Can LLMs draw conversational elicitures and put them to use?

Penny Pan & Andrew Kehler
UC San Diego

CAMP 7 - November 17, 2024

UC San Diego

Large Language Models

and their attested pragmatic abilities

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Large Language Models

and their attested pragmatic abilities

- Large language models (LLMs) achieved high accuracy in various pragmatic tasks and make errors in human-like ways.

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Large Language Models
and their attested pragmatic abilities

- Large language models (LLMs) achieved high accuracy in various pragmatic tasks and make errors in human-like ways.

- Implicatures

    A: “Can you come to my party on Friday?”

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Large Language Models
and their attested pragmatic abilities

- Large language models (LLMs) achieved high accuracy in various pragmatic tasks and make errors in human-like ways.

- Implicatures

  A: “Can you come to my party on Friday?”

  B: “I have work to do.”

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Large Language Models
and their attested pragmatic abilities

- Large language models (LLMs) achieved high accuracy in various pragmatic tasks and make errors in human-like ways.

- Implicatures

  A: “Can you come to my party on Friday?”

  B: “I have work to do.”

  → no, B cannot come to the party.

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Large Language Models
and their attested pragmatic abilities

- Large language models (LLMs) achieved high accuracy in various pragmatic tasks and make errors in human-like ways.

- Implicatures

  A: “Can you come to my party on Friday?”

  B: “I have work to do.”

  → no, B cannot come to the party.

Gricean Maxims

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Large Language Models
and their attested pragmatic abilities

- Large language models (LLMs) achieved high accuracy in various pragmatic tasks and make errors in human-like ways.

- Implicatures

  A: “Can you come to my party on Friday?”

  B: “I have work to do.”

  → no, B cannot come to the party.

  Gricean Maxims

- Indirect speech acts
- Metaphors
- …

Ruis et al. (2024), Hu et al. (2023), *inter alia*

Conversational elicitures

Cohen & Kehler (2021)

Conversational elicitures

Melissa detests the children who are arrogant and rude.

Cohen & Kehler (2021)

Conversational elicitures

Melissa detests the children who are arrogant and rude.

the children are detested **because** they are arrogant and rude.

Cohen & Kehler (2021)

Conversational elicitures

Melissa detests the children who are arrogant and rude.

the children are detested **because** they are arrogant and rude.

Melissa detests the children who live in La Jolla.

Cohen & Kehler (2021)

Conversational elicitures

Melissa detests the children who are arrogant and rude.

the children are detested **because** they are arrogant and rude.

Melissa detests the children who live in La Jolla.

the children are detested **because** … ?

Cohen & Kehler (2021)

Conversational elicitures

Melissa detests the children who are arrogant and rude.

the children are detested **because** they are arrogant and rude.

Melissa detests the children who live in La Jolla.

the children are detested **because** … ?

- A type of pragmatic inferences that are not mandated by any linguistic requirement on utterance felicity.

Cohen & Kehler (2021)

Using elicitures in RC attachment

with human participants

Rohde et al. (2011)

Using elicitures in RC attachment
with human participants

Melissa babysits the children of the musicians who are arrogant and rude.

Rohde et al. (2011)

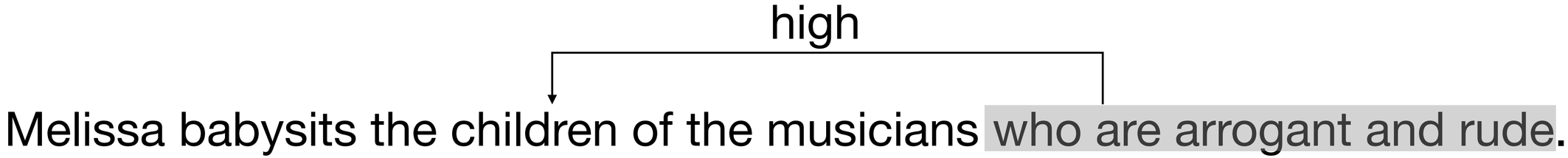
Using elicitures in RC attachment
with human participants

Melissa babysits the children of the musicians who are arrogant and rude.

Rohde et al. (2011)

Using elicitures in RC attachment

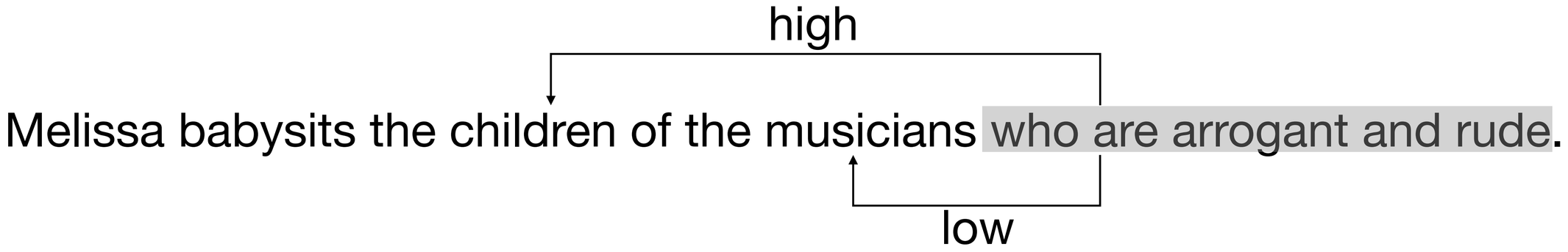
with human participants



Rohde et al. (2011)

Using elicitures in RC attachment

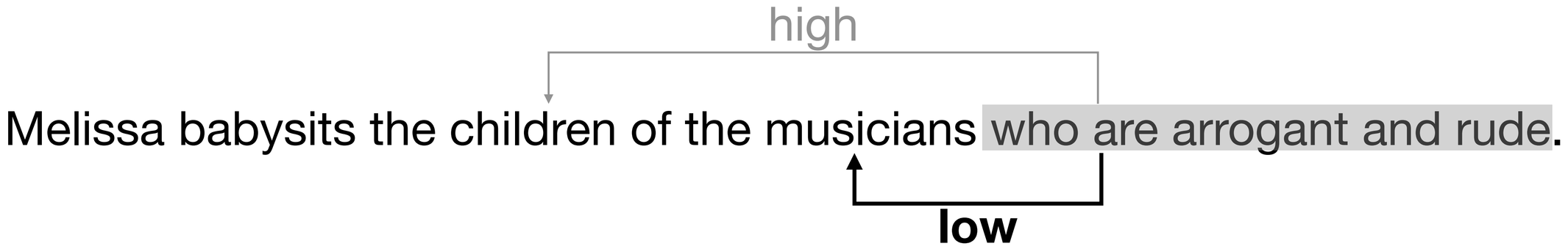
with human participants



Rohde et al. (2011)

Using elicitures in RC attachment

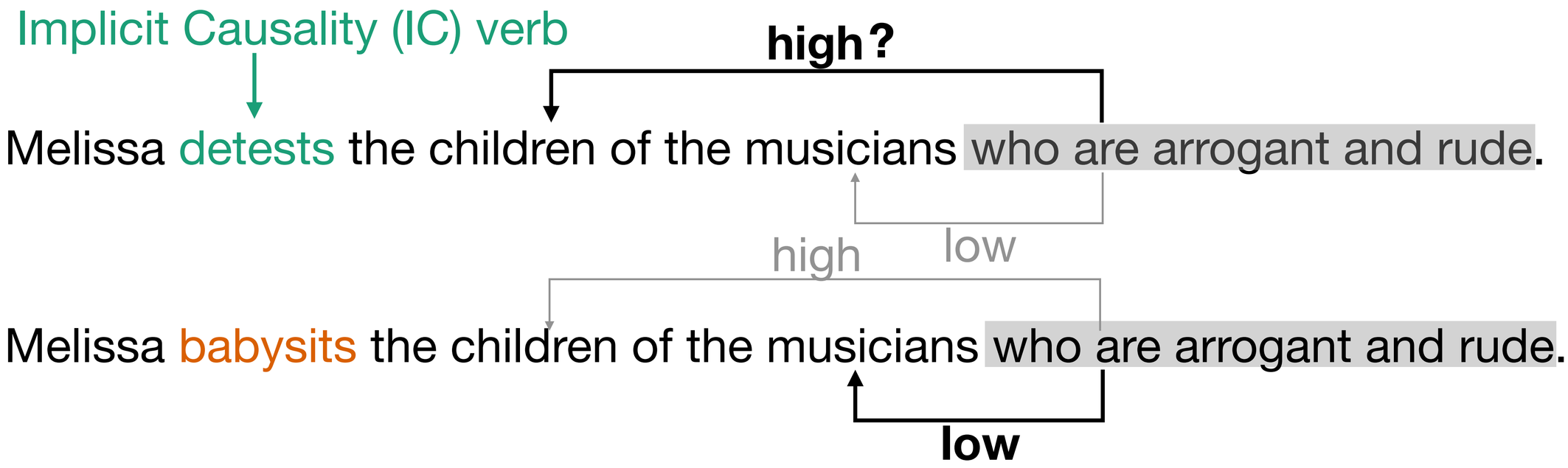
with human participants



Rohde et al. (2011)

Using elicitures in RC attachment

with human participants



Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

Melissa babysits the children of the musicians who are arrogant and rude.

low

Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

- IC verbs create an expectation for an upcoming explanation.

Melissa detests the children.

Melissa babysits the children of the musicians who are arrogant and rude.

Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

- IC verbs create an expectation for an upcoming explanation.
- Explanations can be provided by RCs.

Melissa detests the children who are arrogant and rude.

Melissa babysits the children of the musicians who are arrogant and rude.

Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

- IC verbs create an expectation for an upcoming explanation.
- Explanations can be provided by RCs.
- Certain IC verbs are biased towards the object, such that the explanation is more likely to be about the object.

Melissa detests the children who are arrogant and rude.

Melissa babysits the children of the musicians who are arrogant and rude.

Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

- IC verbs create an expectation for an upcoming explanation.
- Explanations can be provided by RCs.
- Certain IC verbs are biased towards the object, such that the explanation is more likely to be about the object.

Melissa detests the children of the musicians who are arrogant and rude.

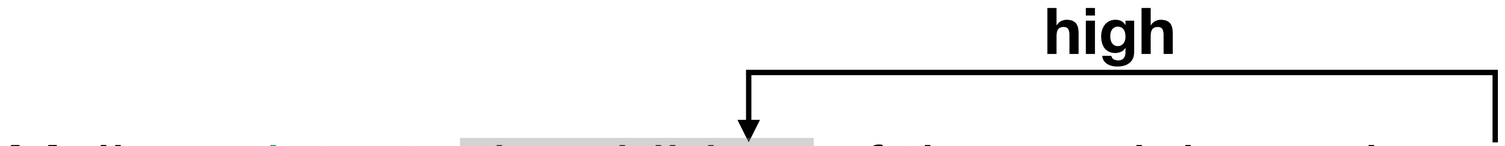
Melissa babysits the children of the musicians who are arrogant and rude.

Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

- IC verbs create an expectation for an upcoming explanation.
- Explanations can be provided by RCs.
- Certain IC verbs are biased towards the object, such that the explanation is more likely to be about the object.



Rohde et al. (2011)

Using elicitures in RC attachment

with human participants

Can LLMs use conversational elicitures?

Rohde et al. (2011)

Tasks and models

Tasks and models

- Study 1: Detecting elicitures
  - Can LLMs draw conversational elicitures?

Tasks and models

- Study 1: Detecting elicitures
  - Can LLMs draw conversational elicitures?
- Study 2: Using elicitures
  - Can LLMs use conversational elicitures to process RCs?

Tasks and models

- Study 1: Detecting elicitures
  - Can LLMs draw conversational elicitures?
- Study 2: Using elicitures
  - Can LLMs use conversational elicitures to process RCs?
- Model selections
  - GPT-3.5-turbo, GPT-4, and GPT-4o
  - Access through OpenAI API

Study 1: Detecting elicitures

- IC verbs (60 sentences = 20 verbs x 3 items)
  Melissa detests the children who are generally arrogant and rude.
- non-IC verbs (60 sentences = 20 verbs x 3 items)
  Melissa babysits the children who are generally arrogant and rude.

Study 1: Detecting elicitures

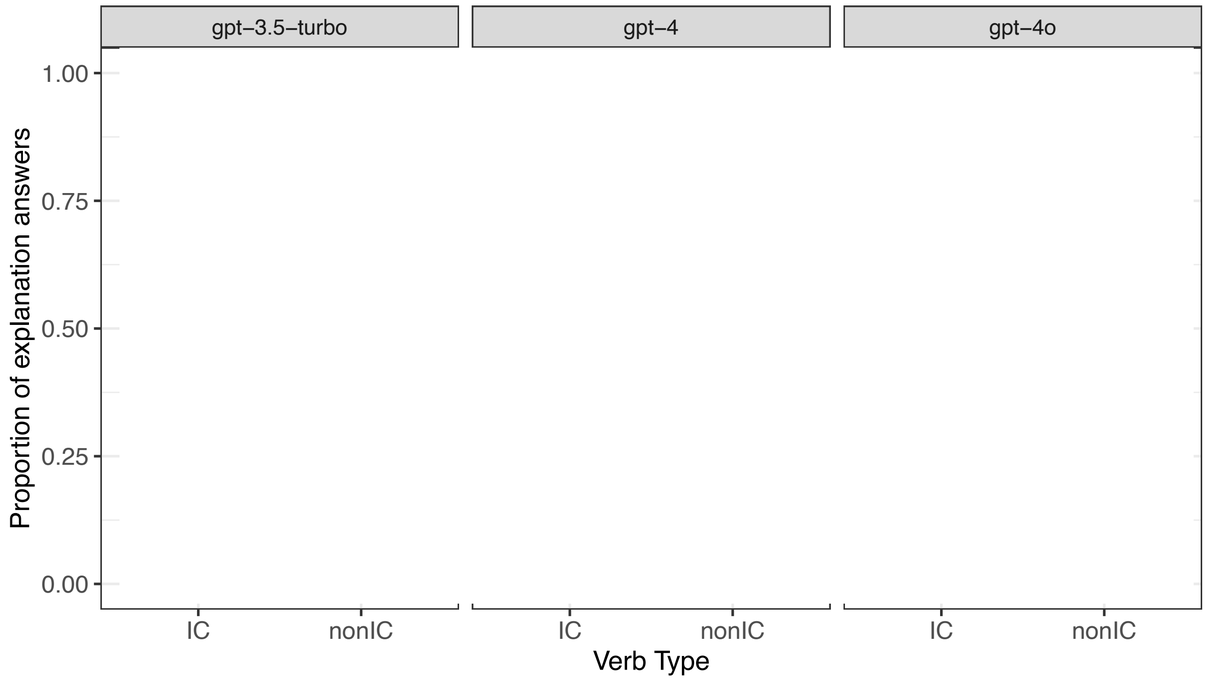
Sentence: Melissa detests/babysits the children who are generally arrogant and rude.

Question: Does the sentence explain why Melissa detests/babysits the children? If yes, please provide an explanation. If not, just say no and you don't need an explanation.

Study 1: Detecting elicitures

Sentence: Melissa detests/babysits the children who are generally arrogant and rude.

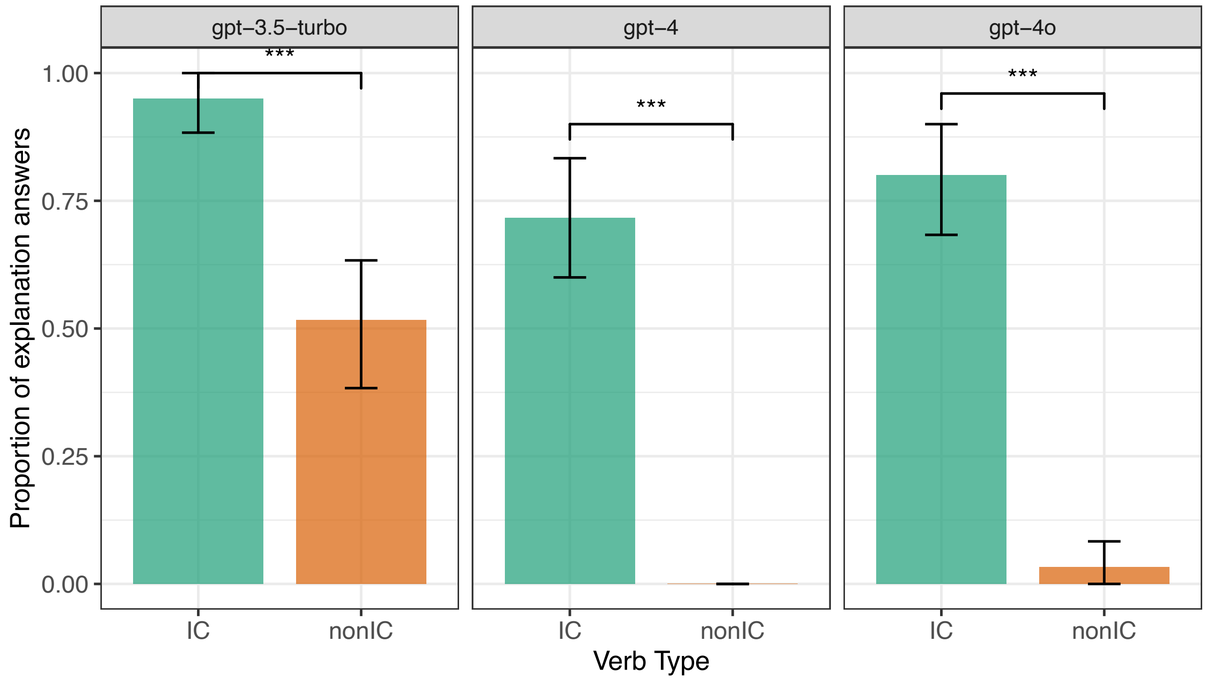
Question: Does the sentence explain why Melissa detests/babysits the children? If yes, please provide an explanation. If not, just say no and you don't need an explanation.



Study 1: Detecting elicitures

Sentence: Melissa detests/babysits the children who are generally arrogant and rude.

Question: Does the sentence explain why Melissa detests/babysits the children? If yes, please provide an explanation. If not, just say no and you don't need an explanation.



Study 2: Using elicitures

Relative clause attachment

- IC verbs (60 sentences = 20 verbs x 3 items)
  Melissa detests the children of the musician who ___
- non-IC verbs (60 sentences = 20 verbs x 3 items)
  Melissa babysits the children of the musician who ___

Study 2: Using elicitures

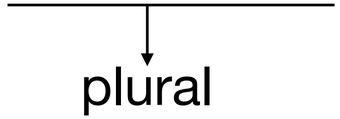
Relative clause attachment

- IC verbs (60 sentences = 20 verbs x 3 items)

  Melissa detests the children of the musician who ___

- non-IC verbs (60 sentences = 20 verbs x 3 items)

  Melissa babysits the children of the musician who ___



Study 2: Using elicitures

Relative clause attachment

- IC verbs (60 sentences = 20 verbs x 3 items)

  Melissa detests the children of the musician who ___

- non-IC verbs (60 sentences = 20 verbs x 3 items)

  Melissa babysits the children of the musician who ___

plural singular

Study 2: Using elicitures

Relative clause attachment

Sentence: Melissa detests/babysits the children of the musician who ___

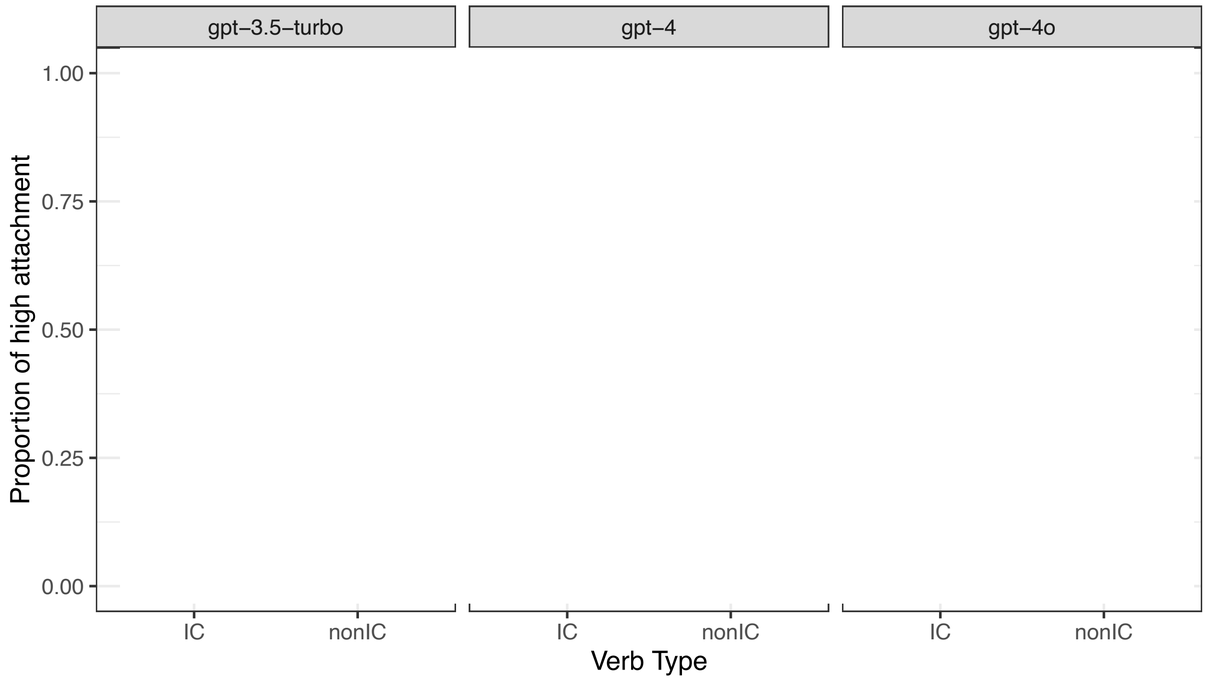
Options: 1) is, 2) are

Study 2: Using elicitures

Relative clause attachment

Sentence: Melissa detests/babysits the children of the musician who ___

Options: 1) is, 2) are

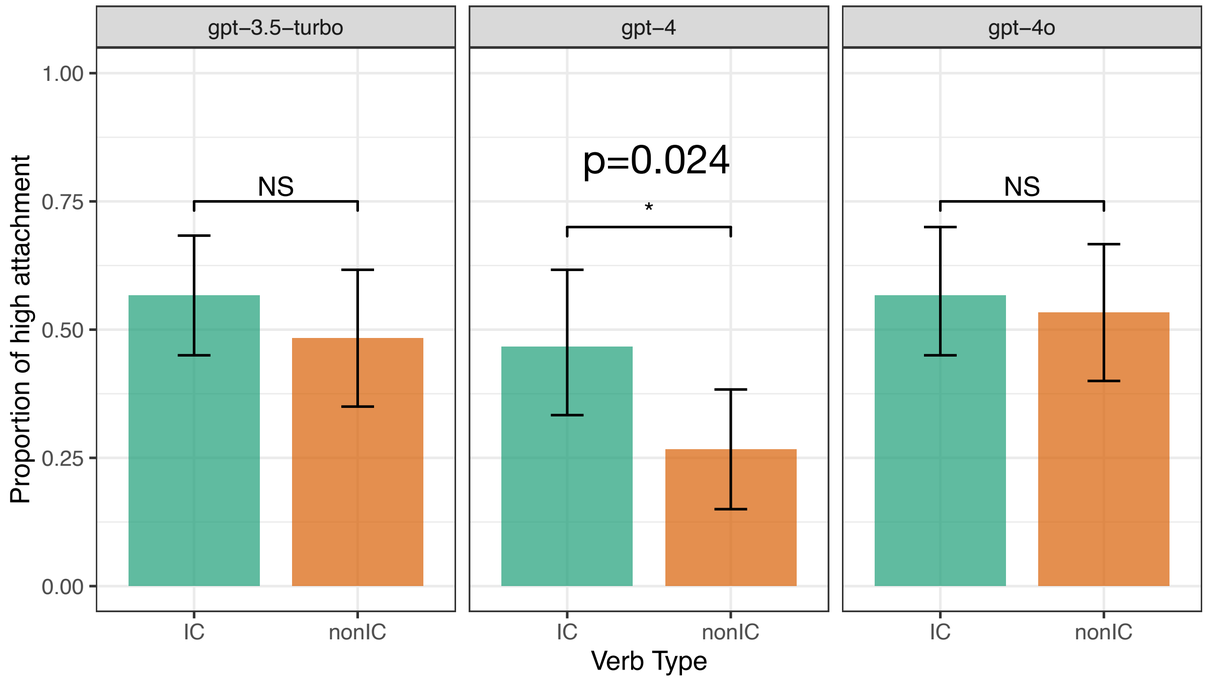


Study 2: Using elicitures

Relative clause attachment

Sentence: Melissa detests/babysits the children of the musician who ___

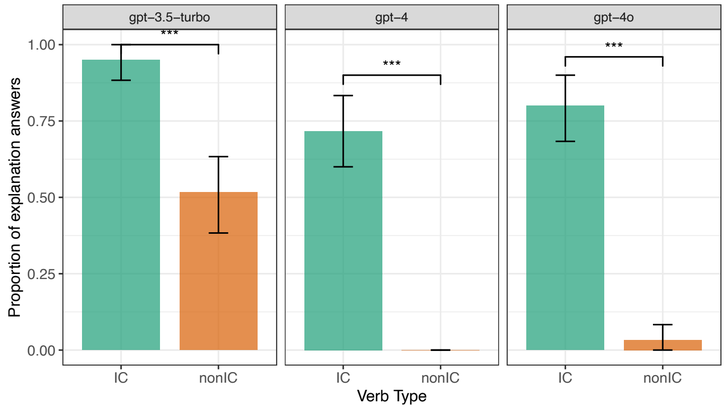
Options: 1) is, 2) are



Summary

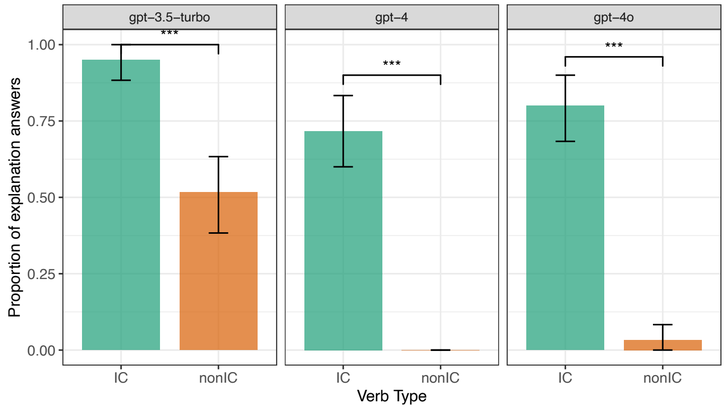
Summary

- Study 1: LLMs can detect elicitures



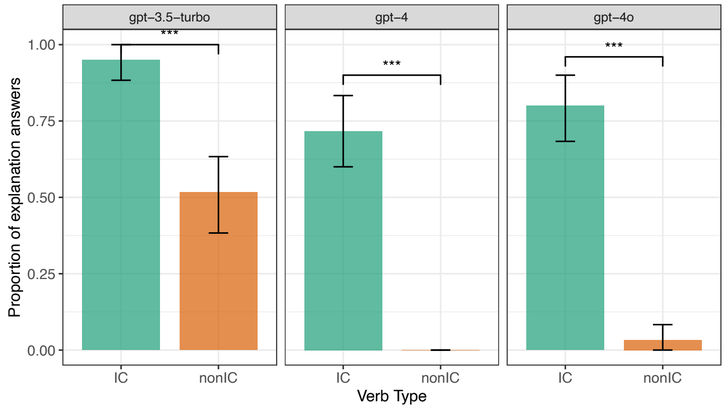
Summary

- Study 1: LLMs can detect elicitures
  - More likely to judge that the explanation RC provides an explanation for IC verbs than for nonIC verbs.



Summary

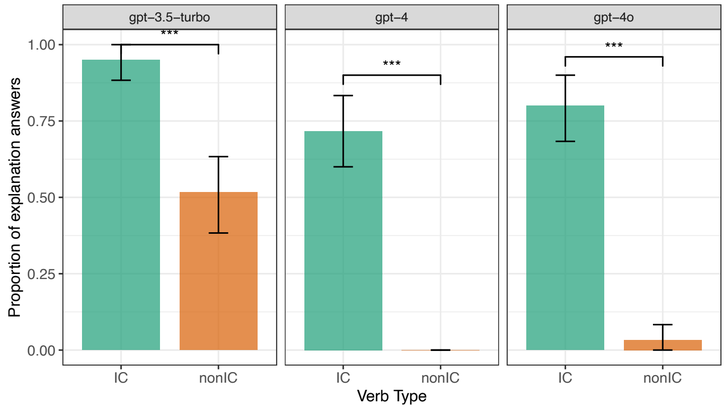
- Study 1: LLMs can detect elicitures
  - More likely to judge that the explanation RC provides an explanation for IC verbs than for nonIC verbs.



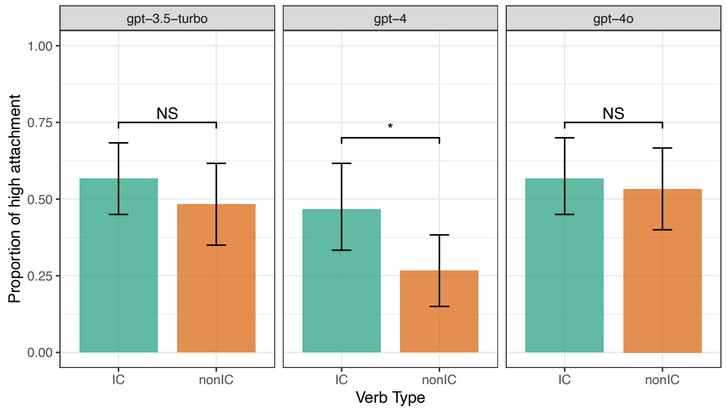
- Study 2: some LLMs (e.g., GPT4) are able to use elicitures to guide RC attachment decision.

Summary

- Study 1: LLMs can detect elicitures
  - More likely to judge that the explanation RC provides an explanation for IC verbs than for nonIC verbs.

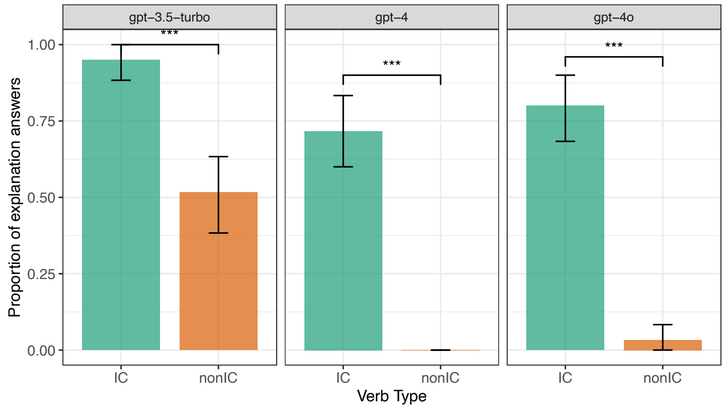


- Study 2: some LLMs (e.g., GPT4) are able to use elicitures to guide RC attachment decision.
  - Significant difference in the high attachment preference between the IC and nonIC verbs.

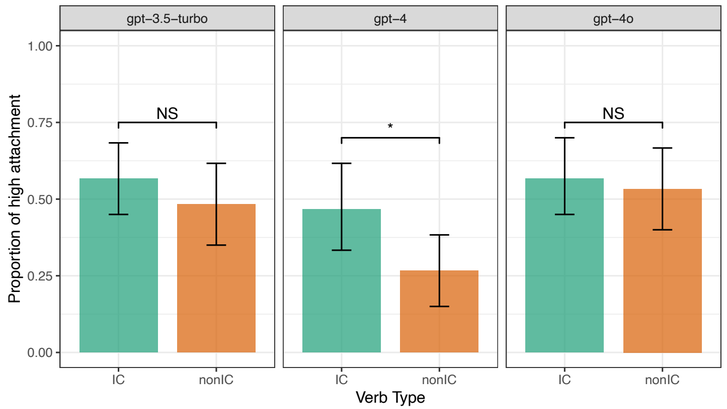


Summary

- Study 1: LLMs can detect elicitures
  - More likely to judge that the explanation RC provides an explanation for IC verbs than for nonIC verbs.



- Study 2: some LLMs (e.g., GPT4) are able to use elicitures to guide RC attachment decision.
  - Significant difference in the high attachment preference between the IC and nonIC verbs.



LLMs have certain pragmatic abilities, and some, but not all, models are using pragmatic inferences in syntactic processing.

Limitations and future directions

Why non-significant results in GPT-3.5-turbo and GPT-4o

Limitations and future directions

Why non-significant results in GPT-3.5-turbo and GPT-4o

- A limited amount of data?
  - 60 items per verb type, and each model as a single participant.

Limitations and future directions

Why non-significant results in GPT-3.5-turbo and GPT-4o

- A limited amount of data?
  - 60 items per verb type, and each model as a single participant.
- Prompting?
  - as a proxy for raw probability distribution, but the results from prompting may not always align with the raw probabilities. Hu & Levy (2023)

Limitations and future directions

Why non-significant results in GPT-3.5-turbo and GPT-4o

- A limited amount of data?
  - 60 items per verb type, and each model as a single participant.
- Prompting?
  - as a proxy for raw probability distribution, but the results from prompting may not always align with the raw probabilities. Hu & Levy (2023)
- Next steps
  - to obtain the raw probabilities in other open source language models (e.g., Llama).
  - to see whether the probability of the models can predict the human reading time.

Acknowledgements

Thanks to

- Sean Trott and Alex Warstadt for discussion

https://github.com/pennydy/llm_eliciture

UC San Diego

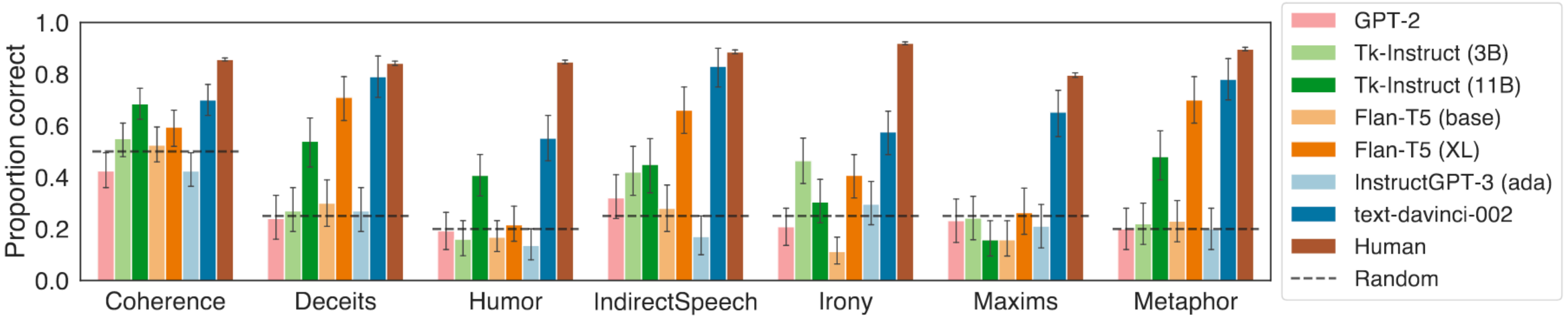
Thank you!

Extra slides

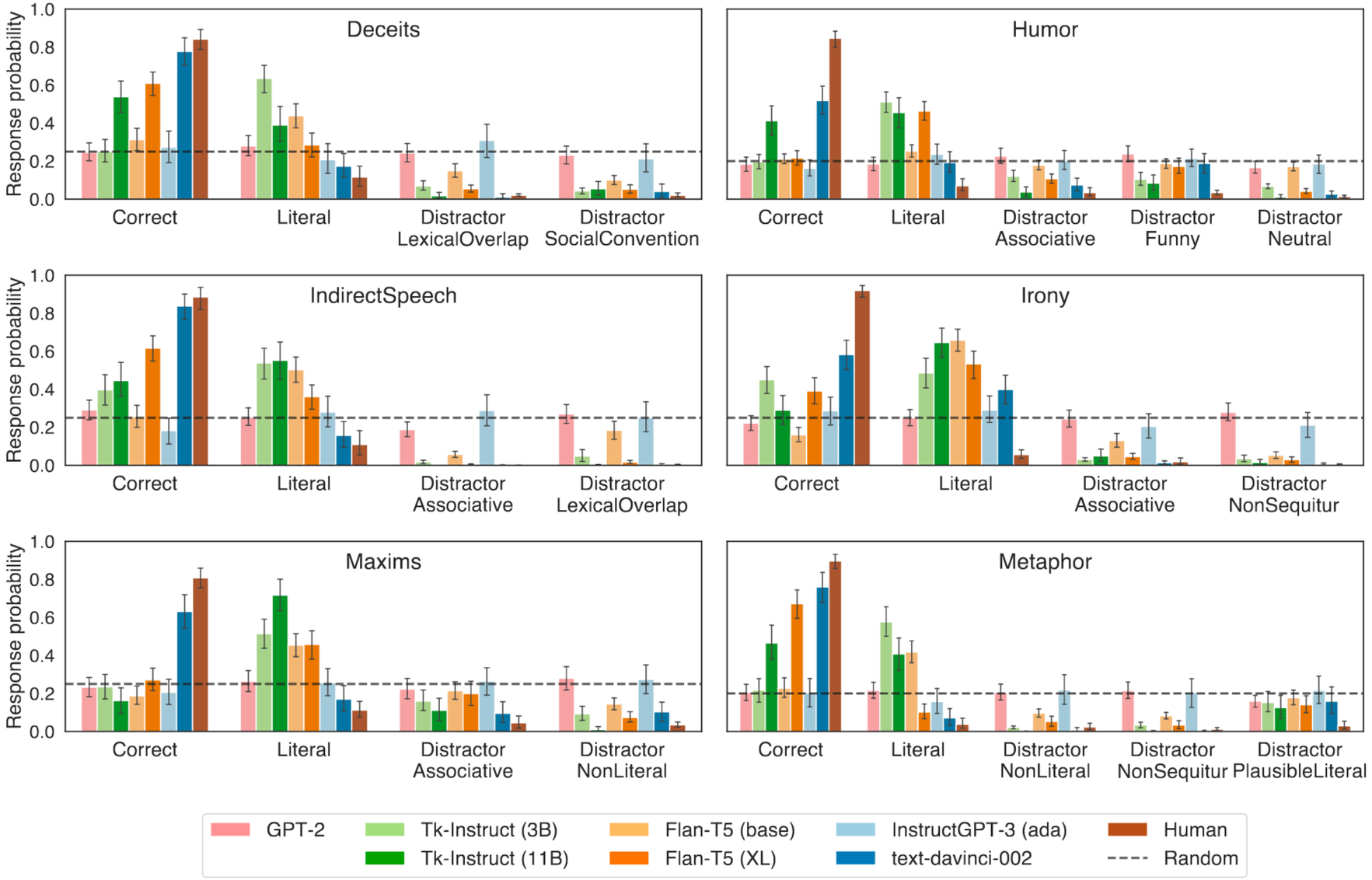
LLMs and their pragmatic abilities

Hu et al. (2023)

pragmatic tasks



Hu et al. (2023)


Deceits
Humor
IndirectSpeech
Irony
Maxims
Metaphor
Response probability
1.0
0.8
0.6
0.4
0.2
0.0
Correct
Literal
Distractor LexicalOverlap
Distractor SocialConvention
Distractor Associative
Distractor Funny
Distractor Neutral
Distractor LexicalOverlap
Distractor NonSequitur
Distractor NonLiteral
Distractor PlausibleLiteral
GPT-2
Tk-Instruct (3B)
Tk-Instruct (11B)
Flan-T5 (base)
Flan-T5 (XL)
InstructGPT-3 (ada)
text-davinci-002
Human
Random

Human experiments

Rohde et al. (2011)

Sentence completion task

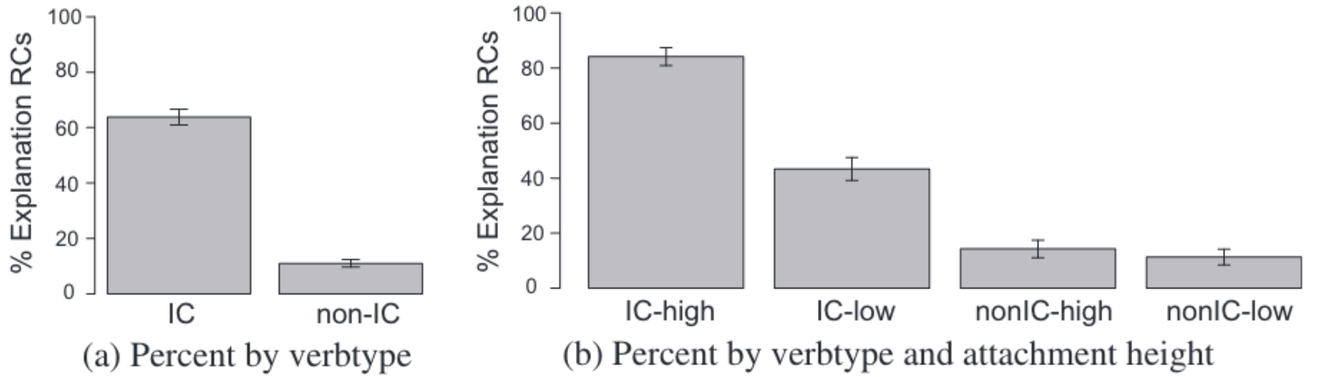


Fig. 2. Percentage of explanation RCs by verbtype and attachment height.

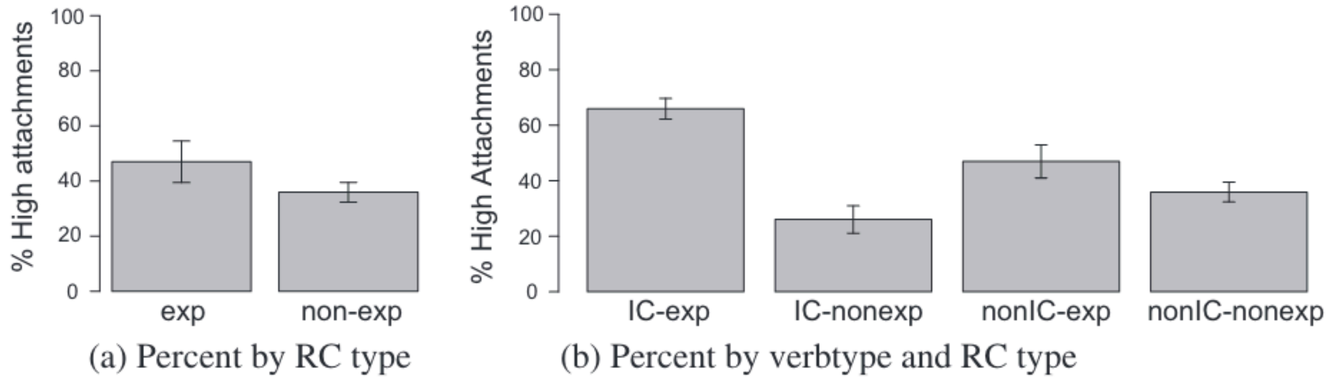
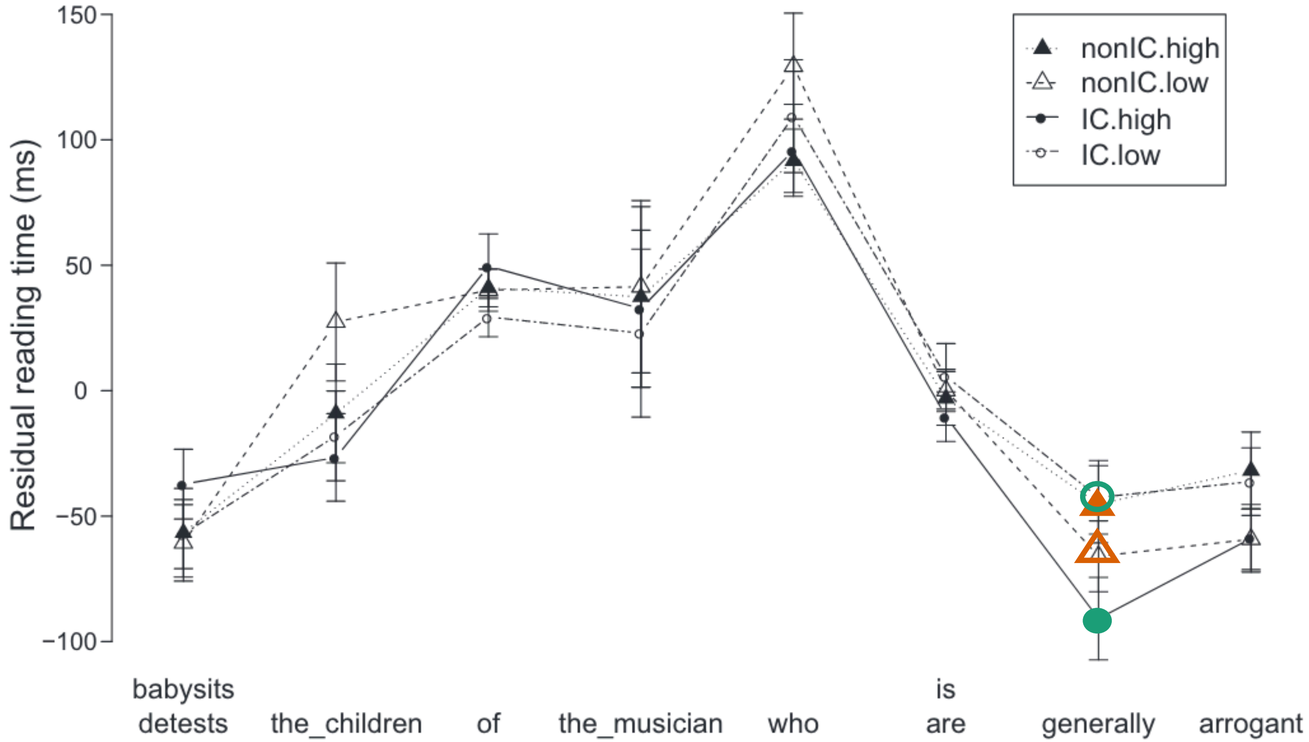


Fig. 3. Percentage of high attachments by verbtype and RC type.

Rohde et al. (2011)

Self-paced reading task

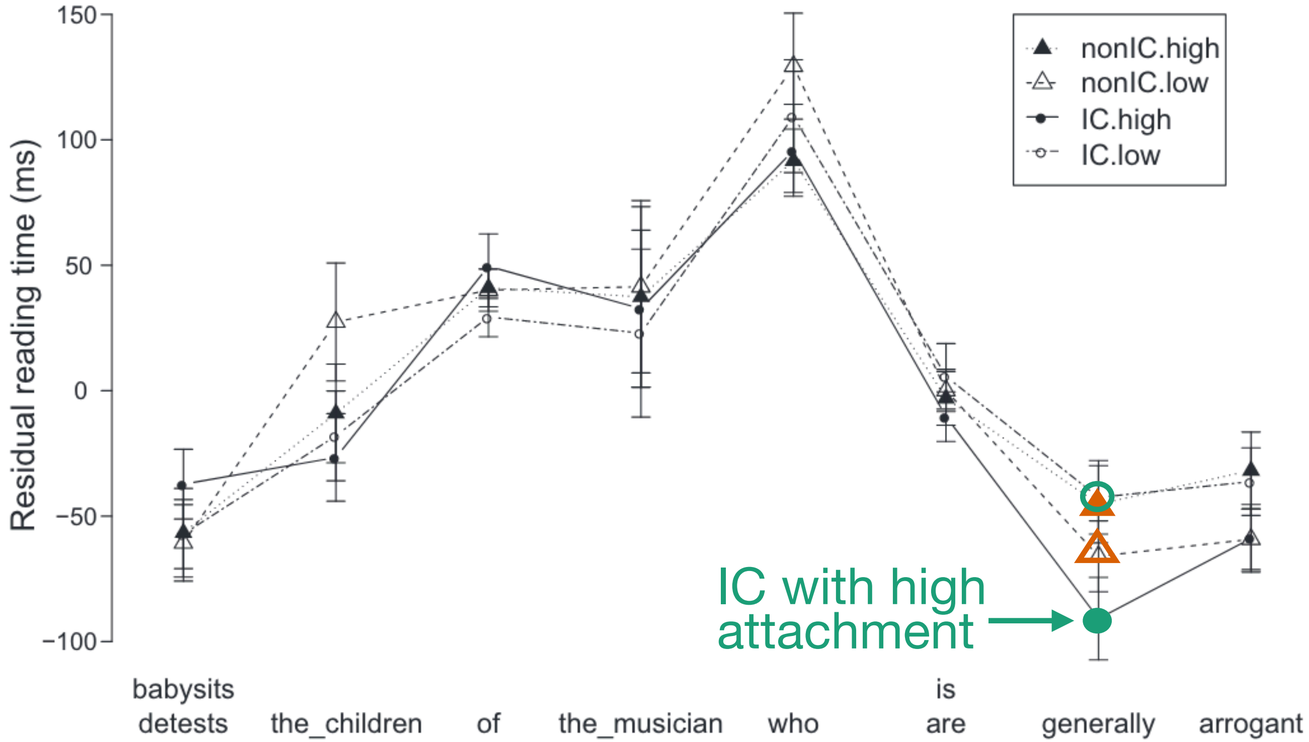
- Humans expect an explanation continuation after an object-based IC verb and show a high-attachment preference for an ambiguous RC



Rohde et al. (2011)

Self-paced reading task

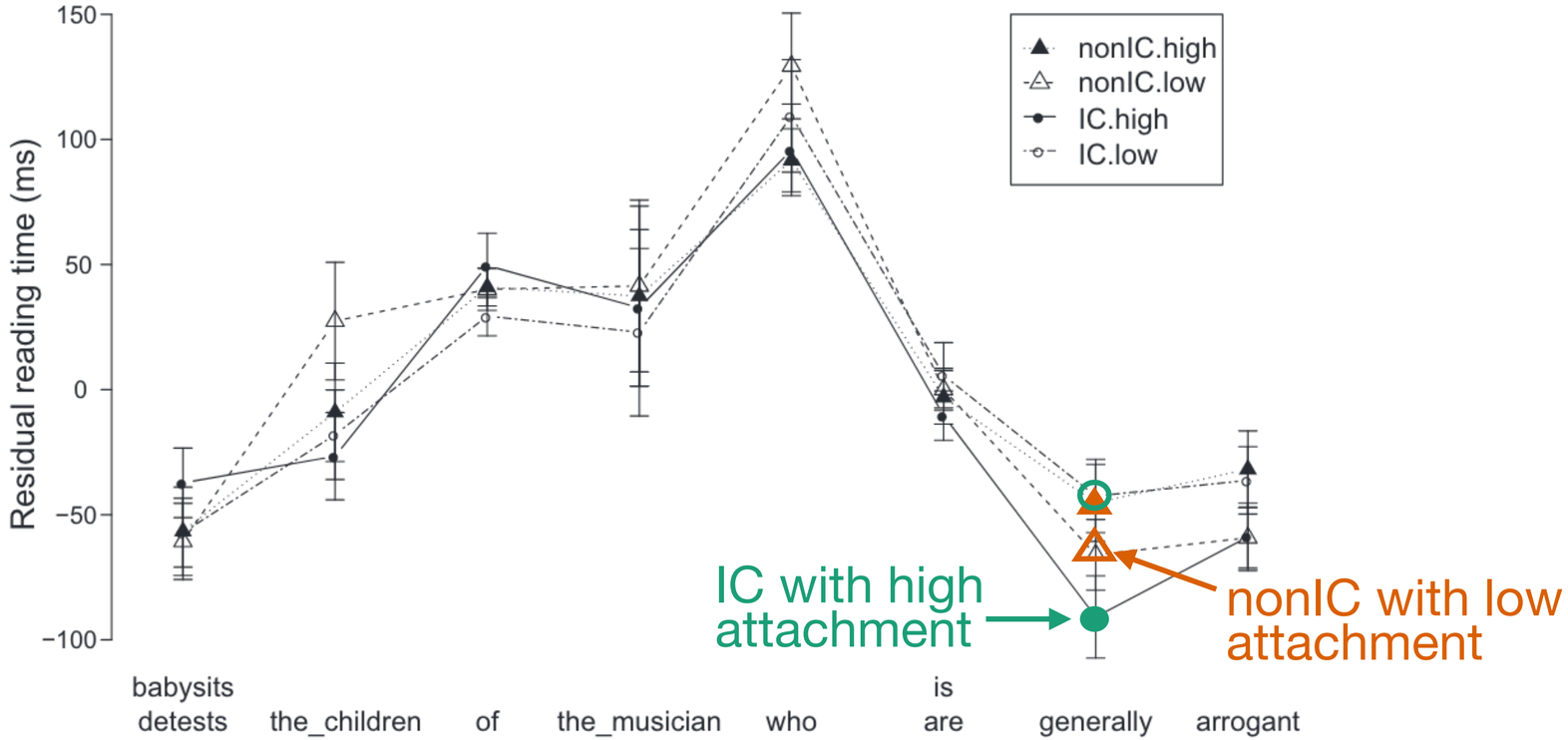
- Humans expect an explanation continuation after an object-based IC verb and show a high-attachment preference for an ambiguous RC



Rohde et al. (2011)

Self-paced reading task

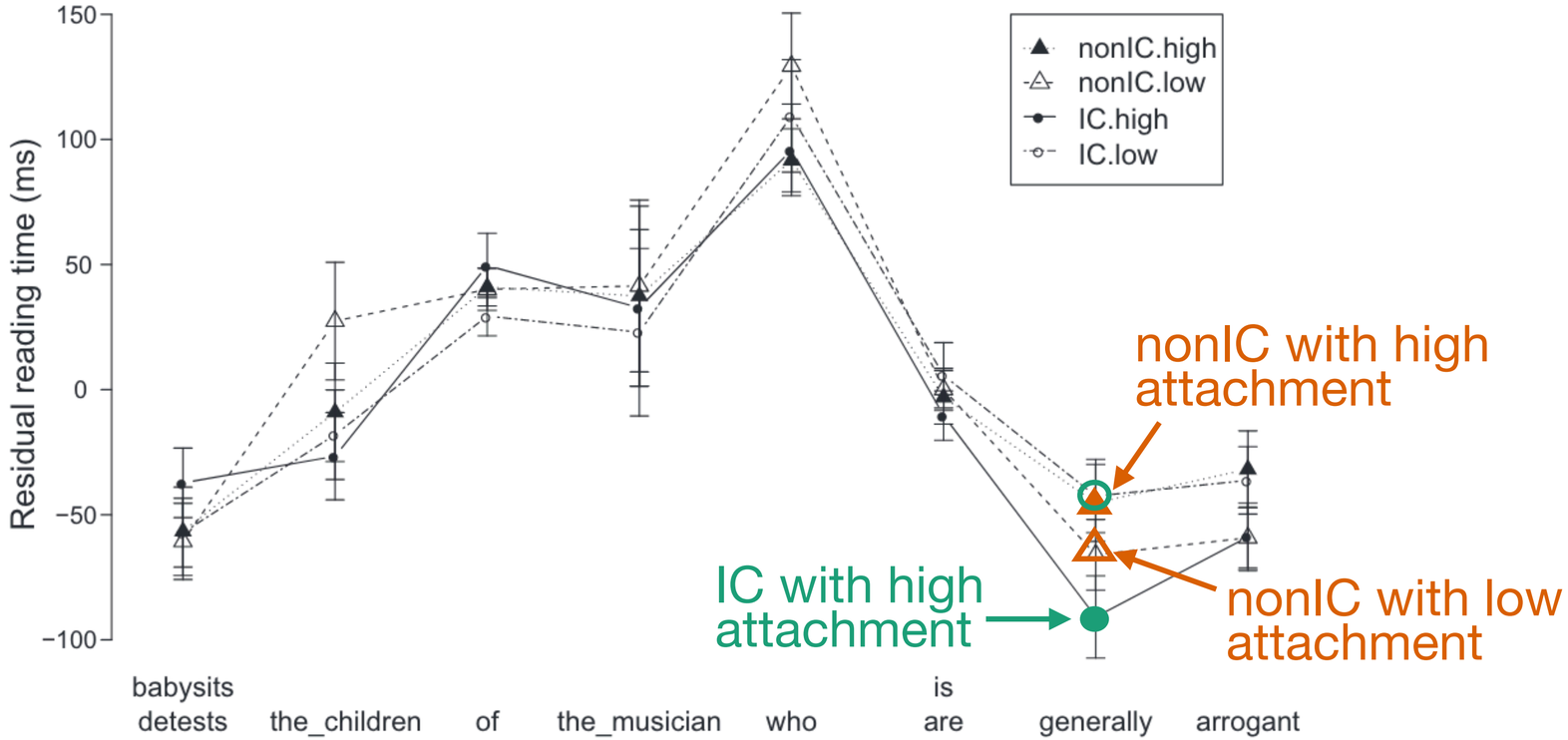
- Humans expect an explanation continuation after an object-based IC verb and show a high-attachment preference for an ambiguous RC



Rohde et al. (2011)

Self-paced reading task

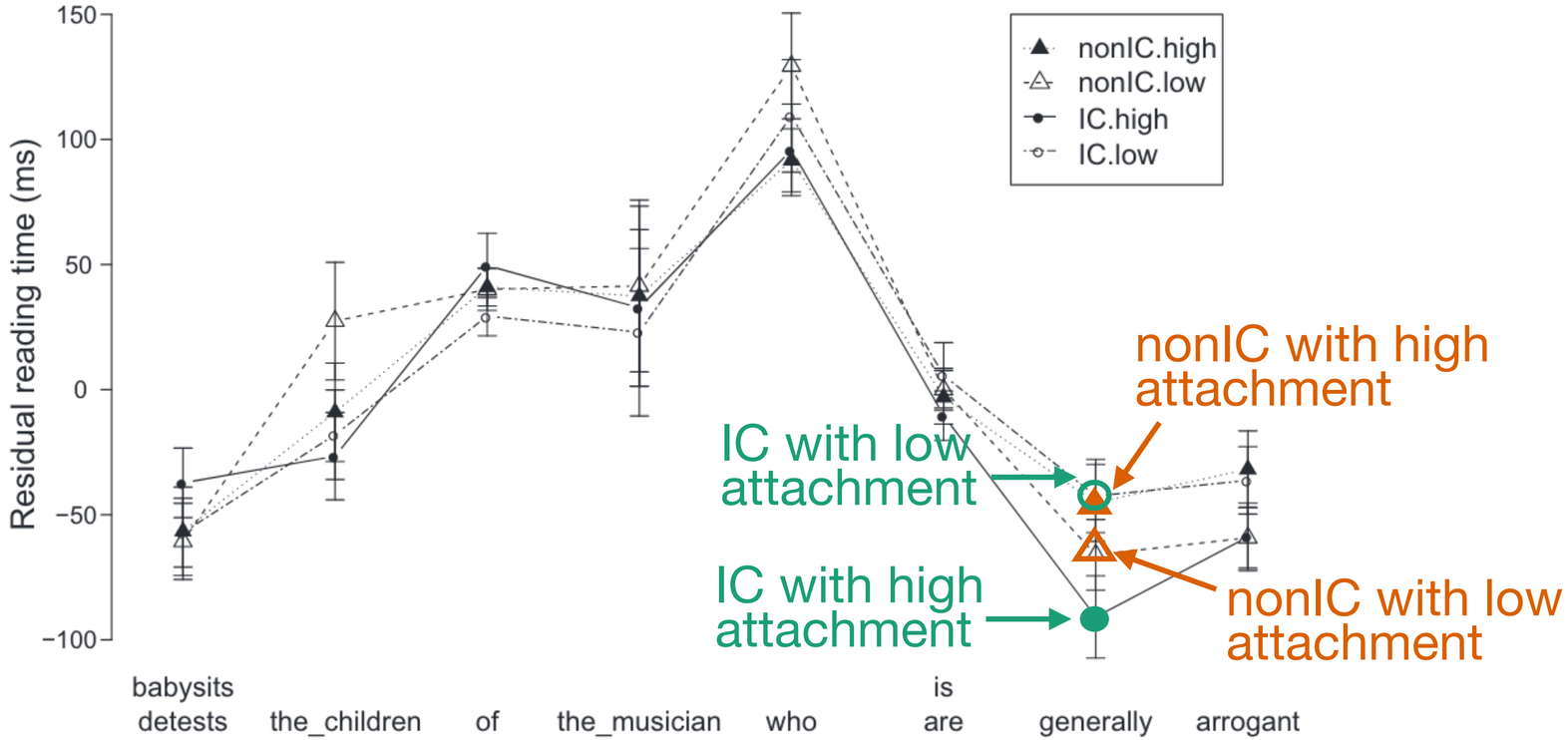
- Humans expect an explanation continuation after an object-based IC verb and show a high-attachment preference for an ambiguous RC



Rohde et al. (2011)

Self-paced reading task

- Humans expect an explanation continuation after an object-based IC verb and show a high-attachment preference for an ambiguous RC



Rohde et al. (2011)

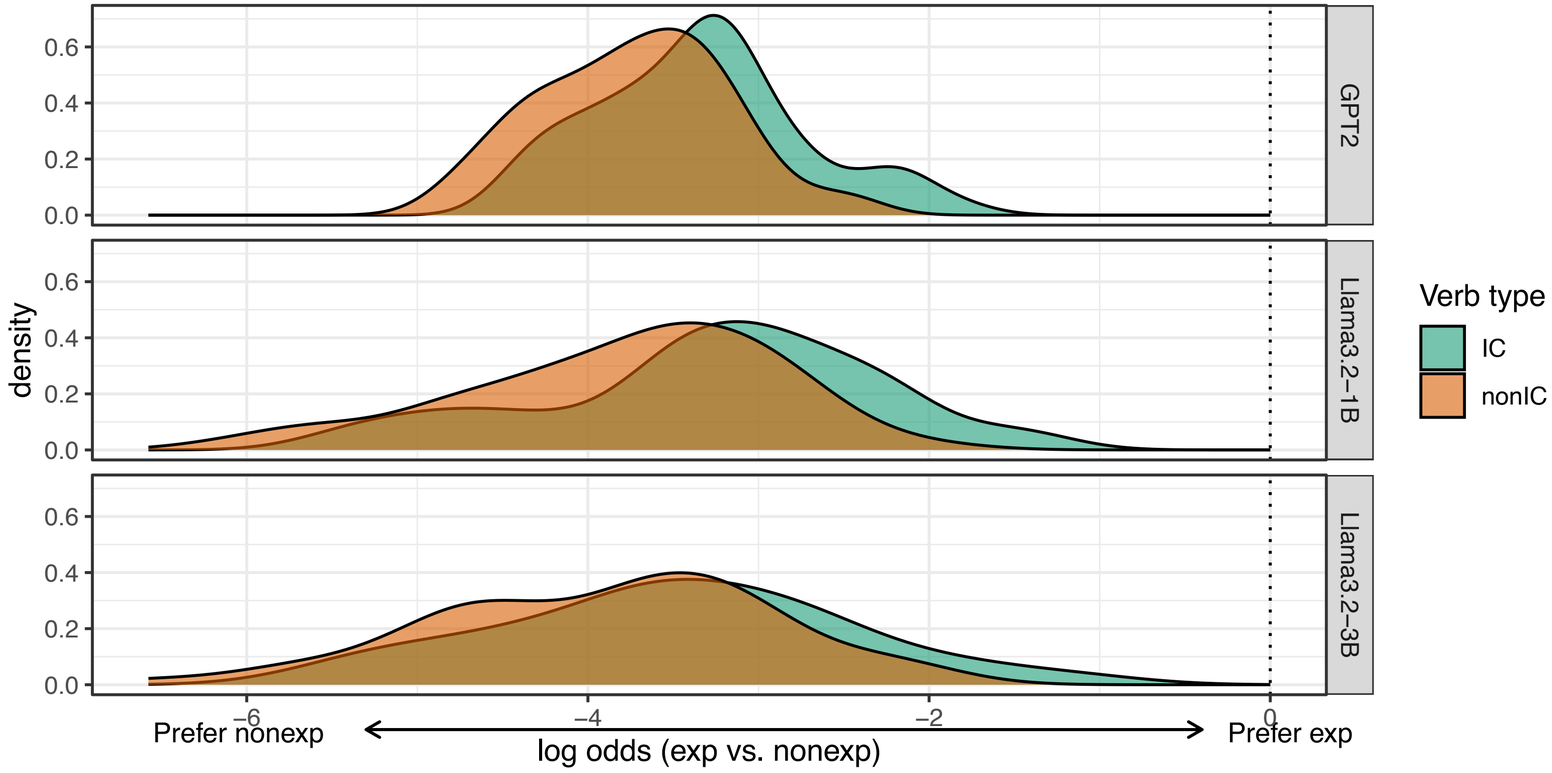
Self-paced reading task

Study 1: Detecting elicitures (raw probability distribution)

1. IC + exp RC + because: Melissa detests the children who are arrogant and rude, **because**
2. IC + exp RC + and: Melissa detests the children who are arrogant and rude, **and**
3. nonIC + exp RC + because: Melissa babysits the children who are arrogant and rude, **because**
4. nonIC + exp RC + and: Melissa babysits the children who are arrogant and rude, **and**

$$\text{log odds} = \log(\frac{P(because)}{P(and)})$$

log odds ~ verb_type + (1|item)



1. IC + exp RC: Melissa detests the children who are arrogant and rude, **because**
2. IC + nonexp RC: Melissa detests the children who live in La Jolla, **because**

$$\text{log odds} = \log\left(\frac{P(because_{\text{exp RC}})}{P(because_{\text{nonexp RC}})}\right)$$

$$\text{log odds} \sim \text{RC\_type} + (1|\text{item})$$

Predictions

- log odds < 0 for IC verbs, i.e., less like to have another explanation with “because” when the RC has already provided one

Study 2: Putting elicitures to use (raw probability distribution)

1. IC high: Melissa detests the children of the musician who **are**
2. IC low: Melissa detests the children of the musician who **is**
3. nonIC high: Melissa babysits the children of the musician who **are**
4. nonIC low: Melissa babysits the children of the musician who **is**

$$\text{log odds} = \log(\frac{P(high)}{P(low)})$$

log odds ~ verb_type + (1|item)