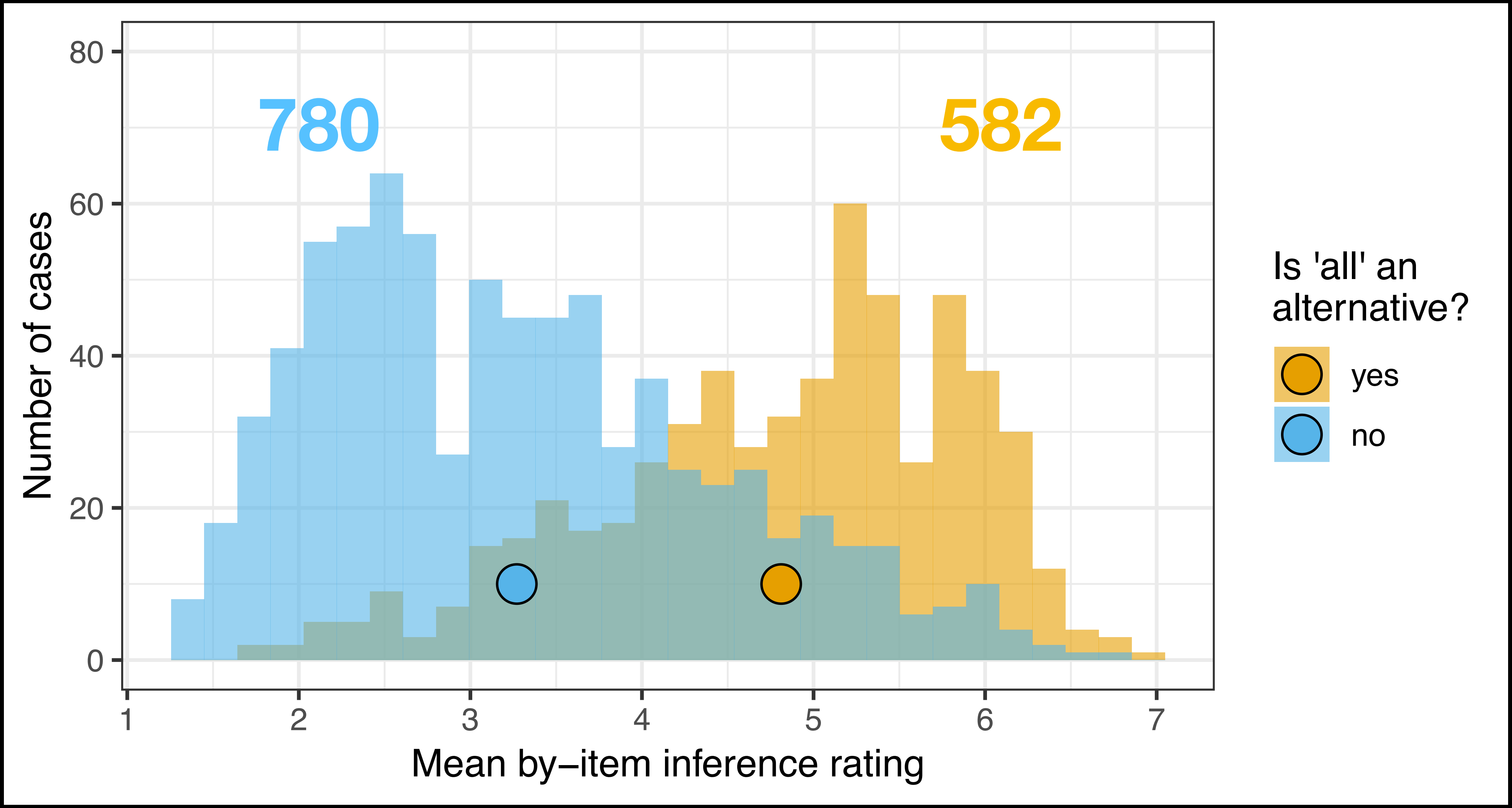
6.4 I think **all parents** go a little bit overboard.

All cases hand-annotated by 2 RAs for whether "some" can be replaced by "all" or only by "a lot (of)"

Variability in inference strength



Variability in inference strength



```
rating ~
  partitive + linguistic mention + subjecthood + …
  + random effects
```

original model

	Coef β	SE(β)	t	p
Intercept	4.01	0.06	68.7	**<.0001**
Partitive	0.91	0.09	9.6	**<.0001**
Strength	−0.50	0.05	−9.5	**<.0001**
Linguistic mention	0.31	0.07	4.4	**<.0001**
Subjecthood	0.41	0.10	4.2	**<.0001**
Modification	0.12	0.06	2.0	**<.05**
Sentence length	0.15	0.05	3.2	**<.01**
Partitive:Strength	0.39	0.10	4.1	**<.0001**
Linguistic mention:Subjecthood	0.17	0.21	0.8	<.44
Linguistic mention:Modification	0.34	0.13	2.6	**<.01**
Subjecthood:Modification	0.27	0.17	1.6	<.12
Linguistic mention:Subjecthood:Modification	0.61	0.42	1.4	<.16

```
rating ~
 (partitive + linguistic mention + subjecthood + …)
 * alternative
 + random effects
```

original model

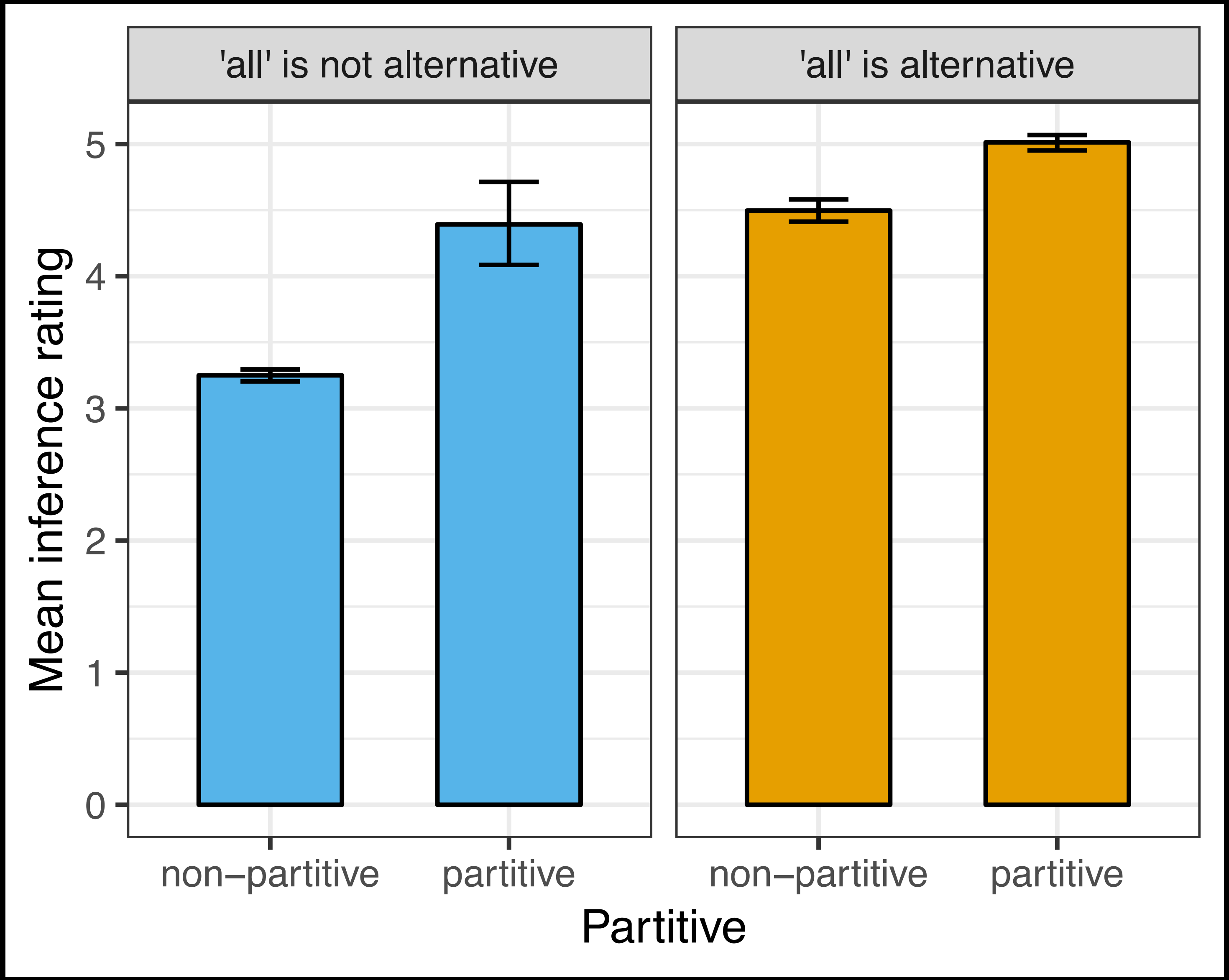
	Coef β	SE(β)	t	p
Intercept	4.01	0.06	68.7	**<.0001**
Partitive	0.91	0.09	9.6	**<.0001**
Strength	−0.50	0.05	−9.5	**<.0001**
Linguistic mention	0.31	0.07	4.4	**<.0001**
Subjecthood	0.41	0.10	4.2	**<.0001**
Modification	0.12	0.06	2.0	**<.05**
Sentence length	0.15	0.05	3.2	**<.01**
Partitive:Strength	0.39	0.10	4.1	**<.0001**
Linguistic mention:Subjecthood	0.17	0.21	0.8	<.44
Linguistic mention:Modification	0.34	0.13	2.6	**<.01**
Subjecthood:Modification	0.27	0.17	1.6	<.12
Linguistic mention:Subjecthood:Modification	0.61	0.42	1.4	<.16

```
rating ~
 (partitive + linguistic mention + subjecthood + …)
 * alternative
 + random effects
```

	original model				original with alternative
	Coef β	SE(β)	t	p	p
Intercept	4.01	0.06	68.7	**<.0001**	**<.0001**
Partitive	0.91	0.09	9.6	**<.0001**	**<.0001**
Strength	−0.50	0.05	−9.5	**<.0001**	**<.0001**
Linguistic mention	0.31	0.07	4.4	**<.0001**	**<.0001**
Subjecthood	0.41	0.10	4.2	**<.0001**	**<.0001**
Modification	0.12	0.06	2.0	**<.05**	**<.0001**
Sentence length	0.15	0.05	3.2	**<.01**	**<.0001**
Partitive:Strength	0.39	0.10	4.1	**<.0001**	**<.0001**
Linguistic mention:Subjecthood	0.17	0.21	0.8	<.44	>0.52
Linguistic mention:Modification	0.34	0.13	2.6	**<.01**	**<.01**
Subjecthood:Modification	0.27	0.17	1.6	<.12	**<.01**
Linguistic mention:Subjecthood:Modification	0.61	0.42	1.4	<.16	**<.01**

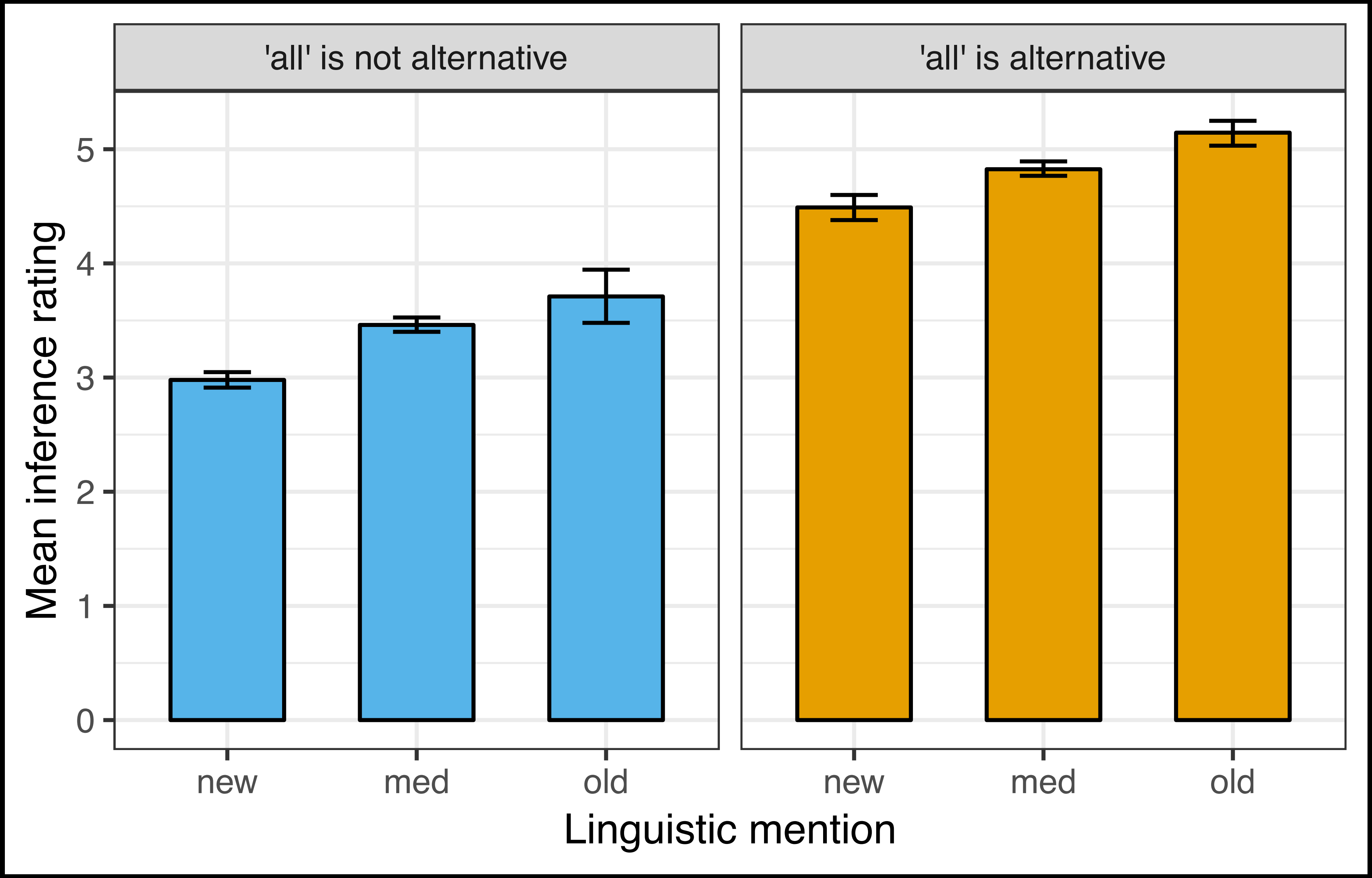
Stronger inferences…

…with **partitive** *some*-NPs.



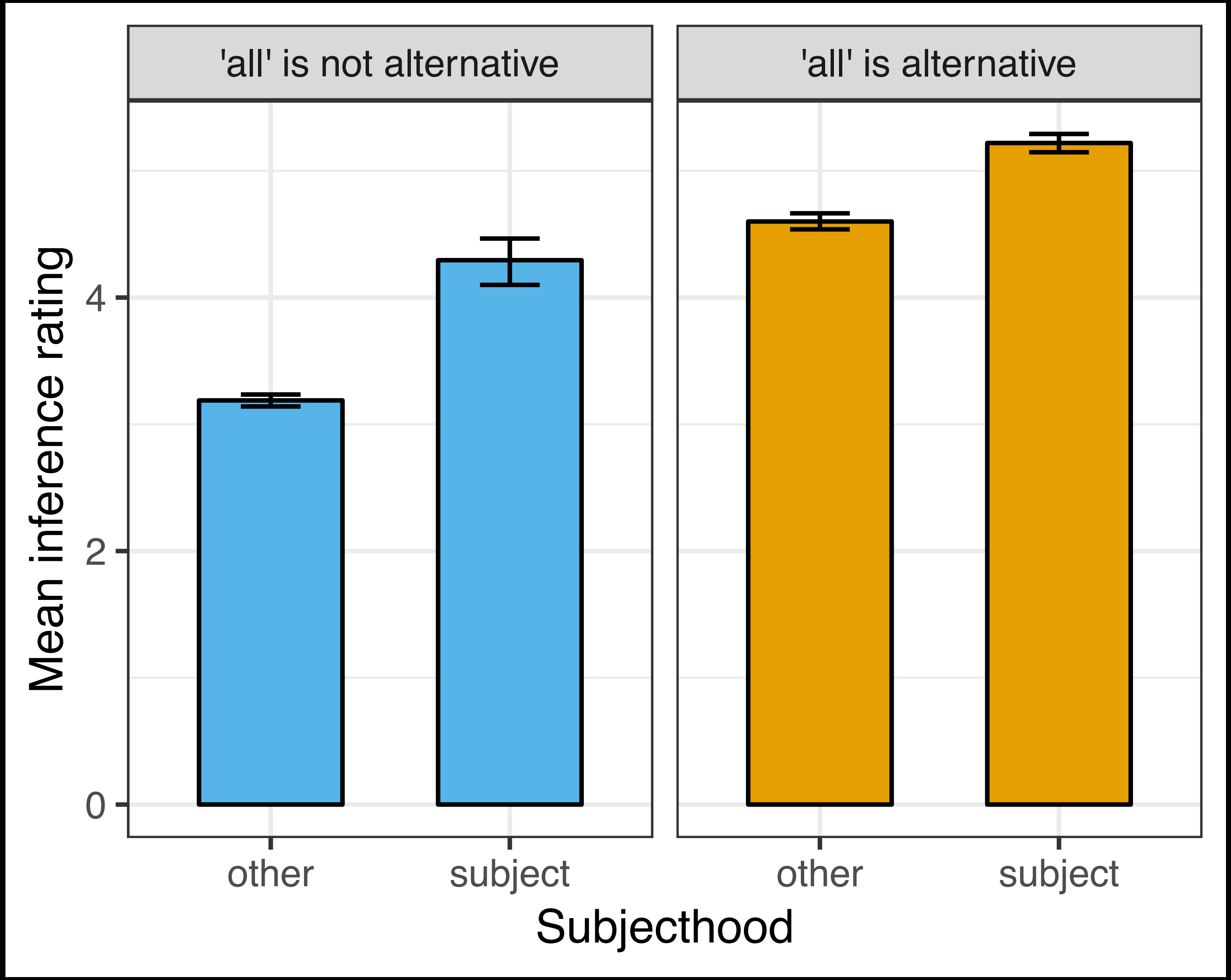
Stronger inferences…

…with **previously mentioned** NP referents.



Stronger inferences…

…with *some*-NPs in **subject** position.



Just noise?

Just noise?

No. Variability in ratings is systematically predicted by syntactic, semantic, and pragmatic features of context.

No. Replication by Eiteljoerge et al 2019 in child-directed speech

Just noise?

No. Variability in ratings is systematically predicted by syntactic, semantic, and pragmatic features of context.

Implications for theories of pragmatic inference

Implications for theories of pragmatic inference

The status of scalar implicatures as GCIs is highly questionable.

How many features? Do they need to be hand-mined?

Predicting inference strength from distributed meaning representations

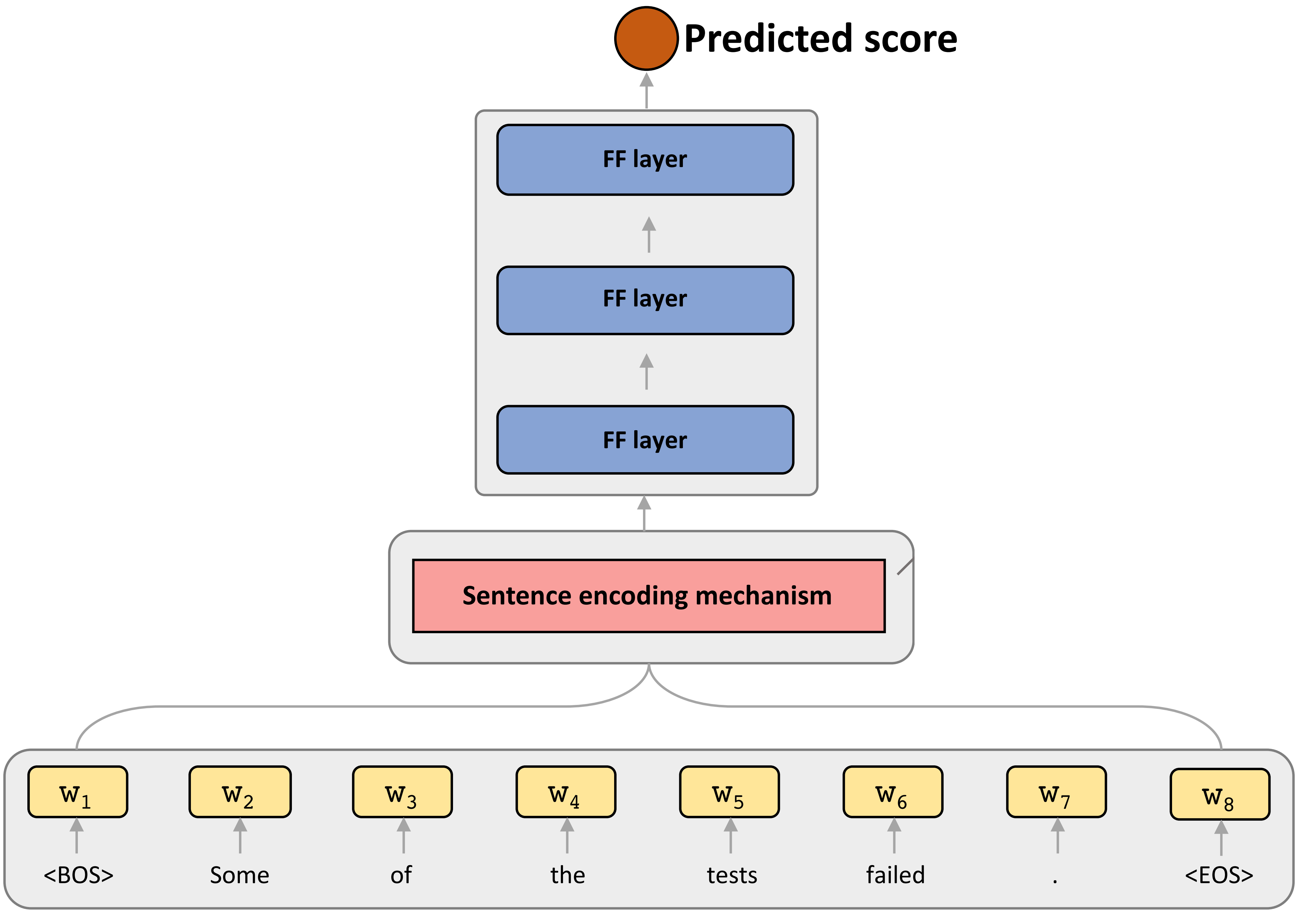

Yuxing Chen
Sebastian Schuster

Ultimate goal:
Use distributed vector-based meaning representation methods from NLP to infer which, if any, linguistically encoded features of context listeners use in drawing inferences, to help inform pragmatic theory.

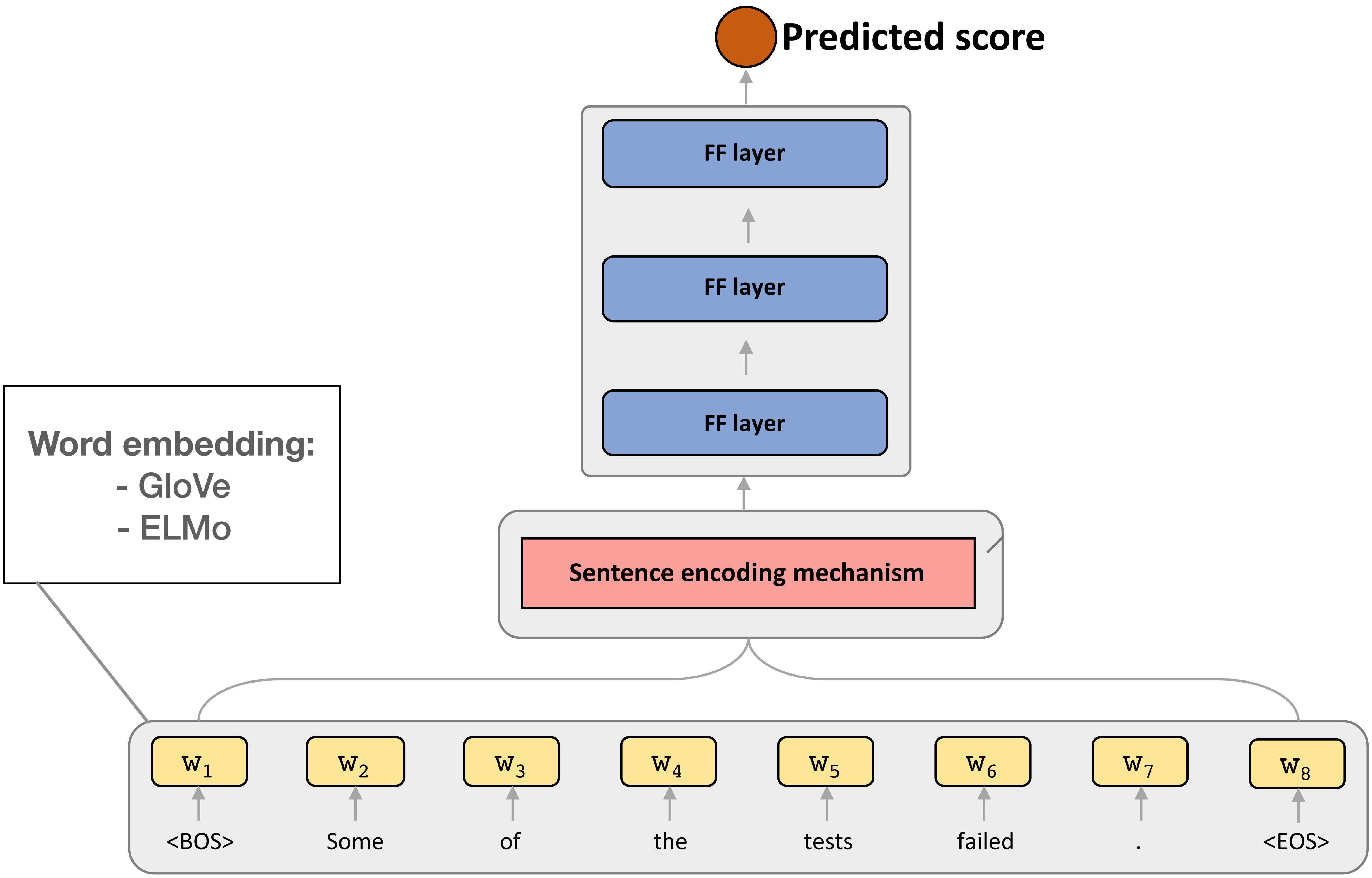
More proximate goal:
Use distributed vector-based meaning representation methods from NLP to test whether any of these methods

- reliably predict inference ratings
- capture the identified context effects

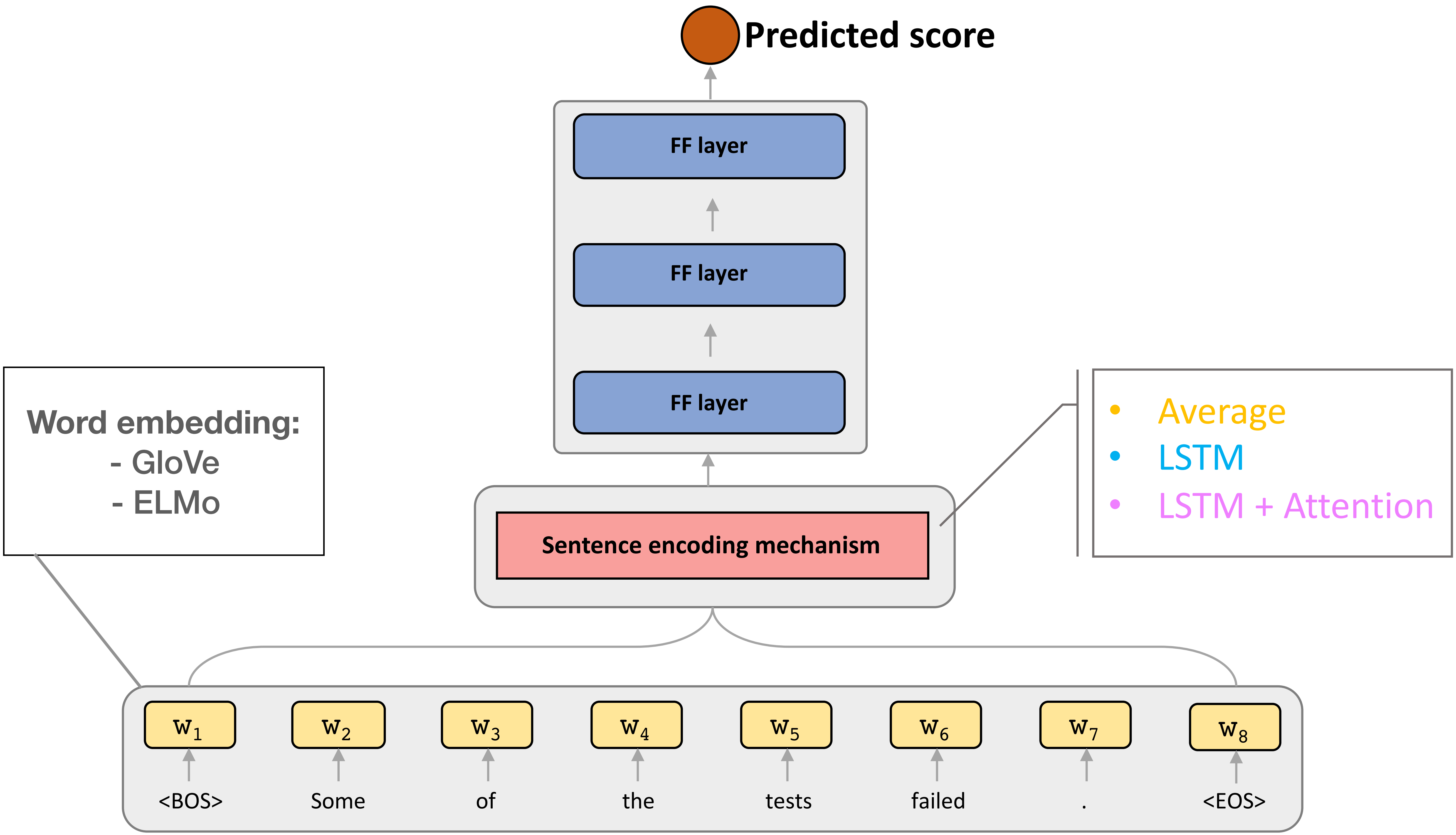
Model architecture



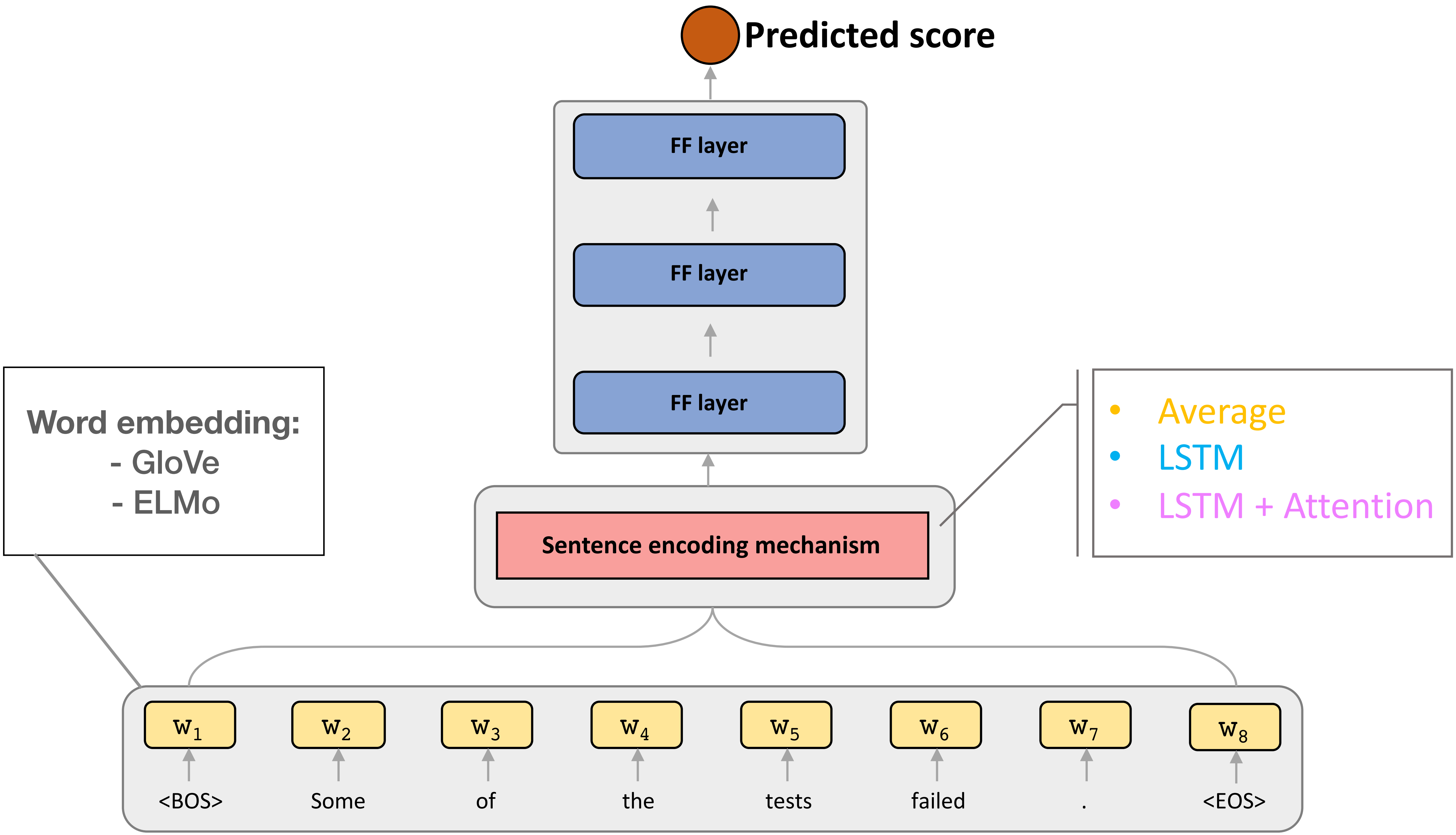
Model architecture



Model architecture



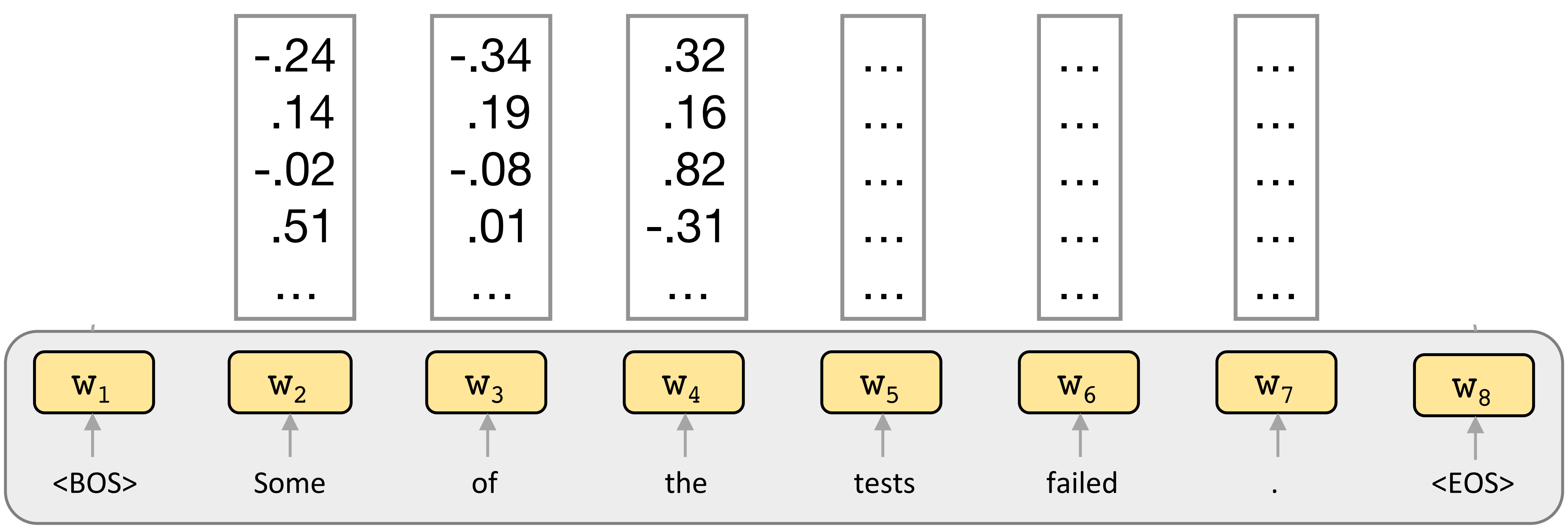
Model architecture



GloVe (Global Vectors for word representation)

Pennington et al 2014

- captures meaning in vector space
- based on co-occurrence statistics of words
- 100-dimensional vector for each word, pre-trained on 6 billion tokens from Wikipedia 2014 and Gigaword 5
- words around “some” encoded in pretrained 100-dimensional GloVe vectors

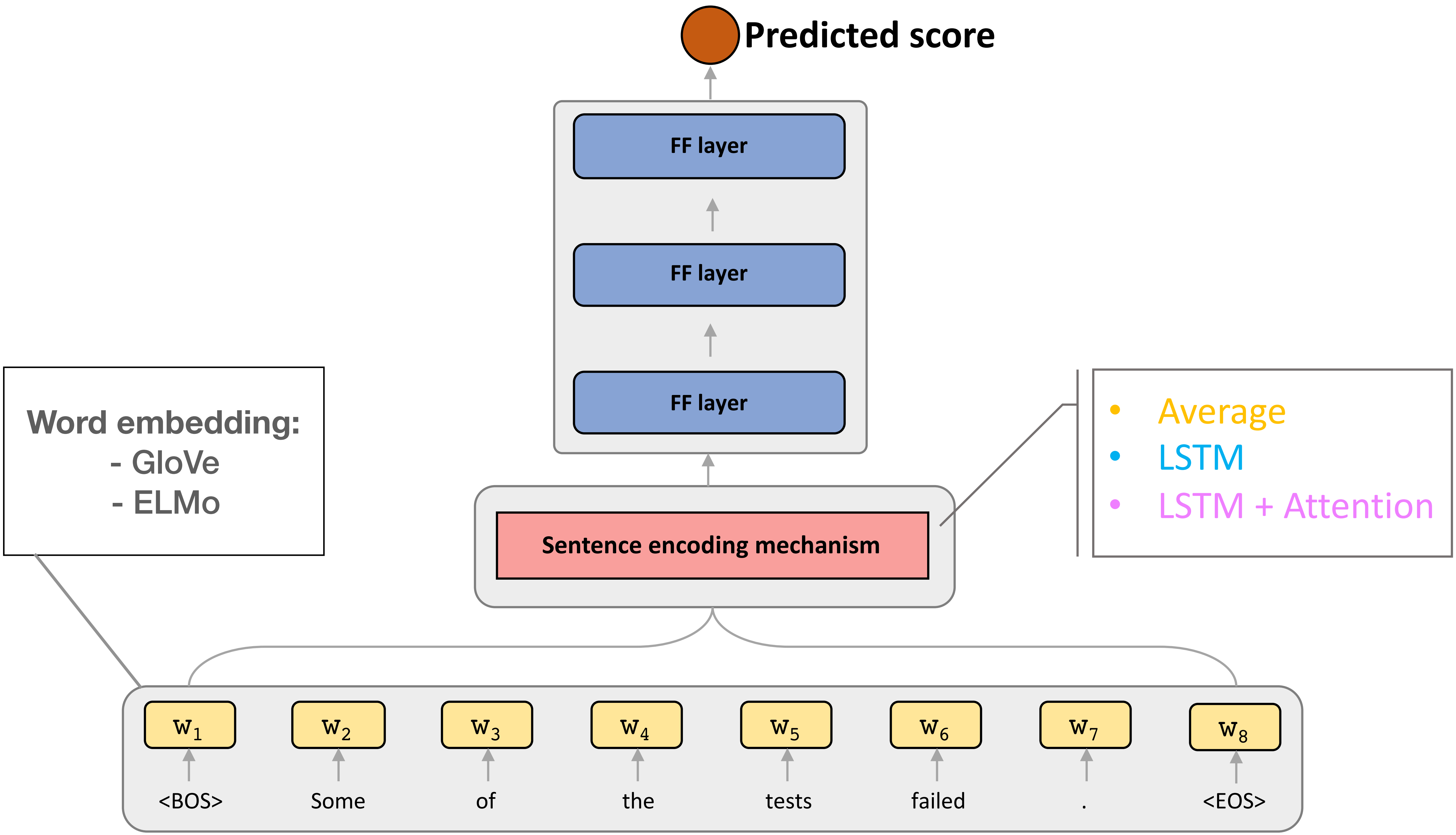


ELMo (Embeddings from Language Models) Peters et al 2018

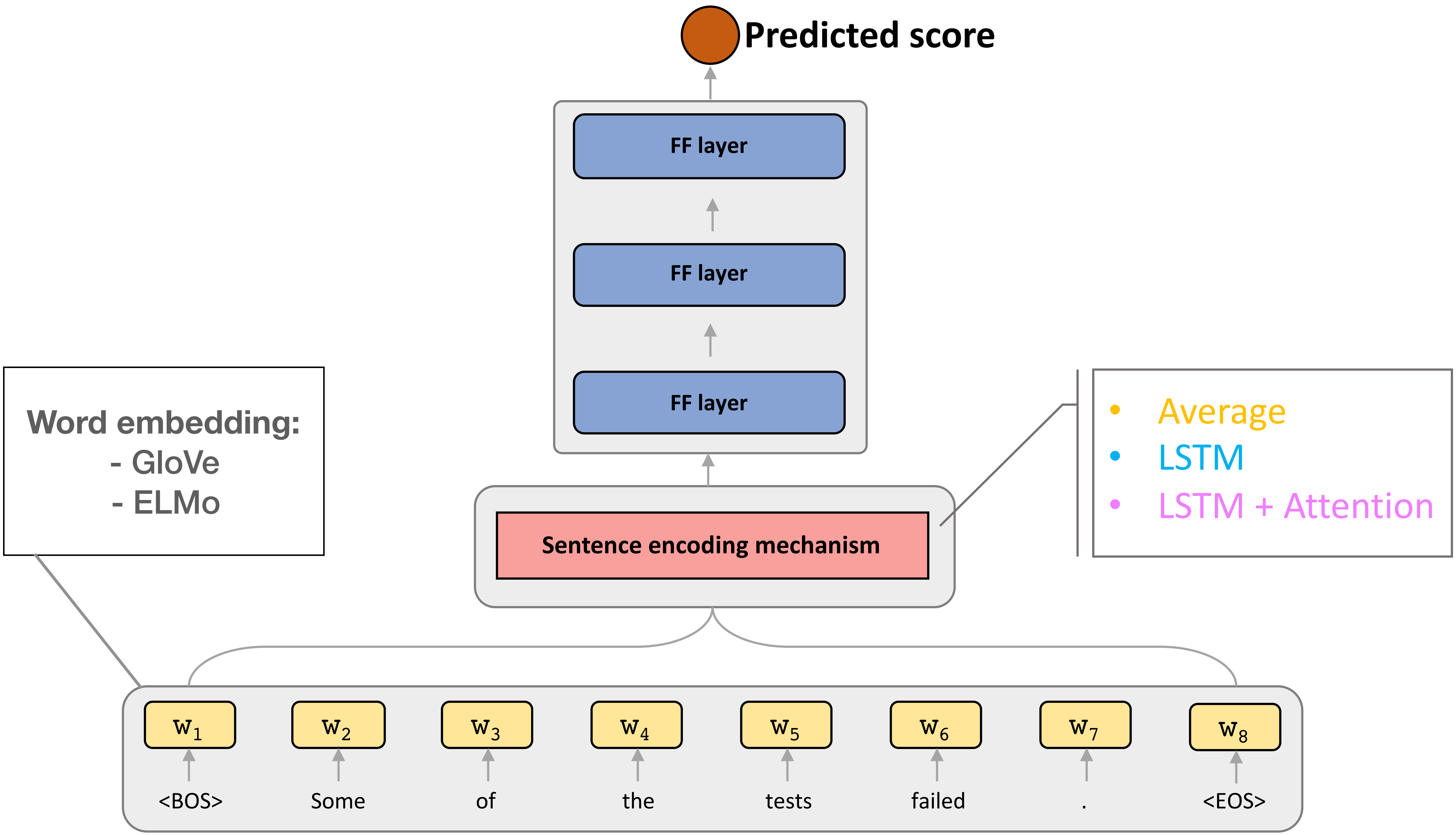
- **contextual** word embeddings (considers entire sentence before assigning a word in it an embedding)
- captures that the same word can have different meanings in different sentences
- based on word sequence modeling (bi-directional LSTM)
- pre-trained on 5.5 billion tokens from Wikipedia and news crawl data from WMT 2008-2012

1. **Apple** announced the new iPhone today.
2. **Google** announced a new browser last week.
3. I ate an **apple** for breakfast.
4. I ate an **orange** after dinner.

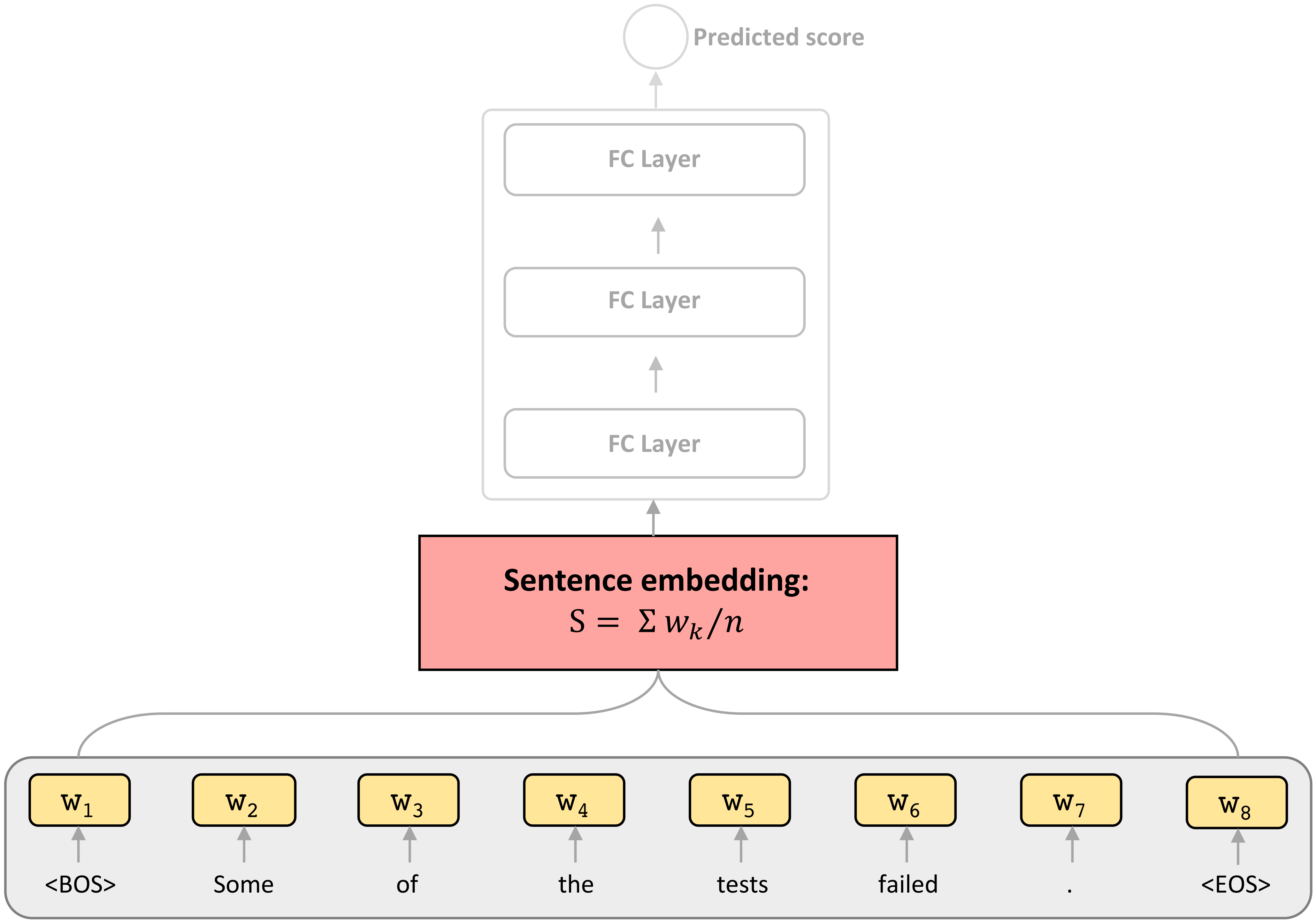
Model architecture



Model architecture

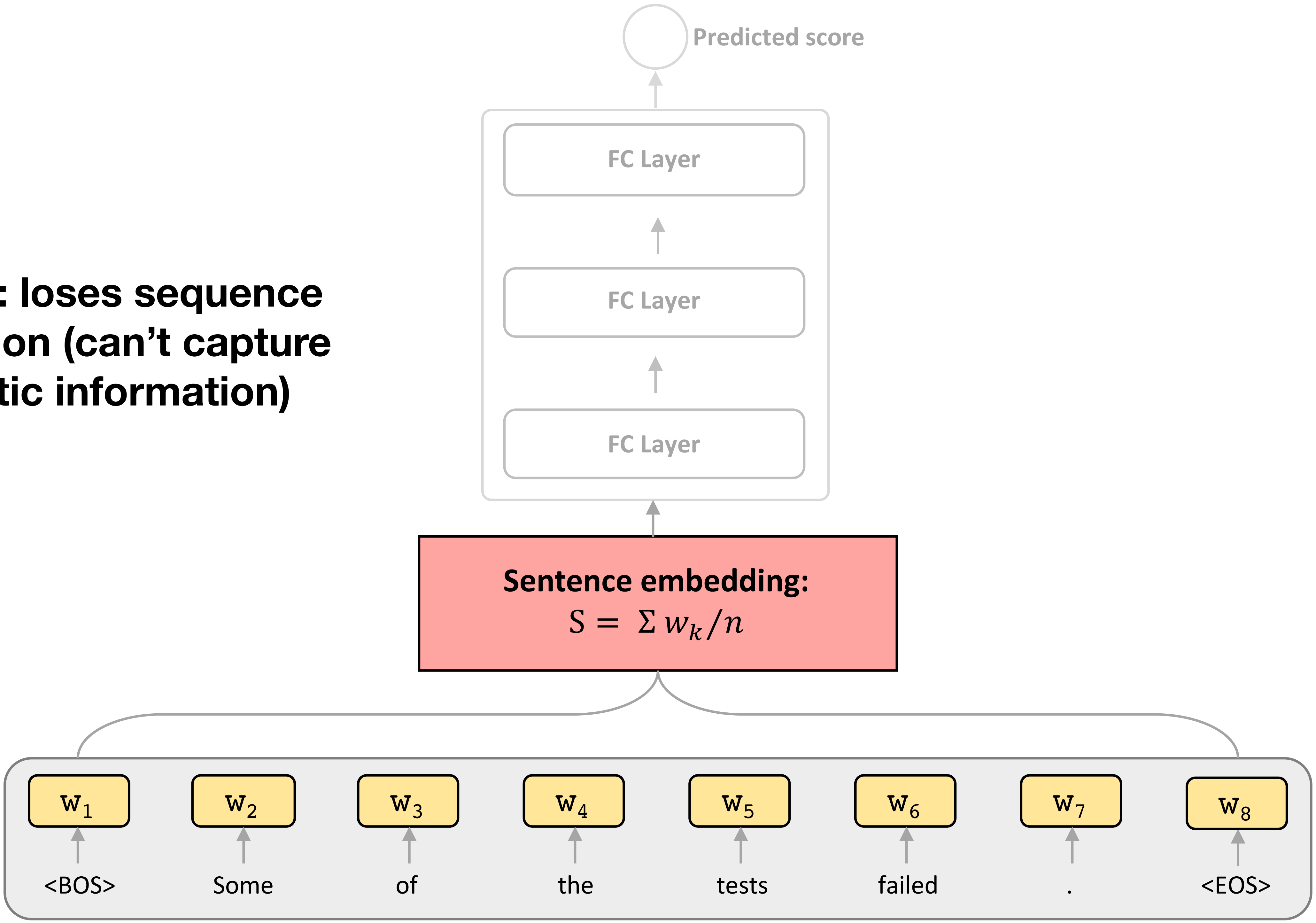


Sentence embedding: average

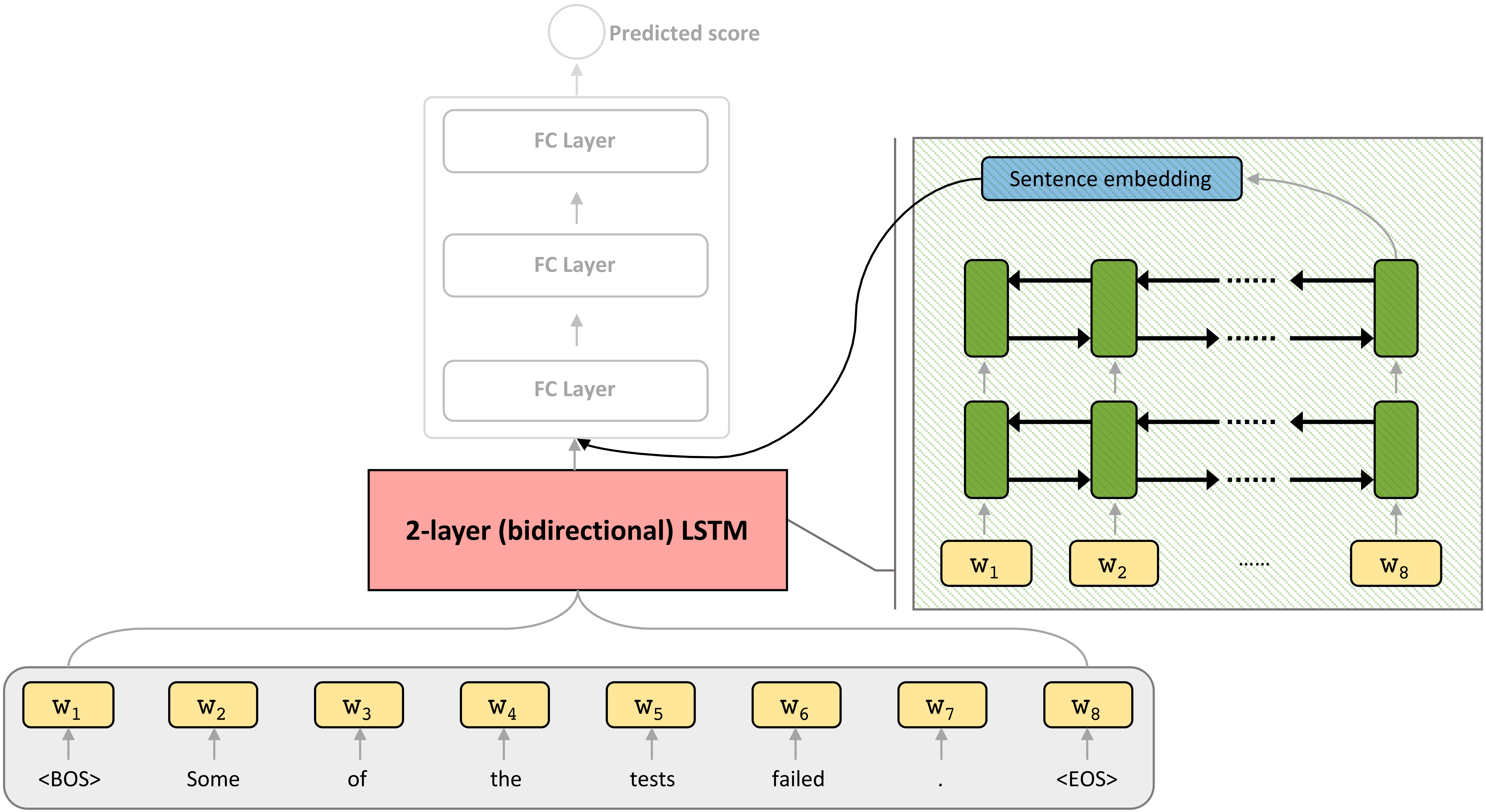


Sentence embedding: average

problem: loses sequence information (can't capture syntactic information)

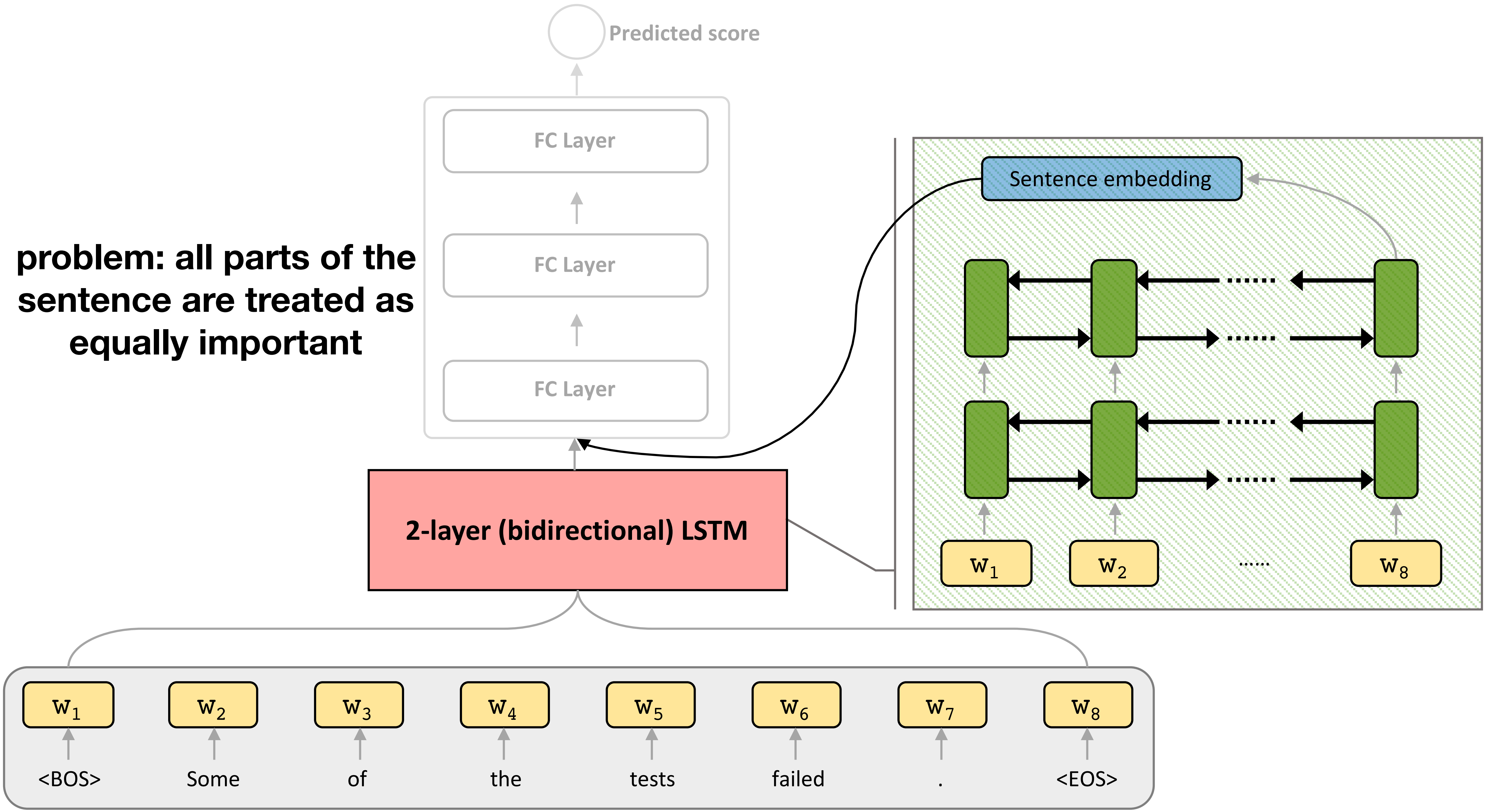


Sentence embedding: LSTM

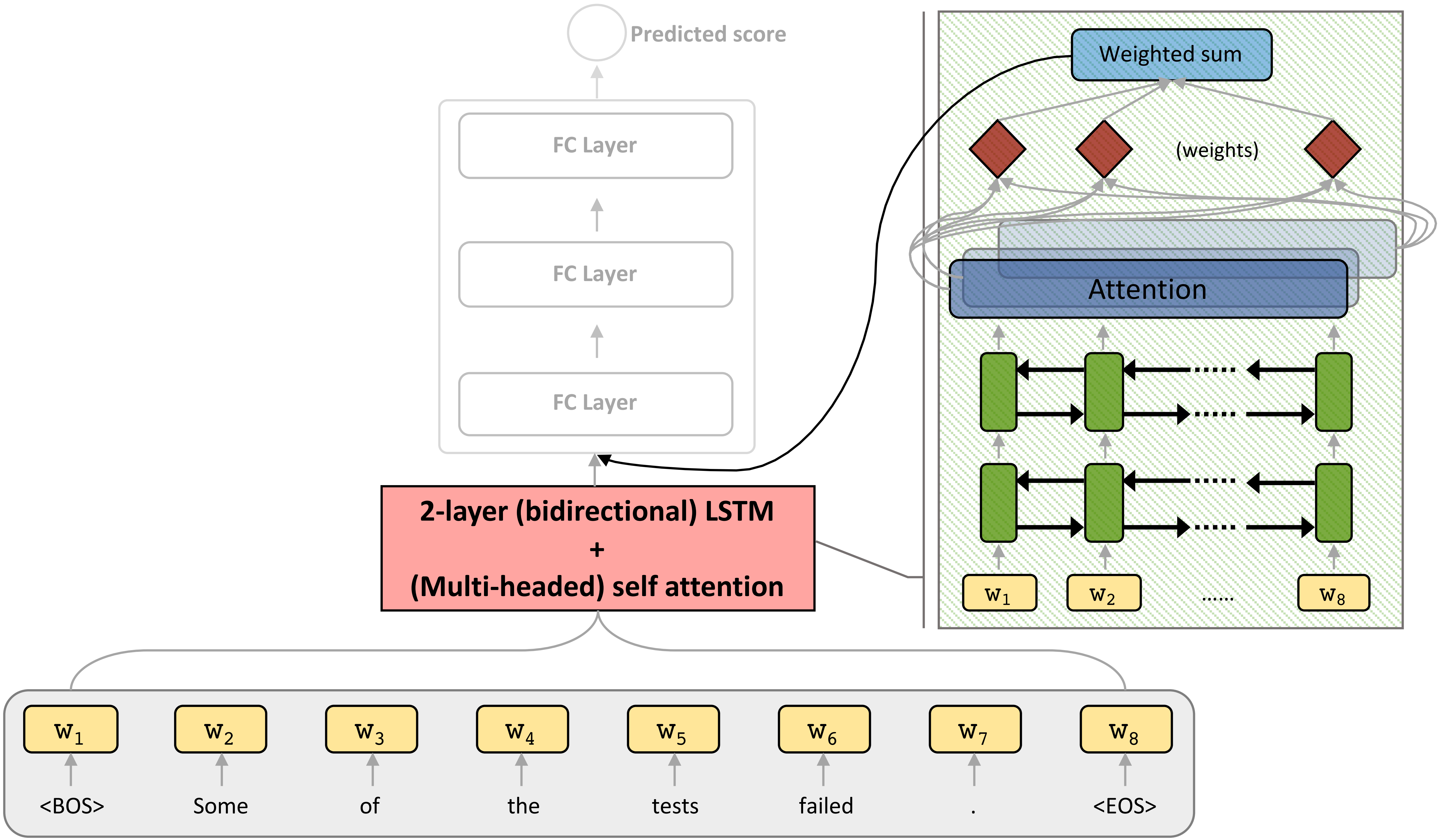


Sentence embedding: LSTM

problem: all parts of the sentence are treated as equally important



Sentence embedding: LSTM+attention



Context preceding sentence

Speaker A: i mean, they just have beautiful, beautiful homes and they have everything. the kids only wear name brand things to school and it's one of these things,

Speaker B: oh me. well that makes it hard for you, doesn't it.

Speaker A: well it does, you know. it really does because i'm a single mom and i have a thirteen year old now and uh, you know, it does.

Speaker B: oh, me.

Speaker A: i mean, we do it to a point but uh, not to where she feels different ,

Speaker B: yeah.

Speaker A:
but some of them are very rich

either did or didn't include context in generating the sentence embedding
(context may be important for capturing factors like linguistic mention)

Context preceding sentence

Speaker A: i mean, they just have beautiful, beautiful homes and they have everything. the kids only wear name brand things to school and it's one of these things,

Speaker B: oh me. well that makes it hard for you, doesn't it.

Speaker A: well it does, you know. it really does because i'm a single mom and i have a thirteen year old now and uh, you know, it does.

Speaker B: oh, me.

Speaker A: i mean, we do it to a point but uh, not to where she feels different ,

Speaker B: yeah.

Speaker A: but some of them are very rich

either did or didn't include context in generating the sentence embedding
(context may be important for capturing factors like linguistic mention)

Context preceding sentence

Speaker A: i mean, they just have beautiful, beautiful homes and they have everything. the kids only wear name brand things to school and it's one of these things,

Speaker B: oh me. well that makes it hard for you, doesn't it.

Speaker A: well it does, you know. it really does because i'm a single mom and i have a thirteen year old now and uh, you know, it does.

Speaker B: oh, me.

Speaker A: i mean, we do it to a point but uh, not to where she feels different ,

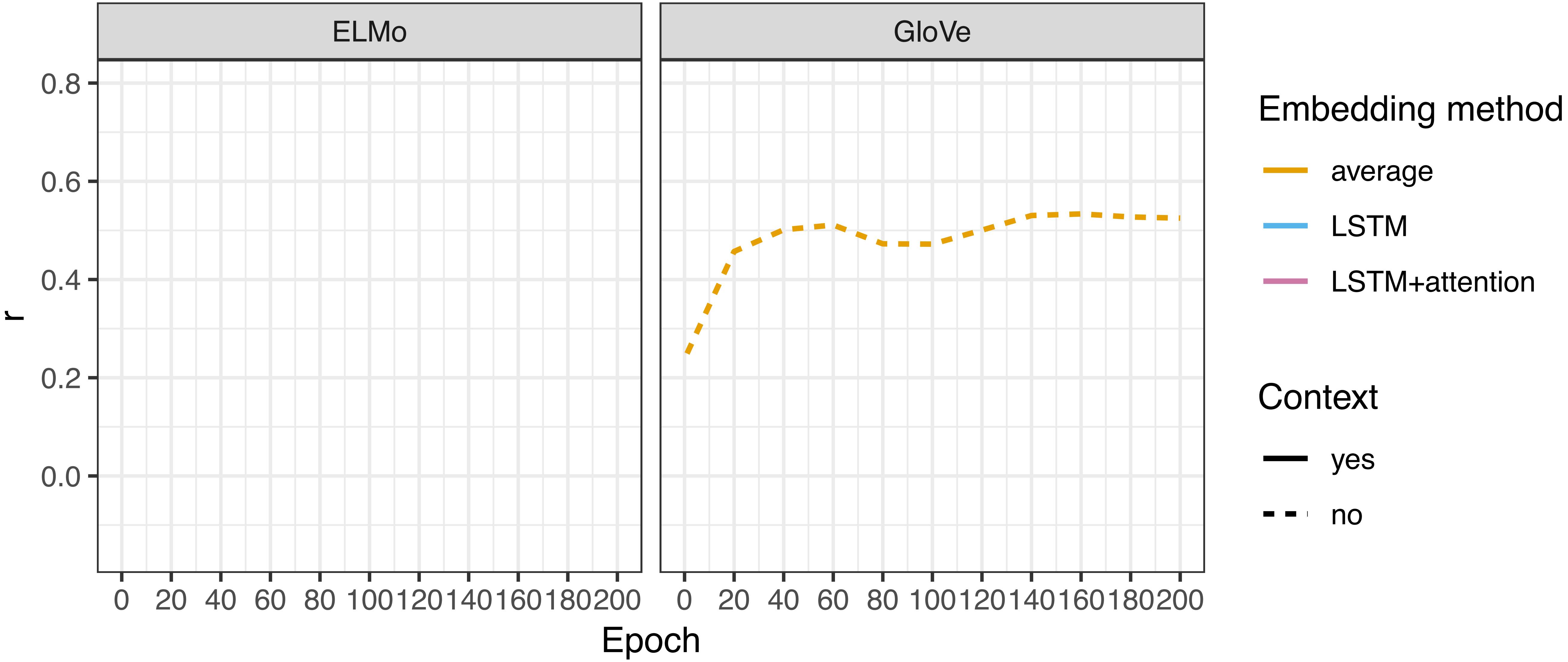
Speaker B: yeah.

Speaker A:

but some of them are very rich

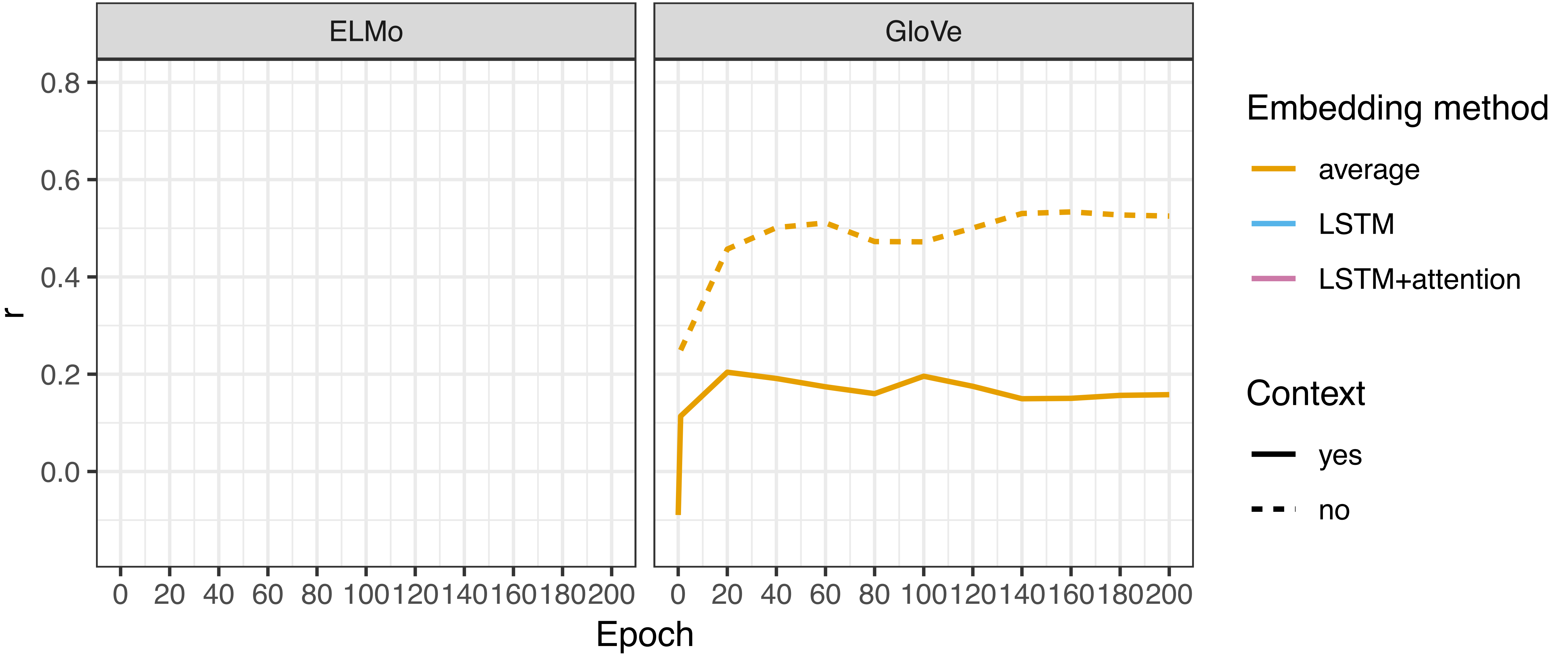
either did or didn't include context in generating the sentence embedding
(context may be important for capturing factors like linguistic mention)

Results



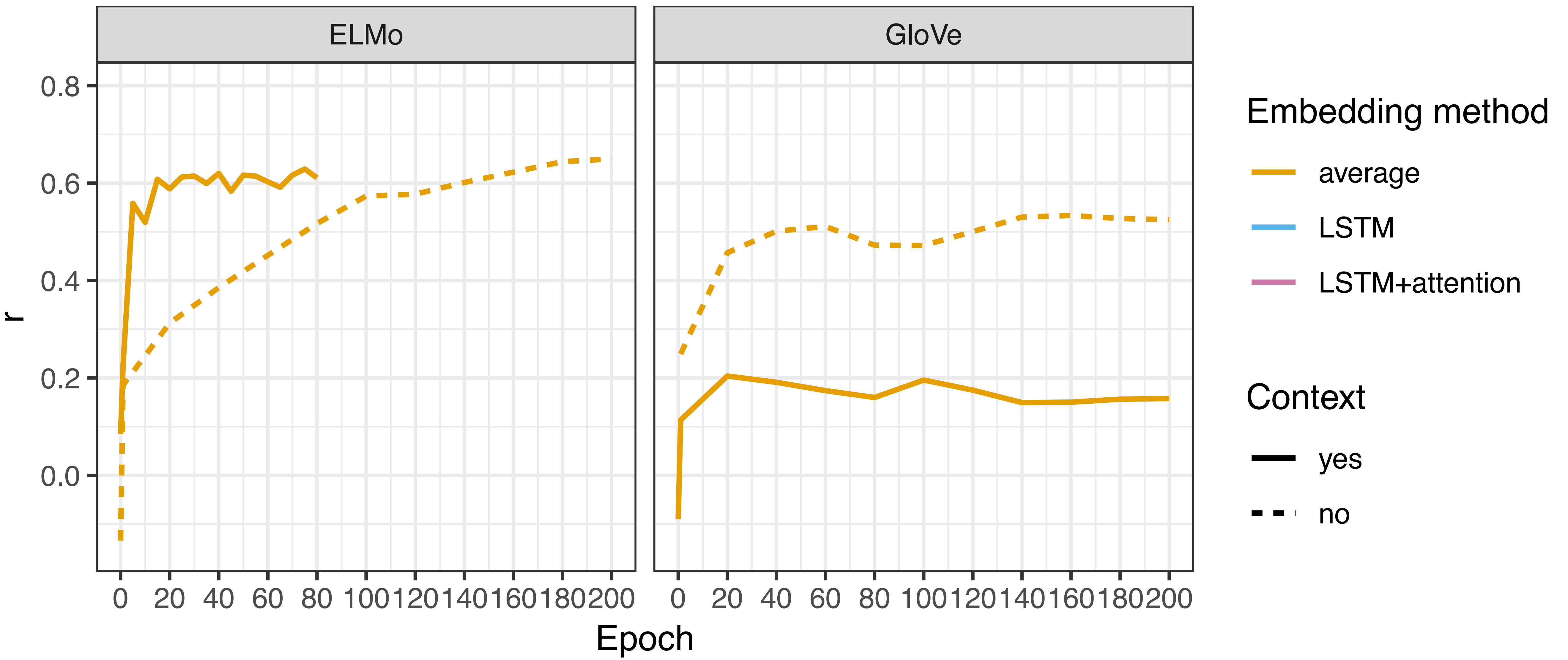
ELMo
GloVe
0.8
0.6
0.4
0.2
0.0
r
0
20
40
60
80
100
120
140
160
180
200
Epoch
Embedding method
average
LSTM
LSTM+attention
Context
yes
no

Results



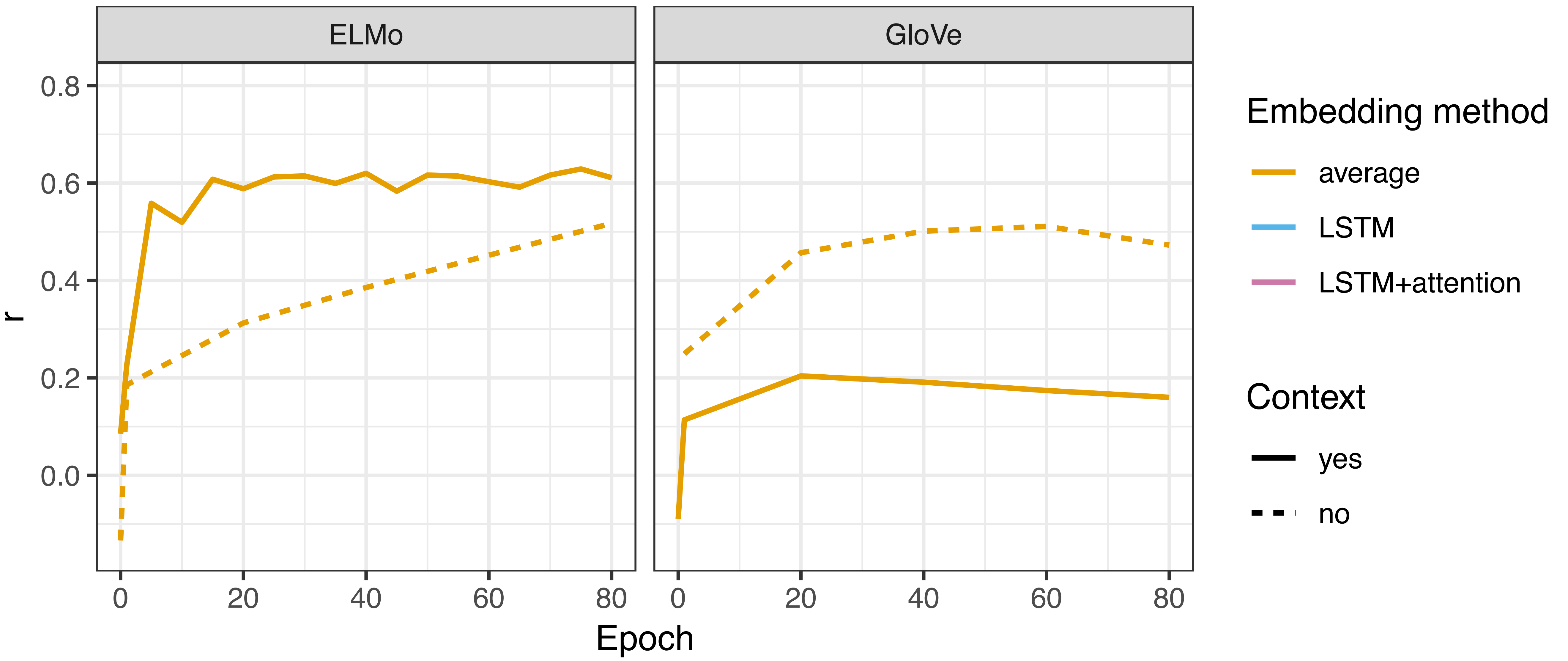
adding context hurts model performance

Results



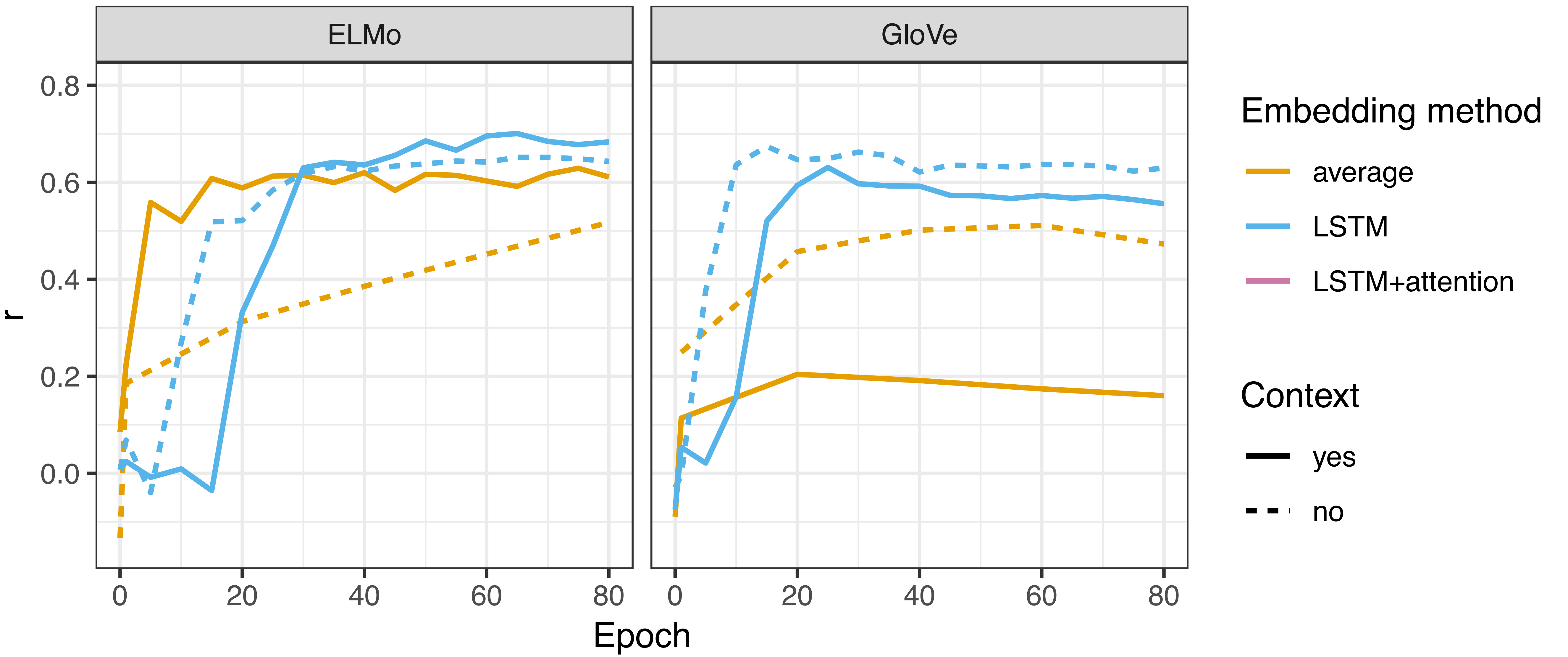
ELMo model learns faster; context helps

Results



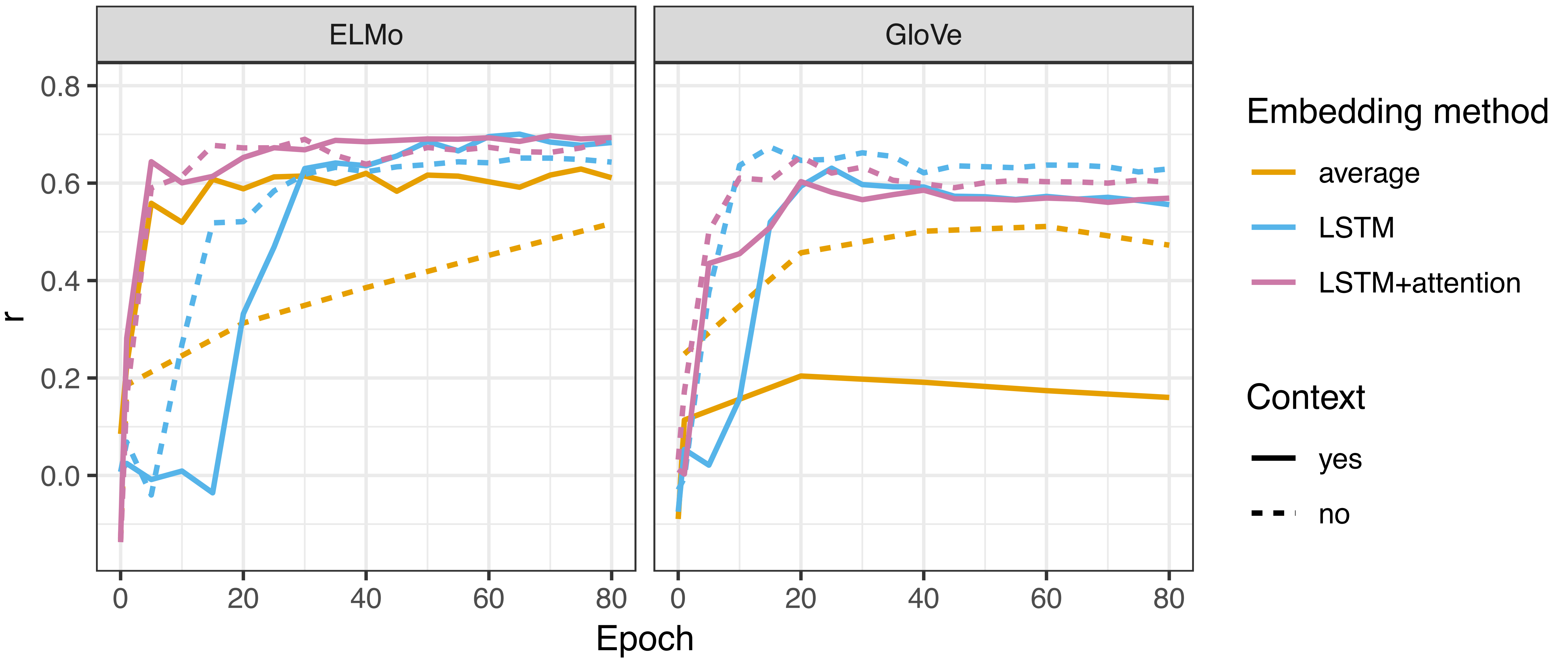
ELMo
GloVe
0.8
0.6
0.4
0.2
0.0
r
0
20
40
60
80
Epoch
Embedding method
average
LSTM
LSTM+attention
Context
yes
no

Results



LSTM does better than simple averaging

Results



attention does not change asymptotic behavior, but learning is faster

Summary

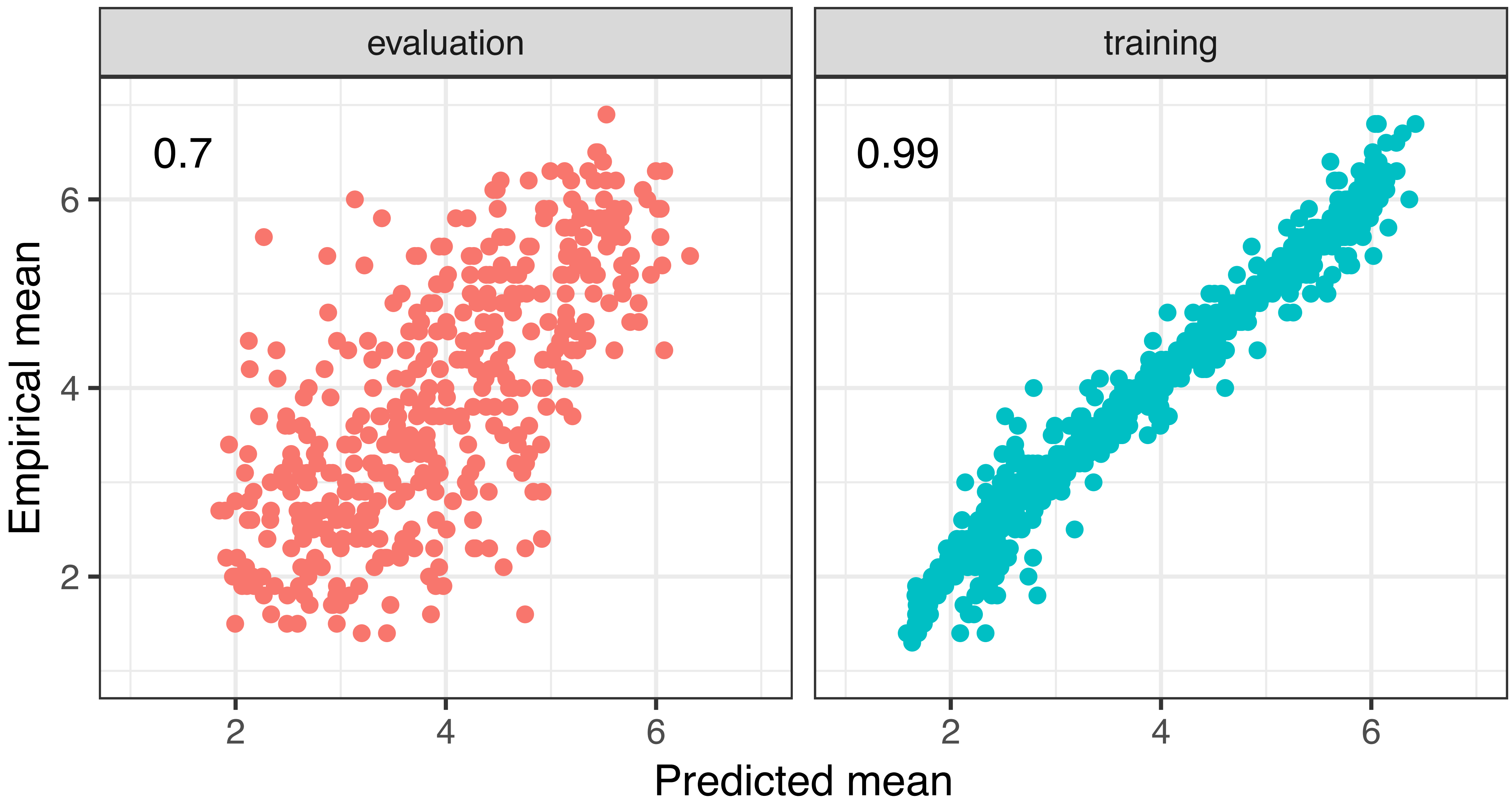
- maximal correlation: .7 (ELMo, LSTM+attention, context)
  —> preliminary but impressive result!

- contextual word embeddings (ELMo) generally outperform non-contextual ones (GloVe)

- context beyond the sentence doesn’t help much in predicting inferences

Model predictions

Best model: ELMo — LSTM + attention — context



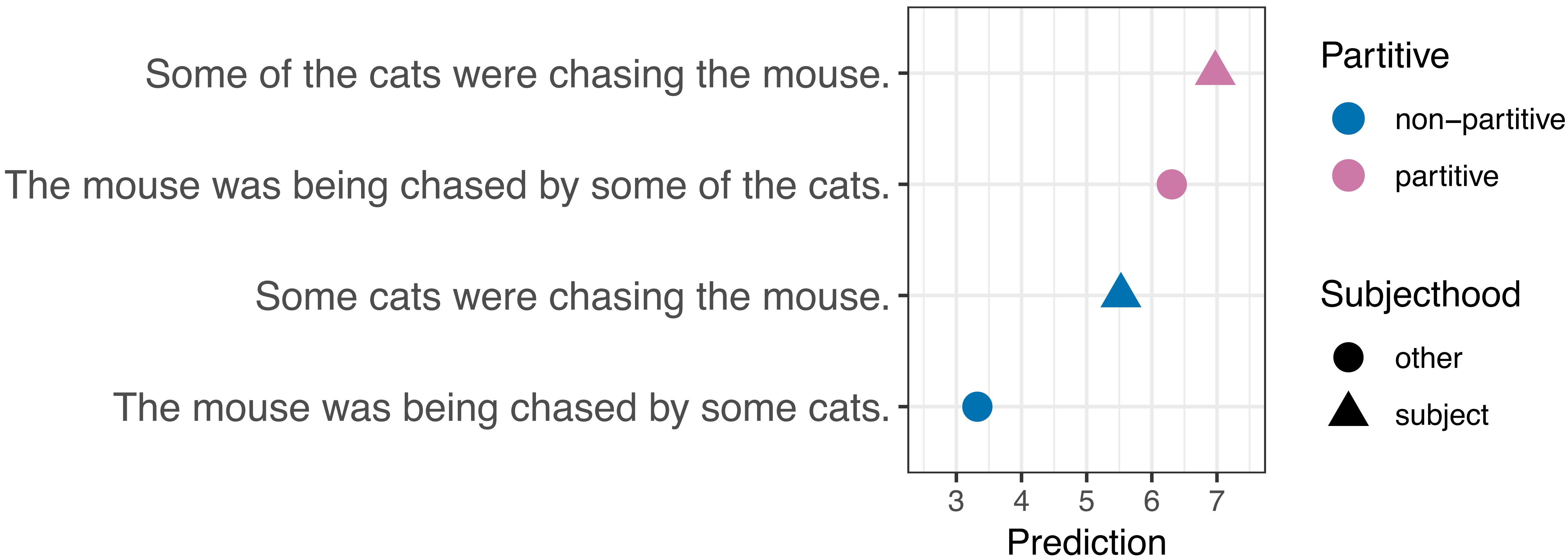
Qualitative model results

Generate sentences that cross factors of interest (e.g., partitive and subjecthood):

1. **Some of the cats** were chasing the mouse.
2. The mouse was being chased by some of the cats.
3. **Some cats** were chasing the mouse.
4. The mouse was being chased by some cats.

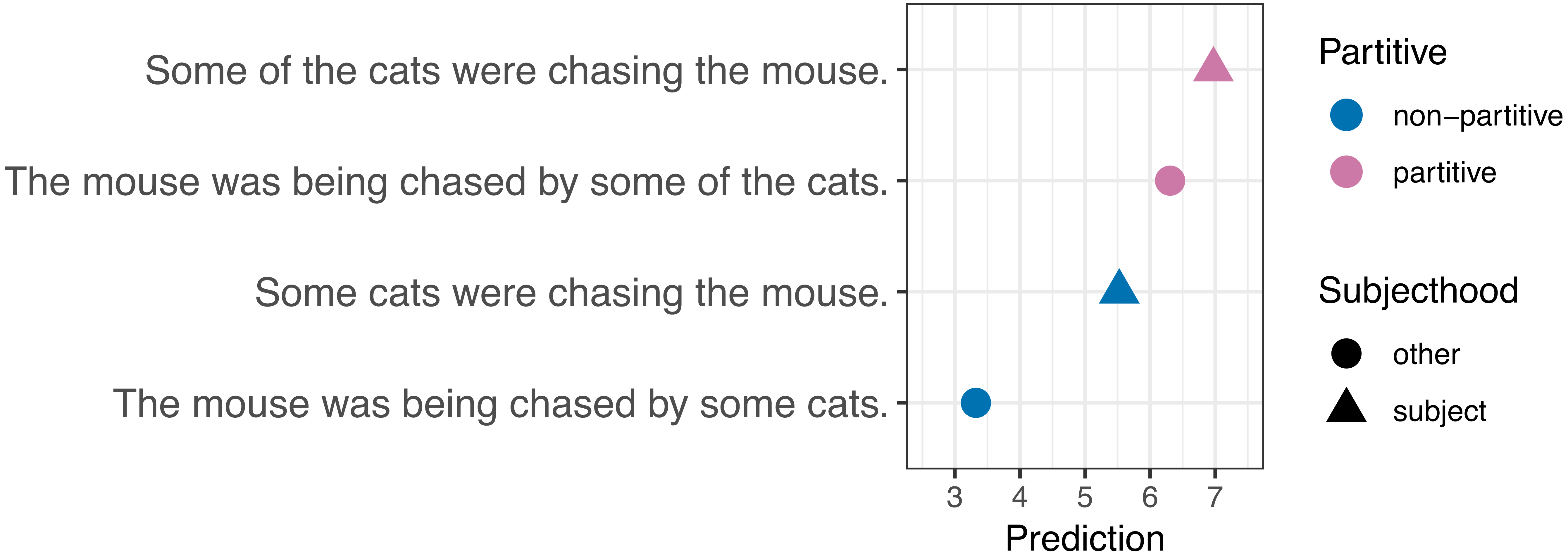
Qualitative model results

Best model: ELMo — LSTM + attention — context



Qualitative model results

Best model: ELMo — LSTM + attention — context



the neural model qualitatively retrieves the partitive and subjecthood effects

Quantitative comparison with original model

	Coef β	SE(β)	**t**	p
Intercept	4.01	0.06	68.7	**<.0001**
Partitive	0.91	0.09	9.6	**<.0001**
Strength	−0.50	0.05	−9.5	**<.0001**
Linguistic mention	0.31	0.07	4.4	**<.0001**
Subjecthood	0.41	0.10	4.2	**<.0001**
Modification	0.12	0.06	2.0	**<.05**
Sentence length	0.15	0.05	3.2	**<.01**
Partitive:Strength	0.39	0.10	4.1	**<.0001**
Linguistic mention:Subjecthood	0.17	0.21	0.8	<.44
Linguistic mention:Modification	0.34	0.13	2.6	**<.01**
Subjecthood:Modification	0.27	0.17	1.6	<.12
Linguistic mention:Subjecthood:Modification	0.61	0.42	1.4	<.16

Quantitative comparison with original model

	Coef β	SE(β)	**t**	p
Intercept	4.00	0.05	81.2	**<.0001**
Neural model score	0.88	0.02	57.7	**<.0001**
Partitive	0.08	0.07	1.2	>0.23
Strength	−0.06	0.04	−1.7	>0.08
Linguistic mention	0.08	0.04	1.9	>0.06
Subjecthood	0.01	0.06	0.1	>0.88
Modification	0.02	0.03	0.6	>0.53
Sentence length	0.03	0.03	1.3	>0.21
Partitive:Strength	0.12	0.06	2.2	**<.05**
Linguistic mention:Subjecthood	0.01	0.12	0.1	>0.91
Linguistic mention:Modification	−0.01	0.07	−0.1	>0.89
Subjecthood:Modification	0.04	0.09	0.4	>0.68
Linguistic mention:Subjecthood:Modification	−0.06	0.23	−0.3	>0.78

Quantitative comparison with original model

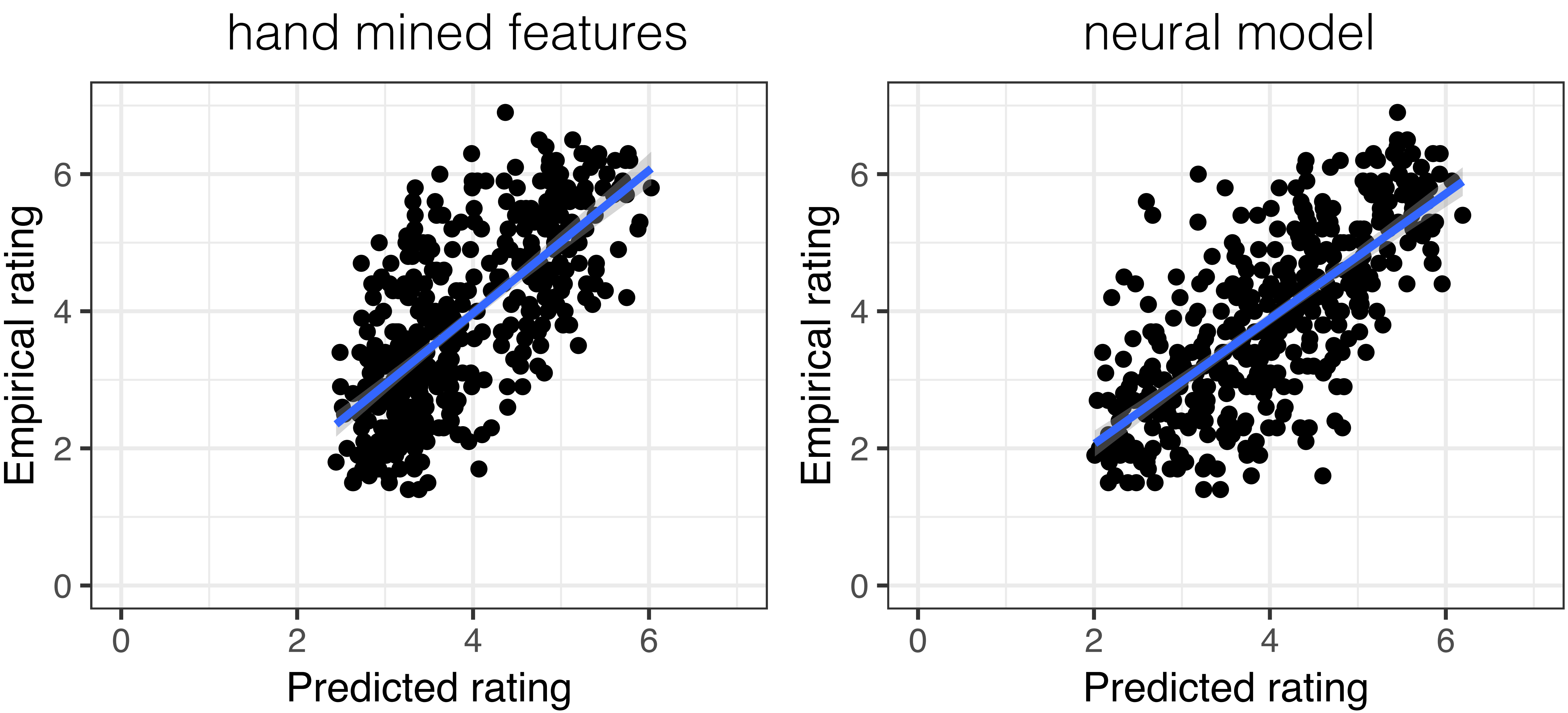
	Coef β	SE(β)	**t**	p
Intercept	4.00	0.05	81.2	**<.0001**
Neural model score	0.88	0.02	57.7	**<.0001**
Partitive	0.08	0.07	1.2	>0.23
Strength	−0.06	0.04	−1.7	>0.08
Linguistic mention	0.08	0.04	1.9	>0.06
Subjecthood	0.01	0.06	0.1	>0.88
Modification	0.02	0.03	0.6	>0.53
Sentence length	0.03	0.03	1.3	>0.21
Partitive:Strength	0.12	0.06	2.2	**<.05**
Linguistic mention:Subjecthood	0.01	0.12	0.1	>0.91
Linguistic mention:Modification	−0.01	0.07	−0.1	>0.89
Subjecthood:Modification	0.04	0.09	0.4	>0.68
Linguistic mention:Subjecthood:Modification	−0.06	0.23	−0.3	>0.78

Quantitative comparison with original model

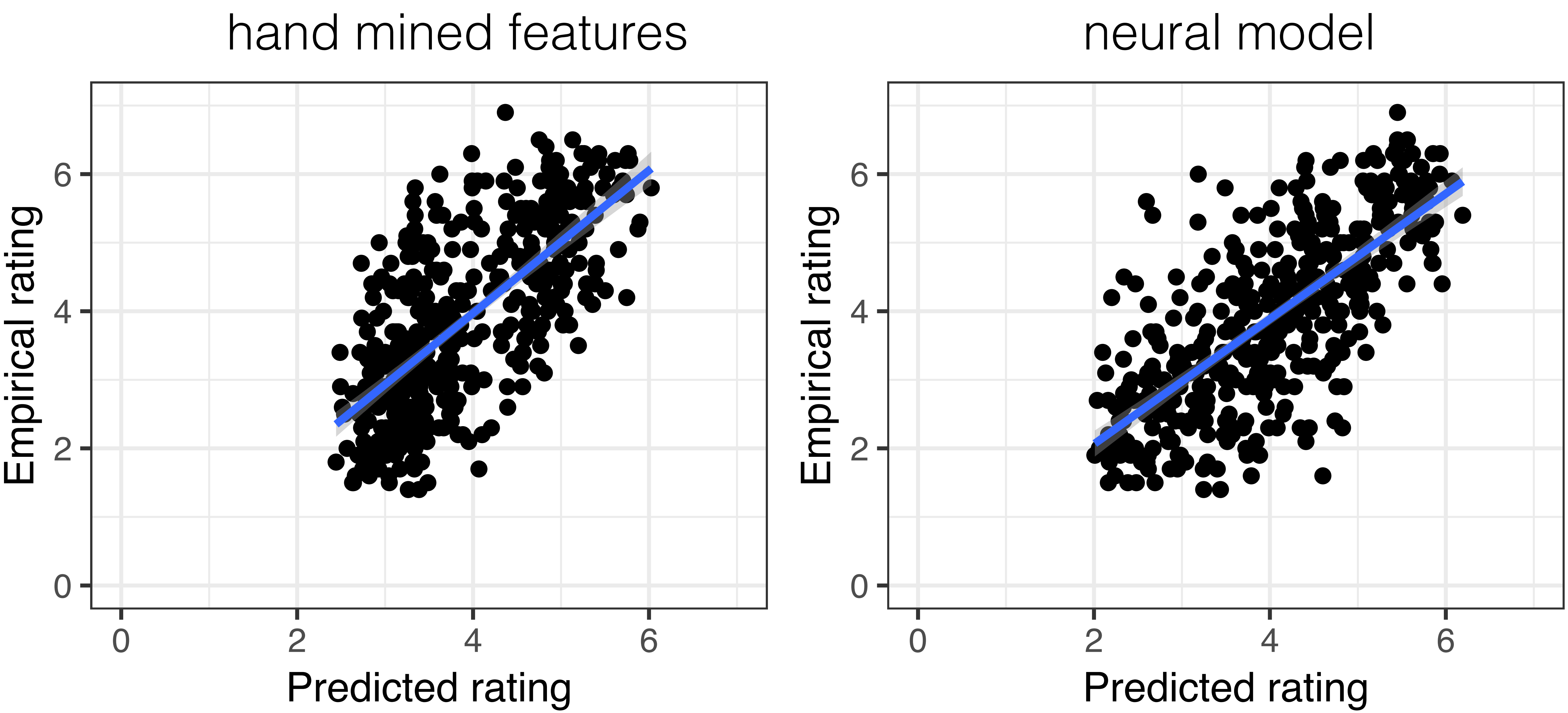
	Coef β	SE(β)	**t**	p
Intercept	4.00	0.05	81.2	**<.0001**
Neural model score	0.88	0.02	57.7	**<.0001**
Partitive	0.08	0.07	1.2	>0.23
Strength	−0.06	0.04	−1.7	>0.08
Linguistic mention	0.08	0.04	1.9	>0.06
Subjecthood	0.01	0.06	0.1	>0.88
Modification	0.02	0.03	0.6	>0.53
Sentence length	0.03	0.03	1.3	>0.21
Partitive:Strength	0.12	0.06	2.2	**<.05**
Linguistic mention:Subjecthood	0.01	0.12	0.1	>0.91
Linguistic mention:Modification	−0.01	0.07	−0.1	>0.89
Subjecthood:Modification	0.04	0.09	0.4	>0.68
Linguistic mention:Subjecthood:Modification	−0.06	0.23	−0.3	>0.78

the neural model subsumes (almost) all hand-mined effects

Quantitative comparison with original model



Quantitative comparison with original model



hand mined features do not explain additional variance

Conclusion

There is much more variability in scalar inferences than commonly assumed — but it's systematically context-dependent, and we can capture a lot of it by inspecting the naturalistic signal.

Recent advances in NLP offer a promising avenue for informing pragmatic theory if we can develop good methods for probing the "black box" neural representations.

Thank you!