Class 8:
What makes a long-distance extraction unacceptable?

Ted Gibson
9.59J/24.905J

Different English constructions involving "fronting" of material probably involve

Wh-questions, Relative clauses, Topicalization

Wh-question: _Who did Anthony ignore?_
Embedded wh-question: Dan asked _who Anthony ignored_.
Relative clause: Dan asked about the man _who Anthony ignored._
Topicalization: Susan, _Anthony ignored._

Wh-question across a second clause:
    Who did Emily notice that Anthony ignored?
Embedded wh-question across a second clause:
    Dan asked _who Emily noticed that Anthony ignored_.
Relative clause across a second clause:
    Dan asked about the man _who Emily noticed that Anthony ignored._
Topicalization across a second clause:
    Susan, _Emily noticed that Anthony ignored._

Different English constructions involving "fronting" of material probably involve

In addition to their superficial similarity, they obey the same kinds of structural constraints, e.g., island phenomena

NP-island, wh-question:
    ? Dan asked who Emily heard the rumor that Anthony ignored [].
NP-island, relative clause:
    ? Dan asked about the man who Emily heard the rumor that Anthony ignored [].
NP-island, topicalization:
    ? Susan, Emily heard the rumor that Anthony ignored [].

Subject-island, wh-question:
    ? Dan asked who Emily heard that some friends of [] ignored Liz.
Subject-island, relative clause:
    ? Dan asked about the man who Emily heard that some friends of [] ignored Liz.
Subject-island, topicalization:
    ? Susan, Emily heard that some friends of [] ignored Liz.

Conjunct-island, wh-question:
    ? Dan asked who Emily heard that Anthony ignored Liz and [].
Conjunct-island, relative clause:
    ? Dan asked about the man who Emily heard that Anthony ignored Liz and [].
Conjunct-island, topicalization:
    ? Susan, Emily heard that Anthony ignored Liz and [].

Different English constructions involving "fronting" of material probably involve

One syntactic "island" for extraction that was discovered by Ross (1967) was the NP-island constraint, whereby it doesn't sound good to have a wh-question out of an NP. For example, while it's ok to extract out of an embedded clause as in (1), it's not so good when the embedding is part of a noun phrase:

(1) What vegetable did the reporter believe that the president disliked?
(2) ?*What vegetable did the reporter believe the fact that the president disliked?

It is thought that relative clauses such as in (3) have a similar structure as wh-questions:

(3) The vegetable that the president dislikes is broccoli.

Provide evidence from your judgments of relative clauses that suggests a similar island effect in relative clauses as in (1) vs (2), and therefore a similar structure in relative clauses as in wh-questions.

Long-distance dependencies
What makes a long-distance extraction unacceptable?
(Ross, 1967; see Ambridge & Goldberg, 2008)

Sometimes long-distance extractions seem ok:

1a. What does Susan think that John bought ___?
1b. What does Sarah believe that Susan thinks that John bought ___?
1c. What does Bill claim that Sarah believes that Susan thinks that John bought ___?

Long-distance dependencies
What makes a long-distance extraction unacceptable?
(Ross, 1967; see Ambridge & Goldberg, 2008)

Complex NPs (both noun complements and relative clauses)
*Who did she see the report that was about _?   (cf. She saw the report that was about x)

Subjects
*Who did that Mary liked _ bother him?   (cf. That Mary liked x bothered him)

Presupposed adjuncts
??What did she leave the movie because they were eating _?
(cf. She left the movie because they were eating x)

Complements of manner-of-speaking verbs
??What did she whisper that he saw_?   (cf. She whispered that he saw x)

Complements of factive verbs
??What did she realize that he saw_?   (cf. She realized that he saw x)

What is the nature of "syntactic islands"?

Purely syntactic accounts, going back to Ross & Chomsky:

E.g., the **Subjacency** Condition (Ross, 1967; Chomsky, 1973; cf. Chomsky, 1962; Ross, 1967; Chomsky, 1986; Chomsky, 1995):

No rule may move a phrase from position Y to position X (or conversely) in:
       ...X...[$_A$...[$_B$...Y...]...]...X...
       Where A and B are cyclic nodes.

(1) *What do [you wonder [whether John bought __]] ?
(2) *Who did [you criticize votes for [the impeachment of __]]?

What is the nature of "syntactic islands"?

BUT: Non-syntactic factors like plausibility and working memory demands affect extraction difficulty (e.g., Deane, 1991; Kluender, 1992; Sag et al., 2007; Hofmeister & Sag, 2010):

    (2) * Who did [you criticize votes for [the impeachment of ___]]?
    (3)   Who did [you obtain votes for [the impeachment of ___]]? (Deane, 1991)

    (4) * Who did Emma doubt the report that we had captured ___ ?
    (5)   Which convict did Emma doubt a report that we had captured ___ ?

Can syntactic island effects be explained by resource complexity or other non-syntactic factors like plausibility or coherence?

Sprouse Wagers & Phillips, 2012
Against WM accounts of extraction unacceptability

Sprouse Wagers & Phillips, 2012 divide the accounts into 3 camps:

Reductionist accounts; E.g., _working memory_ (WM) accounts (Kluender, 1998; Hofmeister & Sag, 2010)

Grammar-based accounts; E.g., _syntax accounts_ (Ross, 1967; Chomsky, 1973, 1986); _Syntax / semantics accounts_ (Goldberg, 1995; Ambridge & Goldberg, 2008)

"Grounded" accounts: WM accounts with _grammaticization_ (e.g., Fodor, 1978; Hawkins, 1999)
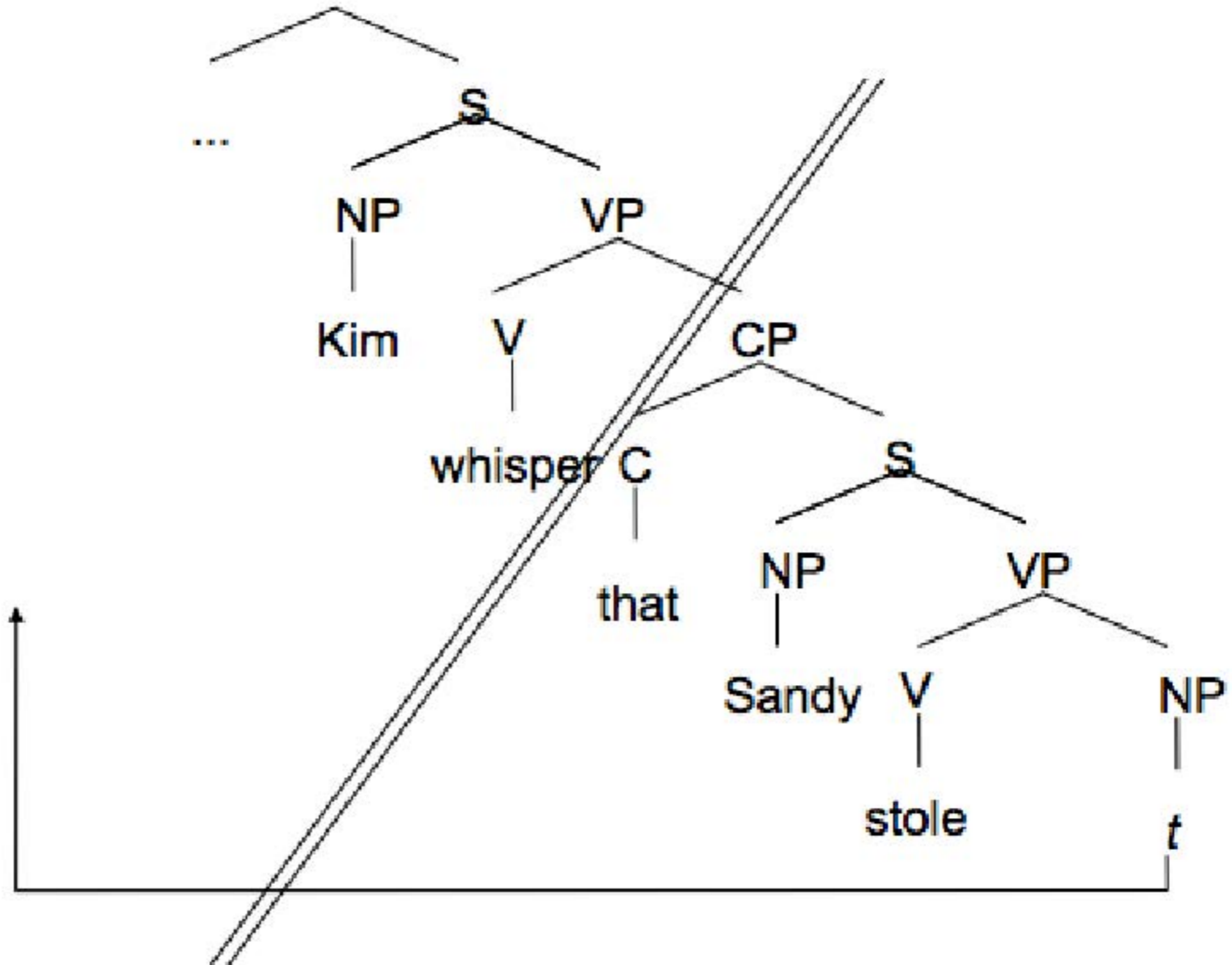
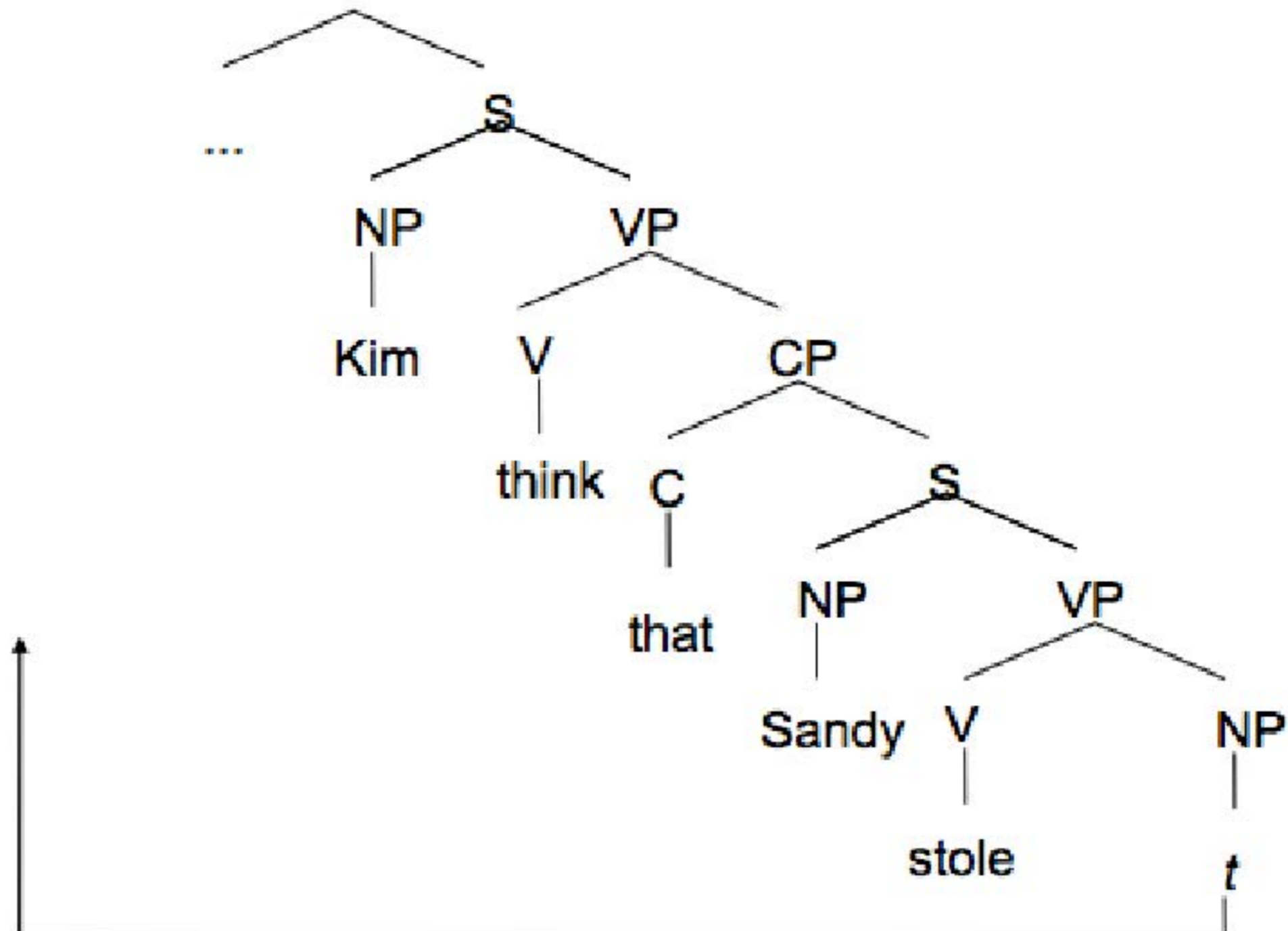# Structure Drives Islands?
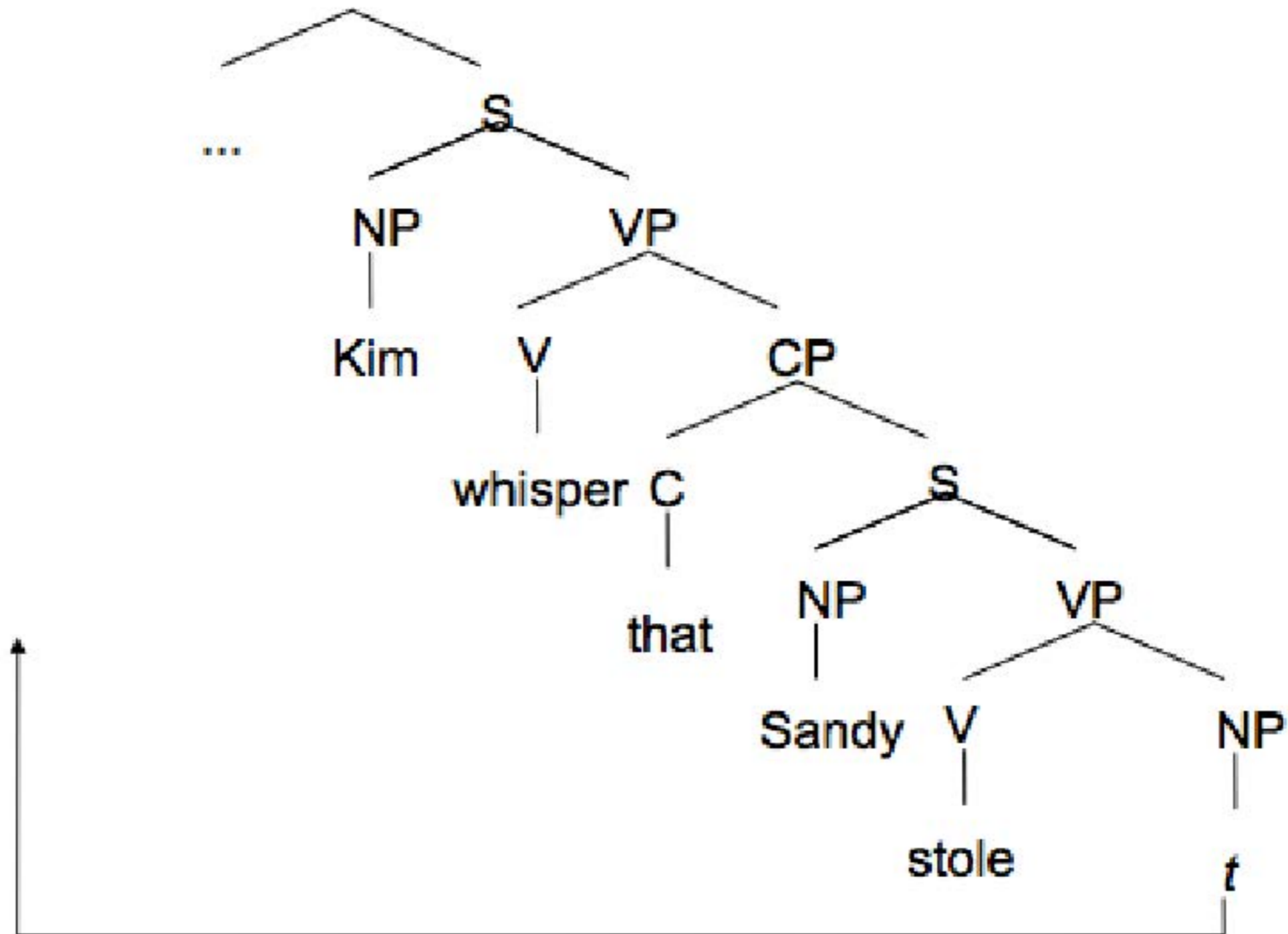
?What did Kim whisper that Sandy stole _?

Structure Drives Islands?

?What did Kim whisper that Sandy stole _?

Structure Drives Islands?

What did Kim think that Sandy stole _?

Information-Structure based accounts
Ambridge & Goldberg, 2008

Perhaps bad extractions are cases where *backgrounded information* is extracted.

Presupposition is one example of backgrounded information:

A test of **backgrounded information**: Consider a sentence S with an embedded clause E. The <u>*main assertion*</u> of S is negated when S is negated.

(1) Mary saw the report that was about Bill.
(2) → (negation) Mary didn't see the report that was about Bill.

The main assertion of the sentence is negated: not seeing the report.

Information-Structure based accounts: BCI account
Ambridge & Goldberg, 2008

(1) Mary saw the report that was about Bill.
(2) → (negation) Mary didn't see the report that was about Bill.

If the meaning of a negative form of S doesn't negate the meaning of E then E *is not the main assertion*: it's presupposed or backgrounded

The meaning of the RC in (2) is still *the report that was about Bill* (not *the report that wasn't about Bill*)

Therefore, a relative clause (RC) is backgrounded information. Then maybe extraction from backgrounded information is disallowed:

*Who did Mary see the report that was about _?

Backgrounded constructions are islands (BCI) *account: Backgrounded constituents may not serve as gaps in filler-gap constructions.*

Information-Structure based accounts: BCI Account
Ambridge & Goldberg, 2008

(3) Mary thought that Bill liked Sue.
(4) → (negation) Mary didn't think that Bill liked Sue.

This can mean: *It is not the case that Bill liked Sue.*

Therefore the embedded clause might be the main assertion in (3). So extraction of the embedded clause is possible:
(5) Who did Mary think that Bill liked _?

(6) Mary whispered that Bill liked Sue.
(7) → (negation) Mary didn't whisper that Bill liked Sue.

This can't mean: *It is not the case that Bill liked Sue.*

Therefore the embedded clause is not the main assertion in (6). So extraction of the embedded clause is not possible:

(8) ?* Who did Mary whisper that Bill liked _?

Information-Structure based accounts
Ambridge & Goldberg, 2008

Experiment: Rate acceptability of long-distance extraction:

(1) Who did Mary think that Bill liked _?
(2) Who did Mary whisper that Bill liked _?

Compare these ratings to ratings of how much one negative sentence implied another, for the same embedding verbs:

embedding verb: *think*
(3) Mary didn't think that Bill liked Sue.
(4) Bill didn't like Sue.

embedding verb: *whisper*
(5) Mary didn't whisper that Bill liked Sue.
(6) Bill didn't like Sue.

3 classes of Verbs:
Factive verbs (realize, remember, notice, know): should not allow extraction
manner-of-speaking verbs (whisper, stammer, mumble, mutter): less likely to allow extraction
'bridge verbs'' (say, decide, think, believe): most likely to allow extraction

Predictions

1. A purely syntactic subjacency account would expect all structurally identical sentences to behave identically, and thus would predict no systematic differences across semantic verb classes.

2. BCI account: a negative correlation between acceptability ratings and how much the embedded clause can receive the negation, in interpretation.

Negation Test Correlates with Unextractability across Bridge Verbs (Ambridge & Goldberg, 2008)
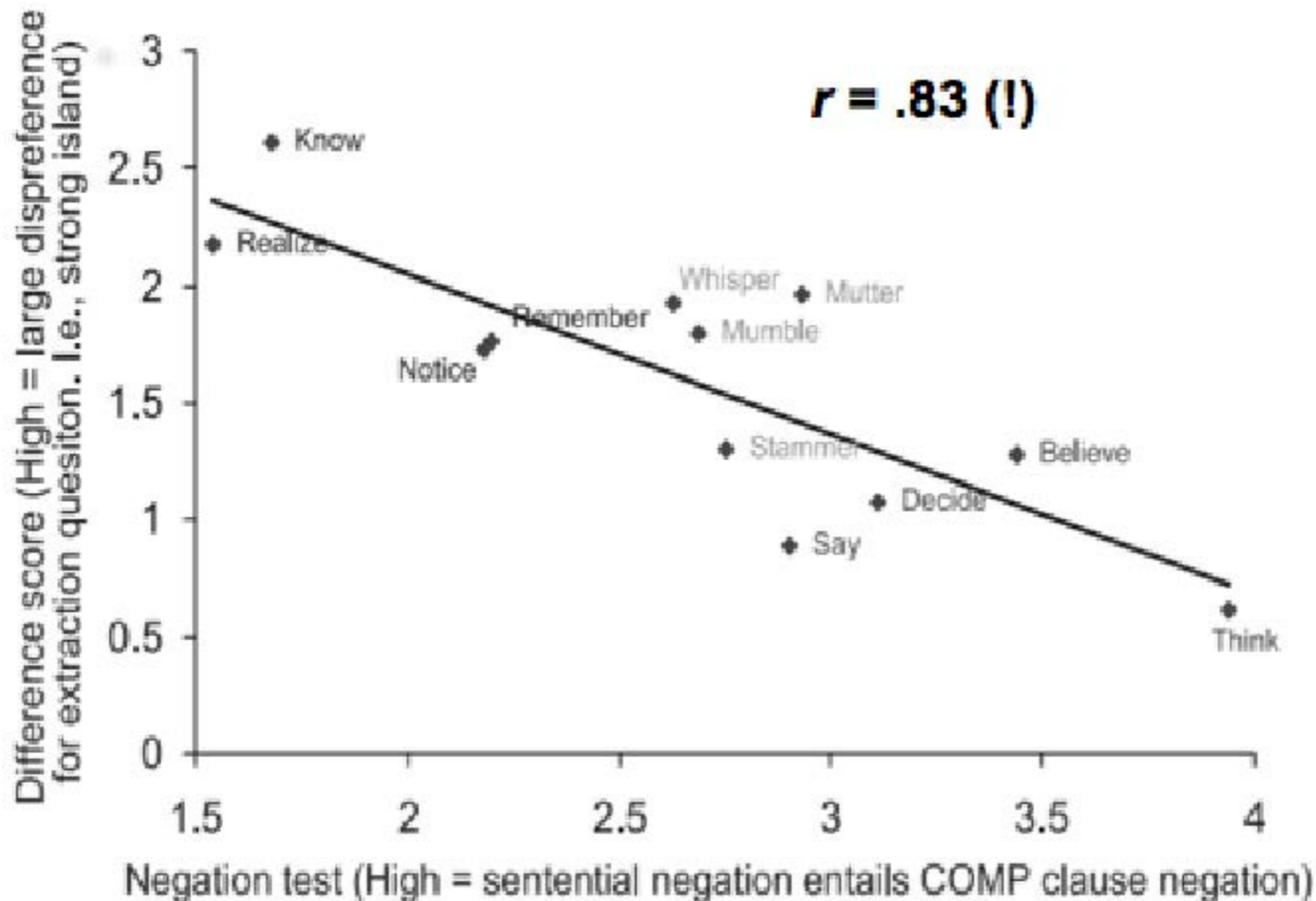

Figure 3. *Correlation between difference scores (dispreference for question scores) and negation test scores*

Sprouse Wagers & Phillips, 2012
Against WM accounts of extraction unacceptability

Sprouse Wagers & Phillips, 2012 divide the accounts into 3 camps:

Reductionist accounts; E.g., _working memory_ (WM) accounts (Kluender, 1998; Hofmeister & Sag, 2010)

Grammar-based accounts; E.g., _syntax accounts_ (Ross, 1967; Chomsky, 1973, 1986); _Syntax / semantics accounts_ (Goldberg, 1995; Ambridge & Goldberg, 2008)

"Grounded" accounts: WM accounts with _grammaticization_ (e.g., Fodor, 1978; Hawkins, 1999)

SWP: Simplest reductionist WM theory

Assumptions of the simplest WM theory of island effects
COMPONENT 1: There is a processing cost associated with the operations necessary to build long-distance WH-dependencies.
COMPONENT 2: There is a processing cost associated with the operations necessary to build the island structures.
LINKING HYPOTHESIS: Processing costs are reflected in acceptability judgments such that higher costs lead to lower acceptability
Prediction: A main effect of each component, **but no interaction**

Test: STRUCTURE × GAP POSITION
a. Who ___ thinks that John bought a car? NON-ISLAND | MATRIX
b. What do you think that John bought ___? NON-ISLAND | EMBEDDED
c. Who ___ wonders whether John bought a car? ISLAND | MATRIX
d. What do you wonder whether John bought ___? ISLAND | EMBEDDED

SWP: Simplest reductionist WM theory

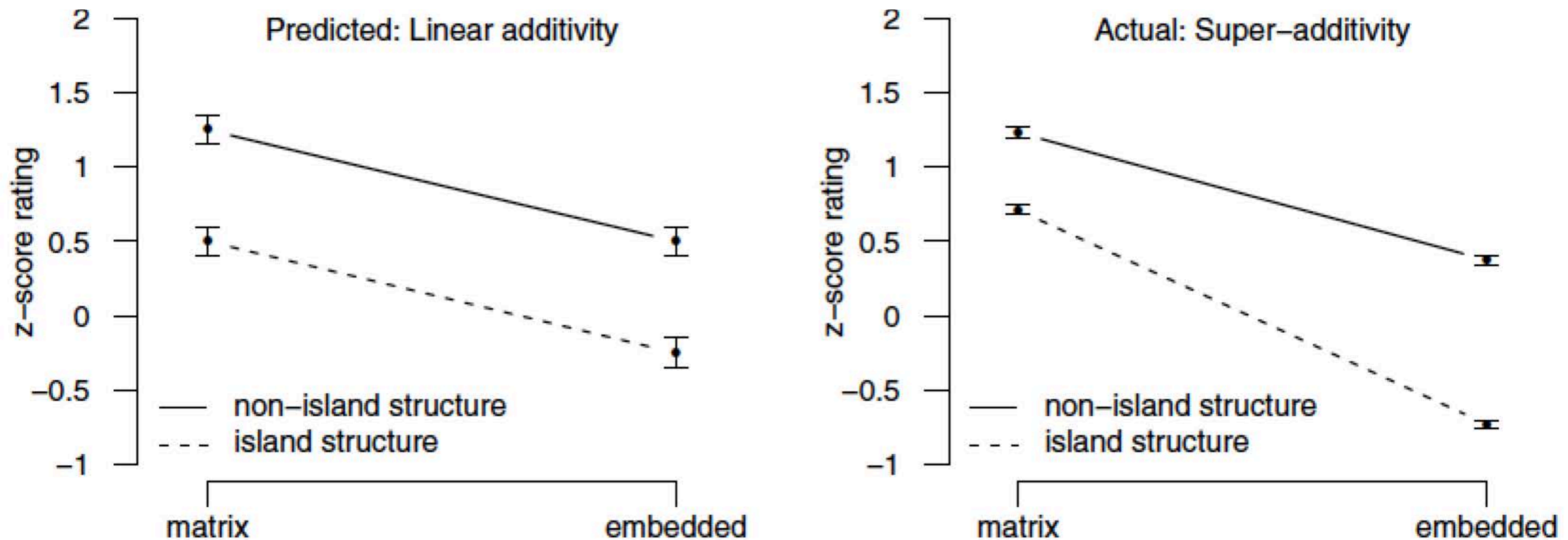This doesn't work out: There is also an interaction

FIGURE 1. The left panel represents the prediction of the simplest reductionist theory. The right panel represents the actual results of using the factorial definition of *whether* islands in 5 in an acceptability-judgment experiment (see §5 for details of the experiment).

SWP: Elaborated WM theory

Core assumptions of the resource-limitation theory (Kluender & Kutas 1993)
COMPONENT 1: There is a processing cost associated with the operations necessary to build long-distance WH-dependencies.
COMPONENT 2: There is a processing cost associated with the operations necessary to build the syntactic structures that we call island structures.
LINKING HYPOTHESIS: Processing costs are reflected in acceptability judgments such that higher costs lead to lower acceptability.

SIMULTANEITY: These two (sets of) processes must be deployed simultaneously in sentences that give rise to island effects.
LIMITED CAPACITY: There is a limited pool of processing resources available that must be shared by all simultaneous processes.
OVERLOAD: Additional unacceptability arises if the simultaneous processes necessary to complete the parse require more resources than are available in the limited pool.

SWP: Elaborated WM theory: Interaction is the difference of differences (DD) score

a. Who ___ thinks that John bought a car? NONISLAND | MATRIX
b. What do you think that John bought ___? NONISLAND | EMBEDDED
c. Who ___ wonders whether John bought a car? ISLAND | MATRIX
d. What do you wonder whether John bought ___? ISLAND | EMBEDDED

Calculating the difference of differences (DD) score with a sample set of mean ratings rating
a. D1 = (NONISLAND/EMBEDDED) − (ISLAND/EMBEDDED) (z-score units)
What do you think that John bought ___?   0.5
What do you wonder whether John bought ___?   −1.5
                                        difference: 2.0
b. D2 = (NONISLAND/MATRIX) − (ISLAND/MATRIX)
Who ___ thinks that John bought a car?   1.5
Who ___ wonders whether John bought a car?   0.7
                                        difference: 0.8
c. DD = D1 − D2 = 2.0 − 0.8 = 1.2

Critical evaluation of WM hypothesis:
Correlate DD score for an individual with their WM score on (a) a serial recall task (Expt 1) and (b) a serial recall task and an n-back task (Expt 2)

SWP: Predictions of two theories
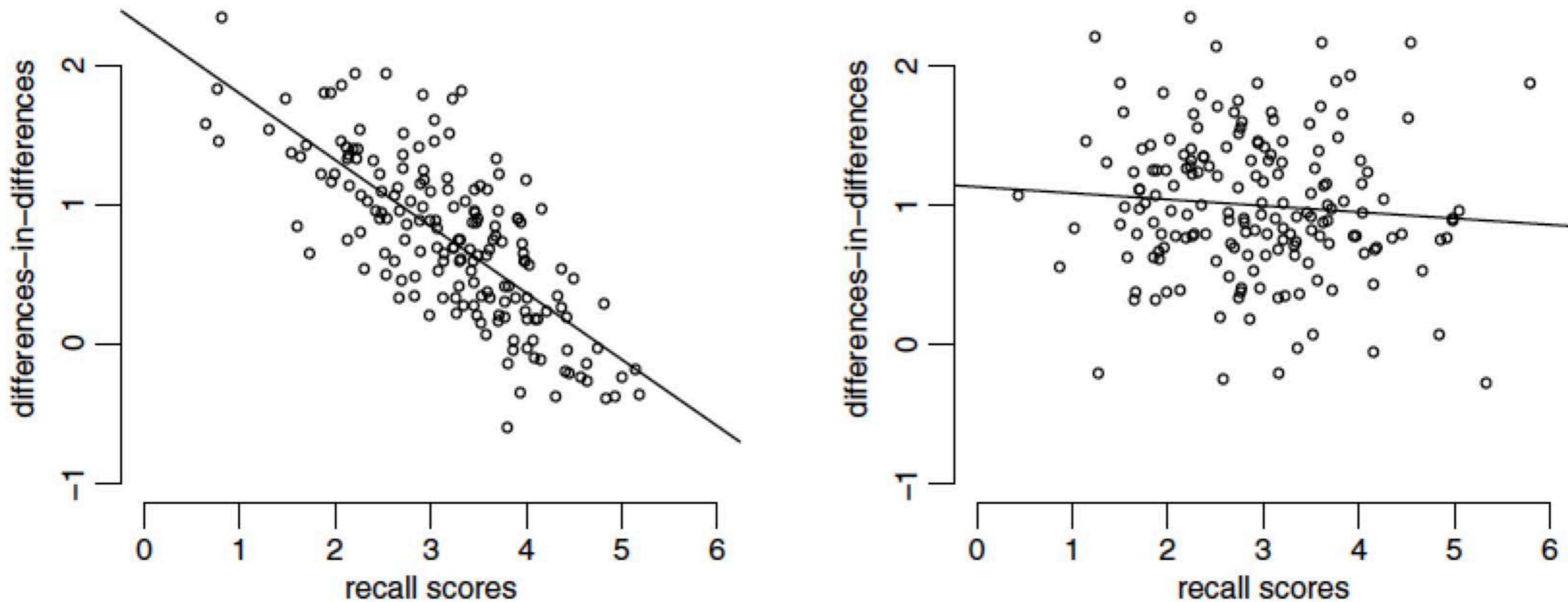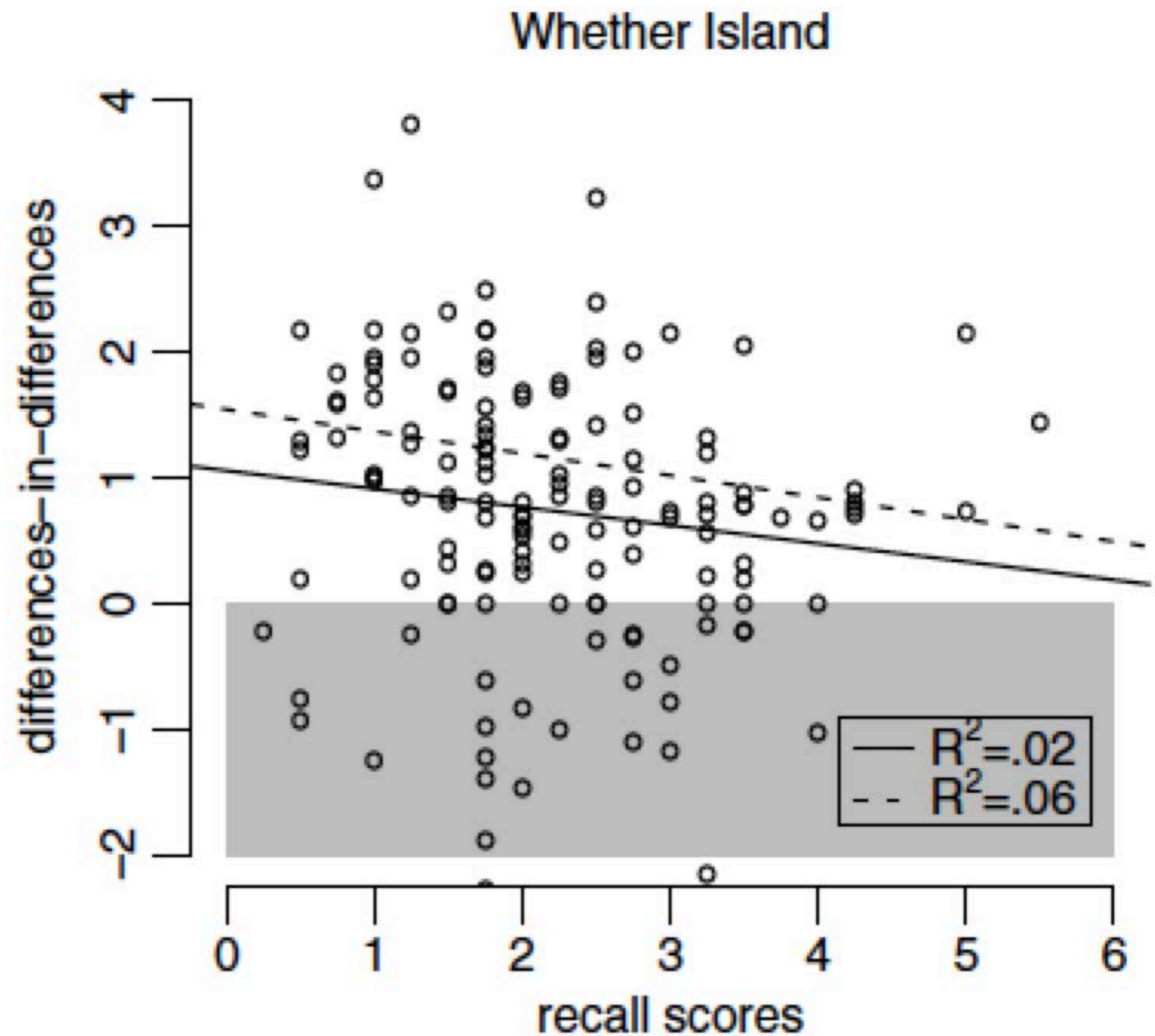
NOTE: these are **not real data!**

FIGURE 2. Predictions of the resource-limitation (left panel) and grammatical theories (right panel).

SWP Expt 1:
Interaction ratings vs. serial recall scores

Whether-island only:

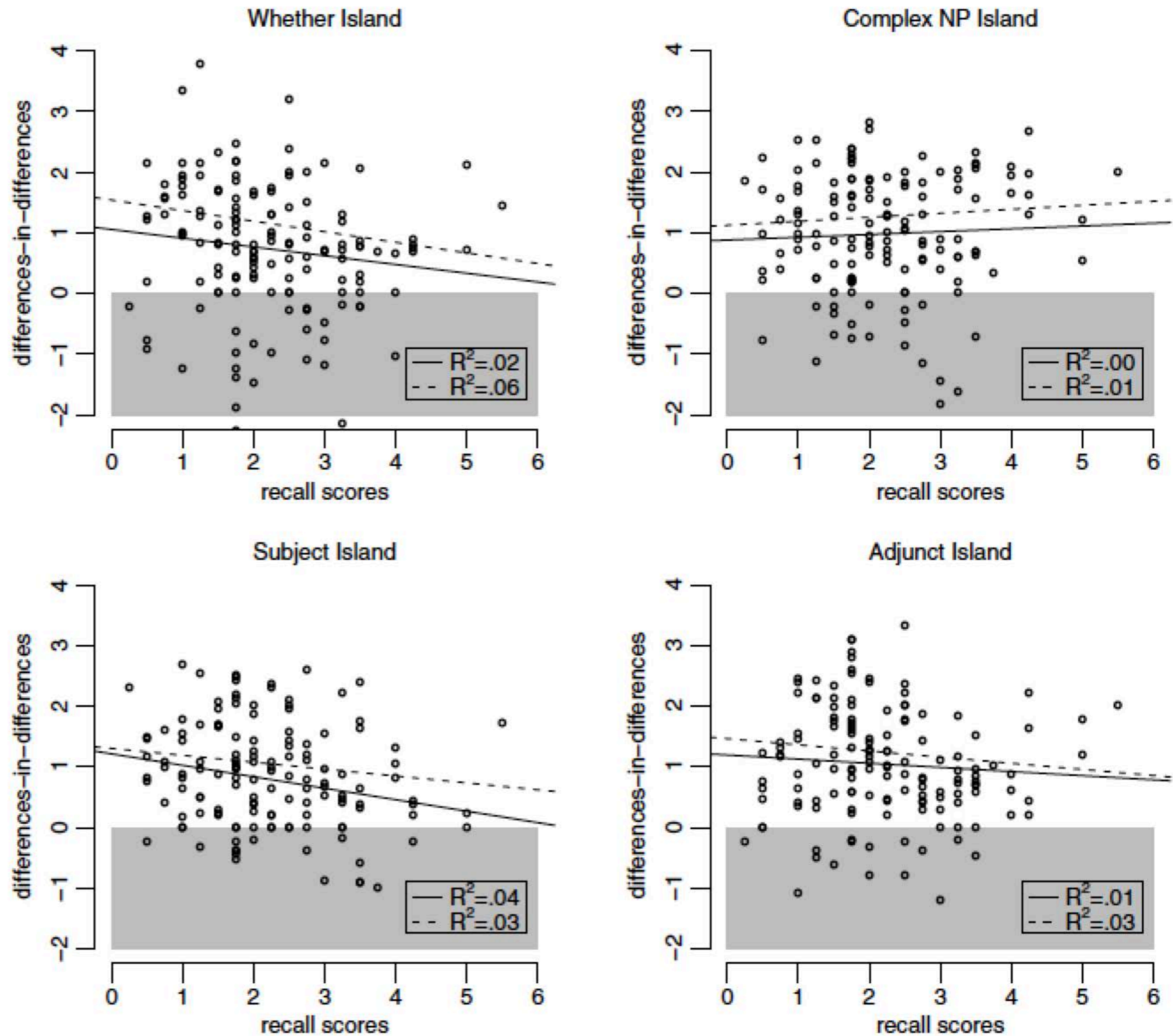# SWP Expt 1: Interaction ratings vs. serial recall scores

FIGURE 4. Experiment 1, differences-in-differences scores plotted as a function of serial-recall scores ($N = 142$). The solid line represents the line of best fit for all of the DD scores. The dashed line represents the line of best fit when DD scores below zero are removed from the analysis (shaded gray). Trend lines were fitted using a least-squares procedure. Adjusted R^2 for each trend line is reported in the legend.

SWP Expt 2:
Interaction ratings vs. 3-back scores

Whether-island only:
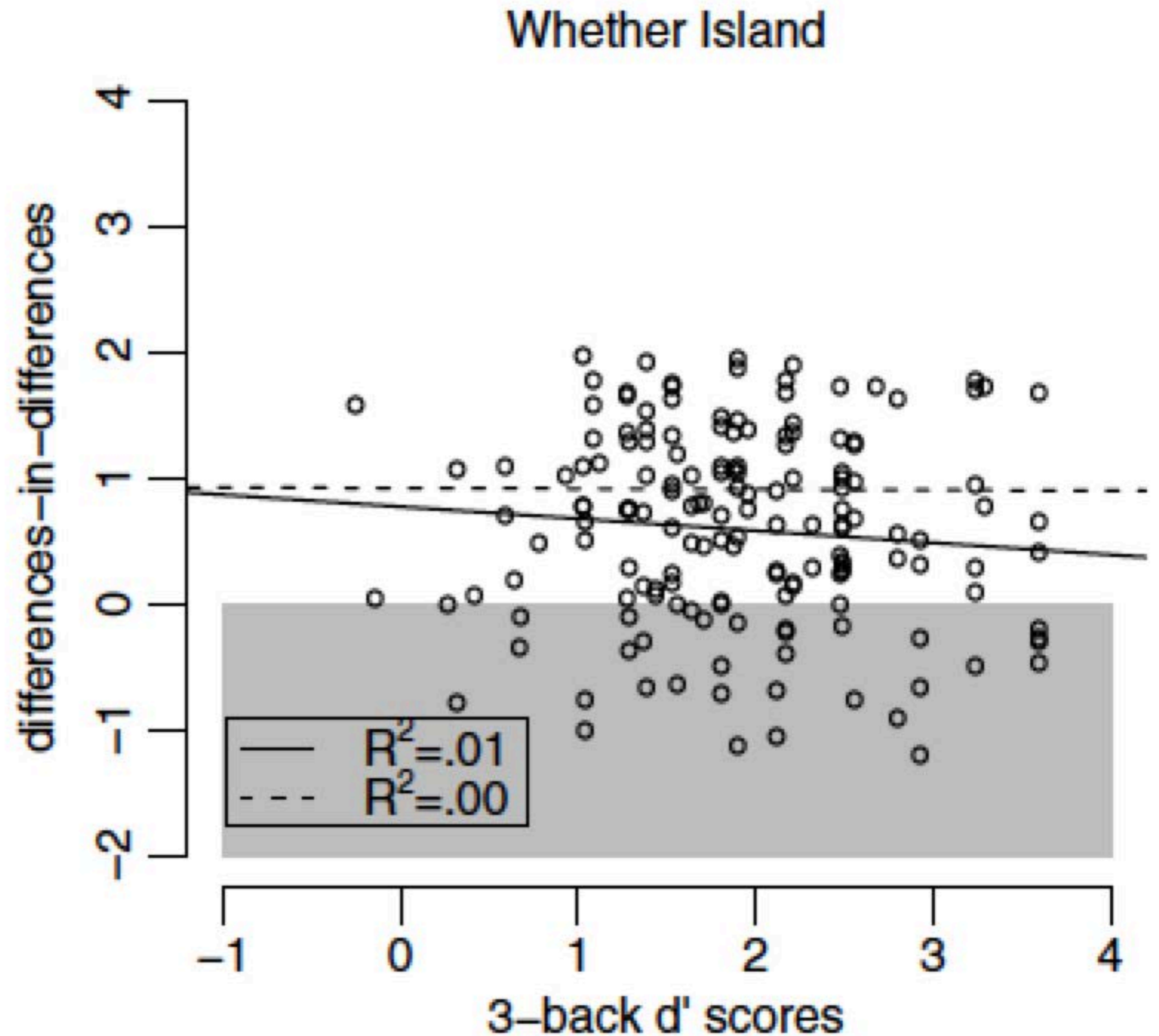
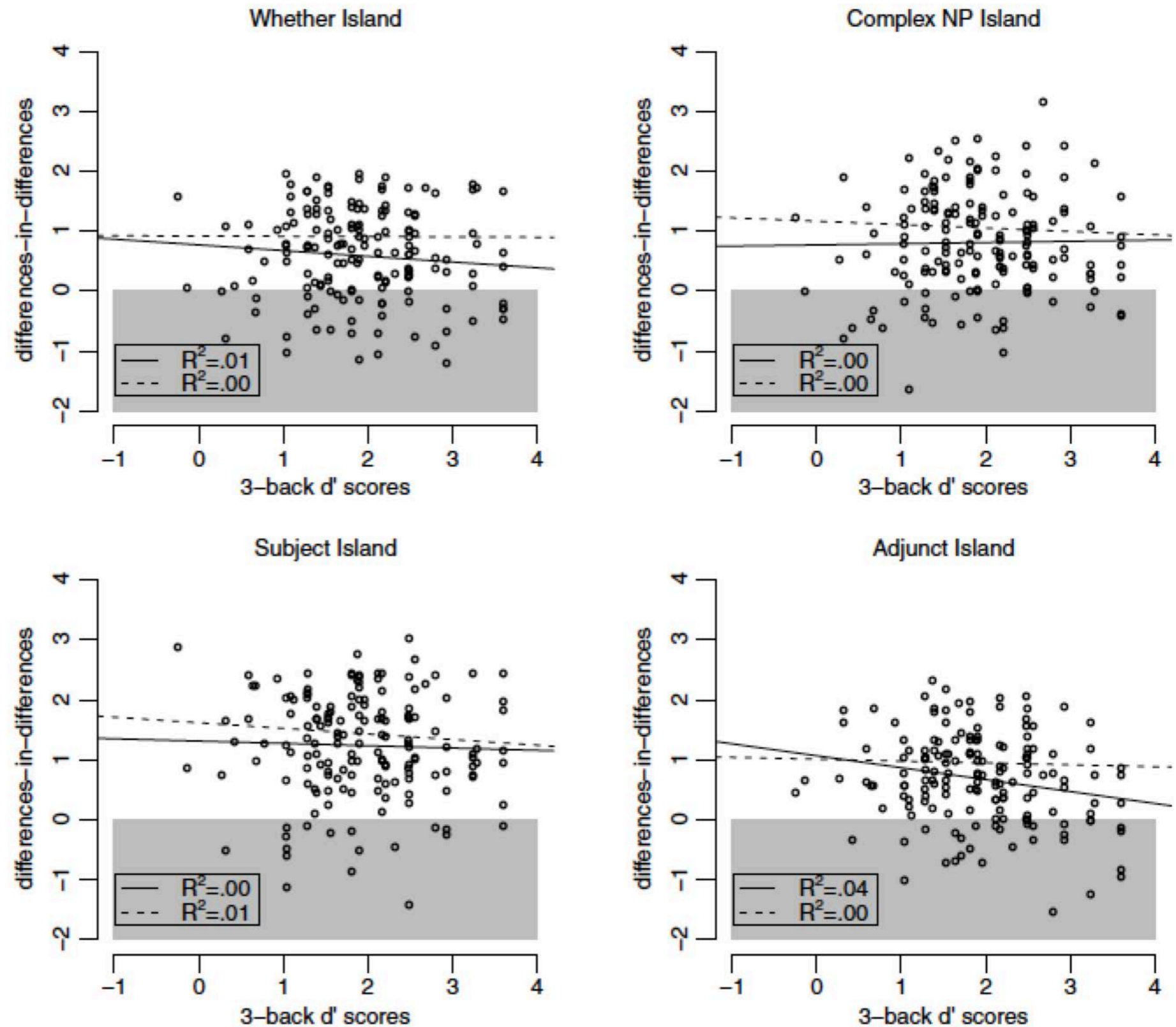# SWP Expt 2: Interaction ratings vs. 3-back scores

FIGURE 8. Experiment 2, differences-in-differences plotted as a function of three-back scores ($N = 161$). The solid line represents the line of best fit for all of the DD scores. The dashed line represents the line of best fit when DD scores below zero are removed from the analysis (shaded gray). Trend lines were fitted using a least-squares procedure.

Open question in the syntax literature:
The **domain-specificity / generality**
of extraction constraints

Sprouse, Wagers & Phillips (2012, Language):
Correlational study, using individual difference measures from two kinds of tasks:
    Acceptability ratings on extractions from islands
    A working memory measure (two tasks: n-back; serial recall)

Result: No correlation whatsoever.

"We believe that the results of the experiments presented in this article provide strong support for grammatical theories of island effects because we can find no evidence of a relationship between processing resource capacity and island effects."

BUT: Absence of evidence is not evidence of absence

Null correlation: how to interpret?
Would we be surprised if we found no correlation between Island ratings and participant height? Probably not.
How is this comparison similar / different?

(Minor) Issues with Sprouse, Wagers & Phillips (2012)

1. Very little data were collected for each participant (2 and 4 trials per condition in Expts 1 and 2, respectively).
(maybe not a critical problem, but worrisome!)

2. No plausibility control:
Island: *What do you sneeze if the dog owner leaves open at night?*
SWP's long-distance extraction control for this example:
What do you hope that the dog owner will leave open at night?

How do we know that the effects aren't due to plausibility differences?

Major Issue with Sprouse, Wagers & Phillips (2012)

No control to show that the complexity of materials that are thought to be complex due to WM considerations correlate with general WM measures

Maybe there is a particular pool of resources just for compositional language comprehension (Caplan & Waters, 1999)

New design (Gibson & Scontras, 2013)

New design: 3 tasks:

Task 1: Sentence completions of WM-complex materials: a general intelligence task, which correlates with IQ tasks (which also correlate with WM tasks)

Task 2: Sentence acceptability of WM-complex materials

Task 3: Sentence acceptability of islands.

Examine the relationship between task 1 and task 2:

If they correlate (as SWP implicitly assume), then we can evaluate task 1's correlation with task 3 (SWP's question)

If not, then SWP's evaluation may not tell us anything about the relationship between WM and island acceptability.

New design (Gibson & Scontras, 2013)

60 participants on M Turk (run twice, with similar results in each run)

(1) a language / fluid intelligence measure that has been shown to correlate strongly with fluid intelligence: complex sentence completion (Gibson, Fedorenko & Mahowald, 2014); (WM correlates highly with fluid intelligence)

(2) a rating study consisting of
nested vs. non-nested sentences, a contrast thought to be related to WM demands; and
adjunct and NP islands, similar to SWP's, but with 8 trials per condition per participant, and with plausibility-matched control conditions.

Participants also answered 2 comprehension questions about each sentence.

Result 1

Complex sentence completion:
Replicating previous results, we found stable completion rates within individuals but varying across individuals. (cf. Gibson & Fedorenko, 2012; Gibson & Thomas, 1999; Frazier, 1978; Christiansen & Chater, 2001; Vasishth et al., 2008)

Lots of incomplete completions: ~50%

Sample completions:
The reporter who the professor who …
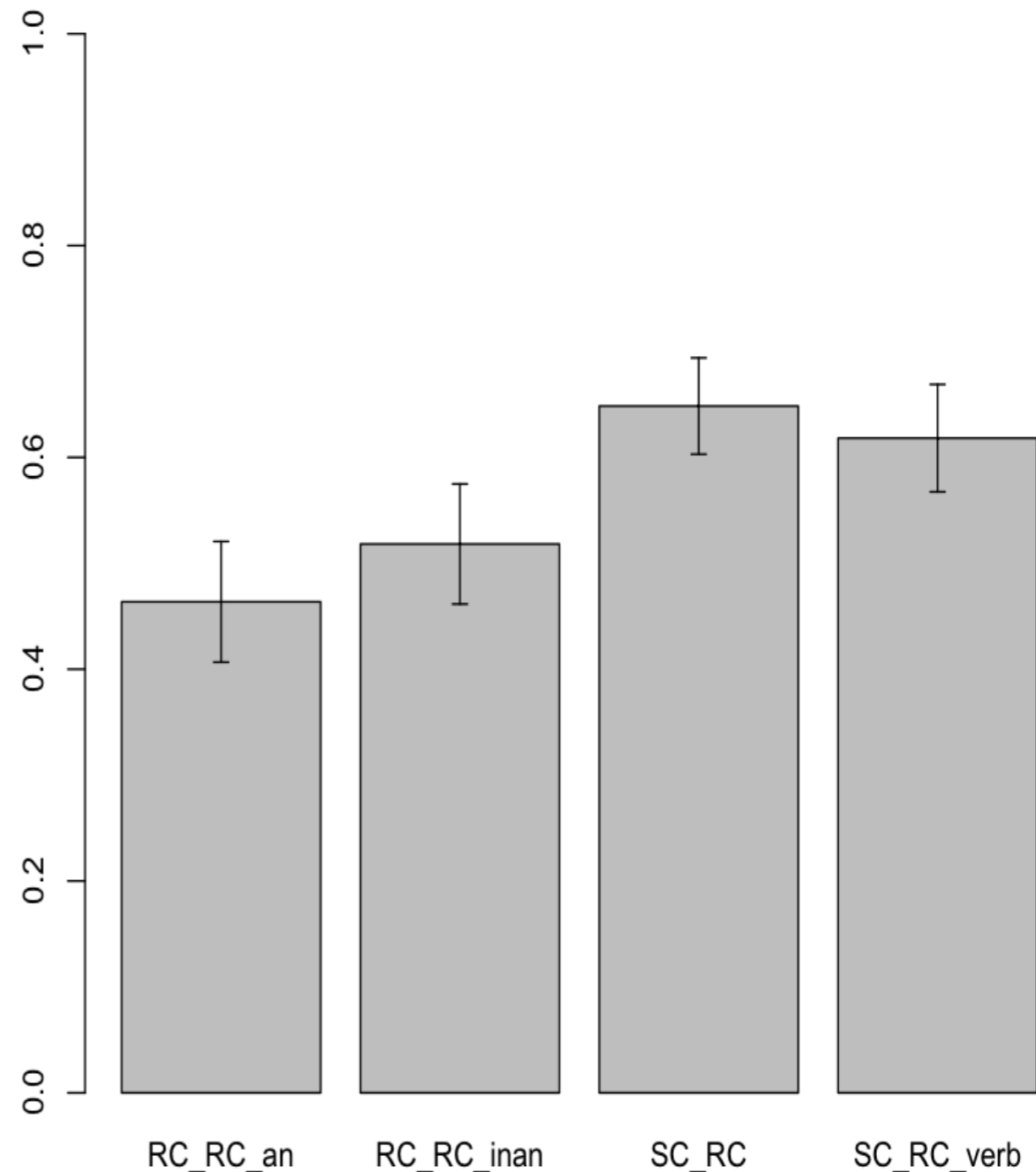
3 VPs:
…taught English called was late to their appointment.
…taught physics, was friends with asked me some questions.

2 VPs:
…was dating him, thought that they were out of line.
…was intelligent, didn't like each other.

Result 2

Complex sentence completion:
This measure correlated reliably with the ability to answer comprehension questions in the rating study (r=.510; p<.001).

The reporter who the professor who …

High WM / IQ participant typically completes with 3 VPs:
…taught English called was late to their appointment.
…taught physics, was friends with asked me some questions.

Low WM / IQ participant typically completes with 2 VPs:
…was intelligent, didn't like each other.
…was dating him, thought that they were out of line.

Result 3

Reliable differences in ratings were observed for each contrast:

nested vs. non-nested:
2.12 vs. 3.17 out of 5; p<.001
The building which the architect who the contractor met designed was commissioned by the city.
The contractor met the architect who designed the building which was commissioned by the city.

adjunct-island vs. pronoun-dependency:
1.72 vs. 2.22; p<.001
The neighbor wonders what you sneeze if the dog owner leaves open at night.
The neighbor wonders what the thing is, such that you sneeze if the dog owner leaves it open at night.

NP-island vs. pronoun-dependency:
1.89 vs. 2.27; p<.001
Joe wondered what the chef heard the statement that Jeff baked.
Joe wondered what the thing was, such that the chef heard the statement that Jeff baked it.

Result 4

Critically, we found no correlation between our IQ / WM measure and the nested vs. non-nested difference score in the ratings (r =.052), suggesting that SWP's assumption about acceptability ratings reflecting WM demands is not warranted.
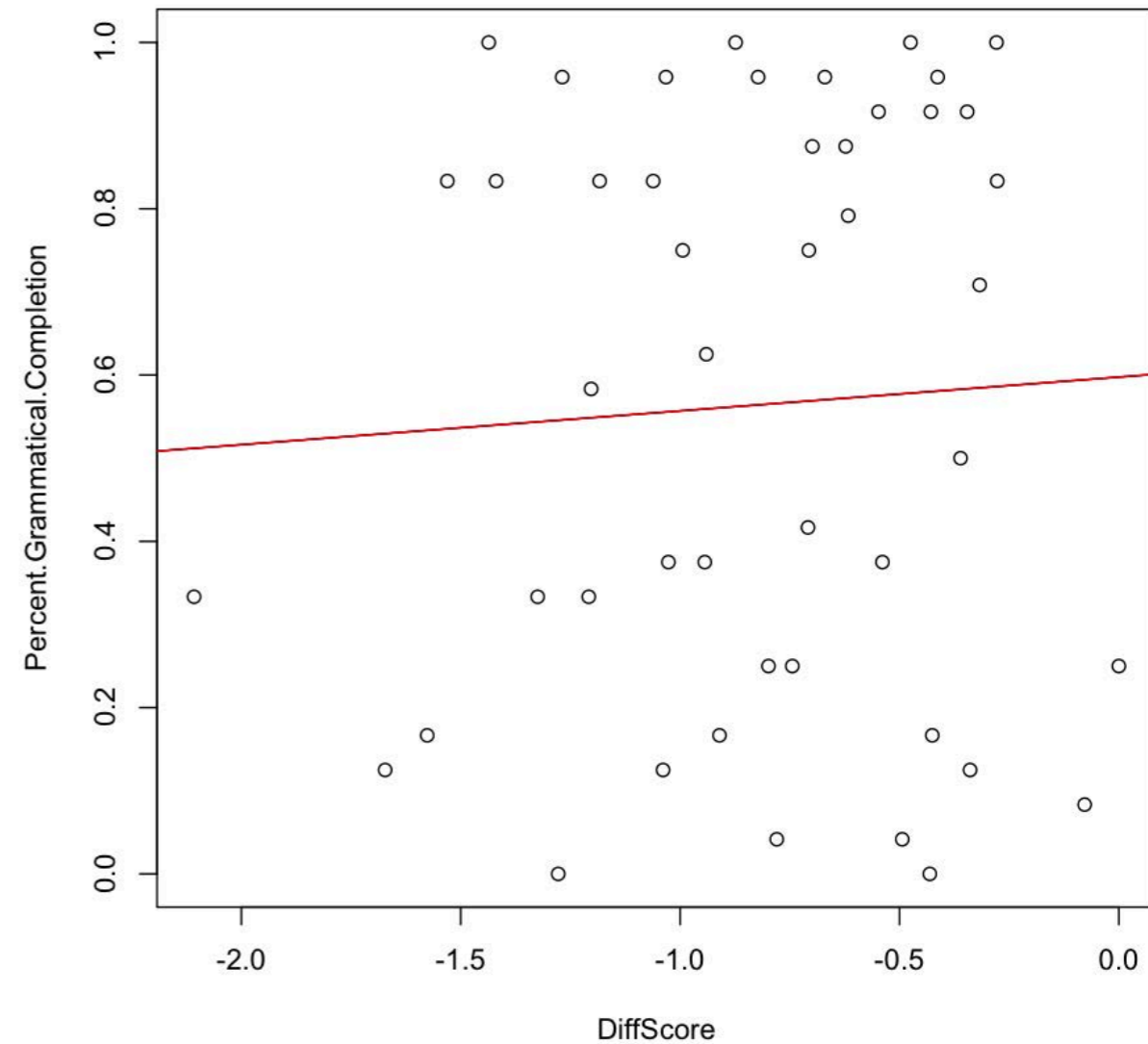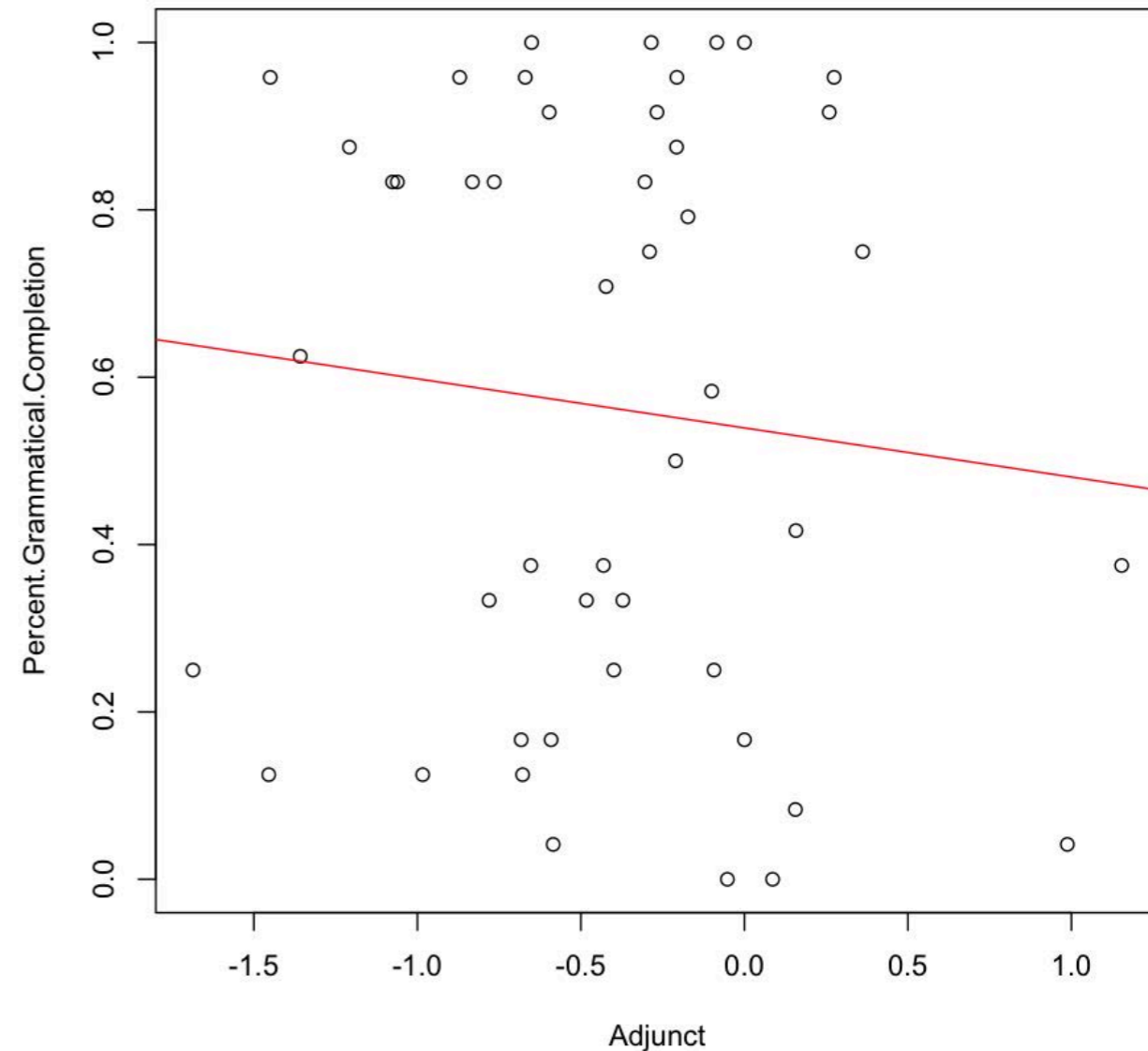
Result 5

Replicating SWP, we found no
correlation between acceptability ratings
for island conditions and our IQ / WM
measure.

r = -.097 for adjunct-islands
r = -.089 for NP-islands

But this is now unsurprising, given that
we also find no correlation with our
WM measure for judgments based on
nested vs. non-nested sentences

Gibson & Scontras (2013):
Summary & Conclusions

There is no correlation between our IQ / WM measure and the nested vs. non-nested difference score in the ratings (r =.052).

Maybe there is a particular pool of resources just for compositional language comprehension (Caplan & Waters, 1999)

7 More Possible Projects

11. Discourse Structure and Constraints on English Relative Clauses (Ambridge & Goldberg, 2008)

The acceptability of certain complex relative clauses in English depends on the verbs present. For example, the question "What did John say that Mary stole from the store?" seems well-formed, but the questions "What did John realize that Mary stole from the store?" and "What did John mumble that Mary stole from the store?" are typically rated as less acceptable.

Ambridge & Goldberg (2008) explore the idea that the unacceptability of the latter two sentences arises because of constraints on discourse structure. In the acceptable sentence, the verb "say" makes it seem that the important information in the sentence is contained in the following clause. In contrast, in the unacceptable sentences, the verbs "realized" and "mumbled" make the following clause seem like background information. They perform ratings studies showing that the degree to which verbs imply that the following clause is background information strongly predicts the acceptability of relative clauses formed from those clauses.

While the results of this study are strong, they have not been extended to other grammatical phenomena in English or other languages.

12. Individual differences in quantifier scope

The sentence "Every student didn't pass the test" is ambiguous between a meaning where none of the students passed the test, and one in which some but not all of the students passed the test. While one might think that the context could disambiguate the meaning, it appears that some people have trouble overcoming a (possibly primed) meaning bias.

Context biased to "none": *Six little-known singers were each getting ready to release their second albums. Their record labels expected that none of the unknown singers' new albums would sell one million copies. As expected, every singer's album didn't sell one million copies.*

Context biased to "some but not all": *Three famous singers and three little-known singers were each getting ready to release their second albums. Their record labels expected that all of the famous singers' new albums would sell one million copies, but that none of unknown singers' new albums would sell one million copies. As expected, every singer's album didn't sell one million copies.*

How many singers' albums sold one million copies?   1) All   2) Some   3)None

Replicate the lack of sensitivity to context that some participants obtain. Extend this in some interesting way, possibly to try to figure out what kind of priming this is, and how you could make it disappear.

7 More Possible Projects

13. Knowledge and Implicature: Context and the interpretation of quantifiers

Goodman & Stuhlmuller (2013) observed that how one interprets "some" depends on the context. In a context where Laura has checked all vs some of the letters to her company, when she says "some" it changes the likely meaning:

Letters to Laura's company almost always have checks inside. Today Laura needs to find out whether 3 of the letters have checks inside. Laura tells you on the phone: I have now looked at 2 (or all 3) of the 3 letters, and given what I saw, I can tell you that some of today's letters have checks inside.

Estimate (in percent) how likely it is that all 3 of the letters have checks inside.

In a context where Laura has looked at all 3, then "some" usually means some but not all. But if she has only looked at 2 of the 3 letters, then this means some and possibly all.

Replicate this, and extend it in some interesting way.

7 More Possible Projects

14. Sedivy (2003) color naming

Sedivy (2003) shows that speakers use color adjectives only when they are typically informative, and that listeners reason on-line about speakers' intentions in using color adjectives by assuming that speakers are trying to be informative. Her paper shows that people use color adjectives to modify nouns referring to objects (such as "the blue cup") when the object can conceivably come in many colors, even when the color adjective is not necessary to disambiguate objects in immediate context. For objects that do not vary freely in color (such as a yellow banana), people do not use color adjectives often. On the comprehension side, in a visual world paradigm, when subjects hear the word "yellow" they are likely to look AWAY from a yellow banana, since they do not think a speaker would use the color adjective for that object. Project: color labeling over M Turk.

15. People are more rational in L2 than in L1.

Keysar et al (2012) observed that using a foreign language reduces decision-making biases. Four experiments show that the framing effect disappears when choices are presented in a foreign tongue. Whereas people were risk averse for gains and risk seeking for losses when choices were presented in their native tongue, they were not influenced by this framing manipulation in a foreign language. Two additional experiments showed that using a foreign language reduces loss aversion, increasing the acceptance of both hypothetical and real bets with positive expected value.

Costa et al (2014): extension to morality judgments.

9.59J/24.905J Lab in Psycholinguistics
Spring 2017