

Weak Cross-over and Informativity

Thomas Wasow & David Clausen
Stanford University

Disclaimer

Disclaimer

- Mindful of Pullum's admonition not to start a talk with an apology, I'm going to start with a disclaimer:



Disclaimer

- Mindful of Pullum's admonition not to start a talk with an apology, I'm going to start with a disclaimer:



This is work in progress. We have some promising results, but more studies are needed.

Disclaimer

- Mindful of Pullum's admonition not to start a talk with an apology, I'm going to start with a disclaimer:



This is work in progress. We have some promising results, but more studies are needed.

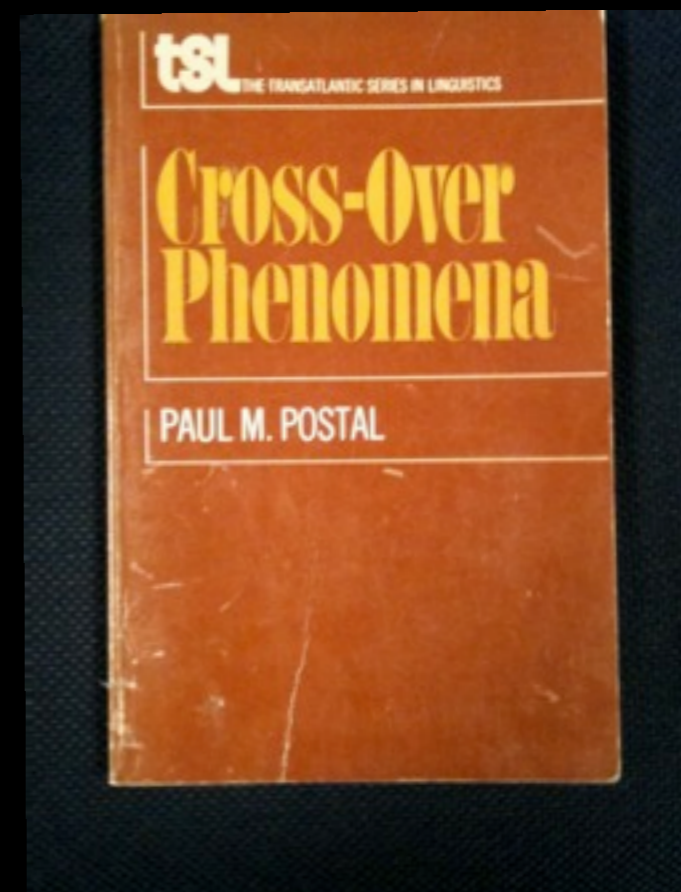
- And now that I'm past the start of the talk, I can apologize for the tentative nature of our findings.

Some Ancient History



In 1971, Postal published a book called *Cross-over Phenomena*,

arguing essentially that no transformation could reverse the relative positions of two coreferential NPs.



Examples from Postal (1971)

- *Charley was stabbed by himself.*
- *Himself was stabbed by Charley.*
- *I was difficult for me to shave.*
- *Myself was difficult for me to shave.*
- *I seem to myself to be clever.*
- *Myself seems to me to be clever.*
- *I talked to Thmug about himself.*
- *I talked to himself about Thmug.*

Examples from Postal (1971)

- *Charley was stabbed by himself.*
- *Himself was stabbed by Charley.*
- *I was difficult for me to shave.*
- *Myself was difficult for me to shave.*
- *I seem to myself to be clever.*
- *Myself seems to me to be clever.*
- *I talked to Thmug about himself.*
- *I talked to himself about Thmug.*

- These are Postal's judgments. Where they seem right, there are more straightforward explanations than his proposed constraint (e.g. eliminating the transformation in question).

Examples from Postal (1971)

- *Charley was stabbed by himself.*
- *Himself was stabbed by Charley.*
- *I was difficult for me to shave.*
- *Myself was difficult for me to shave.*
- *I seem to myself to be clever.*
- *Myself seems to me to be clever.*
- *I talked to Thmug about himself.*
- *I talked to himself about Thmug.*

- These are Postal's judgments. Where they seem right, there are more straightforward explanations than his proposed constraint (e.g. eliminating the transformation in question).
- There were also lots of potential counterexamples, e.g. *The professors; were criticized by their; students.*

More Interesting Cases

More Interesting Cases

- But the most interesting cases involved questions and relative clauses:
 1. **I know who_i Charley thinks he_i hurt.*
 2. **Who_i did you talk to the boy who she_i liked about?*
 3. **The one_i who Charley thinks he_i hurt [is okay].*
 4. **The one_i who you talked to the boy she_i liked about [is here].*

More Interesting Cases

- But the most interesting cases involved questions and relative clauses:
 1. **I know who_i Charley thinks he_i hurt.*
 2. **Who_i did you talk to the boy who she_i liked about?*
 3. **The one_i who Charley thinks he_i hurt [is okay].*
 4. **The one_i who you talked to the boy she_i liked about [is here].*
- No alternative accounts were readily forthcoming

Wasow (1972) questioned Postal's data

Wasow (1972) questioned Postal's data

- Compare (1) & (2):
 1. **I know who_i Charley thinks he_i hurt.*
 2. **Who_i did you talk to the boy who she_i liked about?*

Wasow (1972) questioned Postal's data

- Compare (1) & (2):
 1. **I know who_i Charley thinks he_i hurt.*
 2. **Who_i did you talk to the boy who she_i liked about?*
- (1) is so bad, it's hard to grasp the intended reading:
'I know for which person x, Charley thinks x hurt x'

Wasow (1972) questioned Postal's data

- Compare (1) & (2):
 1. **I know who_i Charley thinks he_i hurt.*
 2. **Who_i did you talk to the boy who she_i liked about?*
- (1) is so bad, it's hard to grasp the intended reading:
'I know for which person x, Charley thinks x hurt x'
- But (2) seems interpretable, especially if *who* is replaced by a longer *wh*-phrase:
??Which eighth-grade girl_i did you talk to the boy who she_i liked about?

Wasow (1972) questioned Postal's data

- Compare (1) & (2):
 1. **I know who_i Charley thinks he_i hurt.*
 2. **Who_i did you talk to the boy who she_i liked about?*
- (1) is so bad, it's hard to grasp the intended reading: 'I know for which person x, Charley thinks x hurt x'
- But (2) seems interpretable, especially if *who* is replaced by a longer *wh*-phrase:

??Which eighth-grade girl_i did you talk to the boy who she_i liked about?
- Wasow (1972) labeled these 'strong' and 'weak' cross-over



Purportedly Acceptable WCO Examples

Purportedly Acceptable WCO Examples

- Invented examples from earlier work

?How many copies of Aspects; does your friend who collects them; own?

?Which well-known actor; did the policeman who arrested him; accuse of being drunk?

Purportedly Acceptable WCO Examples

- Invented examples from earlier work

?How many copies of Aspects; does your friend who collects them; own?

?Which well-known actor; did the policeman who arrested him; accuse of being drunk?

- Examples from usage

He was the type of man with whom; his; work would always come first. [Agatha Christie]

He was the kind of man who; when he; loses his; collar stud bellows the house down. [Agatha Christie]

On December 23rd, the postman brought a large envelope which; when I opened it; at breakfast shed a lot of silvery tinsel into my plate. [Graham Greene]

Old Idea to Handle This: Traces

Old Idea to Handle This: Traces

- Generalization suggested in Wasow (1972): Anaphoric relations involving fillers in filler-gap dependencies are about as good as they would be with the filler in the gap position.

Old Idea to Handle This: Traces

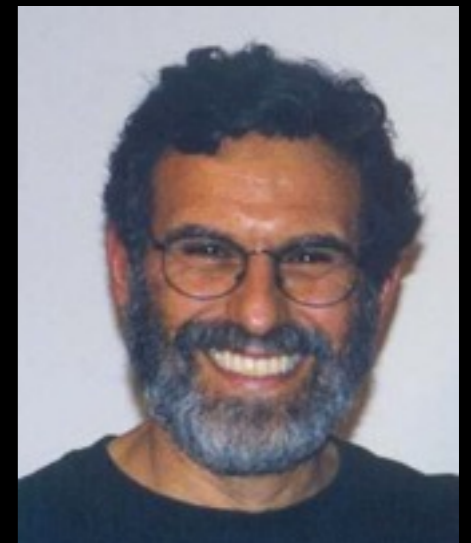
- Generalization suggested in Wasow (1972): Anaphoric relations involving fillers in filler-gap dependencies are about as good as they would be with the filler in the gap position.
- In a transformational theory, this could be handled by ordering rules for anaphora (binding principles) before fronting of fillers.

Old Idea to Handle This: Traces

- Generalization suggested in Wasow (1972): Anaphoric relations involving fillers in filler-gap dependencies are about as good as they would be with the filler in the gap position.
- In a transformational theory, this could be handled by ordering rules for anaphora (binding principles) before fronting of fillers.
- But this led to an ordering paradox:
 - *She_i married one of the men Sue_i had been dating.*
 - Which of the men Sue_i had been dating did she_i marry?*

Old Idea to Handle This: Traces

- Generalization suggested in Wasow (1972): Anaphoric relations involving fillers in filler-gap dependencies are about as good as they would be with the filler in the gap position.
- In a transformational theory, this could be handled by ordering rules for anaphora (binding principles) before fronting of fillers.
- But this led to an ordering paradox:
 - *She_i married one of the men Sue_i had been dating.*
 - Which of the men Sue_i had been dating did she_i marry?*
- Proposed solution (suggested by Peter Culicover, p.c.): Movement rules leave traces, coindexed to the moved element.



More on the Trace Analysis

More on the Trace Analysis

- Strong cross-over is an automatic consequence:
**Who_i does Mary think he_i hurt?* is out for the same reason as **Mary thinks he_i hurt John_i.* is.

More on the Trace Analysis

- Strong cross-over is an automatic consequence:
**Who_i does Mary think he_i hurt?* is out for the same reason as **Mary thinks he_i hurt John_i.* is.
- For WCO, the trace analysis predicts that the sentences should be as good or bad as the corresponding cases of cataphora.
- Various restrictions on cataphora had been proposed (notably by Postal).
- Wasow (1972) claimed these were mirrored in WCO.

Variability of Cataphora

Variability of Cataphora

- Judgments regarding cataphora vary, depending on structure, context, and who is making the judgment

Variability of Cataphora

- Judgments regarding cataphora vary, depending on structure, context, and who is making the judgment
- Things proposed as relevant have included
 - Definiteness
 - Genericness
 - Quantification
 - Certain words that seem to improve it, including *even* and *only*.

Examples of Variability of Cataphora

??*The man who lost it_i needs to find a key_i.*

The man who lost it_i needs to find the master key_i.

??*The fact that he_i lost amused somebody in the crowd_i.*

The fact that he_i lost amused John_i.

??*When they_i are angry, two gorillas_i can be awfully mean.*

When they_i are angry, gorillas_i can be awfully mean.

??*The fact that he_i is being sued should worry some businessman_i.*

The fact that he_i is being sued should worry any businessman_i.

??*The man who designed it_i can understand a computer_i.*

Only the man who designed it_i can understand a computer_i.

??*If you are looking for it_i, you'll never find a unicorn_i.*

Even if you are looking for it_i, you'll never find a unicorn_i.

“Determinateness”

“Determinateness”

- Wasow (1972) used the term “determinate” for those NPs that could serve as antecedents in cataphora, writing (in a footnote in the published version):

“Determinateness”

- Wasow (1972) used the term “determinate” for those NPs that could serve as antecedents in cataphora, writing (in a footnote in the published version):

NP’s admit of degrees of determinateness, depending on how much information is provided regarding the identity of possible referents for the antecedent. This is illustrated by (i), in which increased specificity of the antecedent improves right-to-left anaphora.

(i) *Although it made a loud noise, John ignored*

*something
?*some car
??a car
?a passing car
a certain car

I am grateful to Julius Moravcsik for pointing this fact out to me.



Issues

Issues

- Judgments of cataphora have not been systematically collected (to our knowledge).

Issues

- Judgments of cataphora have not been systematically collected (to our knowledge).
- Other things (e.g. intonation, context) can no doubt influence judgments:

*Although **he_i** doesn't know it yet, **someone_i** is in for a big surprise.*

*In a well-known novel, the architect who designed **it_i** blows up a building_i. For \$10,000 name the novel and its author.*

Issues

- Judgments of cataphora have not been systematically collected (to our knowledge).
- Other things (e.g. intonation, context) can no doubt influence judgments:

*Although **he_i** doesn't know it yet, **someone_i** is in for a big surprise.*

*In a well-known novel, the architect who designed **it_i** blows up a **building_i**. For \$10,000 name the novel and its author.*

- We don't have a good characterization of what factors are relevant even in simple cases.

Literature on WCO

Literature on WCO

- Lots written over the decades -- too much to review

Literature on WCO

- Lots written over the decades -- too much to review
- With the introduction of quantifier raising in transformational theories, contrasts like the following came to be called WCO:

The woman who raised him_i loves John_i.

??The woman who raised him_i loves every man_i.

Literature on WCO

- Lots written over the decades -- too much to review
- With the introduction of quantifier raising in transformational theories, contrasts like the following came to be called WCO:

The woman who raised him_i loves John_i.

??The woman who raised him_i loves every man_i.

- Lasnik & Stowell (1993) summarized conventional wisdom of that time:
“In a configuration where a pronoun P and a trace T are both bound by a quantifier Q, T must c-command P.”



Literature on WCO

- Lots written over the decades -- too much to review
- With the introduction of quantifier raising in transformational theories, contrasts like the following came to be called WCO:

The woman who raised him_i loves $John_i$.

??The woman who raised him_i loves every man_i .

- Lasnik & Stowell (1993) summarized conventional wisdom of that time:
“In a configuration where a pronoun P and a trace T are both bound by a quantifier Q, T must c-command P.”
- They refined this to limit it to “true quantifiers”, claiming this eliminated a number of counterexamples.



Literature on WCO, continued

Literature on WCO, continued

- Postal wrote a reply to Lasnik & Stowell, crediting them with observing that, “the distribution of the WCO effect is inherently linked not to construction types but to types of NPs extracted, regardless of construction”.

Literature on WCO, continued

- Postal wrote a reply to Lasnik & Stowell, crediting them with observing that, “the distribution of the WCO effect is inherently linked not to construction types but to types of NPs extracted, regardless of construction”.
- Postal ends his article saying, “The basic conclusion of these remarks is, I believe, that WCO effects are even more mysterious than they might have seemed previously.”

Literature on WCO, continued

- Postal wrote a reply to Lasnik & Stowell, crediting them with observing that, “the distribution of the WCO effect is inherently linked not to construction types but to types of NPs extracted, regardless of construction”.
- Postal ends his article saying, “The basic conclusion of these remarks is, I believe, that WCO effects are even more mysterious than they might have seemed previously.”
- Although the literature on WCO is large and varied, everyone seems to agree that WCO is a grammatical phenomenon.

An Alternative Approach

An Alternative Approach

- Building on work by Sag, Hofmeister, Clausen, and others, we are exploring the possibility that WCO is not a grammatical phenomenon at all.



An Alternative Approach

- Building on work by Sag, Hofmeister, Clausen, and others, we are exploring the possibility that WCO is not a grammatical phenomenon at all.
- The variable and graded nature of judgments about WCO examples, and the influence of context and the referential properties of the NPs involved are reminiscent of recent discoveries about island constraints.



An Alternative Approach

- Building on work by Sag, Hofmeister, Clausen, and others, we are exploring the possibility that WCO is not a grammatical phenomenon at all.
- The variable and graded nature of judgments about WCO examples, and the influence of context and the referential properties of the NPs involved are reminiscent of recent discoveries about island constraints.
- We have been investigating whether, like some island constraints, WCO effects can be explained in terms of processing demands.



Informativity and Islands

Informativity and Islands

- It has long been known that some island effects are weaker when the filler contains more information.

Informativity and Islands

- It has long been known that some island effects are weaker when the filler contains more information.
- For example,
 - **What did who read?*
 - strikes most people as pretty awful, but
 - ?*Which book did which student read?*
 - is reasonably natural.

Informativity and Islands

- It has long been known that some island effects are weaker when the filler contains more information.
- For example,
 - *What did who read?*
strikes most people as pretty awful, but
 - ?Which book did which student read?*
is reasonably natural.
- This sort of effect has been studied in several contexts, using judgment studies, reading times, and some other methods.

Informativity and Islands

- It has long been known that some island effects are weaker when the filler contains more information.
- For example,
 - *What did who read?*
strikes most people as pretty awful, but
 - ?Which book did which student read?*
is reasonably natural.
- This sort of effect has been studied in several contexts, using judgment studies, reading times, and some other methods.
- While other theories (e.g. Pesetsky's d-linking) have been proposed, the best explanation seems to be in terms of processing.

Hofmeister's Definition of Informativity

Hofmeister's Definition of Informativity

An expression x_1 is more informative than an expression x_2 if the semantic and syntactic information encoded by x_2 is a proper subset of the information encoded by x_1 .

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

- Increased processing difficulty can result in lower judgments of acceptability.

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

- Increased processing difficulty can result in lower judgments of acceptability.
- Filler-gap dependencies make high demands on memory.

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

- Increased processing difficulty can result in lower judgments of acceptability.
- Filler-gap dependencies make high demands on memory.
- When combined with other things that are hard to process, these demands can make filler-gap dependencies unacceptable.

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

- Increased processing difficulty can result in lower judgments of acceptability.
- Filler-gap dependencies make high demands on memory.
- When combined with other things that are hard to process, these demands can make filler-gap dependencies unacceptable.
- More informative fillers take more work to encode and remain more activated.

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

- Increased processing difficulty can result in lower judgments of acceptability.
- Filler-gap dependencies make high demands on memory.
- When combined with other things that are hard to process, these demands can make filler-gap dependencies unacceptable.
- More informative fillers take more work to encode and remain more activated.
- More activated material is easier to retrieve.

Crude Summary of the Hofmeister, Sag, et al Account of Informativity and Island Effects

- Increased processing difficulty can result in lower judgments of acceptability.
- Filler-gap dependencies make high demands on memory.
- When combined with other things that are hard to process, these demands can make filler-gap dependencies unacceptable.
- More informative fillers take more work to encode and remain more activated.
- More activated material is easier to retrieve.
- So more informative fillers can mitigate island effects.

Crude Summary of How a Similar Account of WCO Effects Might Work

Crude Summary of How a Similar Account of WCO Effects Might Work

- Increased processing difficulty can result in lower judgments of acceptability.

Crude Summary of How a Similar Account of WCO Effects Might Work

- Increased processing difficulty can result in lower judgments of acceptability.
- Establishing pronoun-antecedent pairings makes demands on memory (see Almor's work).



Crude Summary of How a Similar Account of WCO Effects Might Work

- Increased processing difficulty can result in lower judgments of acceptability.
- Establishing pronoun-antecedent pairings makes demands on memory (see Almor's work).
- When combined with other things that are hard to process, these demands can make pronoun-antecedent pairings unacceptable.



Crude Summary of How a Similar Account of WCO Effects Might Work

- Increased processing difficulty can result in lower judgments of acceptability.
- Establishing pronoun-antecedent pairings makes demands on memory (see Almor's work).
- When combined with other things that are hard to process, these demands can make pronoun-antecedent pairings unacceptable.
- More informative fillers take more work to encode and remain more activated.



Crude Summary of How a Similar Account of WCO Effects Might Work

- Increased processing difficulty can result in lower judgments of acceptability.
- Establishing pronoun-antecedent pairings makes demands on memory (see Almor's work).
- When combined with other things that are hard to process, these demands can make pronoun-antecedent pairings unacceptable.
- More informative fillers take more work to encode and remain more activated.
- More activated material is easier to retrieve.



Crude Summary of How a Similar Account of WCO Effects Might Work

- Increased processing difficulty can result in lower judgments of acceptability.
- Establishing pronoun-antecedent pairings makes demands on memory (see Almor's work).
- When combined with other things that are hard to process, these demands can make pronoun-antecedent pairings unacceptable.
- More informative fillers take more work to encode and remain more activated.
- More activated material is easier to retrieve.
- So more informative antecedents could mitigate WCO effects.



Key Assumption About Binding

Key Assumption About Binding

- Grammar includes binding principles, roughly along the lines of A, B, & C from GB, except:
 - The relevant command relation is defined in terms of obliqueness or argument structure, not tree configurations

Key Assumption About Binding

- Grammar includes binding principles, roughly along the lines of A, B, & C from GB, except:
 - The relevant command relation is defined in terms of obliqueness or argument structure, not tree configurations
- This assumption is common to various proposals about binding, including those of Pollard & Sag and Bresnan (who attributes the idea to Mohanan).



Consequences of Basing Binding on Obliqueness, not Configurations

Consequences of Basing Binding on Obliqueness, not Configurations

- Strong cross-over examples are just cases of Principle C violations:
 - **Who(m)_i does he_i think we criticized?* is out for the same reason as **He_i thinks we criticized John_i* is.

Consequences of Basing Binding on Obliqueness, not Configurations

- Strong cross-over examples are just cases of Principle C violations:
 - **Who(m)_i does he_i think we criticized?* is out for the same reason as **He_i thinks we criticized John_i* is.
- WCO violations are **not** ruled out by the binding principles:
 - ??*Who(m)_i do people who know him_i criticize?* is allowed for the same reason as *People who know him_i criticize John_i* is.

Other Assumptions

Other Assumptions

- Finding an antecedent given a pronoun is harder than finding a pronoun given an antecedent (i.e., cataphora should be harder than anaphora).

Other Assumptions

- Finding an antecedent given a pronoun is harder than finding a pronoun given an antecedent (i.e., cataphora should be harder than anaphora).
- The easier it is to identify the referent of the antecedent, the easier it is to establish a pronoun-antecedent pairing.

Other Assumptions

- Finding an antecedent given a pronoun is harder than finding a pronoun given an antecedent (i.e., cataphora should be harder than anaphora).
- The easier it is to identify the referent of the antecedent, the easier it is to establish a pronoun-antecedent pairing.
- In processing a sentence, determining whether a given filler can be the antecedent for a given pronoun cannot be determined prior to the filler's associated gap.
 - This is because binding principles are based on obliqueness.
 - Hence, the configuration Filler...Pronoun....Gap should behave like cataphora with respect to the processing of binding.

How to Get Reliable Data on WCO?

How to Get Reliable Data on WCO?

- WCO data in the literature are open to dispute.

How to Get Reliable Data on WCO?

- WCO data in the literature are open to dispute.
- Searching for WCO examples in usage data is not feasible, since it would require a large corpus that is both parsed and annotated for coreference.

How to Get Reliable Data on WCO?

- WCO data in the literature are open to dispute.
- Searching for WCO examples in usage data is not feasible, since it would require a large corpus that is both parsed and annotated for coreference.
- So we needed to run experiments.

How to Get Reliable Data on WCO?

- WCO data in the literature are open to dispute.
- Searching for WCO examples in usage data is not feasible, since it would require a large corpus that is both parsed and annotated for coreference.
- So we needed to run experiments.
- The simplest kind are judgment experiments.

How to Get Reliable Data on WCO?

- WCO data in the literature are open to dispute.
- Searching for WCO examples in usage data is not feasible, since it would require a large corpus that is both parsed and annotated for coreference.
- So we needed to run experiments.
- The simplest kind are judgment experiments.
- But even these are tricky for WCO, because we want judgments of the naturalness sentences **under a particular interpretation** (one where a given pronoun shares its reference with a certain other NP).

Our Experimental Method

Our Experimental Method

- Participants read three-sentence passages, in which the third sentence is of the kind we are testing.

Our Experimental Method

- Participants read three-sentence passages, in which the third sentence is of the kind we are testing.
- At the end, they are asked for two judgments:
 1. How natural is the third sentence (on a 7-point scale)?
 2. Which of two possible referents mentioned in the passage does the critical pronoun refer to?

Our Experimental Method

- Participants read three-sentence passages, in which the third sentence is of the kind we are testing.
- At the end, they are asked for two judgments:
 1. How natural is the third sentence (on a 7-point scale)?
 2. Which of two possible referents mentioned in the passage does the critical pronoun refer to?
- Studies run on Amazon Mechanical Turk, limited to users with IP addresses in the US (to try to get native English speakers).

Our Experimental Method

- Participants read three-sentence passages, in which the third sentence is of the kind we are testing.
- At the end, they are asked for two judgments:
 1. How natural is the third sentence (on a 7-point scale)?
 2. Which of two possible referents mentioned in the passage does the critical pronoun refer to?
- Studies run on Amazon Mechanical Turk, limited to users with IP addresses in the US (to try to get native English speakers).
- 12 experimental items intermingled with 20 fillers, and order randomized.

Our Experimental Method

- Participants read three-sentence passages, in which the third sentence is of the kind we are testing.
- At the end, they are asked for two judgments:
 1. How natural is the third sentence (on a 7-point scale)?
 2. Which of two possible referents mentioned in the passage does the critical pronoun refer to?
- Studies run on Amazon Mechanical Turk, limited to users with IP addresses in the US (to try to get native English speakers).
- 12 experimental items intermingled with 20 fillers, and order randomized.
- Some fillers were ungrammatical, to establish a baseline -- and also to help weed out unreliable participants.

Our Studies

Our Studies

- We ran three studies:
 1. WCO in embedded questions
 2. WCO in restrictive relative clauses
 3. Cataphora

Our Studies

- We ran three studies:
 1. WCO in embedded questions
 2. WCO in restrictive relative clauses
 3. Cataphora
- Each test sentence had four forms (of which any given participant saw only one):
 - Short (uninformative) antecedent with WCO or cataphora
 - Long (informative) antecedent with WCO or cataphora
 - Short antecedent without WCO or cataphora
 - Long antecedent without WCO or cataphora

A Sample Stimulus

A Sample Stimulus

The hiring committee was presented with only a subset of the applicants. Some had been eliminated because their files were incomplete, but the committee chair dropped one applicant because of negative comments from colleagues who knew that applicant well.

A Sample Stimulus

The hiring committee was presented with only a subset of the applicants. Some had been eliminated because their files were incomplete, but the committee chair dropped one applicant because of negative comments from colleagues who knew that applicant well.

THE REST OF THE COMMITTEE WAS NOT TOLD WHO PEOPLE THAT KNEW HIM WELL HAD CRITICIZED.

A Sample Stimulus

The hiring committee was presented with only a subset of the applicants. Some had been eliminated because their files were incomplete, but the committee chair dropped one applicant because of negative comments from colleagues who knew that applicant well.

THE REST OF THE COMMITTEE WAS NOT TOLD WHO PEOPLE THAT KNEW HIM WELL HAD CRITICIZED.

- Participants rated the last sentence on a scale of 1-7, and then chose between the applicant or committee chair as the referent of **HIM**.

A Sample Stimulus

The hiring committee was presented with only a subset of the applicants. Some had been eliminated because their files were incomplete, but the committee chair dropped one applicant because of negative comments from colleagues who knew that applicant well.

THE REST OF THE COMMITTEE WAS NOT TOLD WHO PEOPLE THAT KNEW HIM WELL HAD CRITICIZED.

- Participants rated the last sentence on a scale of 1-7, and then chose between the applicant or committee chair as the referent of **HIM**.
- The “informative” versions replace **WHO** with **WHICH JOB APPLICANT**

A Sample Stimulus

The hiring committee was presented with only a subset of the applicants. Some had been eliminated because their files were incomplete, but the committee chair dropped one applicant because of negative comments from colleagues who knew that applicant well.

THE REST OF THE COMMITTEE WAS NOT TOLD WHO PEOPLE THAT KNEW HIM WELL HAD CRITICIZED.

- Participants rated the last sentence on a scale of 1-7, and then chose between the applicant or committee chair as the referent of **HIM**.
- The “informative” versions replace **WHO** with **WHICH JOB APPLICANT**
- The non-WCO versions replace **PEOPLE THAT KNEW HIM WELL HAD CRITICIZED** with **ELICITED NEGATIVE COMMENTS FROM PEOPLE WHO KNEW HIM WELL.**

Another Sample Stimulus

Another Sample Stimulus

That was from the embedded question WCO study. Each stimulus was modified for the relative clause WCO study and for the cataphora study. Here is one from the RC study.

In the bottom of the fourth inning, a questionable call elicited jeers from the visitors' dugout. One of the umpires evidently heard something sufficiently offensive to stick his head into the dugout and issue a warning.

THE RADIO ANNOUNCER COULDN'T SEE THE PERSON WHO THE UMPIRE HE HAD OFFENDED HAD WARNED.

Yet Another Sample Stimulus

Yet Another Sample Stimulus

And here is one from the cataphora study:

The news media are already beginning to cover the race for the 2012 Republican presidential nomination. At a recent convention of GOP bigwigs, many potential candidates were soliciting endorsements.

EVERYONE HE SOLICITED HAD PURPORTEDLY AGREED TO ENDORSE ONE PERSON.

Some Ungrammatical Stimuli

Some Ungrammatical Stimuli

Jane later expressed amazement at it that the people in the audience had not walked out on him.

The quarterback should have plays that would give advantage to his versatility.

One potential customer wouldn't buy any cookies because she was being dieting.

Nobody realized how hard he was working to make on everyone a good impression.

The therapist asked a colleague for a way how to get the patient to reveal more to him.

The husband believed the problem owing to the previous president's policies and said he couldn't do anything about that.

The priest was in a quandary about what he must not could say without violating the seal of confession.

Some Ungrammatical Stimuli

Jane later expressed amazement at it that the people in the audience had not walked out on him.

The quarterback should have plays that would give advantage to his versatility.

One potential customer wouldn't buy any cookies because she was being dieting.

Nobody realized how hard he was working to make on everyone a good impression.

The therapist asked a colleague for a way how to get the patient to reveal more to him.

The husband believed the problem owing to the previous president's policies and said he couldn't do anything about that.

The priest was in a quandary about what he must not could say without violating the seal of confession.

- In most cases the ungrammaticality is quite subtle, and the acceptability scores reflect this, with means in the 3 studies ranging from 2.72 to 2.96 on a scale of 1-7.

Our Expectations

Our Expectations

- In all three studies, we expected the controls (non-WCO or anaphora) to be judged better than the test sentences (WCO or Cataphora).

Our Expectations

- In all three studies, we expected the controls (non-WCO or anaphora) to be judged better than the test sentences (WCO or Cataphora).
- In all studies, we expected Informative (INF) antecedents to be judged better than the uninformative (BARE) antecedents.

Our Expectations

- In all three studies, we expected the controls (non-WCO or anaphora) to be judged better than the test sentences (WCO or Cataphora).
- In all studies, we expected Informative (INF) antecedents to be judged better than the uninformative (BARE) antecedents.
- We expected WCOs to be judged better than the ungrammatical fillers.

Our Expectations

- In all three studies, we expected the controls (non-WCO or anaphora) to be judged better than the test sentences (WCO or Cataphora).
- In all studies, we expected Informative (INF) antecedents to be judged better than the uninformative (BARE) antecedents.
- We expected WCOs to be judged better than the ungrammatical fillers.
- We expected the embedded question WCOs to be judged worse than the relative clause WCOs.

Our Expectations

- In all three studies, we expected the controls (non-WCO or anaphora) to be judged better than the test sentences (WCO or Cataphora).
- In all studies, we expected Informative (INF) antecedents to be judged better than the uninformative (BARE) antecedents.
- We expected WCOs to be judged better than the ungrammatical fillers.
- We expected the embedded question WCOs to be judged worse than the relative clause WCOs.
- We expected the Cataphora examples to be judged better than the WCOs.

Study Details

Study Details

I. WCO in embedded questions

106 subjects (19 excluded for high ungrammatical ratings)

1111 target ratings (excluding RT outliers, incorrect referents)

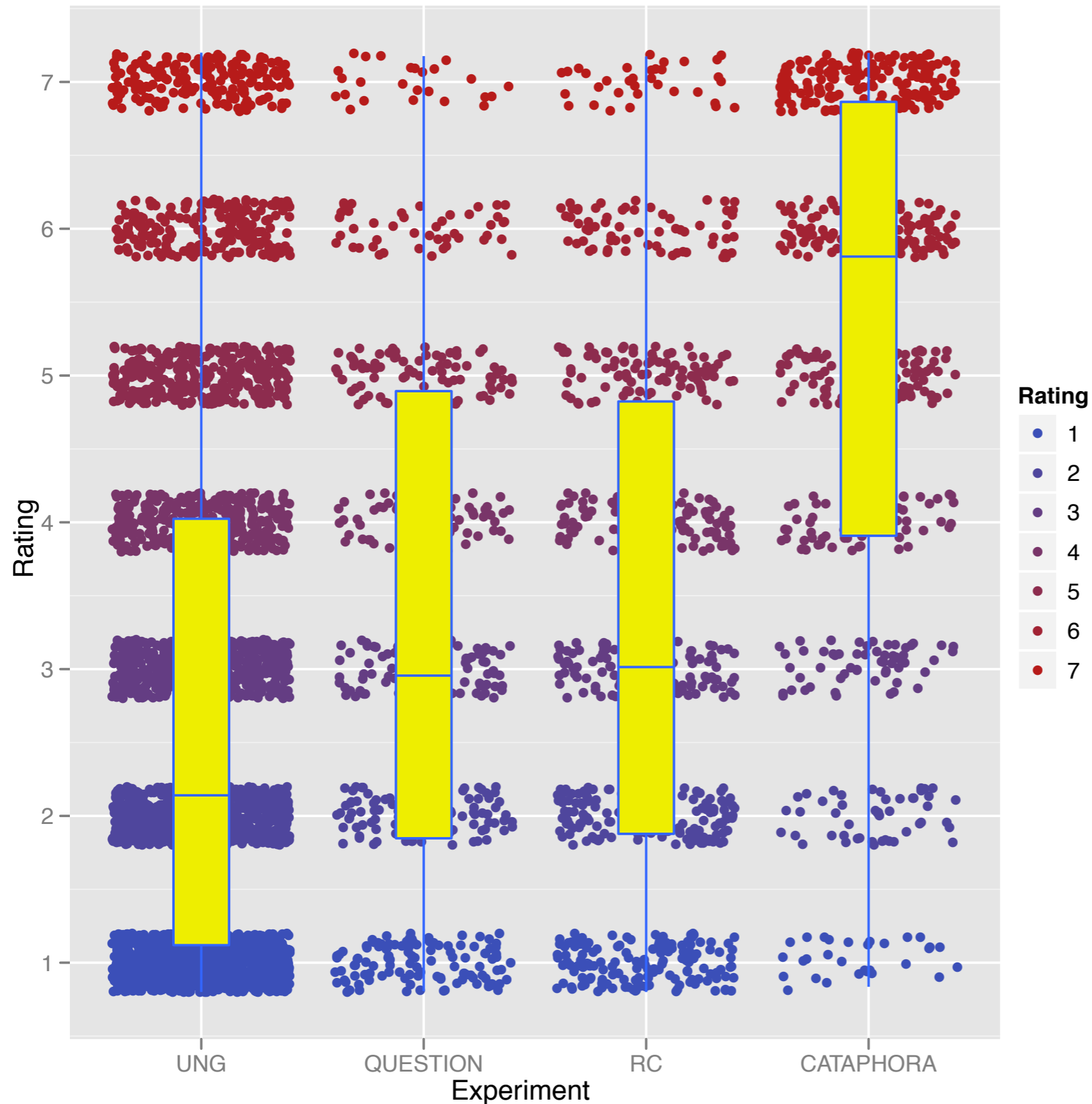
Study Details

1. WCO in embedded questions
106 subjects (19 excluded for high ungrammatical ratings)
1111 target ratings (excluding RT outliers, incorrect referents)
2. WCO in restrictive relative clauses
135 subjects (16 excluded)
1447 target ratings

Study Details

1. WCO in embedded questions
106 subjects (19 excluded for high ungrammatical ratings)
1111 target ratings (excluding RT outliers, incorrect referents)
2. WCO in restrictive relative clauses
135 subjects (16 excluded)
1447 target ratings
3. Cataphora
120 subjects (32 excluded)
1220 target ratings

Experiments

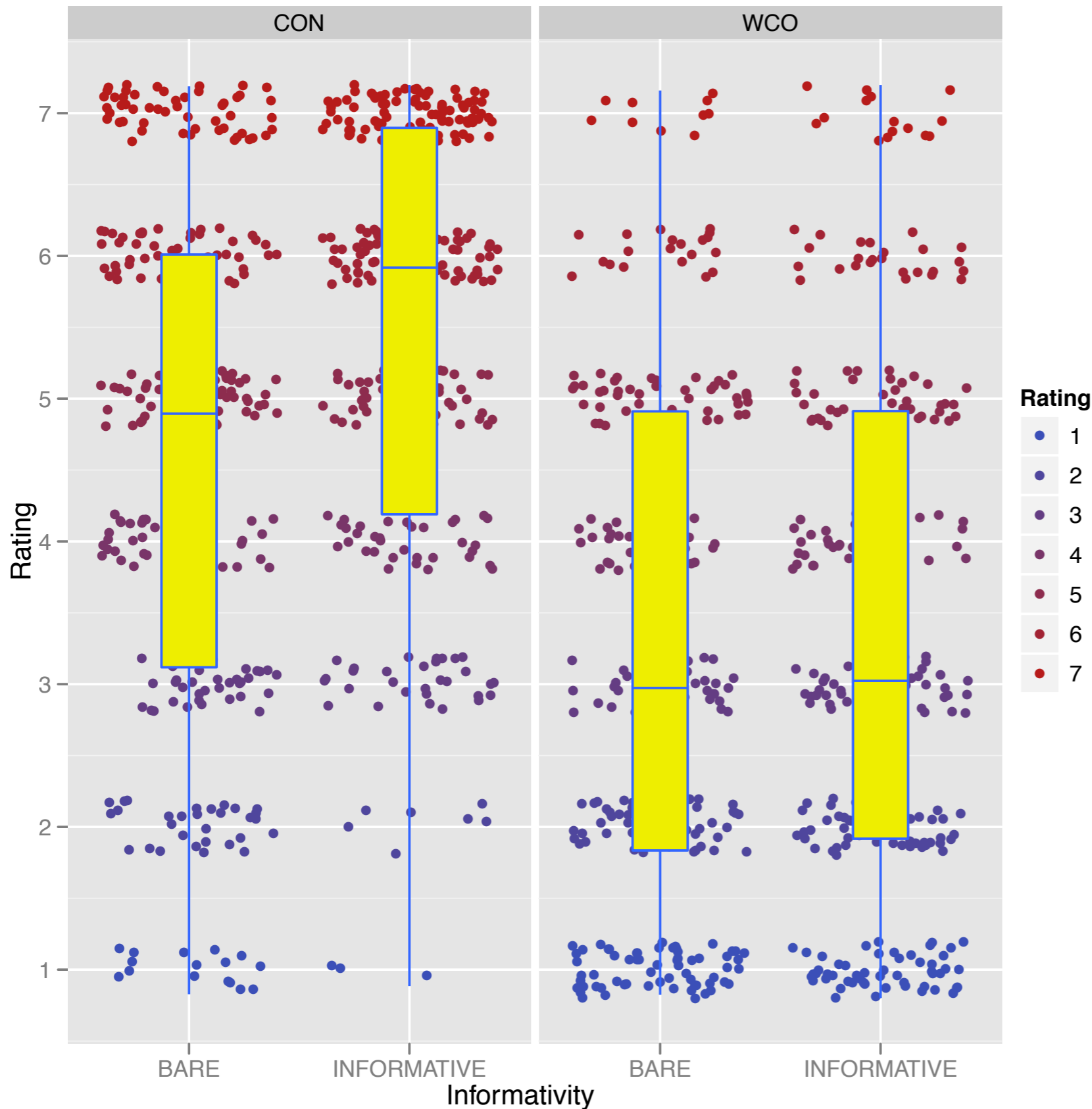


CAT > WCO
W = 179589.5
p < .001

WCO > UNG
W = 1816994
p < .001

QUES > RC
W = 215496
p < .001

Embedded Questions



CON > WCO

W = 234291

p < .001

CON: INF > BARE

W = 30480.5

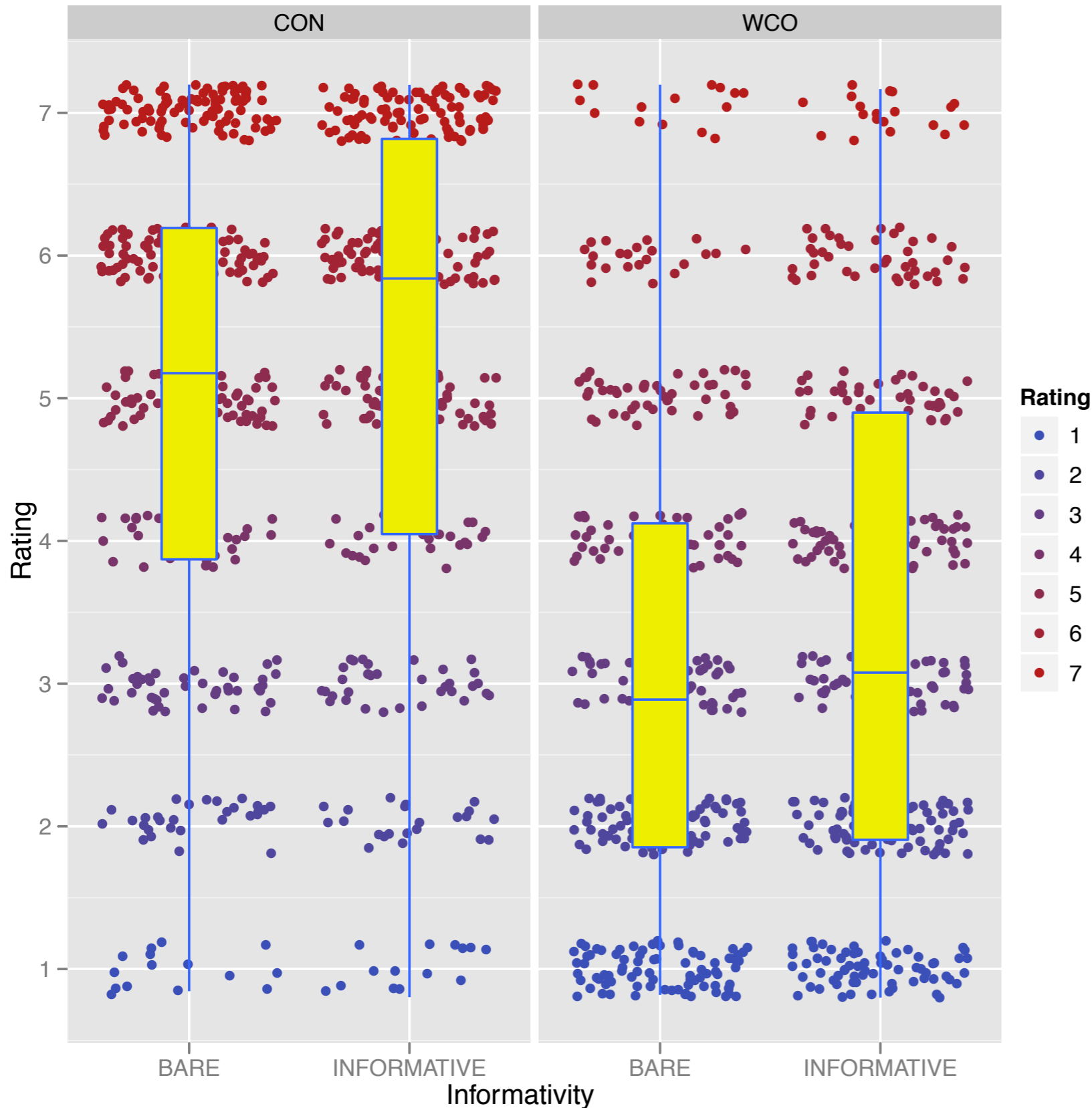
p < .001

WCO: INF = BARE

W = 33630

p = .254

Relative Clauses



CON > WCO

W = 338253.5

p < .001

CON: INF = BARE

W = 52263

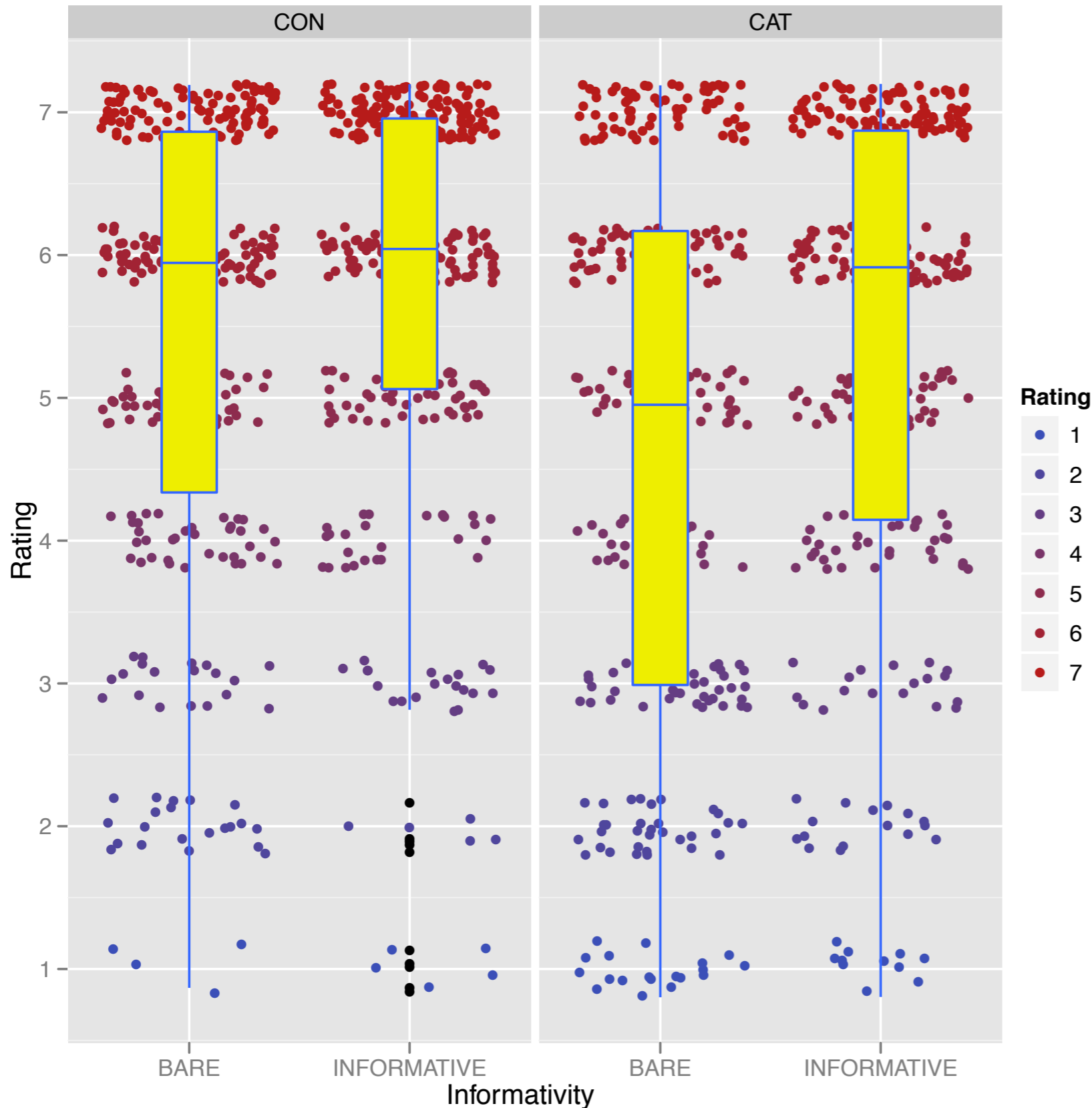
p = 0.315

WCO: INF > BARE

W = 47840

p < 0.05

Cataphora



CON > CAT

W = 215753.5

p < .001

CON: INF > BARE

W = 45337.5

p < .05

CAT: INF > BARE

W = 34889

p < .001

Summary of Findings

Summary of Findings

- As expected the controls (non-WCO or anaphora) were judged better than the test sentences (WCO or cataphora).

Summary of Findings

- As expected the controls (non-WCO or anaphora) were judged better than the test sentences (WCO or cataphora).
- As expected, informative antecedents were generally judged better than the uninformative antecedents, but not in embedded question WCO or in RC controls. This difference is a mystery to us.

Summary of Findings

- As expected the controls (non-WCO or anaphora) were judged better than the test sentences (WCO or cataphora).
- As expected, informative antecedents were generally judged better than the uninformative antecedents, but not in embedded question WCO or in RC controls. This difference is a mystery to us.
- Contrary to our expectations, the embedded question WCOs were judged slightly better than the relative clause WCOs.

Summary of Findings

- As expected the controls (non-WCO or anaphora) were judged better than the test sentences (WCO or cataphora).
- As expected, informative antecedents were generally judged better than the uninformative antecedents, but not in embedded question WCO or in RC controls. This difference is a mystery to us.
- Contrary to our expectations, the embedded question WCOs were judged slightly better than the relative clause WCOs.
- As expected the cataphora examples were judged better than the WCOs.

Summary of Findings

- As expected the controls (non-WCO or anaphora) were judged better than the test sentences (WCO or cataphora).
- As expected, informative antecedents were generally judged better than the uninformative antecedents, but not in embedded question WCO or in RC controls. This difference is a mystery to us.
- Contrary to our expectations, the embedded question WCOs were judged slightly better than the relative clause WCOs.
- As expected the cataphora examples were judged better than the WCOs.
- Even informative WCOs received low acceptability scores, but were rated higher than ungrammatical fillers.

Conclusions

Conclusions

- Processing factors play a role in the low acceptability of WCO. In particular:
 - The processing cost of filler-gap dependencies makes WCO hard.
 - The processing cost of establishing pronoun-antecedent pairings makes WCO hard.
 - WCO involves the same extra processing cost as cataphora.

Conclusions

- Processing factors play a role in the low acceptability of WCO. In particular:
 - The processing cost of filler-gap dependencies makes WCO hard.
 - The processing cost of establishing pronoun-antecedent pairings makes WCO hard.
 - WCO involves the same extra processing cost as cataphora.
- Informative antecedents mitigate the processing costs, at least sometimes.

Conclusions

- Processing factors play a role in the low acceptability of WCO. In particular:
 - The processing cost of filler-gap dependencies makes WCO hard.
 - The processing cost of establishing pronoun-antecedent pairings makes WCO hard.
 - WCO involves the same extra processing cost as cataphora.
- Informative antecedents mitigate the processing costs, at least sometimes.
- Whether the cumulative processing costs are sufficient to account for the low acceptability of WCO examples remains unclear.

Needed Additional Studies

Needed Additional Studies

- Determine the magnitude of each of the three sources of processing difficulties we identified, to see whether they can jointly account for the full unacceptability of WCO.

Needed Additional Studies

- Determine the magnitude of each of the three sources of processing difficulties we identified, to see whether they can jointly account for the full unacceptability of WCO.
- Try other mitigations (e.g., different types of contexts) to reduce the unacceptability of WCO.

Needed Additional Studies

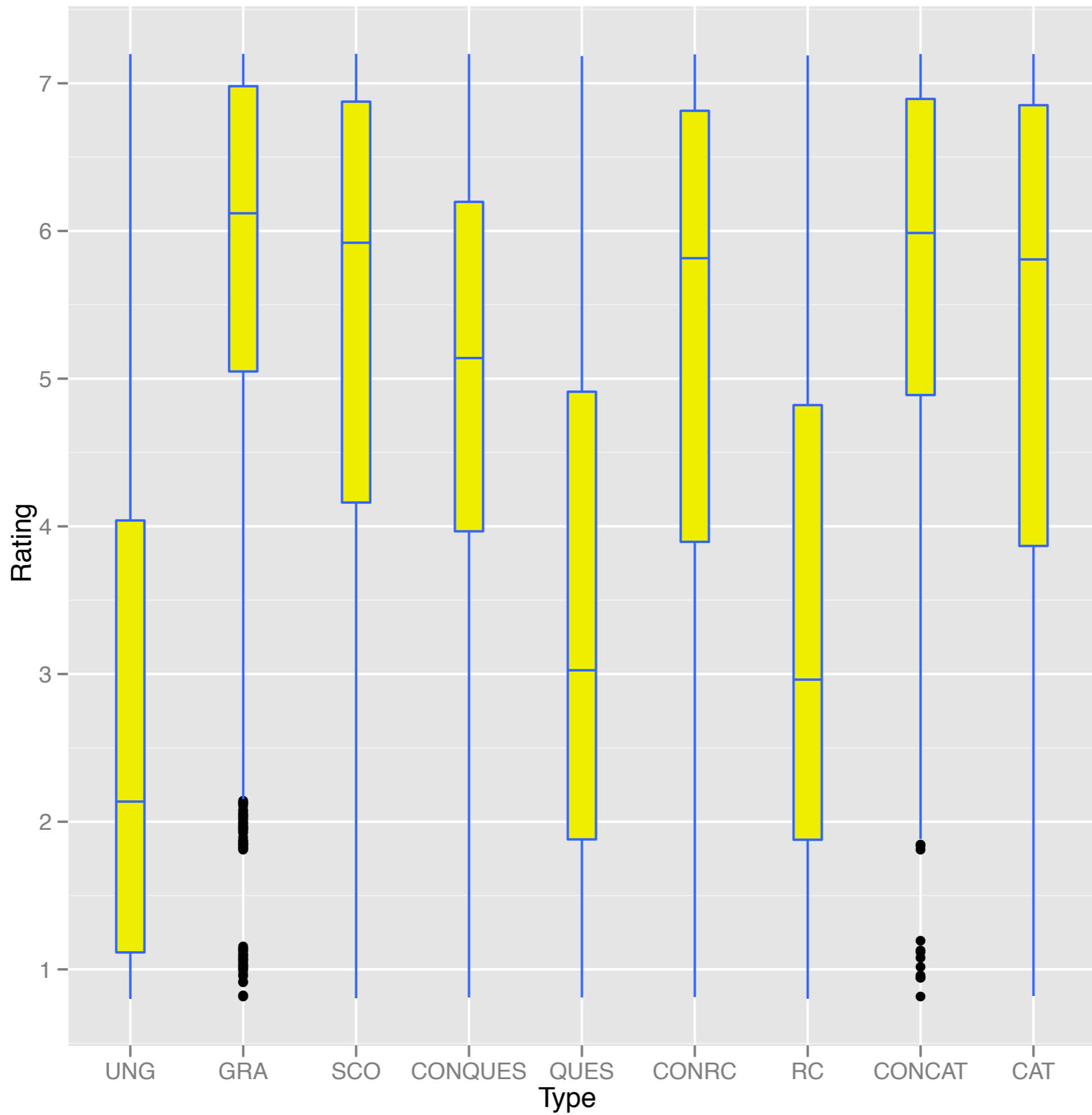
- Determine the magnitude of each of the three sources of processing difficulties we identified, to see whether they can jointly account for the full unacceptability of WCO.
- Try other mitigations (e.g., different types of contexts) to reduce the unacceptability of WCO.
- Perhaps test WCO using on-line methods (e.g., self-paced reading or eye-tracking) to get finer-grained evidence regarding the sources of unacceptability.

Thank you!

Thanks, too, to Mike Frank and
Ivan Sag for helpful discussions.

Choice of Referent

- Contexts were designed to favor one interpretation of the pronoun, and most participants selected that interpretation.
- The exception was with the strong cross-over cases, where the target interpretation is impossible.
- For the test and control sentences in the three experiments, the target interpretation was selected 90-95% of the time.
- For the SCO fillers in the three experiments, the target interpretation was selected only 24-29% of the time.



	coef.est	coef.se
(Intercept)	5.40	0.16
CAT	-0.74	0.18
CONRC	-0.36	0.15
RC	-2.31	0.20
CONQUESTION	-0.82	0.16
QUESTION	-2.26	0.21
INF	0.35	0.18
CAT:INF	0.33	0.25
CONRC:INF	-0.23	0.16
RC:INF	-0.09	0.25
CONQUESTION:INF	0.46	0.17
QUESTION:INF	-0.21	0.25

Error terms:

Groups	Name	Std.Dev.
MD5	(Intercept)	0.84
Item	(Intercept)	0.33
Group	(Intercept)	0.25
Residual		1.45