glowing lecture series

The role of experimental syntax in linguistic theorizing

Sandra Villata

University of Enna Kore

June 6th, 2025

sandra.villata@unikore.it

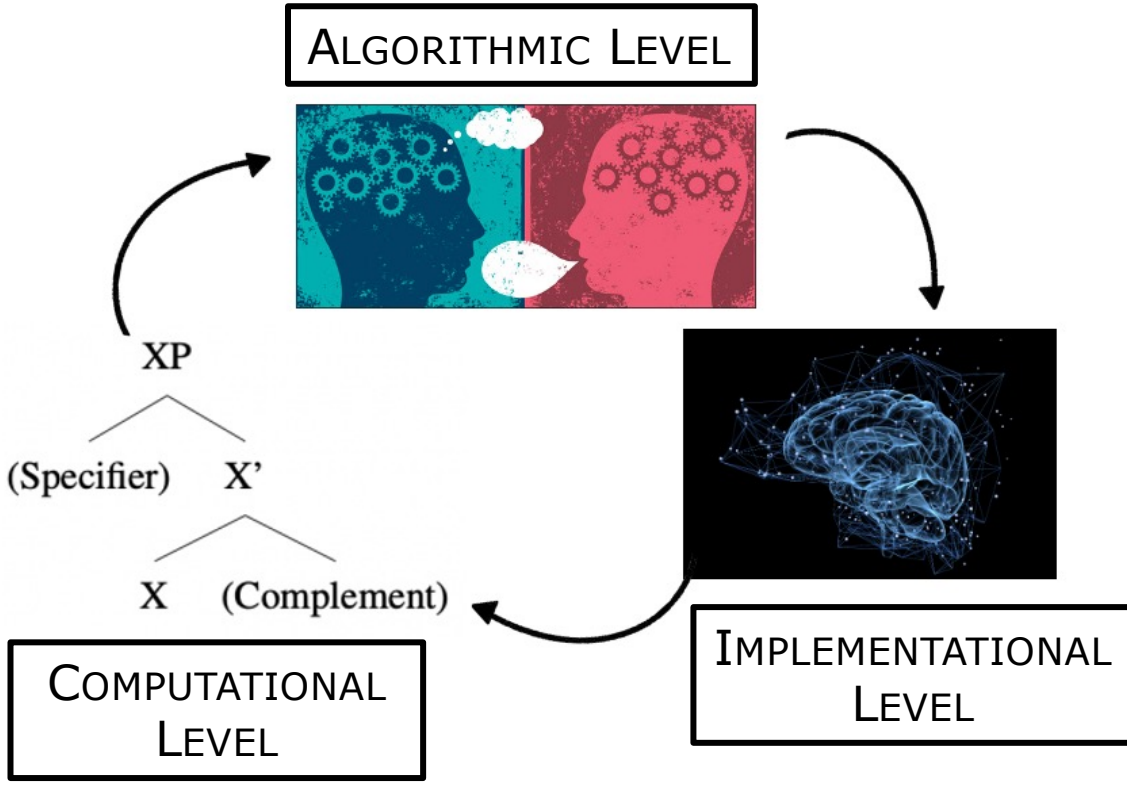
What is experimental linguistics?

Linguistics	The scientific study of language
Experimental linguistics	The scientific study of language through the use of formal experimental methods

Experimental linguistics uses formal, controlled methods to gain insights that go beyond what informal methods alone can reveal

An integrated theory of language

"Most of the phenomena that are central to us as human beings are primarily phenomena of information processing. […] The fundamental point is that in order to understand a device that performs an information-processing task, one needs many different kinds of explanations. […] At one extreme, the top level, is the abstract computational theory of the device […]. In the center is the choice of representation for the input and output and the algorithm to be used to transform one into the other. And at the other extreme are the details of how the algorithm and representation are realized physically". (Marr, 1982)



Experimental syntax operates at the computational level, but methodologically, it bridges syntax and experimentation, thereby reflecting the broader integration challenge on a more tractable scale

The structure of the talk

1. Why do we need experimental syntax?
    i. Definition of an effect
    ii. Isolating the source of an effect
    iii. Augmented precision

2. What are the methods of experimental syntax?
    i. Acceptability judgments
    ii. Self-paced reading / eye-tracking
    iii. EEG/MEG/fMRI

3. What is the future of experimental syntax?

Why do we need experimental syntax?

1. Defining effects

2. Isolating the source of effects

3. Augmented precision

1. Defining effects

Defining effects

- Linguists typically use minimal pairs to detect the **presence of an effect** by comparing a sentence that contains a violation with a grammatical one
- For instance, in each of the pairs below, the sentences in (1) are noticeably degraded compared to those in (2):

(1)*What$_i$ do you wonder whether Mary solved ___$_i$?

(2) What$_i$ do you think that Mary solved ___$_i$?

ISLAND EFFECT

(1)*Mary asked what$_i$ who read ___$_i$.

(2) Mary asked who read what.

SUPERIORITY EFFECT

(1)*Who$_i$ do you think that ___$_i$ saw Mary?

(2) Who$_i$ do you think ___$_i$ saw Mary?

THAT-TRACE EFFECT

Island effects

- **Island effects** arise when the tail of a long-distance dependency is inside certain encapsulated structural domains called *islands* (Ross, 1967)

1) *What$_i$ did you laugh [because Mary forgot _$_i$]? ADJUNCT ISLAND
2) *What$_i$ did you hear [the rumor that Susy discovered _$_i$]? NOUN COMPLEMENT ISLAND
3) *What$_i$ did you think [the speech by _$_i$] interrupted the game? SUBJECT ISLAND
4) *What$_i$ did you wonder [whether Lisa solved _$_i$]? WHETHER ISLAND
5) *What$_i$ did the waiter blame the chef [that overcooked _$_i$]? RELATIVE CLAUSE ISLAND

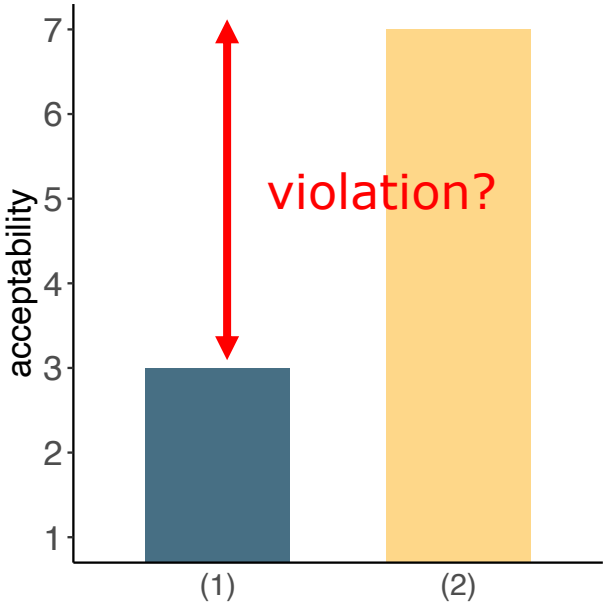
- Why are island effects heavily studied in experimental syntax?

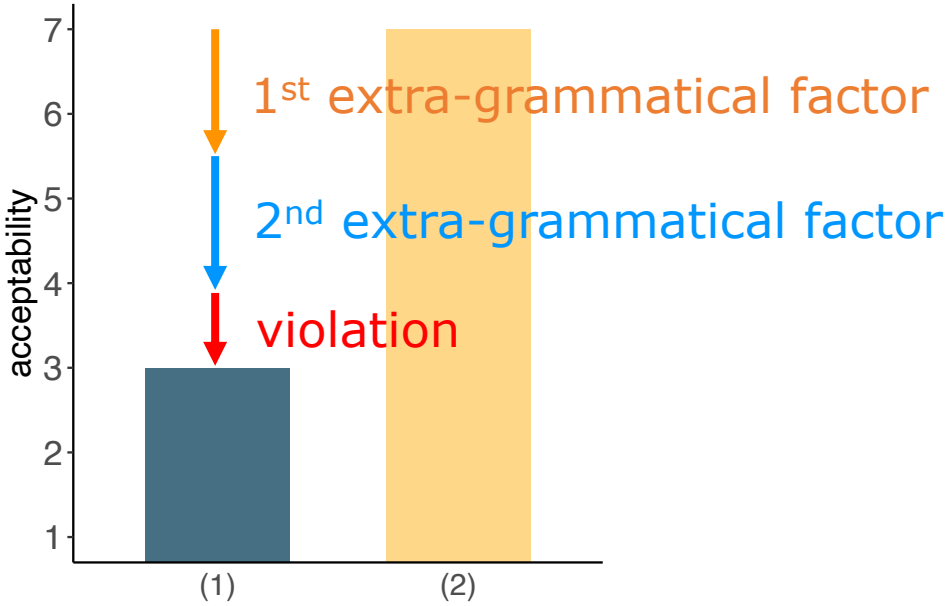
1. Part of the cannon of theoretical linguistics
2. Source of the effect
3. Lack of unified account
4. Cross-linguistic and cross-dependency variation
5. Intersection with sentence processing

Defining effects

(1)*What$_i$ do you wonder whether Mary solved ___$_i$?

(2) What$_i$ do you think that Mary solved ___$_i$?





Experimental syntax allows us to isolate the effect of the violation from extra-grammatical factors that also influence acceptability ratings

Factorial definition of island effects

2x2 factorial design

Factors → **DEPENDENCY LENGTH** x **STRUCTURE TYPE**

Levels → Short, Long | Non-island, Island

1. Who ___ thinks [that Mary solved the problem]?	short	non-island
2. What do you think [that Mary solved ___]?	long	non-island
3. Who ___ wonders [whether Mary solved the problem]?	short	island
4. *What do you wonder [whether Mary solved ___]?	long	island

(Sprouse 2007, Sprouse et al. 2012, Sprouse et al. 2016 and much subsequent work)

Factorial definition of island effects

(2)-(1)= Effect of length

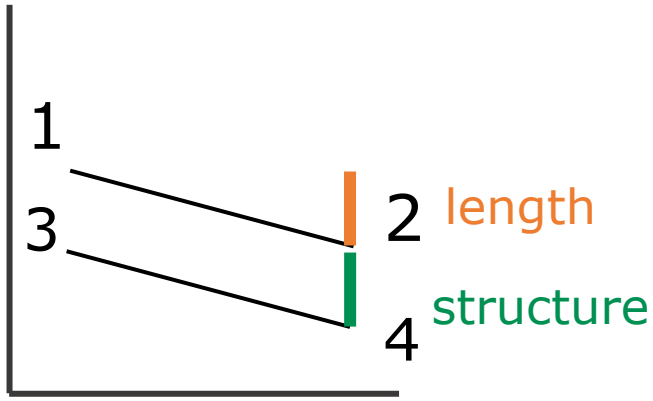
1. Who __ thinks [that Mary solved the problem]?	short	non-island
2. What do you think [that Mary solved __]?	long	non-island
3. Who __ wonders [whether Mary solved the problem]?	short	island
4. *What do you wonder [whether Mary solved __]?	long	island

(3)-(1) = Effect of structure

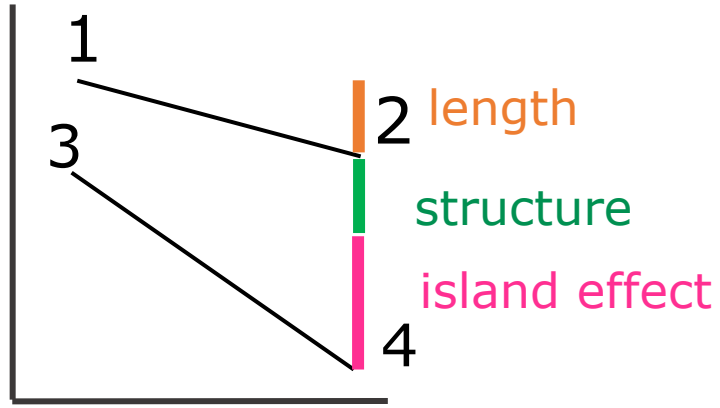
1. Who __ thinks [that Mary solved the problem]?	short	non-island
2. What do you think [that Mary solved __]?	long	non-island
3. Who __ wonders [whether Mary solved the problem]?	short	island
4. *What do you wonder [whether Mary solved __]?	long	island

Factorial definition of island effects

No island effect



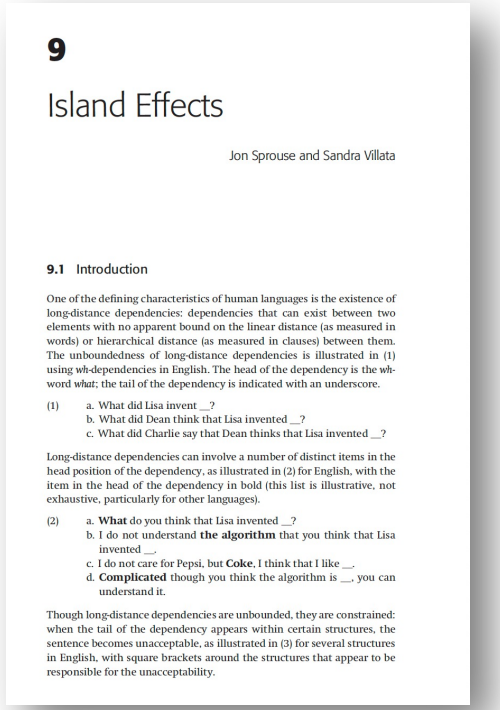
Island effect



The factorial definition of island effects provides us with a way to i) isolate the effect of interest from confounding factors, ii) detect its presence statistically, and iii) precisely quantify its effect size.

study	language	dependency	adj	np	sub	rc	wh	scale
Almeida 2014	Br. Portuguese	bare wh					0.6	z-score
Almeida 2014	Br. Portuguese	topicalization					–	z-score
Lu et al. 2019	Chinese	wh-arg-in-situ				1.5		raw (7)
Lu et al. 2019	Chinese	wh-adj-in-situ				1.6		raw (7)
Christensen et al. 2013	Danish	bare wh					1.2	raw (5)
Poulsen 2008	Danish	topicalization	4.0					raw (5)
López-Sancio 2015	Spanish	bare wh					1.75	z-score
López-Sancio 2015	Spanish	rel. clause					1.25	z-score
Sprouse et al. 2016	English	bare wh	0.7	1.1	0.6		1.2	z-score
Sprouse et al. 2016	English	complex wh	0.8	0.5	0.5		0.6	z-score
Sprouse et al. 2016	English	rel. clause	–	0.5	0.5		0.4	z-score
Almeida 2014	English	topicalization					–	z-score
Sprouse et al. 2011	English	wh-arg-in-situ	–	–	–		–	z-score
Sprouse et al. 2016	Italian	bare wh	1.3	0.9	1.4		1.7	z-score
Sprouse et al. 2016	Italian	rel. clause	1.1	0.6	–		0.7	z-score
Omaki et al. 2019	Japanese	np scrambling			–			z-score
Sprouse et al. 2011	Japanese	wh-arg-in-situ	–	–	–		–	z-score
Kim & Goodall 2016	Korean	wh-arg-in-situ	–				0.3	z-score
Kim & Goodall 2016	Korean	wh scrambling	–				0.7	z-score
Ko et al. 2019	Korean	np scrambling	–			–	–	z-score
Tucker et al. 2019	MS Arabic	complex wh	0.8	0.5			0.4	z-score
Kush et al. 2018	Norwegian	bare wh	1.1	1.7	1.3	1.4	0.4	z-score
Kush et al. 2018	Norwegian	complex wh	1.3	1.2	1.2	1.4	0.3	z-score
Kush et al. 2019	Norwegian	topicalization	0.2	0.5	1.7	0.7	–	z-score
Stepanov et al. 2018	Slovenian	bare wh			0.6		–	z-score
Ortega-Santos et al. 2018	English	bare wh					1.36	z-score
Ortega-Santos et al. 2018	Spanish	bare wh					1.15	z-score
Bondevik et al. 2021	Norwegian	topicalization	0.397		1.375		0.375	z-score
Bondevik et al. 2021	Norwegian	topicalization	0.485					z-score
Bondevik et al. 2021	Norwegian	topicalization	1.032					z-score
Stigliano & Xiang 2021	Spanish	rel. clause					0.67	z-score
Stigliano & Xiang 2021	Spanish	rel. clause					0.95	z-score
Pañeda & Kush 2022	Spanish	complex wh					0.22	z-score
Pañeda & Kush 2022	Spanish	complex wh					0.38	z-score
Pañeda & Kush 2022	Spanish	complex wh					1.09	z-score
Pañeda & Kush 2022	Spanish	complex wh					1.39	z-score
Pañeda et al. 2024	English	simple wh					0.87	z-score
Pañeda et al. 2024	English	simple wh					1.09	z-score
Pañeda et al. 2024	English	simple wh					1.2	z-score
Pañeda et al. 2024	Spanish	simple wh					1.19	z-score
Pañeda et al. 2024	Spanish	simple wh					1.3	z-score
Pañeda et al. 2024	Spanish	simple wh					1.55	z-score

Numbers indicate the size of the island effect—higher numbers reflect stronger effect sizes. A dash means no island effect and an empty cell indicates that the effect hasn't been tested in that study.



9

Island Effects

Jon Sprouse and Sandra Villata

9.1 Introduction

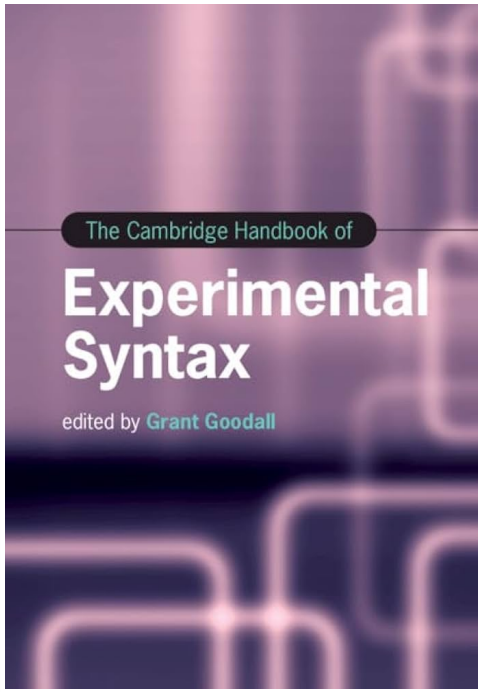
One of the defining characteristics of human languages is the existence of long-distance dependencies: dependencies that can exist between two elements with no apparent bound on the linear distance (as measured in words) or hierarchical distance (as measured in clauses) between them. The unboundedness of long-distance dependencies is illustrated in (1) using *wh*-dependencies in English. The head of the dependency is the *wh*-word *what*; the tail of the dependency is indicated with an underscore.

(1) a. What did Lisa invent __?
b. What did Dean think that Lisa invented __?
c. What did Charlie say that Dean thinks that Lisa invented __?

Long-distance dependencies can involve a number of distinct items in the head position of the dependency, as illustrated in (2) for English, with the item in the head of the dependency in bold (this list is illustrative, not exhaustive, particularly for other languages).

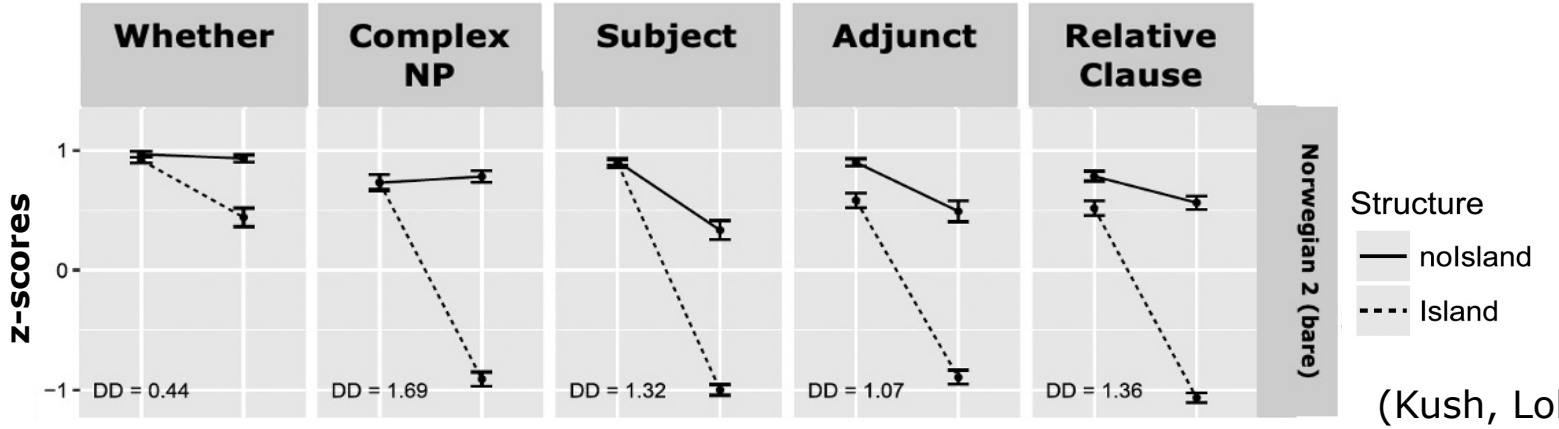
(2) a. **What** do you think that Lisa invented __?
b. I do not understand **the algorithm** that you think that Lisa invented __.
c. I do not care for Pepsi, but **Coke**, I think that I like __.
d. **Complicated** though you think the algorithm is __, you can understand it.

Though long-distance dependencies are unbounded, they are constrained: when the tail of the dependency appears within certain structures, the sentence becomes unacceptable, as illustrated in (3) for several structures in English, with square brackets around the structures that appear to be responsible for the unacceptability.



An extended (still non-exhaustive) list of published studies using the factorial design for island effects in adjunct, complex noun phrase, subject, relative clause, and wh-islands (adapted from Sprouse & Villata 2021)

Islands in Norwegian



(Kush, Lohndal & Sprouse 2018)

Results across 3 different experiments (only experiment 2 reported here) show that:

1. All 5 structures are islands in Norwegian (including those previously not thought to be islands)
2. *Whether*-islands fall within the acceptable spectrum and exhibit substantial variation across-speakers

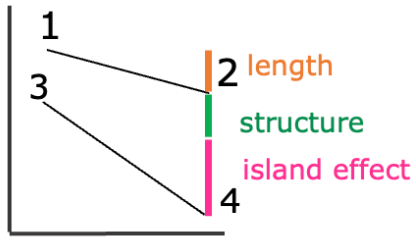
Experimental syntax can reveal patterns that only emerge when testing large participant samples across multiple items and using inferential statistics

2. Isolating the source of effects

The source of island effects (part 1)

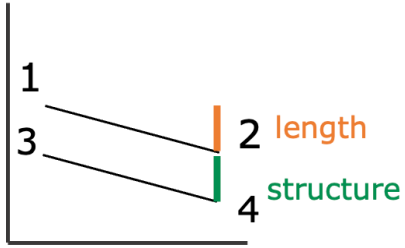
Are islands effects in the grammar or are they grounded in the processing system?

Grammar-based theories:



1. Syntax-based theories:
   a. Impenetrability theories (e.g. Subjacency, Barriers, Phases)
      (e.g. Chomsky 2001)
   b. Intervention-based theories (Relativized Minimality)
      (e.g. Rizzi 1990)
2. Semantic-based theories
   (e.g. . Szabolcsi & Zwart 1993, 1997, Abrusàn 2014)
3. Information-structure theories
   (e.g. Erteshick Shir 1973, Goldberg 2006, Abeille et al. 2020)

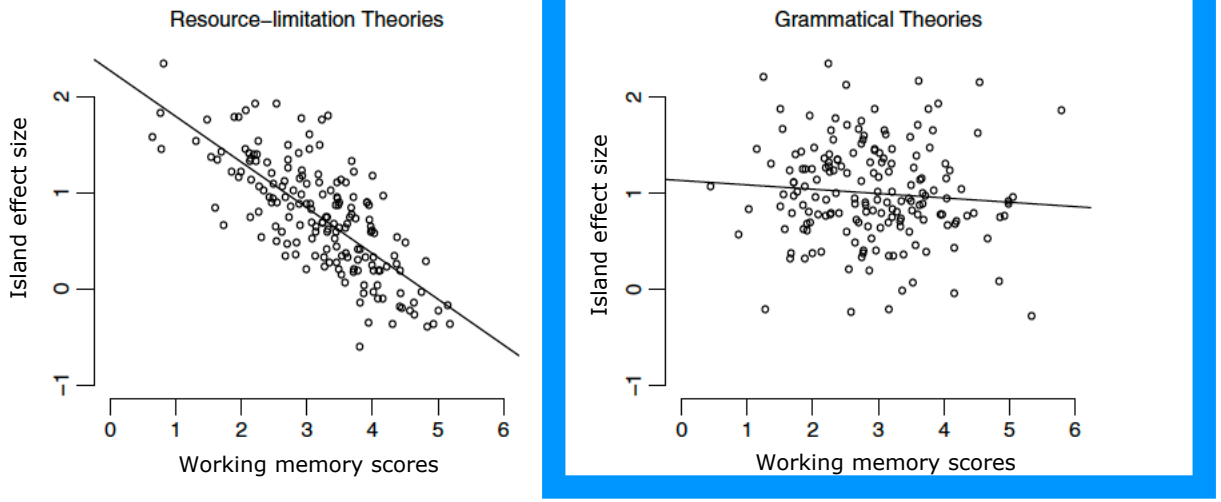
Processing-based theories:



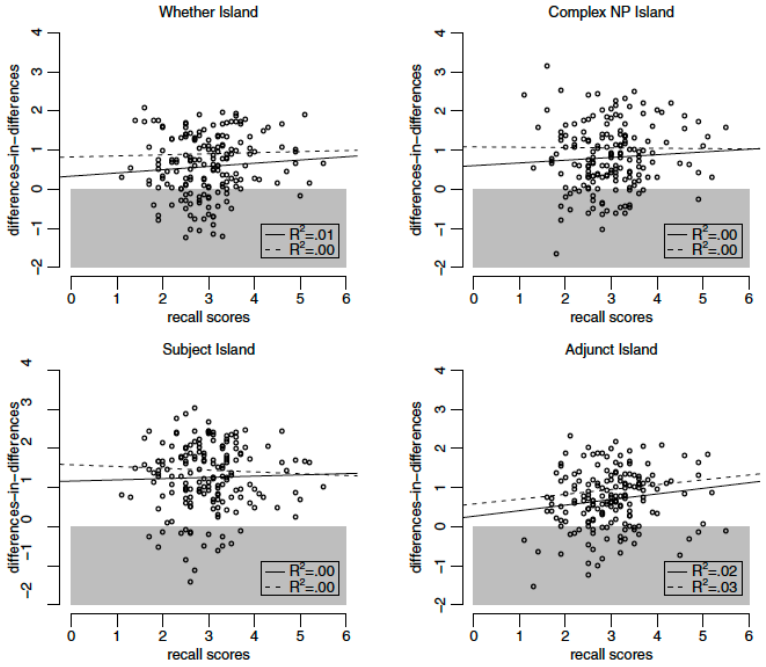
Island violating sentences are **grammatical**, but perceived as **unacceptable** due to the combined processing costs of 1) a long dependency and 2) the presence of an island structure, which together exceed working memory capacity
(e.g. Kluender & Kutas 1993, Hofmesiter & Sag 2010)

Measuring working-memory/island effects relationship

Predictions



Results



No evidence of a relationship between working-memory capacity and island effects

(Figures from Sprouse, Wagers & Phillips (2012); see Michel (2014) for converging evidence)

Strong/weak islands distinction

- Islands have been classified into two varieties, strong and weak, depending on their selectivity to extraction

Strong islands: block all dependencies

simple wh: *What did John frown [because you asked_]? ADJUNCT ISLAND
complex wh: *Which question did John frown [because you asked_]?

Weak islands: block some dependencies, but not others

simple wh: *What did John wonder [whether you asked _]? WHETHER ISLAND
complex wh: Which question did John wonder [whether you asked _]?

The source of island effects (part 2)

- It has been claimed that the **semantic fit** between the extracted element and the embedded verb eases **memory retrieval**, resulting in higher acceptability ratings (e.g. Hofmeister & Sag, 2010)
- This has been proposed as a potential explanation for why complex wh-phrases sometimes yield higher ratings than simple ones when extracted from islands

(1) ?**Which book** do you wonder whether the student read ___ ? Easy retrieval → higher acceptability

(2) ***What** do you wonder whether the student read ___ ? Difficult retrieval → lower acceptability

How can we test this memory-based hypothesis?

We can use **pseudowords**!

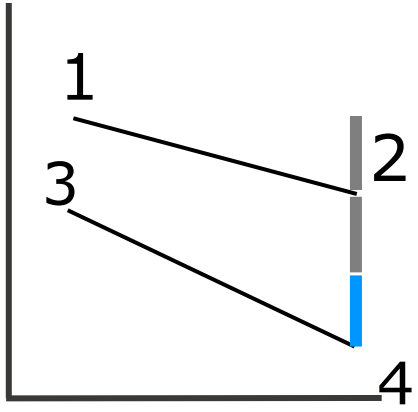
Prediction: If the higher acceptability sometimes observed with complex wh-phrases stems from increased semantic coherence, replacing the lexical element in complex wh-phrases with a **pseudoword** should remove the advantage

(1) ***Which sping** do you wonder whether the student read __ ?

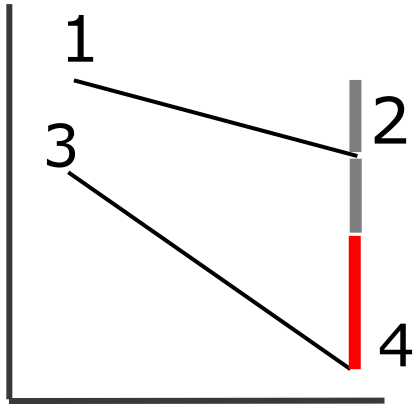
(2) ***What** do you wonder whether the **sping** read __ ?

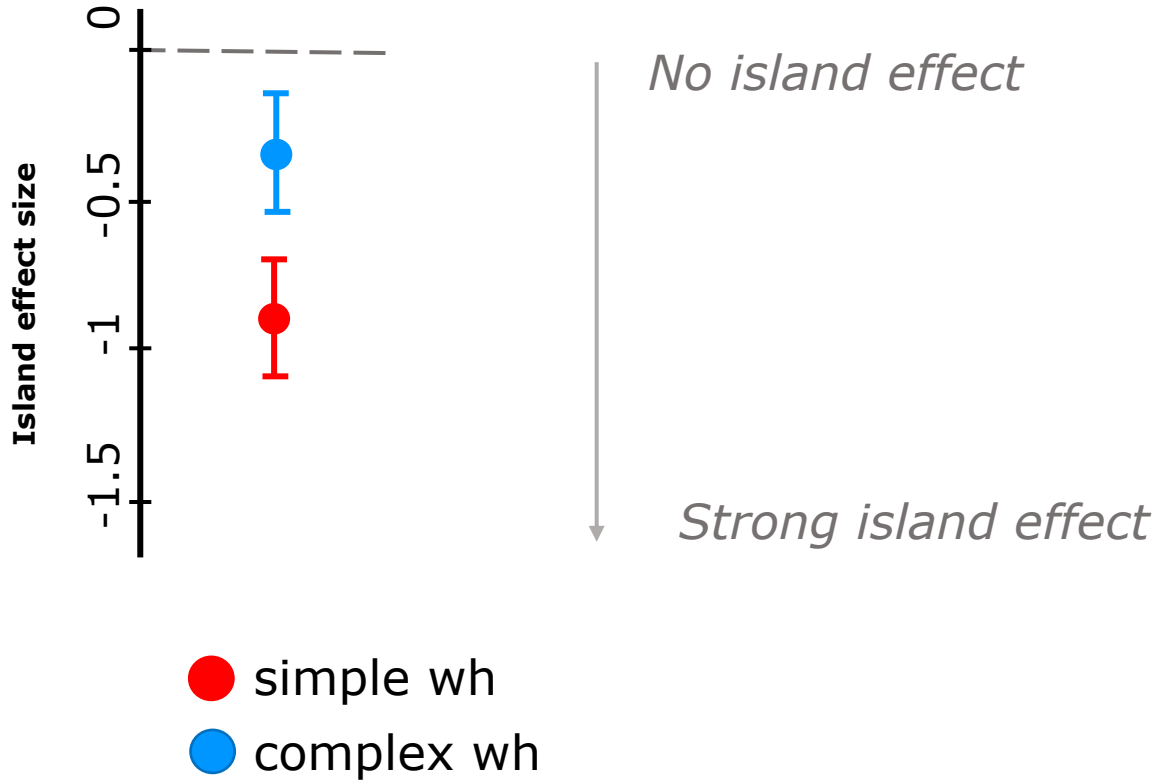
Measuring effect sizes

Complex wh



Simple wh

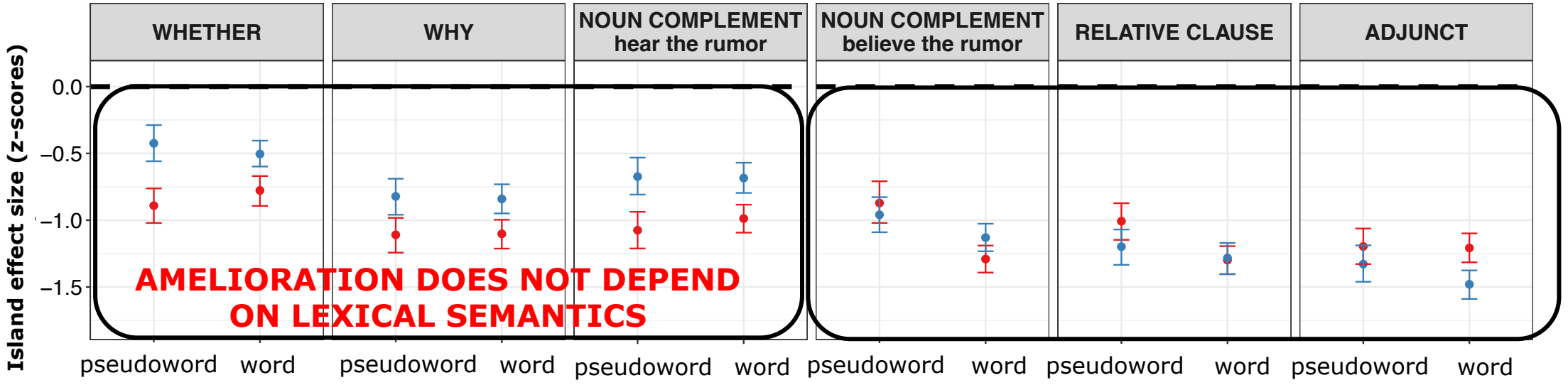




Same pattern for words and pseudowords

Results for pseudowords and words are nearly identical:

1. When **no increase** in acceptability is observed with complex wh-phrases, it holds **regardless of** whether the element is a **word** or a **pseudoword**
2. When there **is** an increase in acceptability, the effect occurs **regardless of lexical status** either (word vs. pseudoword)



Villata & Sprouse (in prep.)

3. Augmented precision

Why is the strong/weak island distinction interesting?

1. The theoretical literature does not agree on the judgments for the extraction of complex wh-phrases, and the experimental literature shows indications of **partial amelioration** (e.g. Sprouse et al. 2016, Kush et al. 2018 among many others)

?Which question did John wonder [whether you asked _]?

2. However, **partial amelioration** suggests **different effect sizes**, which in turn raises deep architectural questions about the nature of the grammar itself (categorical vs. gradient)

A large-scale experiment

- We tested **7 island types** which formed a mix of strong and weak islands
- Inside each island type we tested **several island tokens**
- Each island was tested with both simple and complex wh (~60 experiments total)
- ~ 200 self-assessed English native speakers per experiment

ISLAND TYPES

- Wh-islands (whether, if, who, why…)
- Factive (communication V, discovery V….)
- Negative (not, n't)
- Noun Complement (make the claim, believe the rumor…)
- Subject (definite, indefinite …)
- Relative clause (that, who)
- Adjunct (causal, temporal, conditional)

x

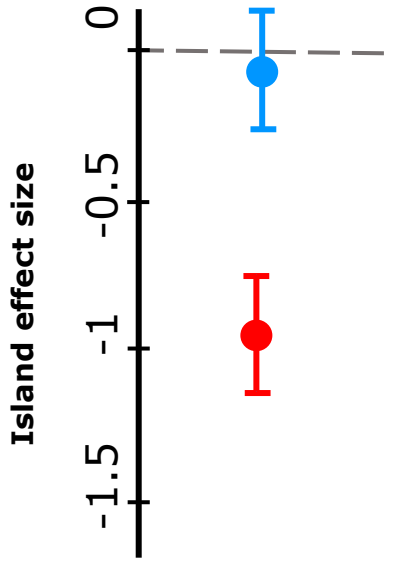
WH-TYPES

- Simple
- Complex

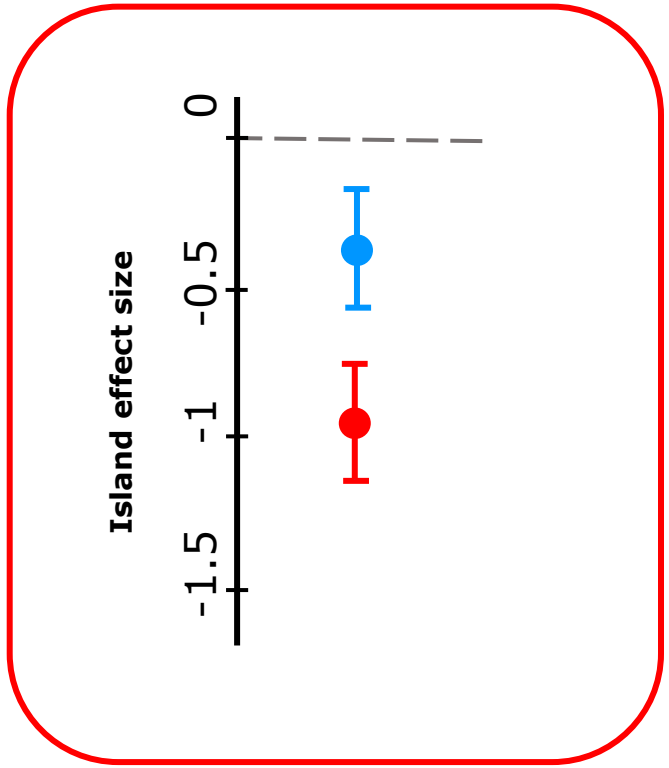
Villata & Sprouse (in prep.)

Measuring effect sizes

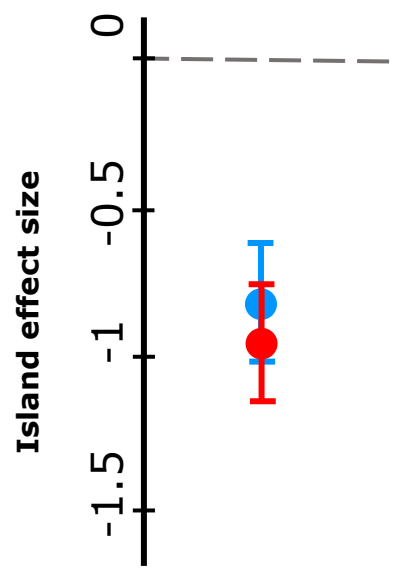
FULL AMELIORATION



PARTIAL AMELIORATION



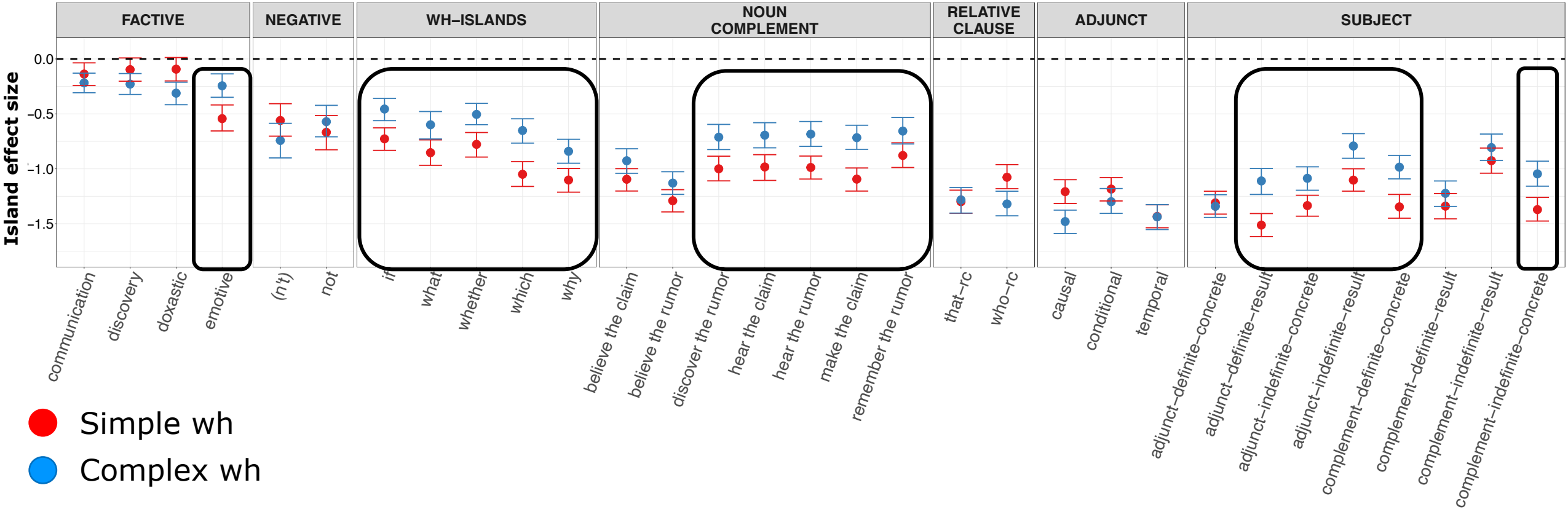
NO AMELIORATION



simple wh

complex wh

Different effect sizes!



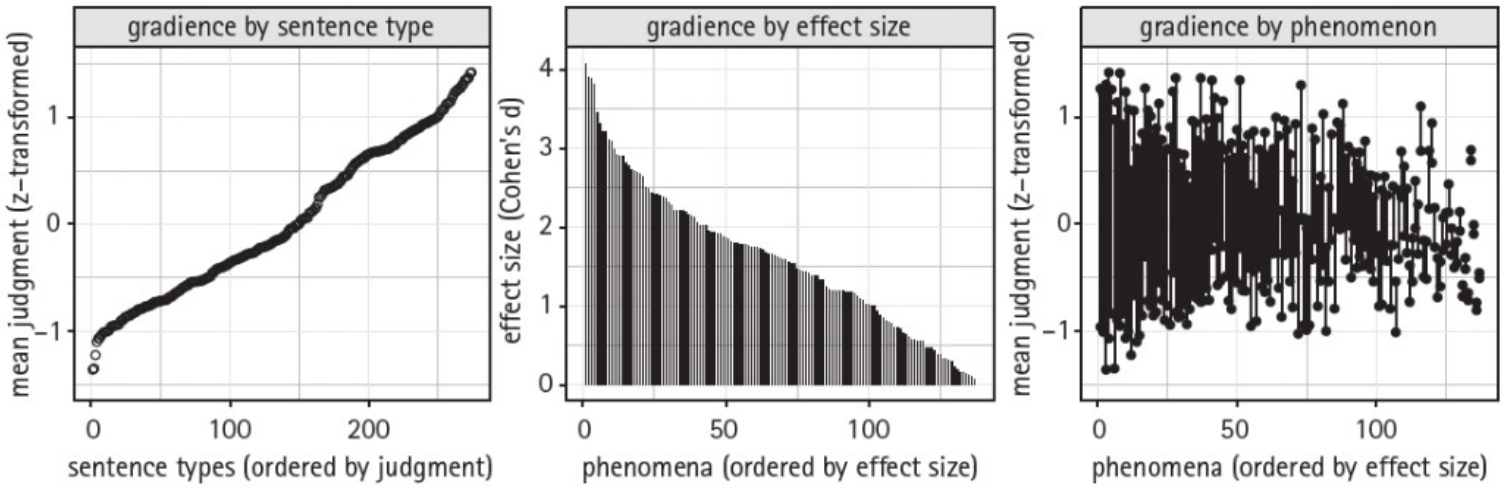
Simple wh

Complex wh

Villata & Sprouse (in prep.)

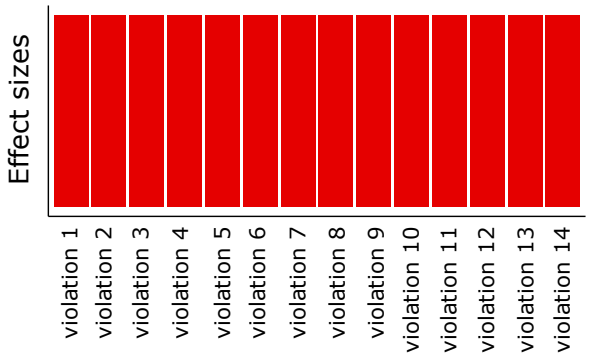
We observe **partial amelioration** across several island tokens, suggesting the presence of **different effect sizes**

Gradient effects

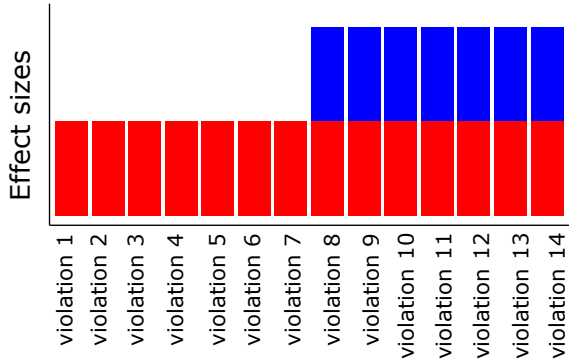


Sprouse (2018)

CATEGORICAL GRAMMAR

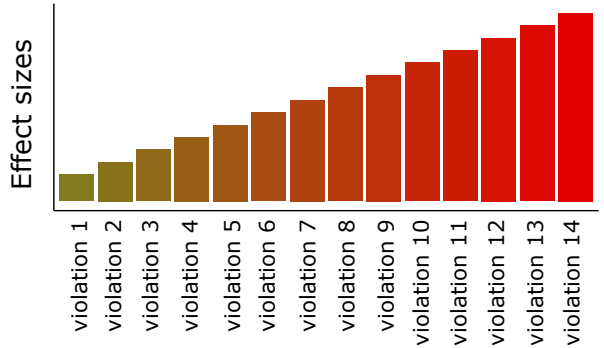


1 effect size



violations stacking

CONTINUOUS GRAMMAR



multiple effect sizes on a continuum

What are the methods of experimental syntax?

Beyond acceptability judgments:
Online measures

1. Behavioral
2. Neuroimaging

Online consequences of islands

- **Active gap filling** : the parser's anticipation of the gap location before encountering unambiguous evidence for the gap
(e.g. Crain & Fodor 1985, Stowe 1986 a.o.)

  1) My brother wanted to know **who** Ruth will bring us home to ___ at Christmas.
  2) My brother wanted to know **if** Ruth will bring us home to Mom at Christmas.

  (Stowe 1986)

- One processing consequence of island is the suppression of **active-gap filling** inside of island domains

  (e.g. Stowe 1986, Pickering et al., 2004)

  3) The teacher asked **what** the team laughed about Greg's brother fumbling ___.
  4) The teacher asked **what** *[the silly story about Greg's brother]* was supposed to mean ___.

  (Stowe 1986)

Active-gap filling suppression inside of islands suggests that the parser respects island constraints in real time

Parasitic gaps

- However, when there is a second gap in the sentence (**parasitic gap**), the subject island gap is licensed (Engdahl 1983, Culicover & Postal 2001) and active-gap filling is observed inside the subject island (Phillips 2006)

3) *The outspoken environmentalist worked to investigate what [the local campaign to preserve ___] had harmed the annual migration.
4) The outspoken environmentalist worked to investigate what [the local campaign to preserve $\_\_\__{PG}$] had harmed ___.

(Phillips 2006)

- Since at the time of active-gap filling in (4) the parser does not know that another gap will appear, this cannot be due to a general difficult of the parser to posit a gap inside of islands, contra processing approaches of island effects

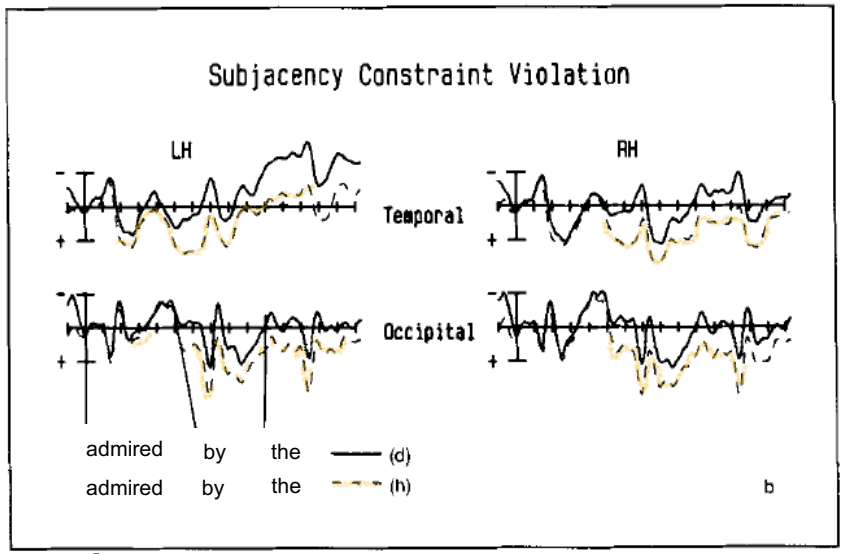
Online measures can be helpful to provide additional evidence to adjudicate between **competing theories** about the **source of the effect**

Why are neuroimaging data helpful?

- Acceptability judgments are unidimensional data
- Neuroimaging data (e.g. EEG, MEG, fMRI) are multidimensional

MEG	EEG	EEG + MEG	fMRI
			
Proudfoot M, et al (2014) *Pract Neurol;* 14:336 – 343	Papadelis & Perry (2021). *Semin Pediatr Neurol*. October;39: 100919 doi:10.1016/j.spen.2021.100919	Papadelis & Perry (2021). *Semin Pediatr Neurol*. October;39: 100919 doi:10.1016/j.spen.2021.100919	

EEG and island effects



Neville et al. (1991): subject island violation detection

Was a sketch of the landscape **admired** by the man?

*What was [a sketch of __] **admired** by the man?

P600

McKinnon & Osterhout (1996): adjunct island boundary detection

*I wonder which of his staff members the candidate was annoyed [**when** his son was questioned by __]?

Kluender & Kutas (1993): wh-island boundary detection

*What do you wonder [**who** they caught at __ **by** accident?)

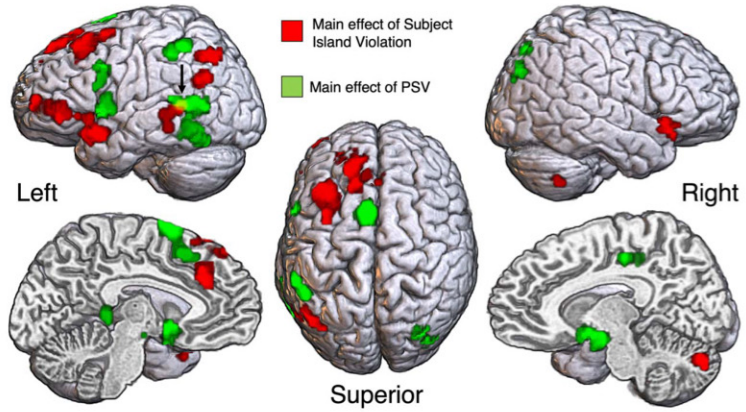
N400 LAN

fMRI and island effects

Phrase structure violation (PSV):	*Which candidate does the moderator of the panel think ___ avoided the debate's about questions healthcare
Subject island violation:	*Which candidate does the moderator think [the speech by ___] ruined the debate's questions about healthcare

The two violations produced non-overlapping maps:

1. PSV activates working memory brain networks
2. Subject islands activate networks linked to semantic processing



(Matchin et al. 2025)

These results suggest that the source of island effects is unlikely to be found in memory-related processes

What is the future of experimental syntax?

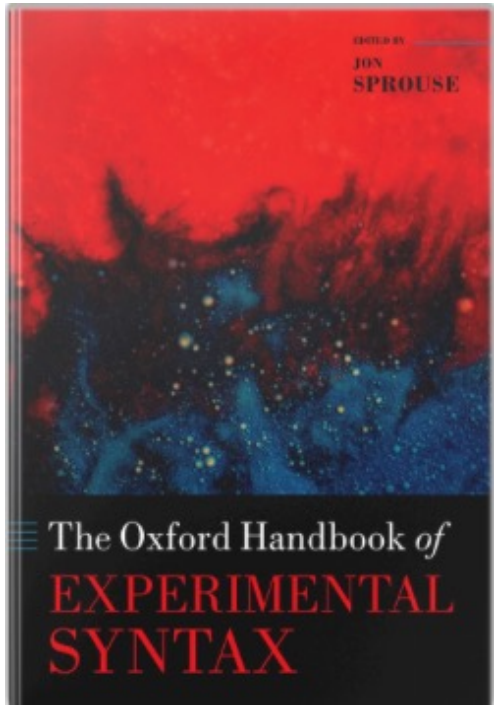
- First, recent advances in experimental syntax, combined with platforms that enable large-scale data collection, place us in a uniquely strong position to tackle increasingly pressing questions in the field, like the gradient nature of acceptability judgments and their implications for the nature of grammar
- Second, across-level integration is more important than ever to foster substantial theoretical progress and to build an integrated theory of language.
- Third, we need to keep expanding the range of languages we investigate experimentally.
- Lastly, we need to keep nurturing a new generation of linguists with experimentation in their toolkit.

The next generation

"Designing experiments that answer interesting syntactic questions is hard, as it typically requires some syntactic sophistication as well as some understanding of language processing. But most students do not get training in both syntax and psycholinguistics. They can do theoretical syntax, or they can do psycholinguistics, but they cannot do both.

The answer may lie in teamwork. […] we can bring people with these disparate sets of interests to the same room and have them ask questions of each other."

(Maria Polinsky, "The future of experimental syntax", 2023, p.655)



I believe it *is* possible to raise a new generation of linguists who are truly bilingual in both syntax and experimentation if we create the right training conditions for that to happen: building solid institutional bridges across departments, building dedicated programs!

Thank you!

glowing lecture series