Quantifying weak and strong crossover

HAYLEY ROSS, GENNARO CHIERCHIA AND KATHRYN DAVIDSON

SINN UND BEDEUTUNG 27, 2022

https://tinyurl.com/SuB-crossover

Structure of this talk

1. Designing experiments for bound pronouns
2. Experiment 1: *wh*-crossover

   We'll use *wh*-crossover to set a baseline and showcase the experiment design
3. Experiment 2: proper name cataphora

   We'll gather data on these disputed judgements and highlight an open theoretical question

Binding vs. crossover

Binding: pronoun co-varies with a c-commanding antecedent

wh … [gap] … pronoun

(1) The teacher wondered **which$_i$ of the students** __ enjoyed the essay topic **they$_i$** had chosen.

each student chose their own topic

Crossover: bound reading of the pronoun is blocked

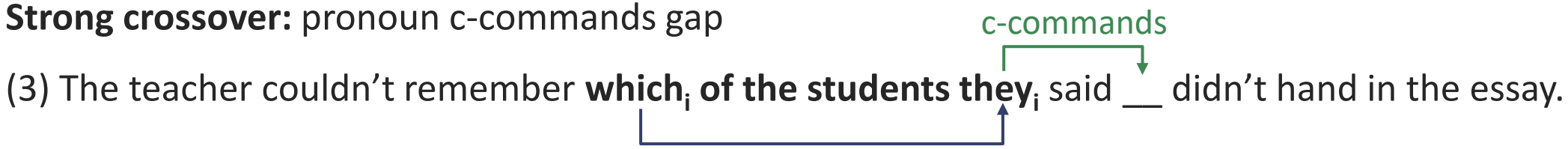
wh … pronoun … [gap]

(2) The teacher couldn't decide **which$_i$ student**'s poem topic **they$_i$** liked __ the most.

not available: each student likes their own topic

Strong vs. weak crossover

Strong crossover: pronoun c-commands gap



Secondary strong crossover: pronoun c-commands gap, *wh*-word in Spec

Weak crossover: pronoun in Spec of XP, XP c-commands gap

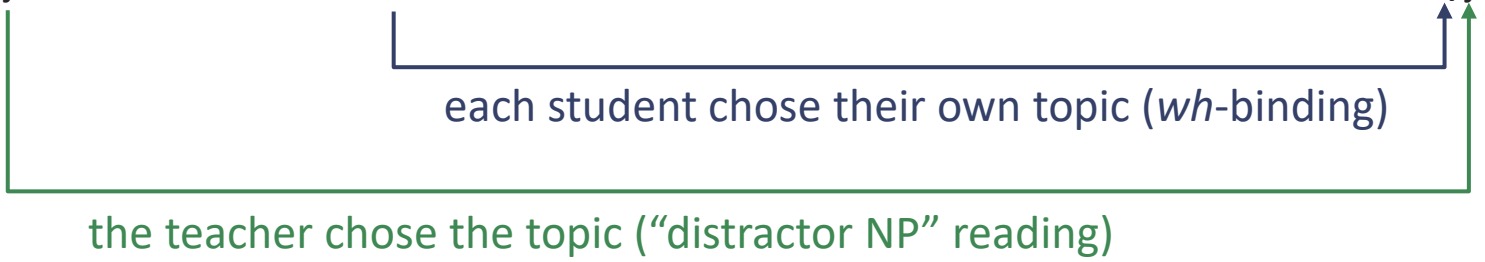
Disentangling pronoun reference

We want to study a particular reading (co-indexation) of the sentence

- Pronoun could be bound – or it could *corefer* with some other salient entity
- Our experiment design takes advantage of this ambiguity!

Each sentence has two possible readings:

(1) The **teacher**$_j$ wondered **which**$_i$ **of the students** __ enjoyed the essay topic **they**$_{i/j}$ had chosen.



Experiment design pilot

Rating **meaning availability** gave more intuitive results on binding than **sentence acceptability**

(A) Sentence acceptability (e.g. Kush 2013)

In English literature class, this week's assignment was for each student to write an essay on a topic of their choice. The students had each chosen an essay topic, but not all of them were enjoying their topic after all. The teacher wondered which students were enjoying it.

How acceptable is the following sentence in this situation?

	Completely unacceptable	Not really	Unsure	Mostly	Completely acceptable
The new teacher wondered which of the students enjoyed the essay topic they had chosen.	○	○	○	○	○

(B) Meaning availability

The new teacher wondered which of the students enjoyed the essay topic they had chosen.

To what degree can this mean...

	Definitely no	Not really	Unsure	Mostly	Definitely yes
The new teacher had chosen an essay topic, and wondered which of the students enjoyed it.	○	○	○	○	○
The students had each chosen an essay topic, and the new teacher wondered which of them enjoyed their topic.	○	○	○	○	○

Experiment design pilot

Rating **meaning availability** gave more intuitive results on binding than **sentence acceptability**

(A) Sentence acceptability (e.g. Kush 2013)

Context for bound reading

In English literature class, this week's assignment was for each student to write an essay on a topic of their choice. The students had each chosen an essay topic, but not all of them were enjoying their topic after all. The teacher wondered which students were enjoying it.

How acceptable is the following sentence in this situation?

Target sentence

The new teacher wondered which of the students enjoyed the essay topic they had chosen.

	Completely unacceptable	Not really	Unsure	Mostly	Completely acceptable
The new teacher wondered which of the students enjoyed the essay topic they had chosen.	○	○	○	○	○

(B) Meaning availability

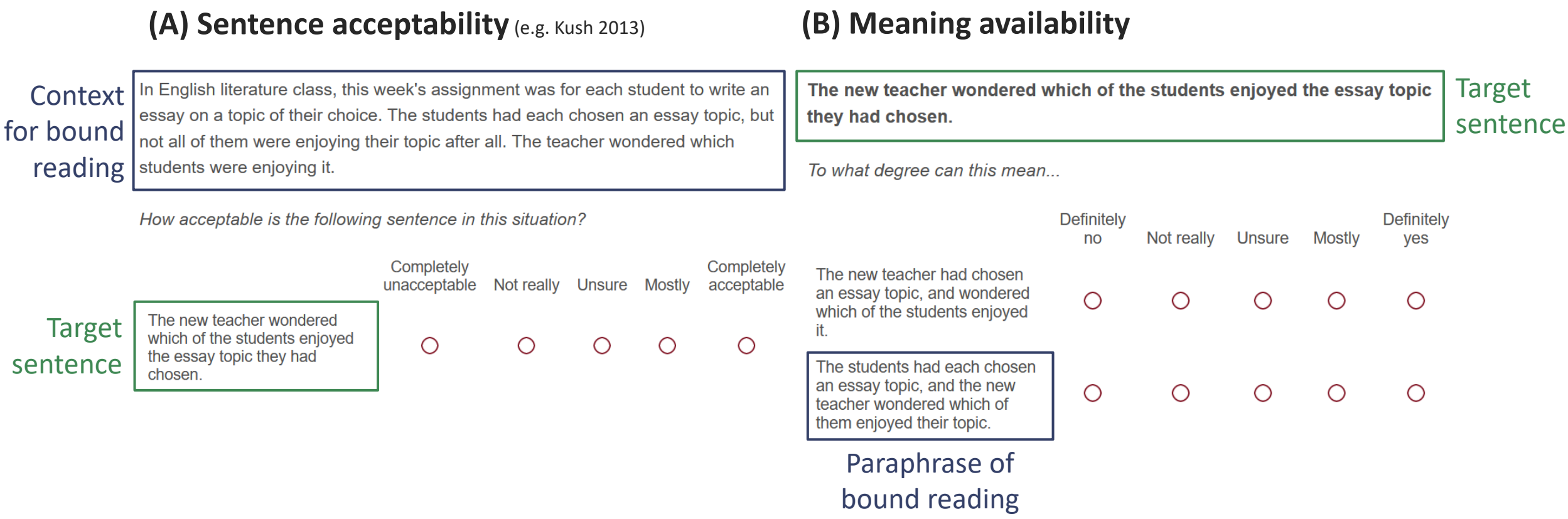
The new teacher wondered which of the students enjoyed the essay topic they had chosen.

To what degree can this mean...

	Definitely no	Not really	Unsure	Mostly	Definitely yes
The new teacher had chosen an essay topic, and wondered which of the students enjoyed it.	○	○	○	○	○
The students had each chosen an essay topic, and the new teacher wondered which of them enjoyed their topic.	○	○	○	○	○

Experiment design pilot

Rating **meaning availability** gave more intuitive results on binding than **sentence acceptability**



Experiment 1: *wh*-crossover

Experiment setup

We use a **2x3x2 factor design** with a 5-point Likert scale

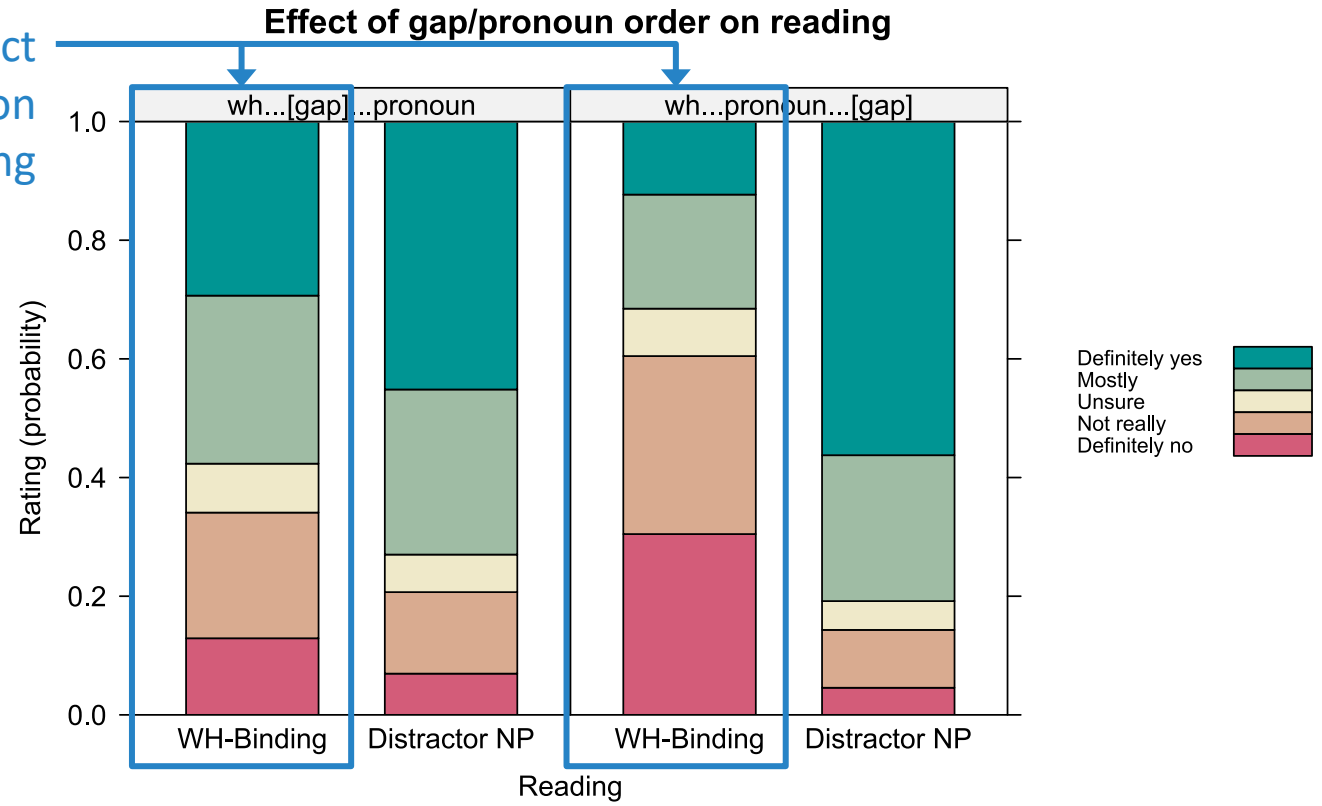
- 144 native English speakers recruited on Prolific (of which 8 excluded)
- Latin square design: each participant saw 6 target items and 6 fillers

Gap / pronoun order		**C-command configuration**		**Reading**
wh … [gap] … pronoun *(Binding)*	×	Pronoun c-commands gap *(Strong)*	×	wh-binding reading
wh … pronoun … [gap] *(Crossover)*		Pronoun c-commands gap, wh in Spec *(Secondary Strong)*		Distractor NP reading
		Pronoun in Spec of XP, XP c-commands gap *(Weak)*		

Effect of crossover vs. binding

We find a **significant crossover effect** using an ordinal mixed effects model

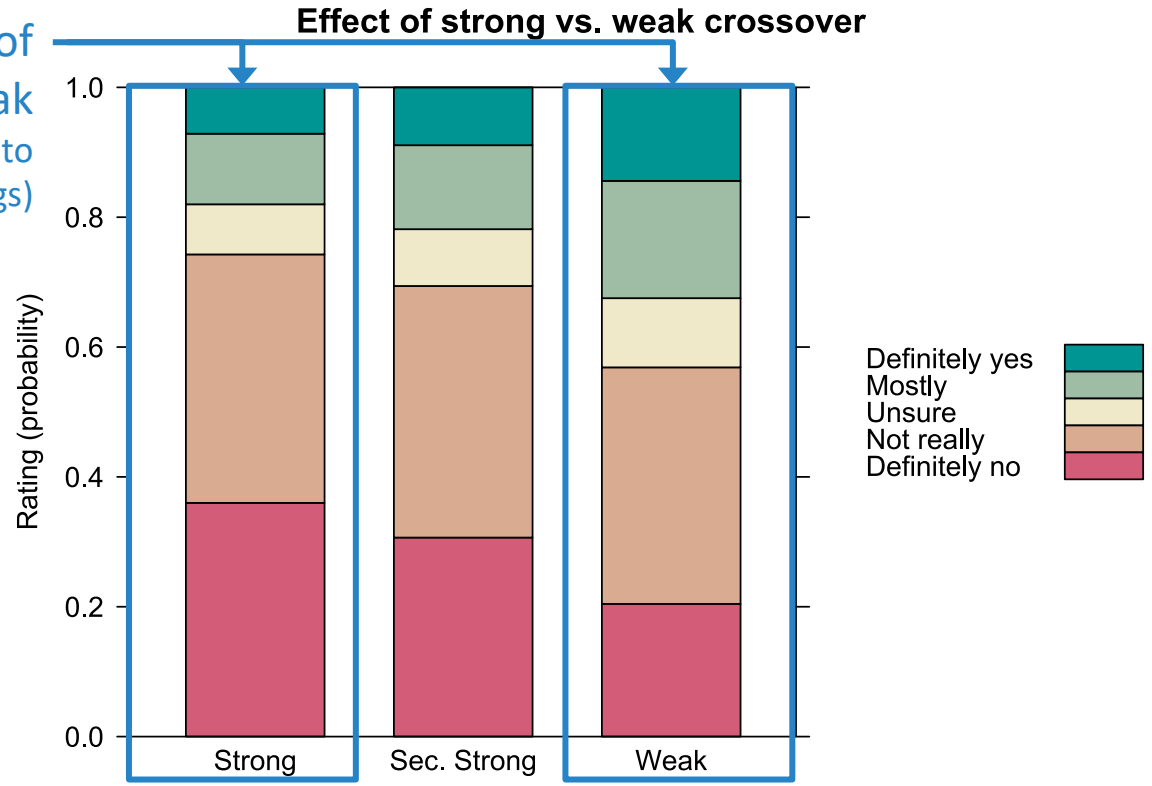
Significant negative effect of pronoun … [gap] on *wh*-binding reading



Effect of strong vs. weak crossover

We find a **significant difference between strong and weak crossover** using an ordinal mixed effects model (on just the bound reading of pronoun … [gap])

Significant effect of strong vs. weak
(weak twice as likely to have high ratings)

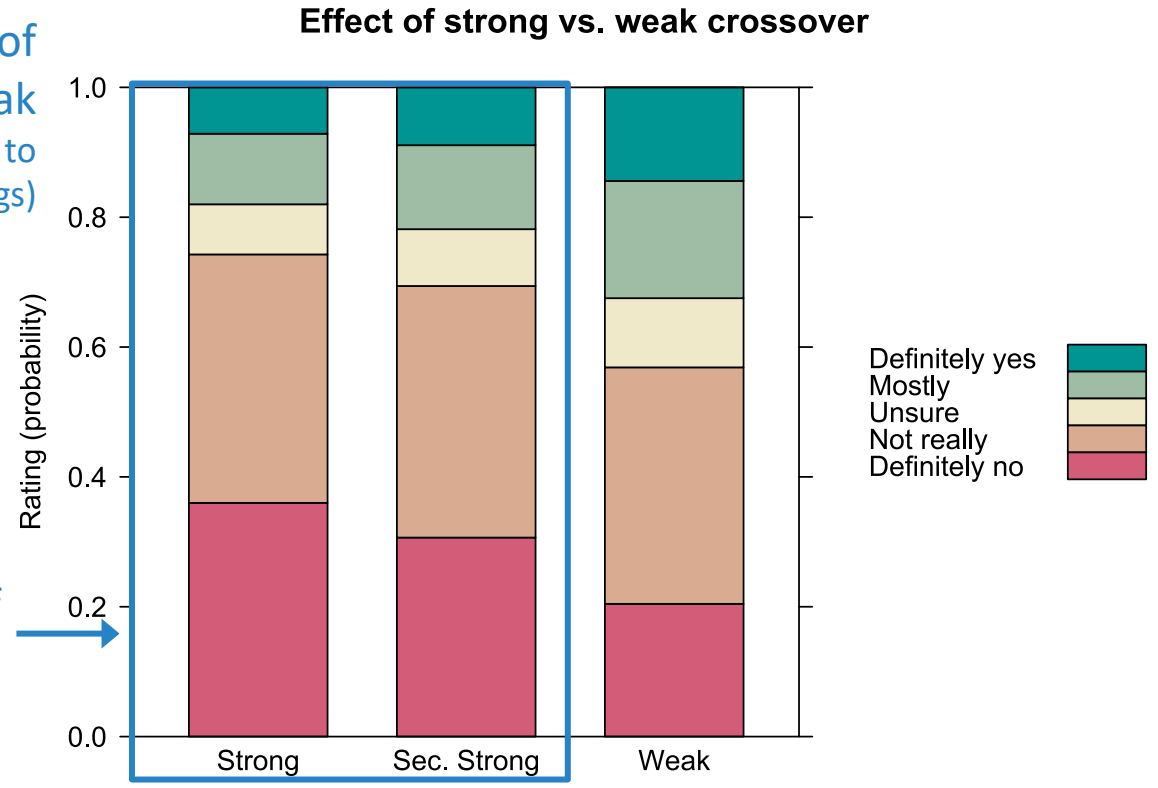


Effect of strong vs. weak crossover

We find a **significant difference between strong and weak crossover** using an ordinal mixed effects model (on just the bound reading of pronoun … [gap])

Significant effect of strong vs. weak (weak twice as likely to have high ratings)

No significant effect of strong vs. secondary strong



Impact on the theory

Our results **favour accounts which distinguish strong and weak crossover**

Account	**Distinguishes strong/weak?**
Koopman & Sportiche (1982), Safir (1984), Ruys (2000) i.a.	✓
Reinhart (1983), Safir (2004) i.a.	✗

Experiment 2: proper names

Anaphora vs. cataphora for proper names

"Strong": first mention c-commands second

c-commands

(8) The chef knew that **Daniel**$_i$ was disappointed by the soup **he**$_i$ made.

(9) The chef knew that **he**$_i$ was disappointed by the soup **Daniel**$_i$ made. ← Unacceptable

"Weak": first mention in Spec of XP, XP c-commands second

(10) The chef knew that [$_{XP}$ **Daniel**$_i$**'s** soup] had disappointed **him**$_i$.

(11) The chef knew that [$_{XP}$ **his**$_i$ soup] had disappointed **Daniel**$_i$. ← Acceptability is disputed!
(Lasnik & Stowell, 1991)

Experiment setup

We use a **2x2x2 factor design** with a 5-point Likert scale

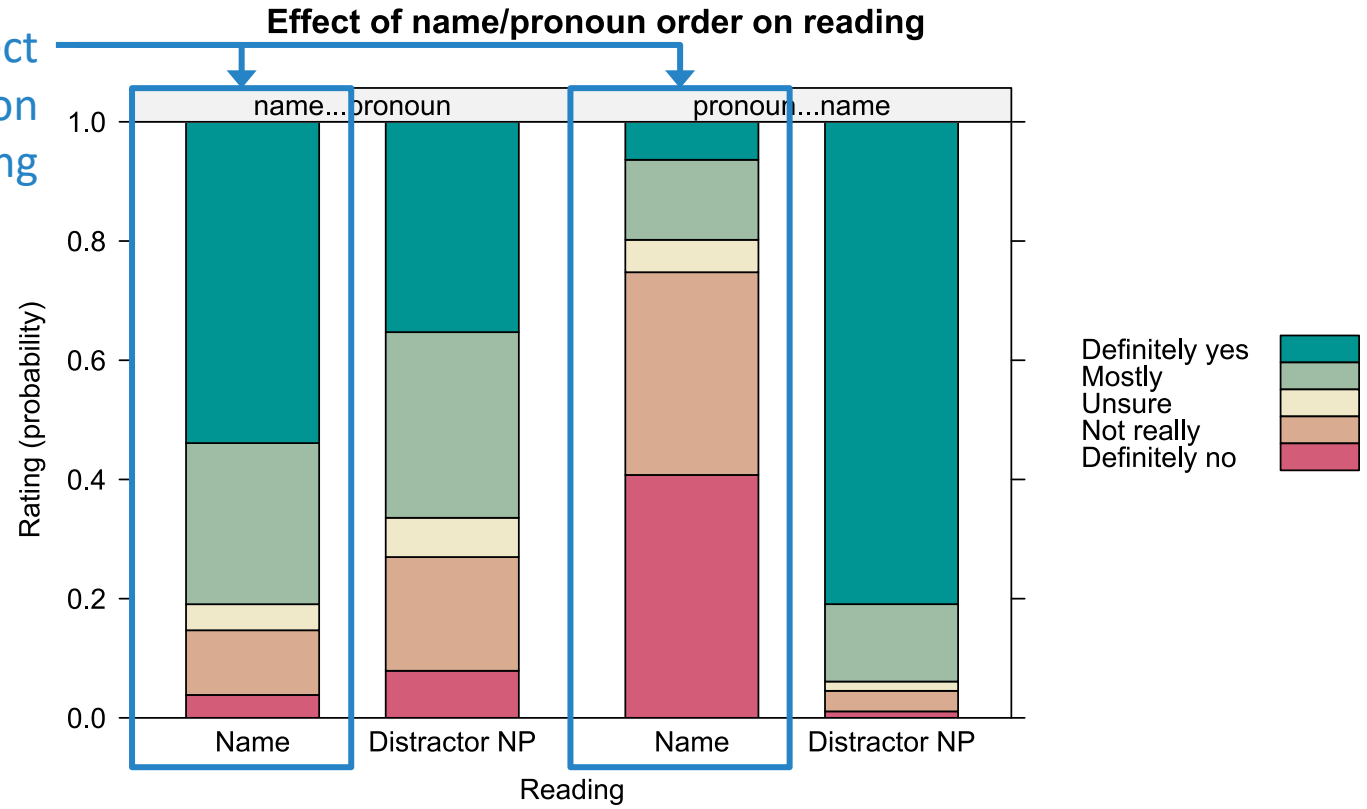
- 48 native English speakers recruited on Prolific (of which 1 excluded)
- Latin square design: each participant saw 6 target items and 6 fillers

Name / pronoun order		**C-command configuration**		**Reading**
name … pronoun	×	First mention c-commands second *(“Strong”)*	×	Name reading
pronoun … name		First mention in Spec of XP, XP c-commands second *(“Weak”)*		Distractor NP reading

Effect of cataphora

We find a **significant cataphora effect** similar to crossover using an ordinal mixed effects model

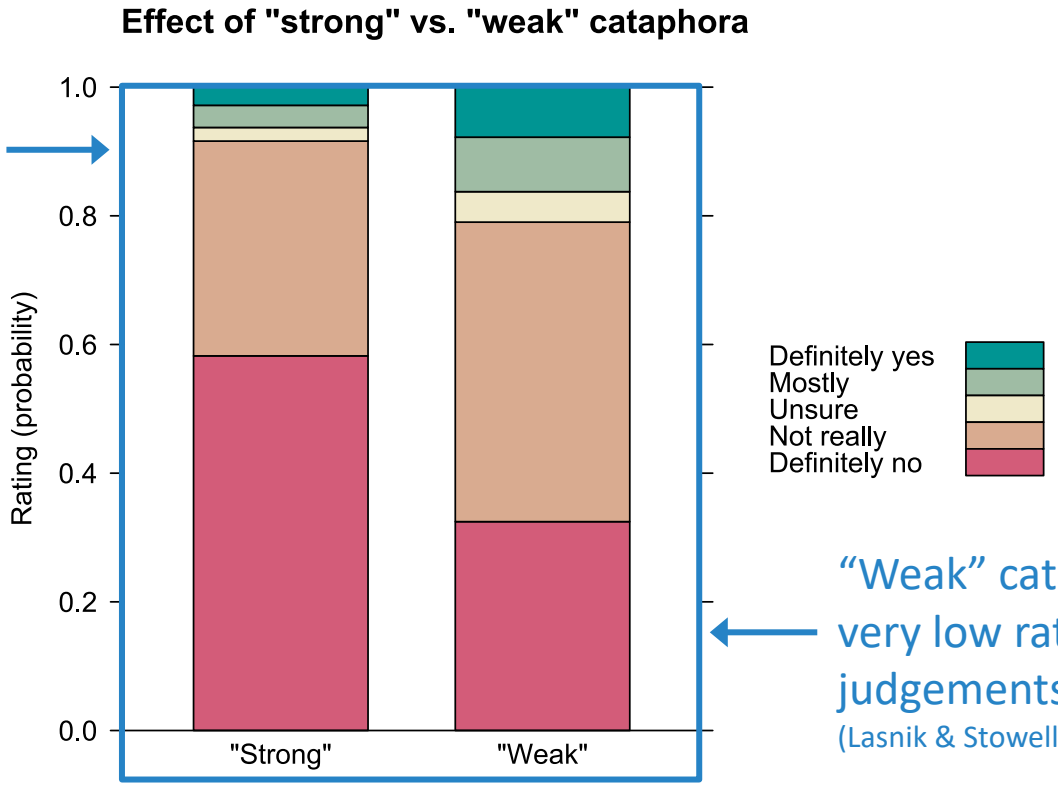
Significant negative effect of pronoun … name on Name reading



Effect of “strong” vs. “weak” cataphora

Proper names also show a **strong / weak crossover-style effect** with an ordinal mixed effects model (on just the Name reading of pronoun … name)

Significant effect of strong vs. weak
(weak three times as likely to have high ratings)



“Weak” cataphora still receive very low ratings overall – contra judgements in the literature
(Lasnik & Stowell 1991, Heim 2007, i.a.)

Impact on the theory (I)

Our “strong” results **support Rule I** and similar principles (Grodzinsky & Reinhart, 1993)

Rule I: “You can’t have coreference if you could have got that meaning with binding”

c-commands → could bind

(9) * The chef knew that **he**$_i$ was disappointed by the soup **Daniel**$_i$ made.

Impact on the theory (II)

But **Rule I doesn't account for "weak" cataphora** – should we extend it to indirect binding?

> **Rule I:** "You can't have coreference if you could have got that meaning with binding"

no c-command → no binding

(11) * The chef knew that [$_{XP}$ **his$_i$** soup] had disappointed **Daniel$_i$**.

(12) The article claimed that [$_{XP}$ every **city$_i$**'s mayor] enjoyed governing **it$_i$**. (adapted from Büring, 2004)

covaries ("indirectly binds"), despite lack of c-command

Prediction: cataphora are ruled out if and only if some kind of binding is possible – is that true?

This still doesn't really explain *why* "weak" cataphora are less bad!

Summary

We develop a **new experimental method to quantify crossover** effects

1. Significant effect of *wh*-crossover and strong/weak distinction in English
2. Similar pattern for proper name cataphora in English – open theoretical question!

Next steps:

- Compare quantificational crossover to *wh*-crossover and proper names
- Other languages and configurations, e.g.
  - weak crossover in Chinese (Lyu, 2017)
  - relative clauses in French (Postal, 1993)

References (I)

Balhorn, Mark. 2004. The Rise of Epicene They. *Journal of English linguistics* 32(2). Thousand Oaks, CA: SAGE Publications, 79–104.

Bjorkman, Bronwyn M. 2017. Singular they and the syntactic representation of gender in English. *Glossa: a journal of general linguistics* 2(1). Open Library of Humanities.

Büring, Daniel. 2004. Crossover situations. Natural Language Semantics 12(1). Springer, 23–62.

Felser, Claudia & Janna-Deborah Drummer. 2017. Sensitivity to Crossover Constraints During Native and Non-native Pronoun Resolution. *Journal of Psycholinguistic Research* 46(3), 771–789.

Grodzinsky, Yosef & Tanya Reinhart. 1993. The Innateness of Binding and Coreference. *Linguistic Inquiry* 24(1). The MIT Press, 69–101.

Heim, Irene R. 2007. Forks in the Road to Rule I. (Ed.) M Abdurrahman, A Schardl & M Walkow. *NELS 38 : Proceedings of the Thirty-Eighth Annual meeting of the North East Linguistic Society*. University of Ottawa.

Koopman, H. & D. Sportiche. 1982. Variables and the Bijection Principle. *Linguistic review* 2(2). Walter de Gruyter, 139–160.

Kush, Dave W. 2013. Respecting relations: Memory access and antecedent retrieval in incremental sentence processing. University of Maryland, College Park PhD Thesis.

References (II)

Lasnik, Howard & Tim Stowell. 1991. Weakest crossover. *Linguistic inquiry* 22(4). JSTOR, 687–720.

Lyu, Jun. 2017. Weak Crossover in Chinese—now you see it, now you don't. *Proceedings of the 17th Texas Linguistic Society*, 54--64.

Marty, Paul, Emmanuel Chemla & Jon Sprouse. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: a journal of general linguistics* 5(1), 72.

Postal, Paul M. 1993. Remarks on weak crossover effects. *Linguistic Inquiry* 24(3). JSTOR, 539–556.

Reinhart, Tanya. 1983. *Anaphora and semantic interpretation* (Croom Helm Linguistics Series). London: Croom Helm.

Ruys, E. G. 2000. Weak Crossover as a Scope Phenomenon. *Linguistic inquiry* 31(3). One Rogers Street, Cambridge, MA 02142-1209, USA: The MIT Press, 513–539.

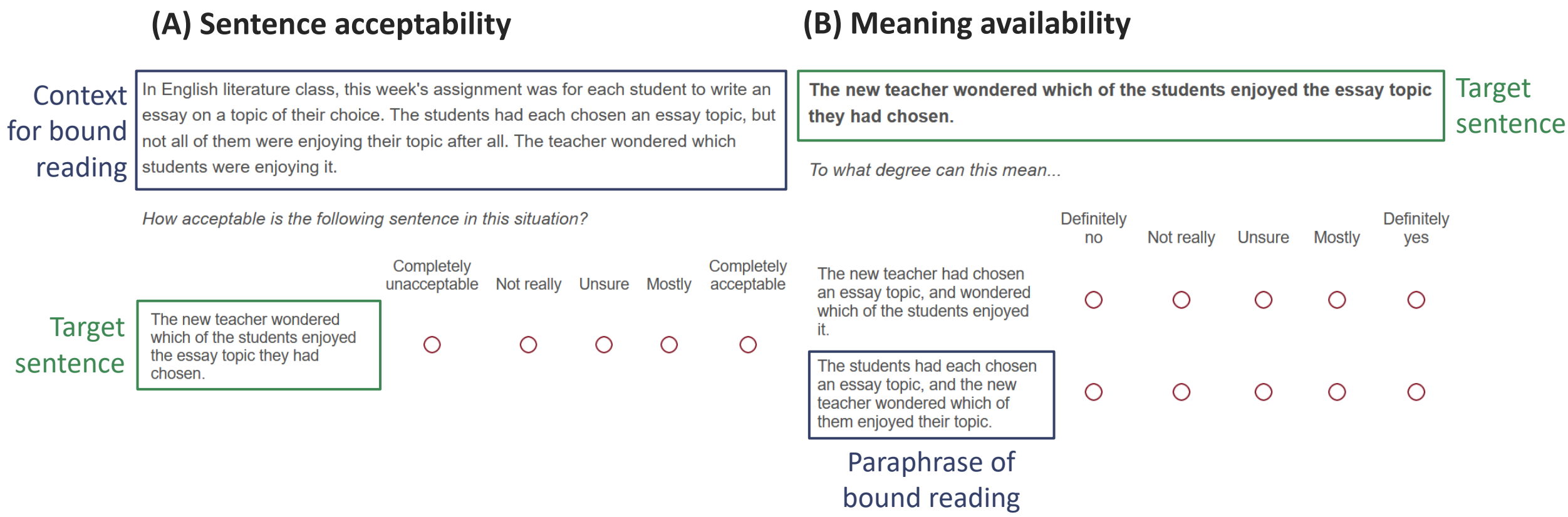
Safir, Kenneth. 1984. Multiple variable binding. *Linguistic Inquiry*. JSTOR, 603–638.

Safir, Kenneth. 2004. *The Syntax of (In)dependence*.

Appendix

Experiment design pilot

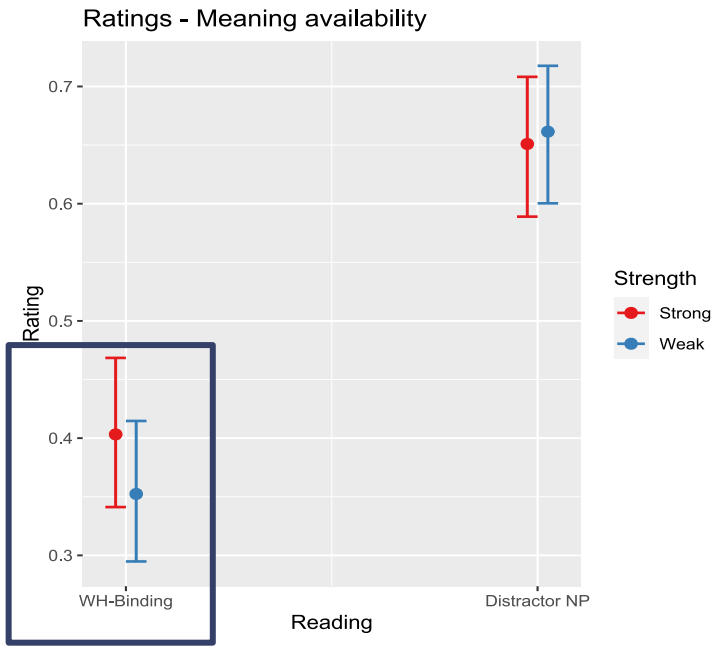
Rating the **meaning availability** produced crisper results on binding than **sentence acceptability**



Results from the design pilot

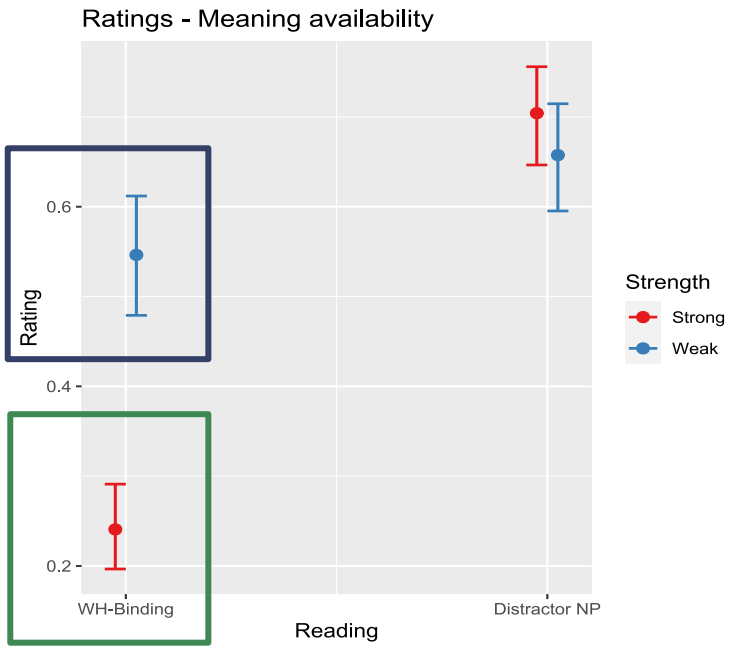
Rating **meaning availability** gave more intuitive results on binding than **sentence acceptability**
(shown here: mixed effects beta regression on just wh…[gap]…pronoun sentences)

Binding rated rather low, very different from Distractor NP



Binding rated higher, close to Distractor NP

This was a separate issue with the specific sentences used



A note on singular *they*

Singular ("epicene") *they* has been in use for antecedents of unknown or irrelevant gender since the 1400s (Balhorn, 2004; Bjorkman, 2017 i.a.)

(1) There's not a man I meet but doth salute me
As if I were **their** well-acquainted friend.

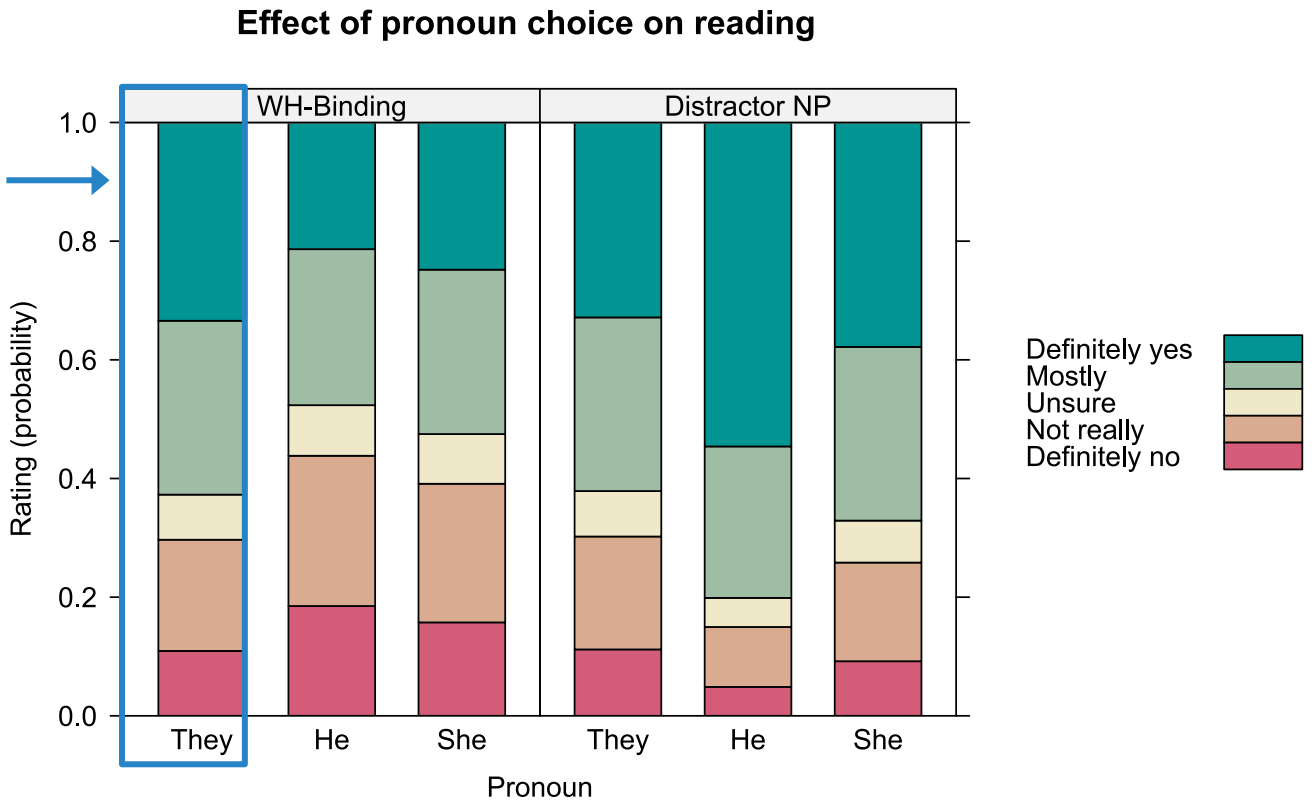
— Shakespeare, The Comedy of Errors (late 1500s)

(2) Somebody called while you were out and **they** said **they**'d call back later.

(3) Everyone left **their** lunch at home today.

Binding ratings by pronoun

Singular *they* is most likely to have bound reading
(difference between they/he is statistically significant)



Theory: Ruling in “weak” cataphora

Cataphora dispreference may be sufficient to account for the “badness” of weak cataphora

If so, Rule I is sufficient – or is it?

> **Rule I:** “You can’t have coreference if you could have got that meaning with binding”

(13) **The train$_i$** arrived and the passengers boarded **it$_i$**.

no c-command ⟶ no binding

(14) **It$_i$** arrived and the passengers boarded **the train$_i$**.

Prediction: (14) should be as acceptable as our “weak” cases – is that really true?

Model parameters

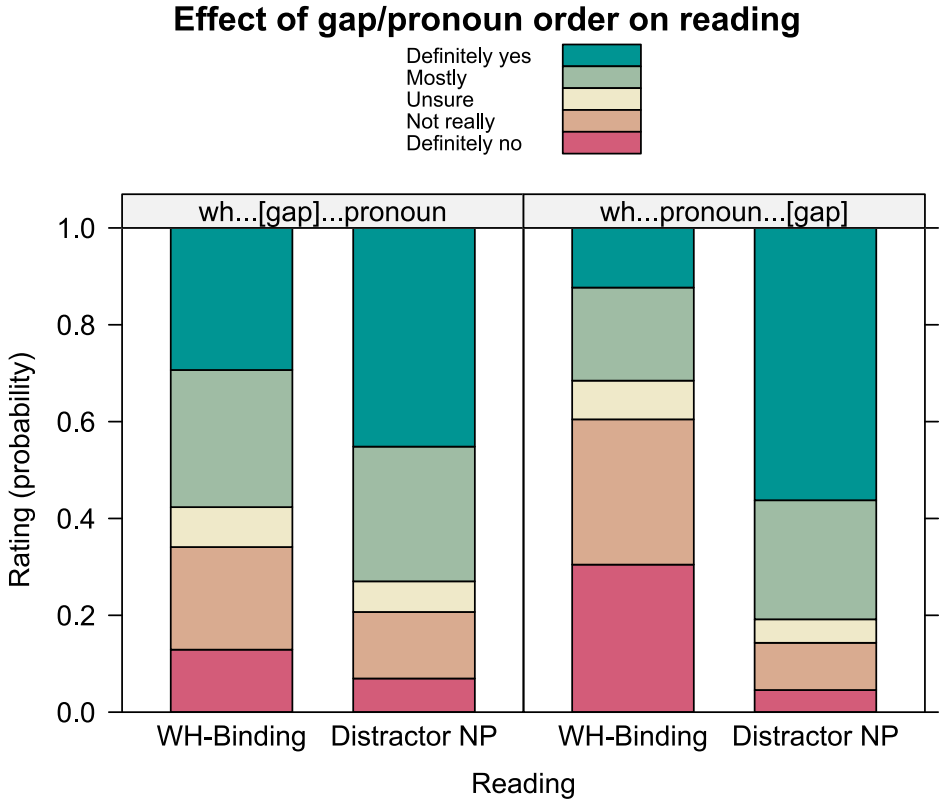
Parameter	Odds ratio	*p*-value
***wh*-crossover**		
Distractor NP (reading)	–	$p = 0.14$
wh … pronoun … [gap]	0.33	$p < 0.05$
wh … pronoun … [gap] * Distractor NP	4.61	$p < 0.05$
Strong vs. weak	2.19	$p < 0.05$
Strong vs. secondary strong	–	$p = 0.30$
Proper name cataphora		
Distractor NP (reading)	–	$p = 0.08$
pronoun … name	0.06	$p < 0.05$
pronoun … name * Distractor NP	133.76	$p < 0.05$
“Strong” vs. “weak”	2.90	$p < 0.05$

Model: crossover vs. binding

Fit an ordinal mixed effects model

```
Rating ~ Reading * Order
         + (1 + Reading | Scenario)
         + (1 + Reading | AmbiguityGroup)
         + (1 | ParticipantID)
```

`AmbiguityGroup`: tendency of participant to notice multiple available meanings in fillers (always / sometimes / never)

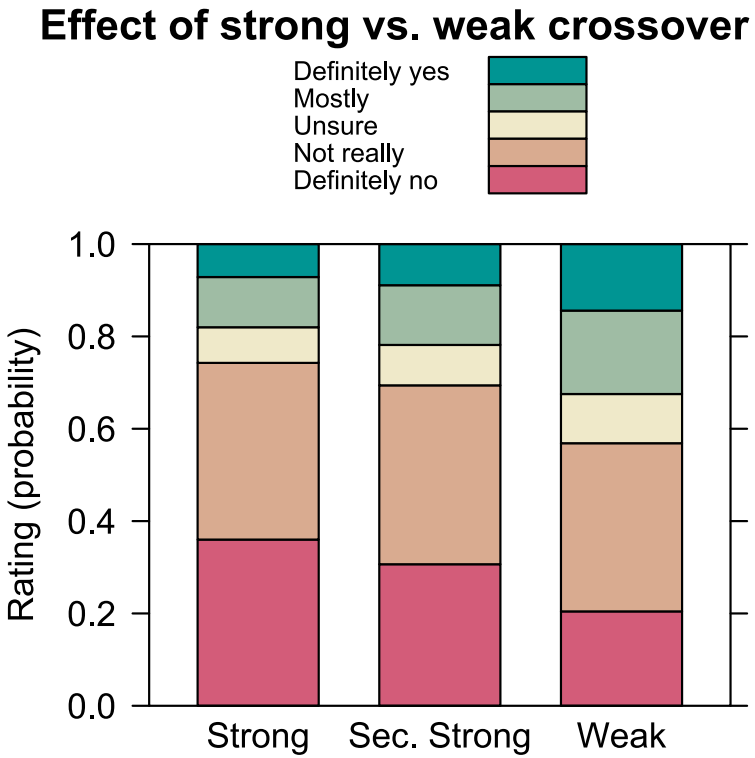


Model: weak vs. strong crossover

Fit an ordinal mixed effects model on just bound reading of pronoun … [gap]

```
Rating ~ Strength
         + (1 |Scenario)
         + (1 | AmbiguityGroup)
         + (1 | ParticipantID)
```

`AmbiguityGroup`: tendency of participant to notice multiple available meanings in fillers (always / sometimes / never)

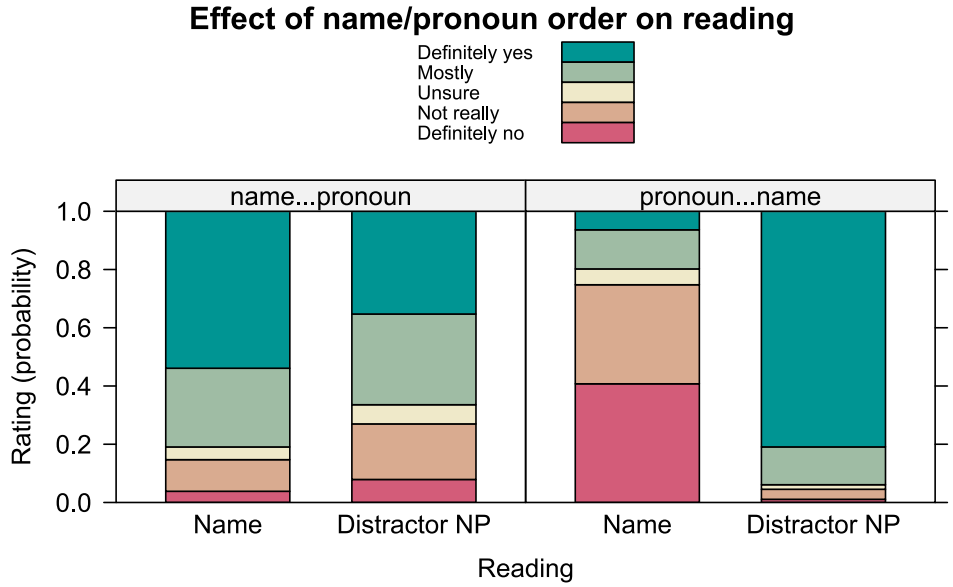


Model: name/pronoun order

Fit an ordinal mixed effects model

```
Rating ~ Reading * Order
         + (1 + Reading | Scenario)
         + (1 + Reading | AmbiguityGroup
         + (1 | ParticipantID)
```

`AmbiguityGroup`: tendency of participant to notice multiple available meanings in fillers (always / sometimes / never)



Model: “strong” vs. “weak” cataphora

Fit an ordinal mixed effects model on just bound reading of pronoun … name

```
Rating ~ Strength
         + (1 |Scenario)
         + (1 | AmbiguityGroup)
         + (1 | ParticipantID)
```

`AmbiguityGroup`: tendency of participant to notice multiple available meanings in fillers (always / sometimes / never)