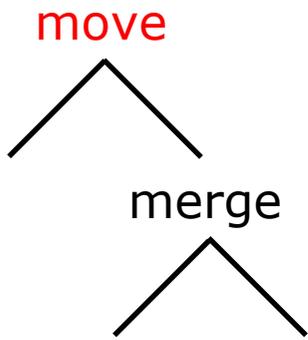


A large-scale test of "d-linking" as a diagnostic of the strong/weak island distinction in English



Jon Sprouse & Sandra Villata
NYU Abu Dhabi

Penn Locality Workshop
04.01.23



A typical description of the strong/weak island distinction

strong islands: Strong islands block all dependencies.

weak islands: Weak islands block some dependencies, and allow others through.

Adjunct island

simple: * **What** did the waiter frown because the chef burned ___?

complex: * **Which dish** did the waiter frown because the chef burned ___?

Whether island

simple: * **What** did the waiter wonder whether the chef burned ___?

complex: * **Which dish** did the waiter wonder whether the chef burned ___?

Why is the strong/weak distinction interesting?

This is an **interaction** of island type and dependency type.

	strong islands	weak islands
simple wh	*	*
complex wh	*	✓

This means that it is **not** just a **main effect of complex-wh**, which would be comparatively easy to explain:

~~Complex wh-phrases are base generated (or otherwise island-insensitive).~~

~~Complex wh-phrases convey a retrieval advantage during processing that increases acceptability (or eliminates the island).~~

Instead, it might be telling us something deep **about the nature of different island types**:

It could indicate a difference in the source of different island types.

It could indicate a difference in the process of acquisition of island types.

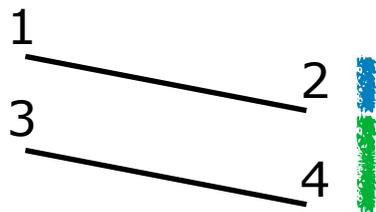
Question 1: Partial vs total amelioration

The theoretical literature does not agree on the judgments for the d-linked sentences. And the experimental literature shows only partial amelioration:

[* ? ✓] **Which dish** did the waiter wonder whether the chef burned ___?

	<u>dependency</u>	<u>structure</u>
1. Who ___ thinks [that Mary invented the app]?	short	non-island
2. What do you think [that Mary invented ___]?	long	non-island
3. Who ___ wonders [whether Mary invented the app]?	short	island
4. What do you wonder [whether Mary invented ___]?	long	island

no island effect



island effect



Island effects are defined as a decrease in acceptability over and above the decreases for the dependency and the island structure itself.

Expanding the design to complex wh

We can use the same design for complex wh-arguments. This is fundamentally a 2x2x2, but we will conceptualize it as comparing two 2x2s to each other, and asking **how the size of their interactions compare to each other**:

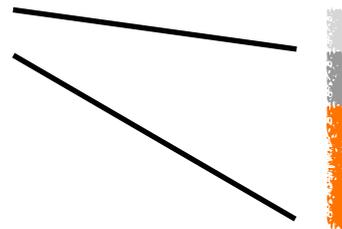
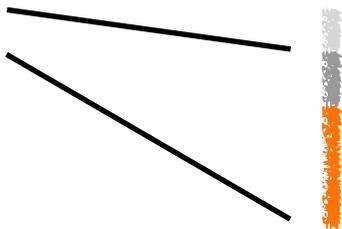
	<u>dep</u>	<u>structure</u>
1. Who ___ thinks [that Mary invented the app]?	short	non-island
2. What do you think [that Mary invented ___]?	long	non-island
3. Who ___ wonders [whether Mary invented the app]?	short	island
4. What do you wonder [whether Mary invented ___]?	long	island
1. Which coder ___ thinks [that Mary invented the app]?	short	non-island
2. Which app do you think [that Mary invented ___]?	long	non-island
3. Which coder ___ wonders [whether Mary invented the app]?	short	island
4. Which app do you wonder [whether Mary invented ___]?	long	island

Question 1: Partial vs total amelioration

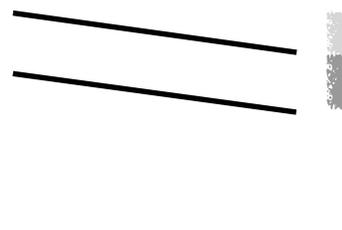
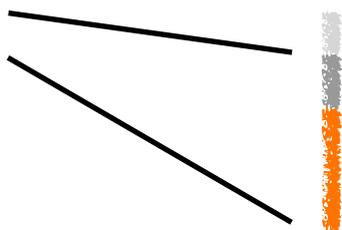
We can use the same design for complex wh-arguments. This is fundamentally a 2x2x2, but we will conceptualize it as comparing two 2x2s to each other, and asking **how the size of their interactions compare to each other**:

simple wh

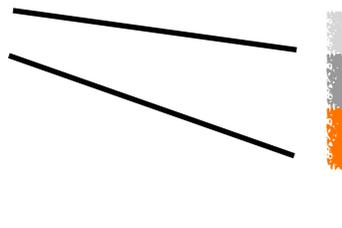
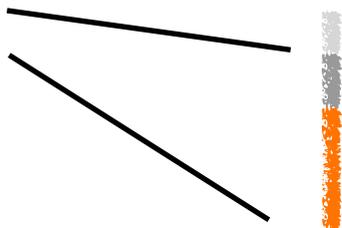
complex wh



This is the **no amelioration** pattern we expect from **strong islands**. Both wh-types show the same size interaction.



This is the **total amelioration** pattern that is typically discussed for **weak islands**. The complex-wh shows no island effect.



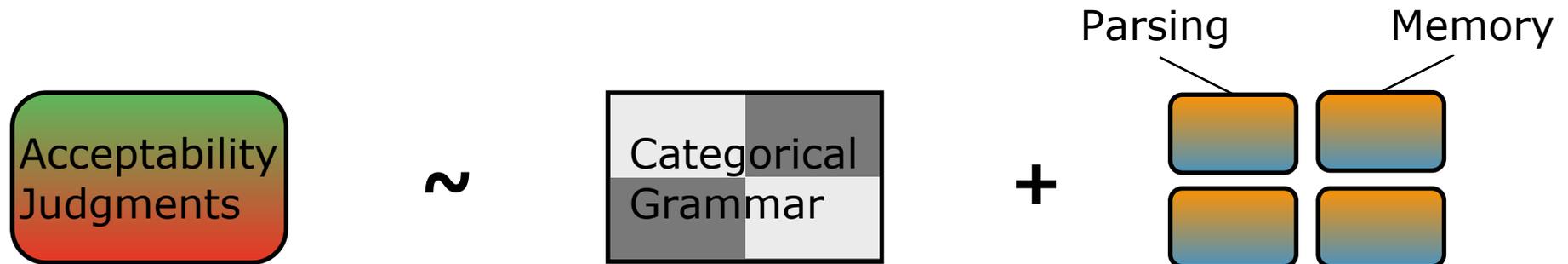
This is a **partial amelioration** pattern. The complex-wh shows an island effect, but it is **smaller**.

The **partial amelioration** challenge



For binary categorical grammars, constraints are either satisfied or violated. Strings/trees are either grammatical or ungrammatical. There should be **one effect size**.

Differences in effect size present a challenge for binary categorical grammars. The typical approach to explaining effect size difference is to point to extra-grammatical properties, like sentence processing:



But lexical and structural matching in the 2x2x2 design eliminates the obvious candidates for this. So we need a **grammatical explanation** for the effect size.

simple: * **What** did you wonder whether Lisa invented ___?

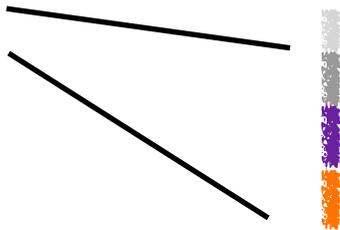
complex: ? **Which app** did you wonder whether Lisa invented ___?

The **partial amelioration** challenge

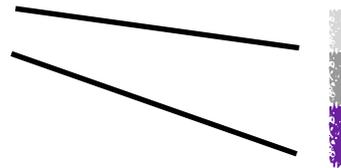


For binary categorical grammars, the only option we can see is to postulate that there are two constraint violations for simple-wh, with one constraint satisfied by complex-wh.

simple wh



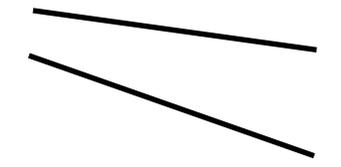
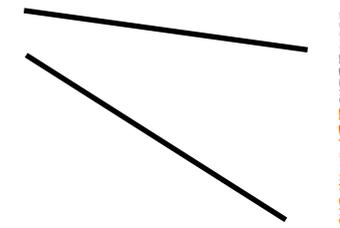
complex wh



The question here is **what this second constraint could be**, and why complex-wh satisfies it.



Another possibility is that the grammar is continuous: constraints have continuous values (like weights, or probabilities) that lead strings/trees to have continuous values.

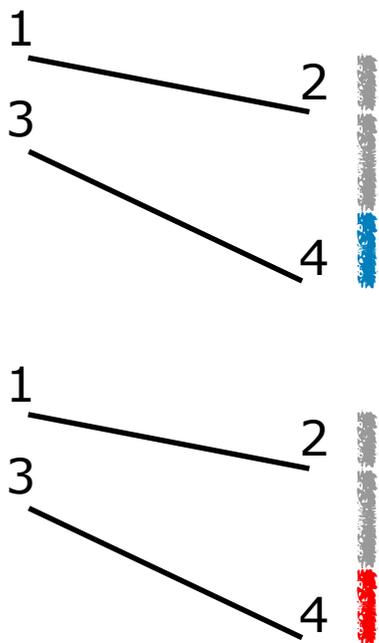


The question here is **how these weights are acquired**. It is not clear what sort of evidence children could receive to set constraint weights.

Therefore our empirical focus will be on estimating the island effect sizes

First, for each island type and wh-type, we recruited **approximately 200 participants** using CloudResearch's pre-screened participant list for Amazon Mechanical Turk. Based on previous sensitivity studies, this gives us good (>80%) power for even very small effect sizes ($d=.05$). For the 28 island types, this is **over 11,200 participants**.

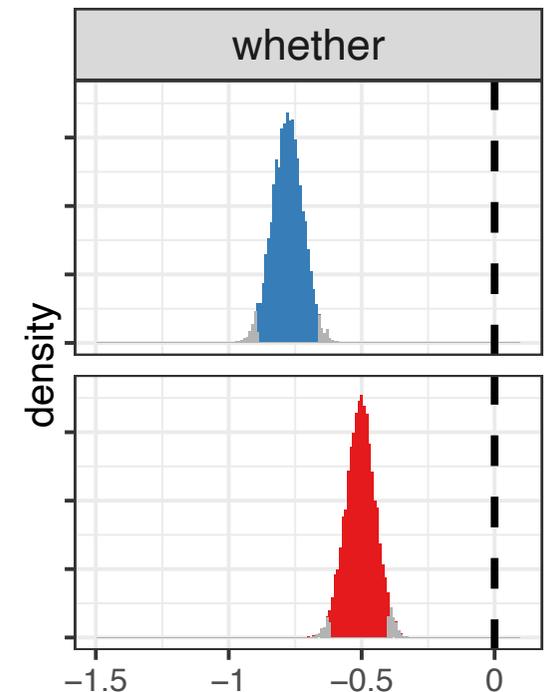
We then followed the Bayesian regression approach in Kruschke 2014 adapted by Kurz 2021 using the **brms** package (Bürkner 2017, 2018, 2021) to estimate a posterior distribution for the size of the interaction term in z-scores (the island effect), along with **95% credible intervals**.



We report results assuming uninformative priors and a Gaussian likelihood function.

Bayesian linear modeling

And our Bayesian hypothesis test is whether the 95% CrIs overlap.



We also focus on wh-arguments

Though we tend to talk about two classes of islands - strong/weak - the number of classes is an empirical question based on the number of dependency types that behave distinctly. In principle, selectivity could be very fine grained.

For the current study we will focus only on simple vs complex wh-arguments. This means that we can distinguish three classes of structures: those that block both dependencies (**strong**), those that block only simple (**weak**), and everything else (which will look like **not-an-island** in our study).

In this table, I have added wh-adjuncts to illustrate how our narrow focus will inevitably group together possible distinctions. Wh-adjuncts are central so semantic approaches to islands, so this distinction matters!

		strong	weak	weaker	weakest	not island
wh-arg	complex	*	✓	✓	✓	✓
	simple	*	*	✓	✓	✓
wh-adj	complex	*	*	*	✓	✓
	simple	*	*	*	*	✓

Question 2: Classifying islands

There is not as much certainty as we might like about the classification of different island types. So we decided to test 7 island types, and several island “tokens” within each type, so that we can (i) classify each island token, and (ii) look for uniformity within island types.

Factive islands

communicative
discovery
doxastic
emotive

Negative islands

n't
not

Wh-islands

whether
if
what
who
why
which
what NP

Noun Complements

make the claim
believe the claim
believe the rumor
hear the claim
hear the rumor

Subject islands

Adjunct def (by, for, with, from)
Adjunct indef (by, for, with, from)
Complement def (personal of)
Complement indef (personal of)
Complement def (of)

Relative Clause

that
who

Adjunct islands

causal (because)
conditional (if)
temporal (after)

Materials: Negative and Factive islands

Negative islands appear straightforward in principle - there is **not** and there is **contraction**. But negative questions may be odd for some speakers without context, so we will perform some subset analyses to account for this by only including participants who accept negative questions in a condition that does not have the island violation (short/island)::

n't: **Which dish** **didn't** the waiter think that the chef burned ___?

not: **Which dish** did the waiter **not** think that the chef burned ___?

For factives, we test four types of predicates discussed in Karttunen 2016:

communicative: **Which dish** did the waiter **acknowledge** that the chef burned ___?

discovery: **Which dish** did the waiter **realize** that the chef burned ___?

doxastic: **Which dish** did the waiter **forget** that chef burned ___?

emotive: **Which dish** was the waiter **sad** that the chef burned ___?

Materials: Wh-island sub-types

Wh-islands are central to theories of weak islands, so we wanted to test as many as possible. But, we only tested **tensed** embedded clauses:

- whether:** Which dish did the waiter wonder **whether** the chef burned ___?
- if:** Which dish did the waiter wonder **if** the chef burned ___?
- why:** Which dish did the waiter wonder **why** the chef burned ___?

For wh-arguments, **who** creates a double-name penalty in the short/island condition (who... who). **What** eliminates this but requires **psych verbs**:

- who:** Which dish did the waiter wonder **who** burned ___?
- what:** Which chef did the waiter wonder **what** **angered** ___?

Similarly, **which** could cause a double-name penalty in the long/island condition (2x which). So we also ran **what NP** with **psych verbs**:

- which:** Which dish did the waiter wonder **which chef** burned ___?
- what NP:** Which chef did the waiter wonder **what order** **angered** ___?

Materials: Noun Complement islands

For noun complements, we test the special light verb construction make the claim, because some have claimed it is not an island at all (e.g., Ross 1967):

make/claim: Which dish did the waiter make the claim that the chef burned ___?

We also test two verbs and two nouns to try to further tease apart the role of structure (all noun complements) and the role of lexical semantics, or even just collocation issues (like “hear the rumor”):

believe/claim: Which dish did the waiter believe the claim that the chef burned ___?

believe/rumor: Which dish did the waiter believe the rumor that the chef burned ___?

hear/claim: Which dish did the waiter hear the claim that the chef burned ___?

hear/rumor: Which dish did the waiter hear the rumor that the chef burned ___?

Materials: Subject islands

There are a number of different theories of noun types and how they may impact extraction (e.g., Davies and Dubinsky 2003) independently of subjects. We tried to target three dimensions that might matter: structural relationship, semantic type (result/concrete), and definiteness.

Adjunct prepositions (by, for, with, from), result, indefinite/definite

adj/def: Which doctor did the nurse think the report by ___ modified the procedure?

adj/indef: Which doctor did the nurse think a report by ___ modified the procedure?

Complement preposition (of), concrete (personal relationships), not unique (so context may be an issue), indefinite/definite:

comp/def: Which doctor did the nurse think the student of ___ modified the procedure?

comp/indef: Which doctor did the nurse think a student of ___ modified the procedure?

Complement preposition (of), result, not unique, only definite so far (because indefinite is not felicitous, so we need to rethink the construction).

comp/def: Which doctor did the nurse think the insight of ___ improved the procedure?

Materials: Relative Clause and Adjunct islands

For relative clauses, we tested two relative pronouns (that, who). But who creates a double name penalty in the short conditions with simple-wh.

that: Which dish did the waiter blame the chef **that** overcooked ___?

who: Which dish did the waiter blame the chef **who** overcooked ___?

For adjuncts, we tested three types: causal (because), conditional (if), and temporal (after):

causal: Which dish did the waiter frown **because** the chef burned ___?

conditional: Which dish did the waiter frown **if** the chef burned ___?

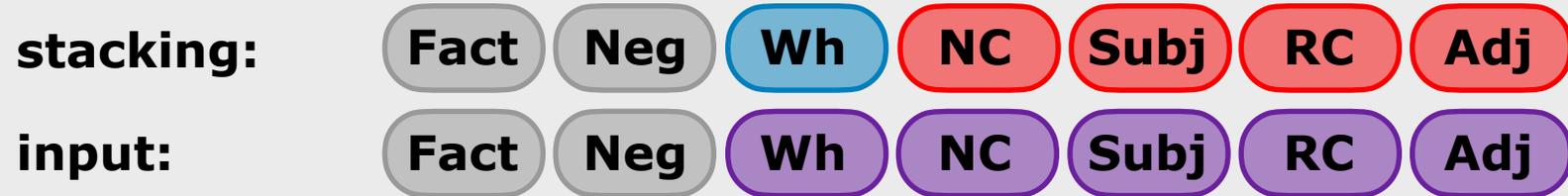
temporal: Which dish did the waiter frown **after** the chef burned ___?

Question 3: Theories of islands

It is very unlikely that there is a single theory that will explain all island effects. Instead there are several active in the literature. But, they vary quite a bit with respect to the strong/weak distinction.

- Phases:** There are certain syntactic domains that are **impenetrable** (Chomsky 2000 et seq.). This is a theory of **strong islands**.
- Relativized Minimality:** An item with the same feature(s) cannot **intervene** between the head and tail of the dependency (Rizzi 1990, 2004). This is explicitly an account of **weak islands**.
- Information Structure:** The IS properties of the dependency must match the IS properties of the constituent that they are extracted from. (Erteschik-Shir 1973 et seq., Goldberg 2006, Abeille et al. 2020). I would say this was intended as **strong islands**.
- Semantic Approaches:** Weak islands arise because of an incompatibility between the semantics of wh-questions and “island” clause types (Szabolcsi and Zwarts 1993, 1997, Abrusán 2014). These are explicitly accounts of **weak islands**, but focused on a wh-argument/wh-adjunct asymmetry.

Impenetrability theories (e.g., phases)



Though there are different approaches to phase-based locality, they all share the core idea that there are certain syntactic domains that are **impenetrable**. (Chomsky 2000 and many many others.)

Strong: Phase theory is only about extraction domains. It makes no reference to the head of the dependency. In that sense, **it captures strong islands by default**, but also predicts **all islands to be strong**.

Weak: There are at least two options for weak islands. One would be for simple and complex-wh dependencies to be **acquired separately** such that **differences in the input** drive the grammar (but this challenges the architectural assumptions of phases).

Another would be **stacking two violations**, like Superiority/MLC/ Attract-Closest and Phase Impenetrability (e.g., this would make wh-islands weak).

(featural) Relativized Minimality

fRM:

Fact

Neg

Wh

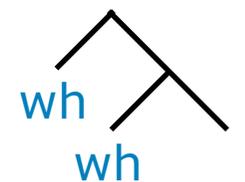
NC

Subj

RC

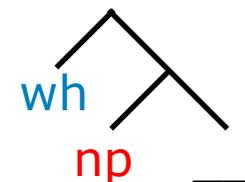
Adj

An item with the same feature(s) cannot **intervene** between the head and tail of the dependency. (Rizzi 1990, 2004, a.o.)

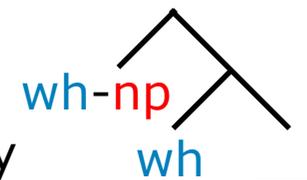


Strong: RM is not intended for strong islands. It only holds for islands with **interveners**.

Weak: RM explains amelioration through featural distinctness: if the intervener has **totally distinct** features, there will be **no island effect**.



If the intervener has **partially distinct** features, there will be **a smaller island effect**. This featural approach to RM predicts that the number of effect sizes will be related to the number of grammatically active features in dependencies.



Information Structure-based approaches

IS default:

Fact

Neg

Wh

NC

Subj

RC

Adj

exception:

Fact

Neg

Wh

NC

Subj

RC

Adj

IS approaches propose that the IS properties of the dependency must match the IS properties of the constituent that they are extracted from. Island effects arise when there is a clash. Wh-dependencies are focus constructions, so they must be extracted from focused constituents. (Erteschik-Shir 1973 et seq., Goldberg 2006, Abeille et al. 2020)

Strong: A strong island is a constituent whose **IS properties clash with every (A') dependency type** in a language.

Weak: A weak island is a constituent whose **IS properties clash with some (A') dependency types** and not others.

Simple and complex-wh likely have the same IS properties, so all islands would be strong. But **Erteschik-Shir 1973 explicitly exempts the “list” meaning of complex wh** (d-linking) from the clash, so all would be weak.

Semantic approaches

semantic:

Fact

Neg

Wh

NC

Subj

RC

Adj

I will unfairly group the Szabolcsi and Zwarts (1993, 1997) and Abrusán (2014) semantic approaches together because I am not smart enough to speak eloquently about their differences. But the shared idea is that weak islands arise because of an incompatibility between the semantics of certain wh-questions and the semantic operators that introduce certain islands.

Strong: These approaches are not intended to explain strong islands.

Weak: Crucially, these approaches are designed to explain an asymmetry roughly between **wh-argument questions** (individuals) and **wh-adjunct questions** (not individuals).

We do not test wh-adjuncts, so we do not test these theories directly. That said, these theories tend to assume that extraction of **simple** and **complex wh-arguments** are **both acceptable out of weak islands**. This would appear as **no-island effect** for both simple and complex in our experiments.

(Abrusán assumes **tensed wh-islands are strong**.)

Question 3: Theories of islands

This is just a single slide with all of the predictions in one place.

stacking:

Fact

Neg

Wh

NC

Subj

RC

Adj

input:

Fact

Neg

Wh

NC

Subj

RC

Adj

fRM:

Fact

Neg

Wh

NC

Subj

RC

Adj

IS default:

Fact

Neg

Wh

NC

Subj

RC

Adj

exception:

Fact

Neg

Wh

NC

Subj

RC

Adj

semantic:

Fact

Neg

Wh

NC

Subj

RC

Adj

Some experimental details

A 7-point scale judgment task

31 total items to rate.

3 anchor items
(1, 4, 7)

These appear in the instructions with a rating already specified.

9 practice items
(7, 1, 4, 6, 2, 5, 3, 7, 1)

These are unannounced: not marked as distinct from the experiment.

8 target items
1 island type/1 wh-type
2 tokens of each condition

We created 16 lexically matched sets for each island, distributed into lists using a latin square procedure.

14 filler items
2 each of ratings 1 through 7
from massive LI replication

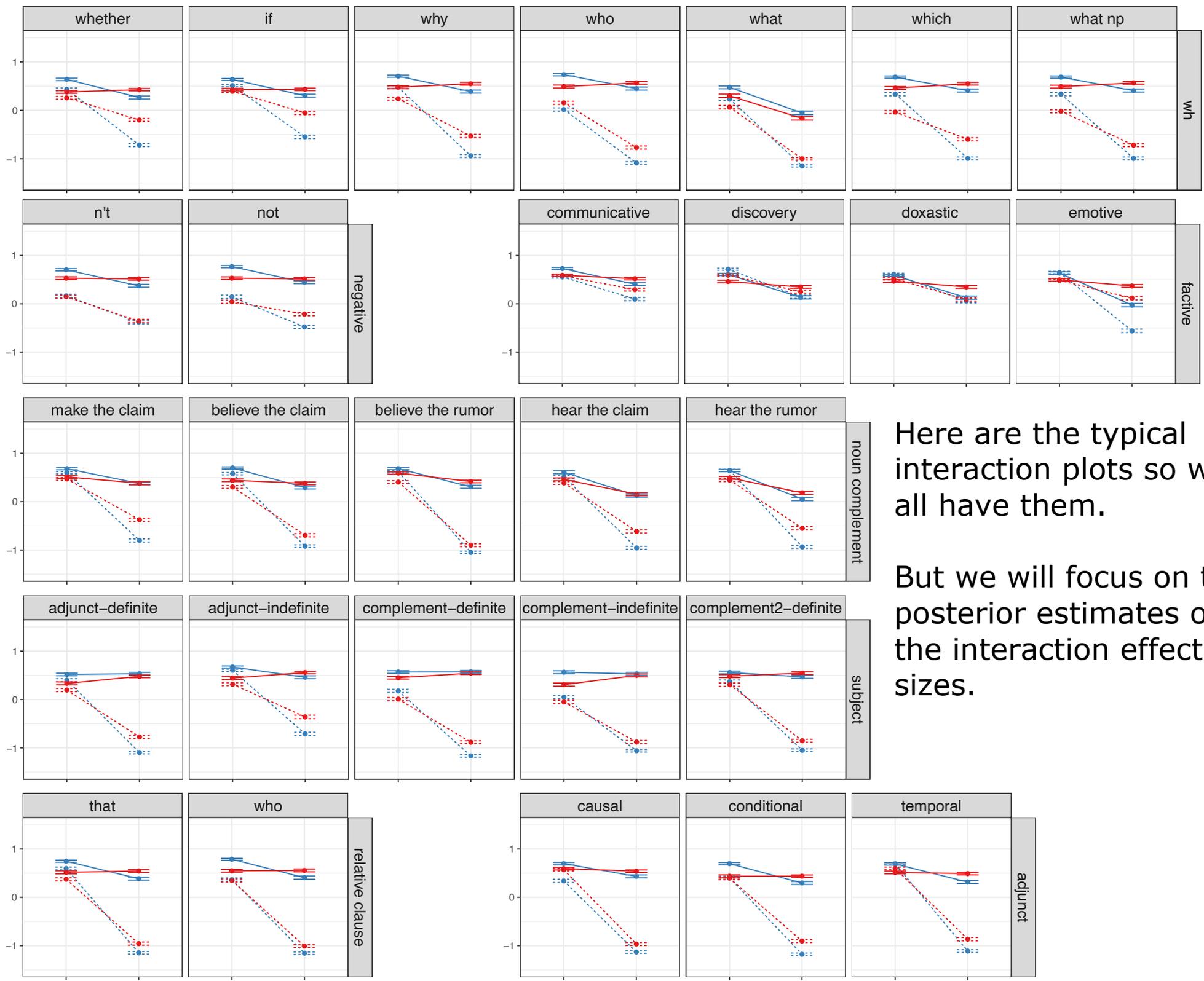
The same for all participants. The ratio of fillers/practice items to target items is nearly 3:1.

2 morality questions
(answer in complete sentence)

To identify bots and non-native speakers.

~200 participants through
CloudResearch.

28 islands x 2 wh-items
= ~11,200 participants

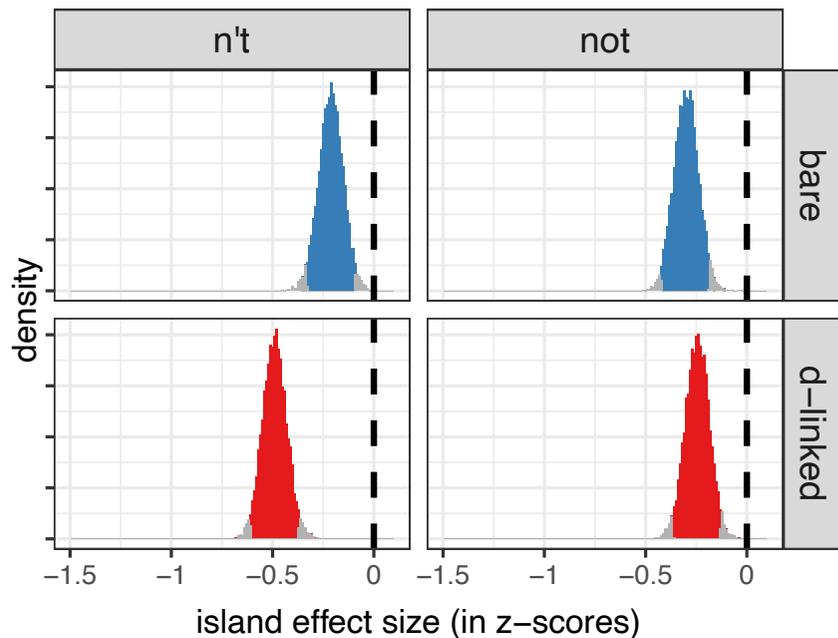


Here are the typical interaction plots so we all have them.

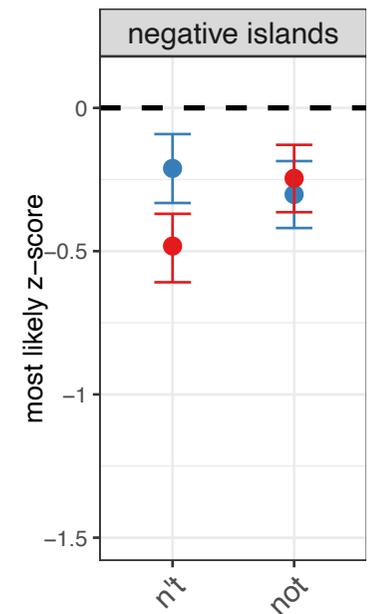
But we will focus on the posterior estimates of the interaction effect sizes.

Posterior estimates of the island effect size (95% credible intervals)

The posterior estimates that we calculated are probability distributions over possible island effect sizes. We can calculate 95% credible intervals from these distributions. Here is an example using negative islands:



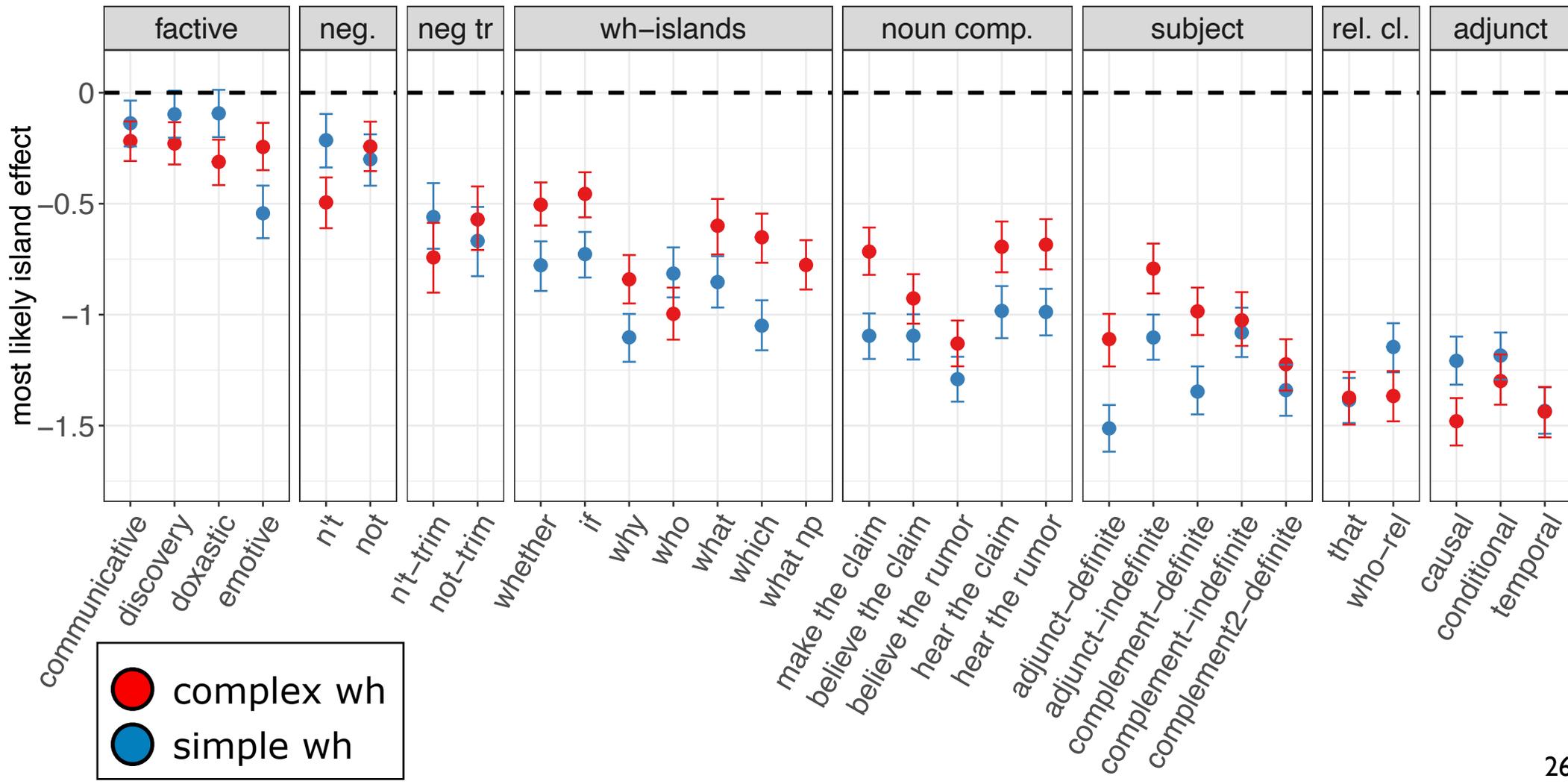
We can more compactly represent the mean of the distribution and the 95% interval in a dot plot with 90° change in orientation.



And, now, we are ready to look at the full 28 island tokens...

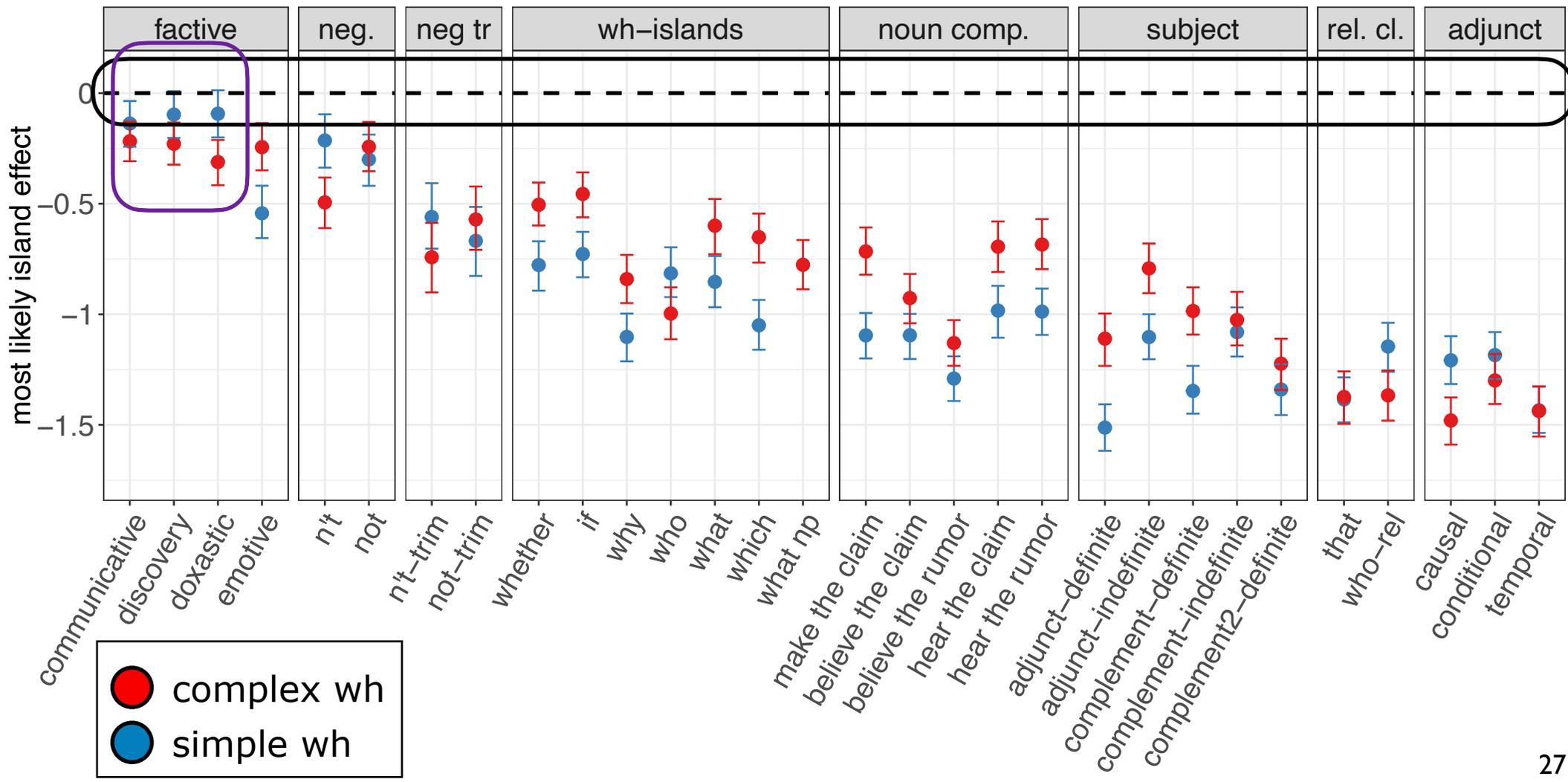
Posterior estimates of the island effect size (95% credible intervals)

There is a lot of information here (28 island tokens x 2 wh-types). It is frankly a bit overwhelming. So let's move systematically through it.



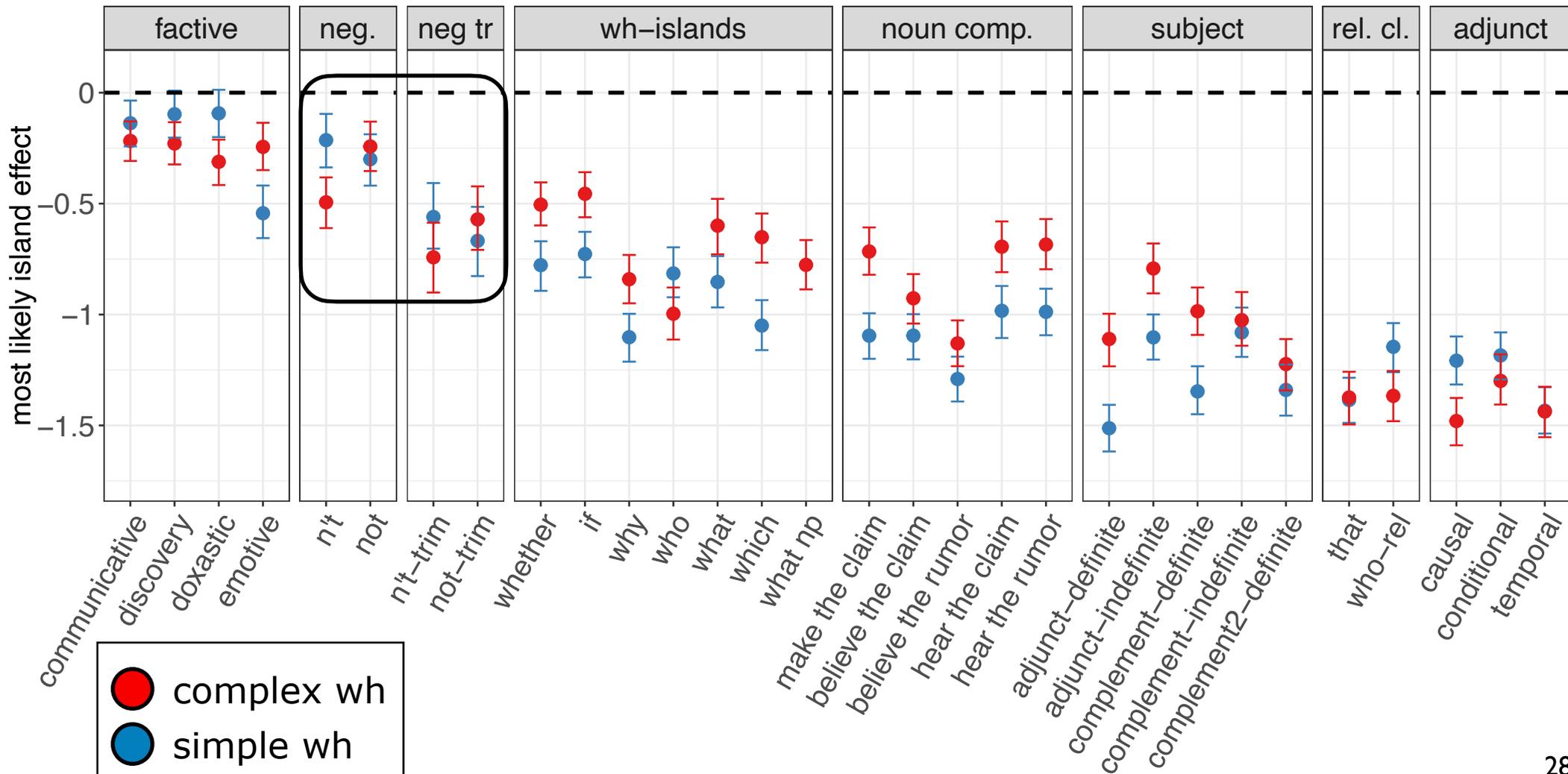
Some quick issues: Which are islands?

The dashed line represents no island effect. 25 out of 28 **simple-wh** credible intervals are below the line, suggesting an island effect (a reduction in acceptability), except for 3 types of the factives. Karttunen 2016 argued that these 3 did not carry factive presuppositions, so these results align with that.



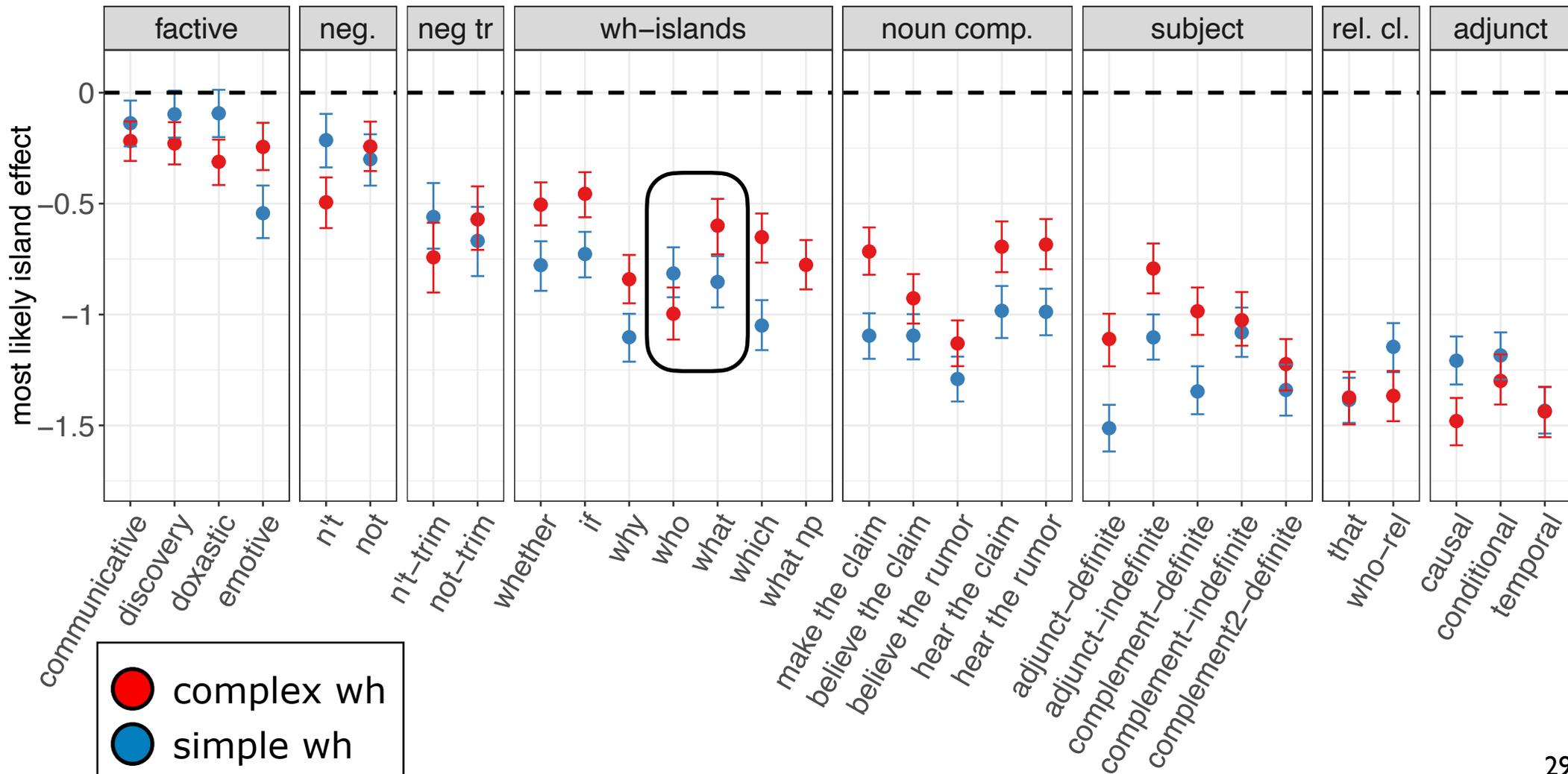
Some quick issues: Trimming Neg islands?

The negative questions in **neg-islands** may require context to be completely felicitous. To make sure this was not confounding our analysis, we removed participants who rated the short/island condition (a negative question) below 0. This is presented in “neg tr(immed)”. These samples are ~100 participants.



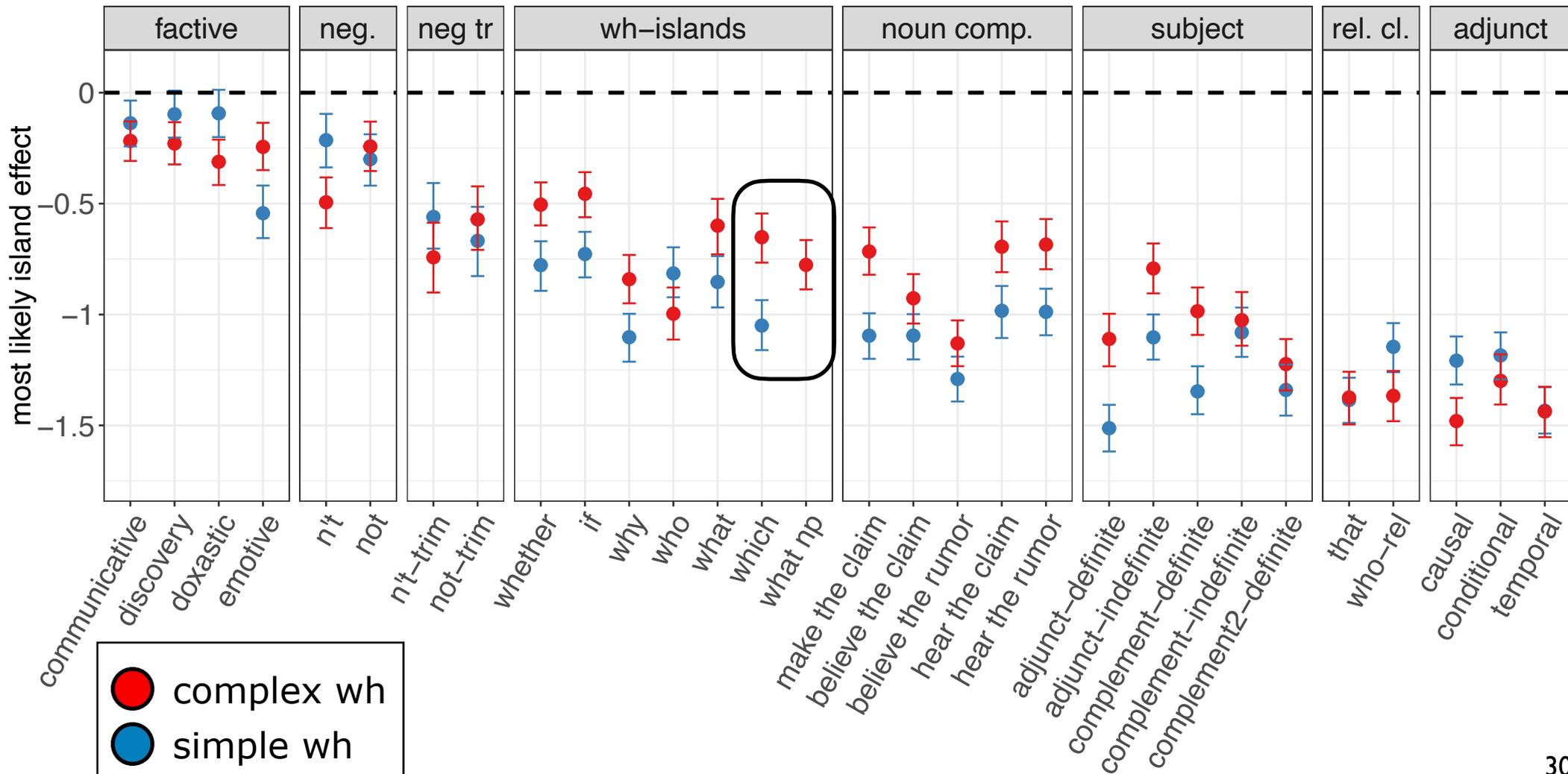
Some quick issues: who double name penalty

In *wh-islands*, the *who-island* patterns differently. This is because in the *simple-wh* design, the short/island condition has a double-name penalty (2x who) that makes the island effect smaller. The *what-island* resolves this, but uses psych verbs.



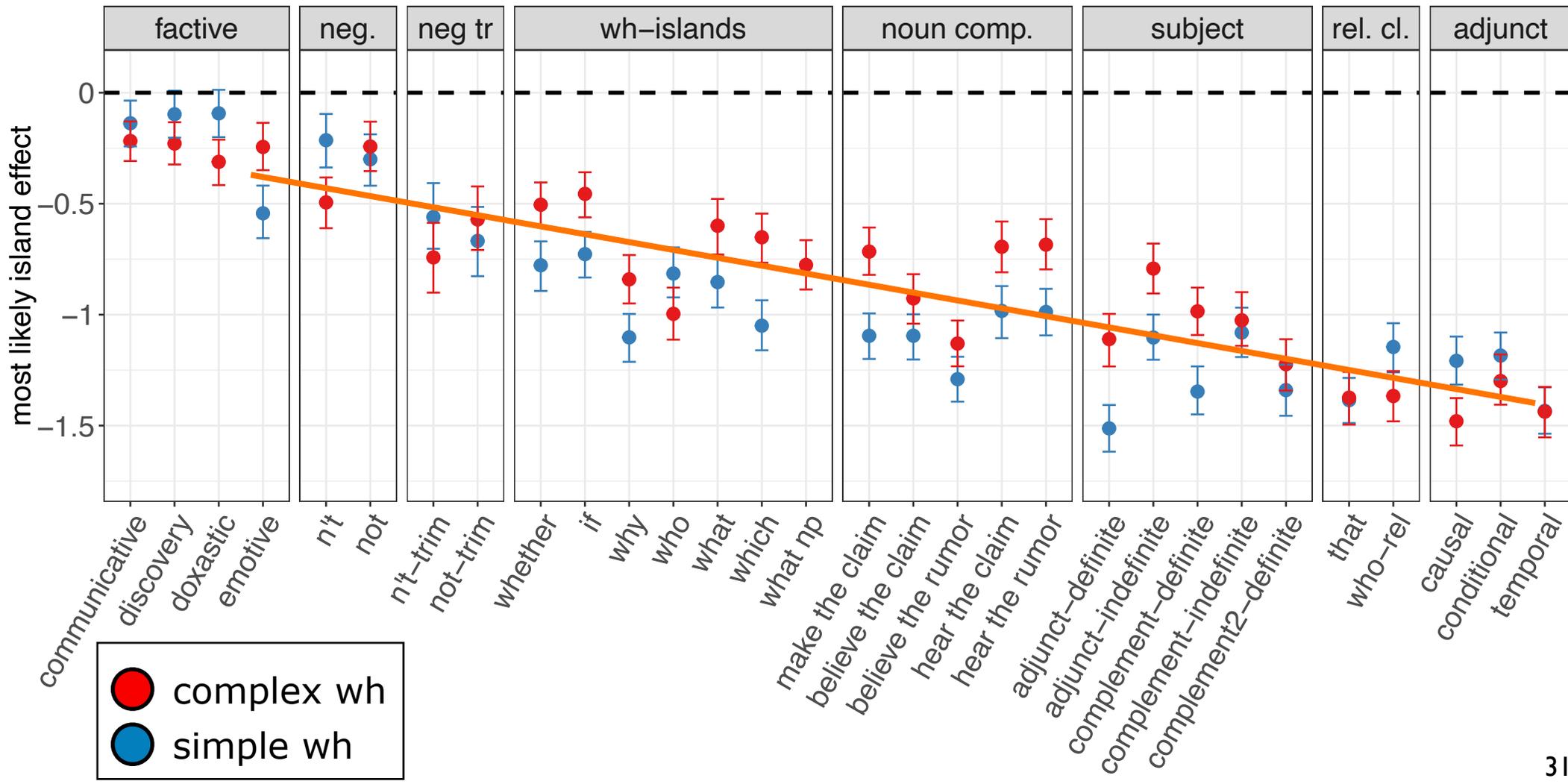
Some quick issues: which is ok!

In *wh-islands*, the *which-island* could also potentially contain a double-name penalty (2x which), this time in the *complex-wh* long/island condition (the island violation), potentially inflating the island effect. To check this, we also tested *what np* as the island. It shows roughly the same island effect, suggesting no double name penalty.



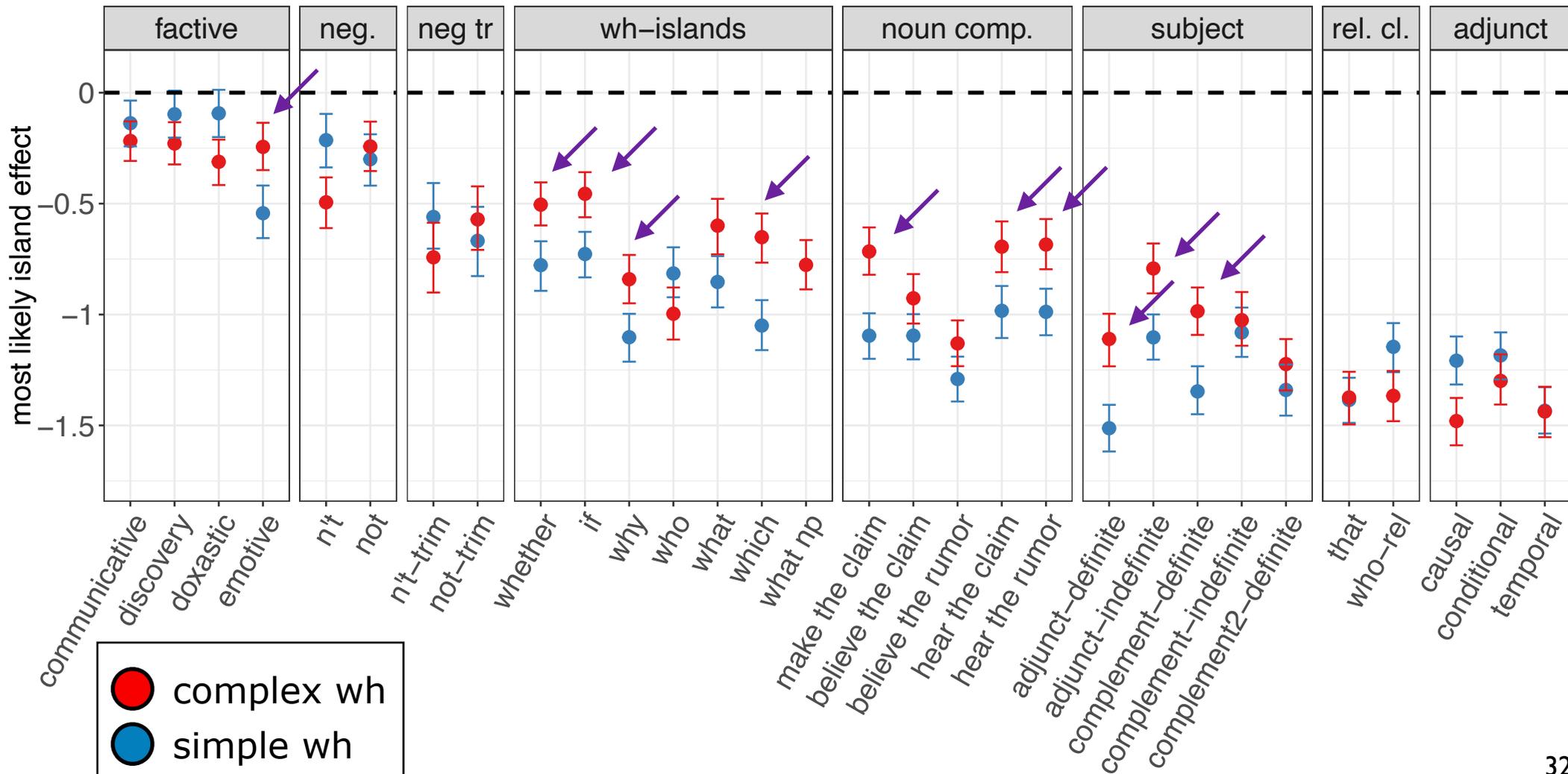
Caveat - selectivity versus effect size

The terms **strong/weak** are used in two senses in the field. The precise sense that we use in journal articles is about **selectivity to different dependencies**. But, there is also an informal sense sometimes used when we talk about the **relative size of the island effect**. There is something to this informal sense visible in the plot. But today's talk is about the precise sense - **selectivity**.

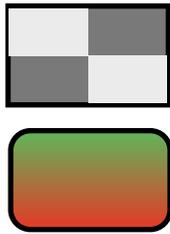


Question 1: partial amelioration

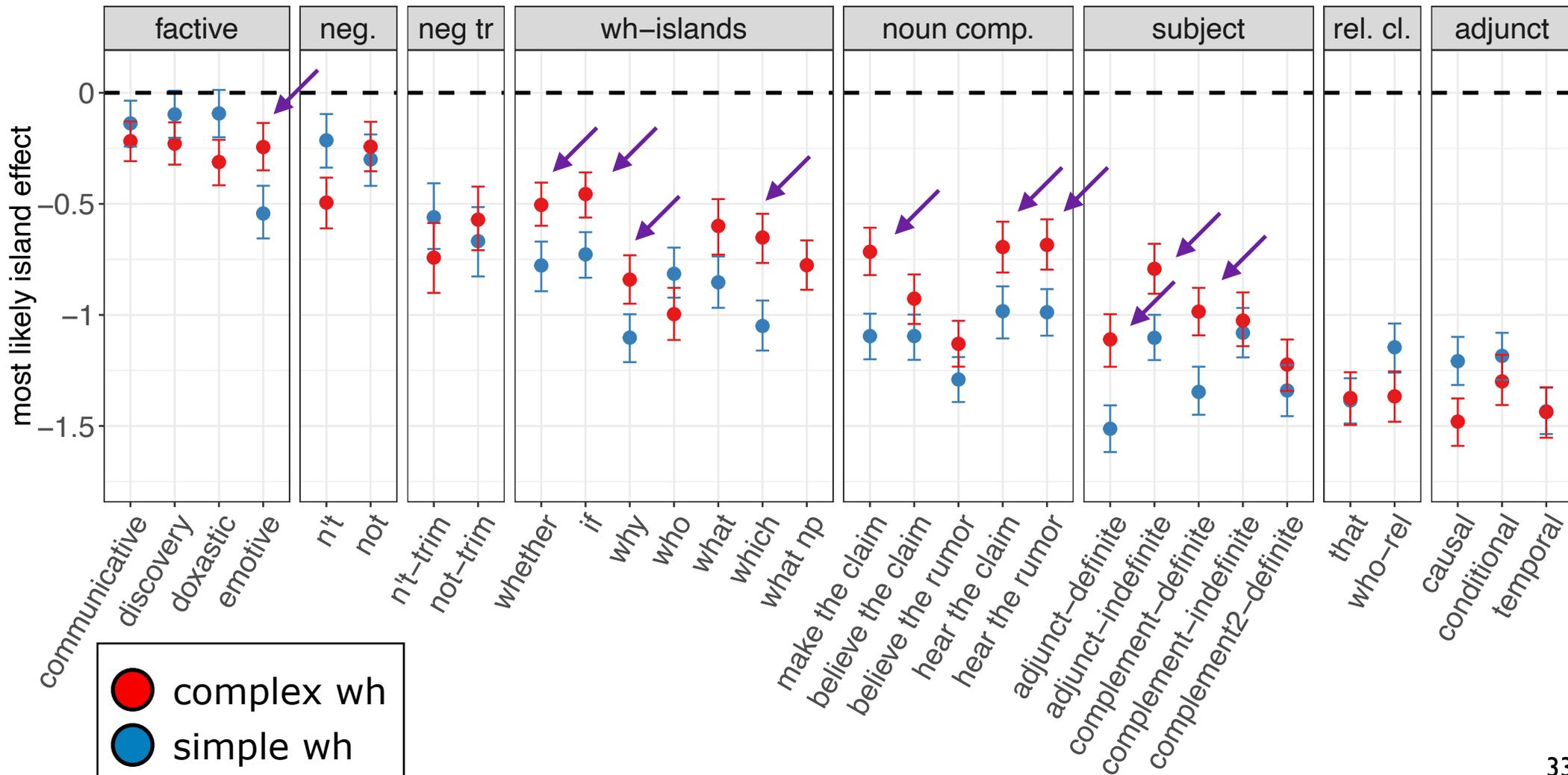
The amelioration effect of **complex-wh** shows up as separation in the intervals, crucially with **complex-wh** above (smaller decrease) than **simple-wh**. Several island tokens show this pattern. But, crucially, in these cases, the complex-wh interval does not overlap the zero line. This suggests that **there is still an island effect with complex-wh**. That is **partial amelioration**.



Question 1: partial amelioration

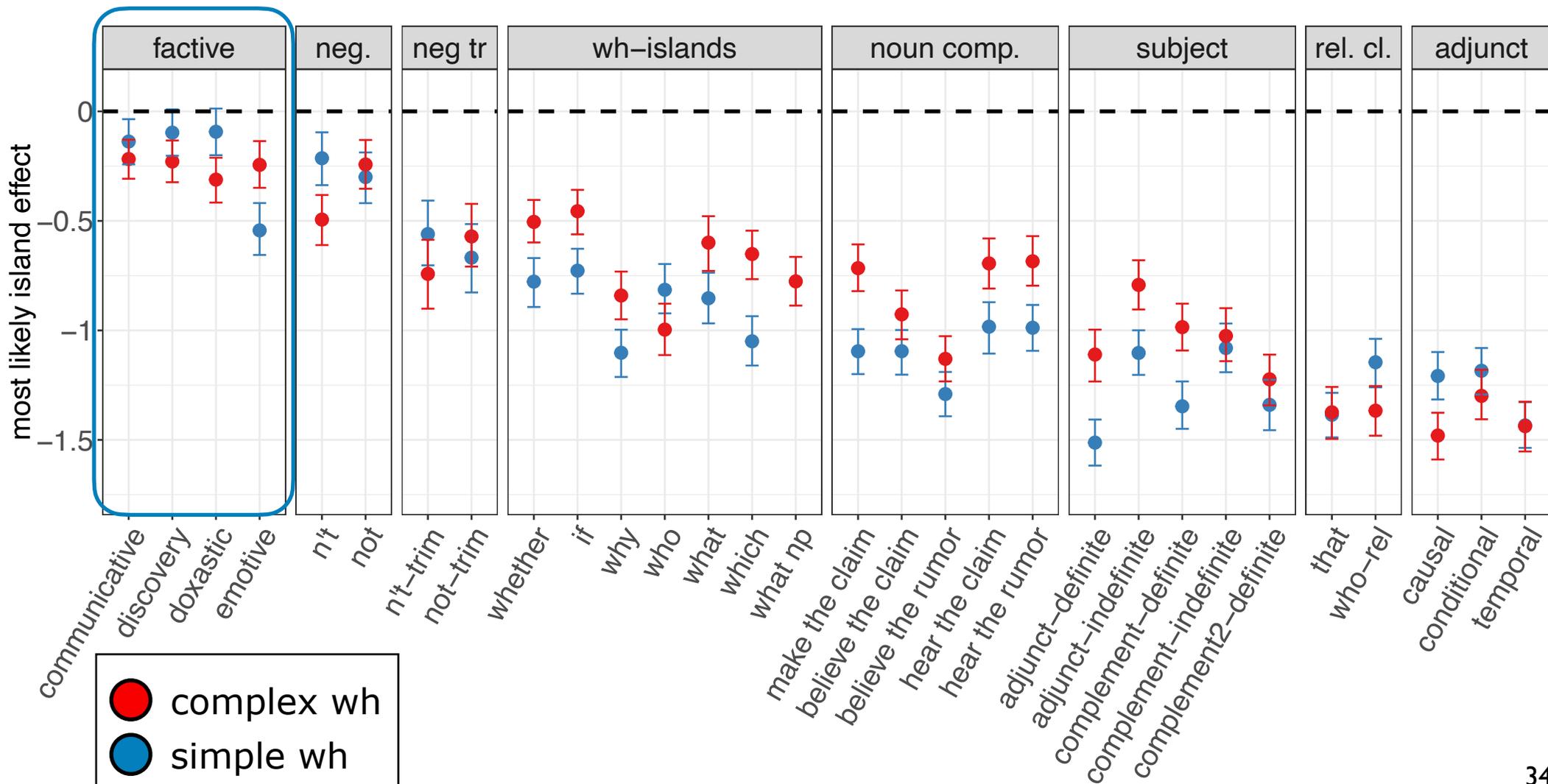


This is potentially disruptive. We must now explore either a **categorical grammar with two violations**, or a **continuous grammar with (difficult-to-acquire) constraint weights**. I will return to these in our discussion of Question 3 - island theories!



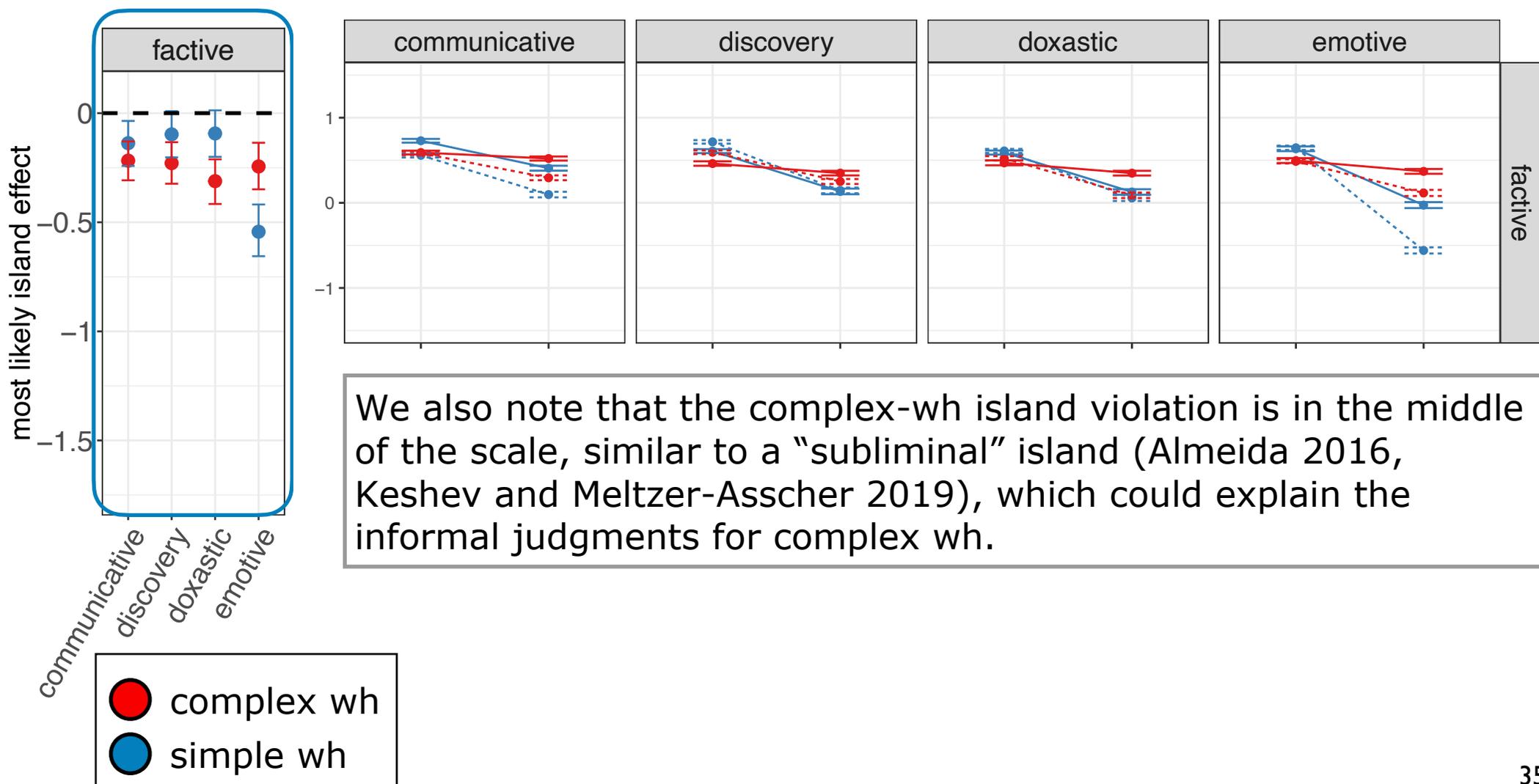
Question 2: Classifying islands

Emotive factive islands are *weak by our diagnostic*. They are islands with both simple and complex wh, contra the assumption of semantic approaches, though again, this does not preclude an additional (partial/gradient) argument/adjunct difference (with wh-adjuncts rated even lower).



Question 2: Classifying islands

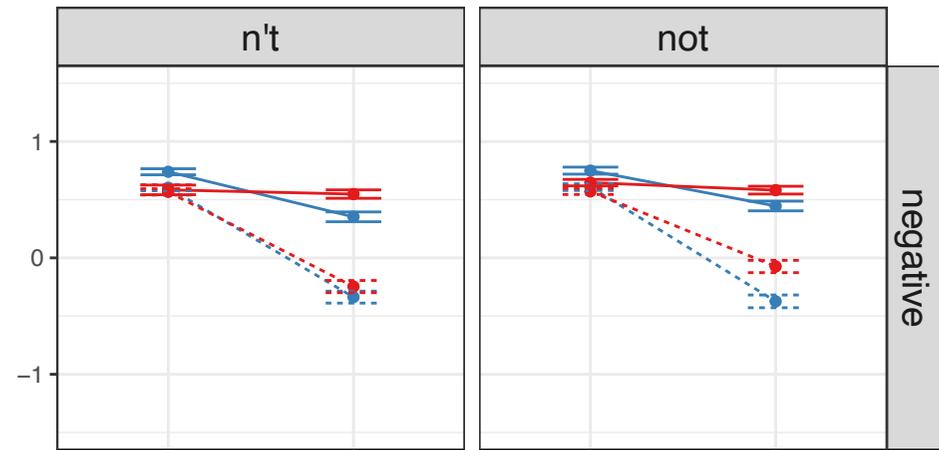
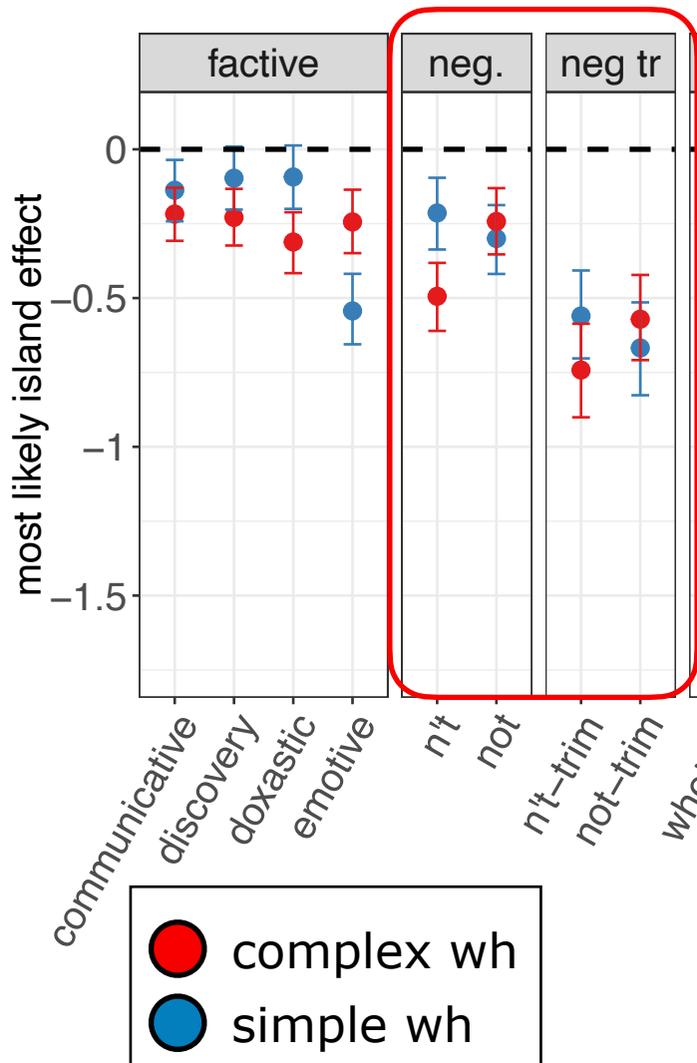
Emotive factive islands are *weak by our diagnostic*. They are islands with both simple and complex wh, contra the assumption of semantic approaches, though again, this does not preclude an additional (partial/gradient) argument/adjunct difference (with wh-adjuncts rated even lower).



We also note that the complex-wh island violation is in the middle of the scale, similar to a “subliminal” island (Almeida 2016, Keshev and Meltzer-Asscher 2019), which could explain the informal judgments for complex wh.

Question 2: Classifying islands

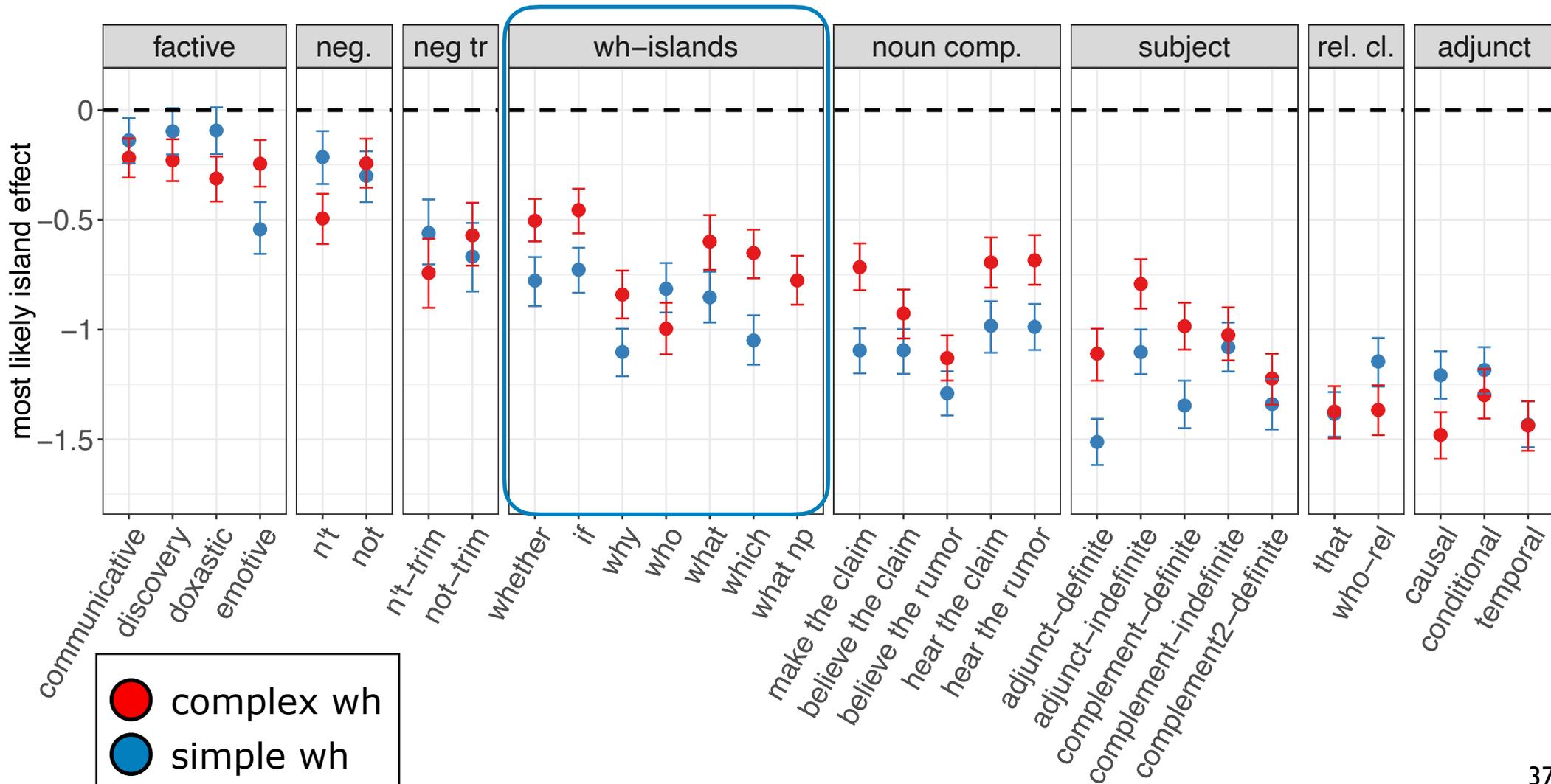
Neg-islands appear to be **strong by this diagnostic**, regardless of trimming. We didn't test wh-adjuncts, so this does not preclude a partial/gradient weak effect for wh-adjuncts, but it does mean that the effect is not completely eliminated by wh-arguments as semantic approaches tend to assume.



We also note that the complex-wh island violation is in the middle of the scale, similar to a “subliminal” island (Almeida 2016, Keshev and Meltzer-Asscher 2019), which could explain the informal judgments for complex wh.

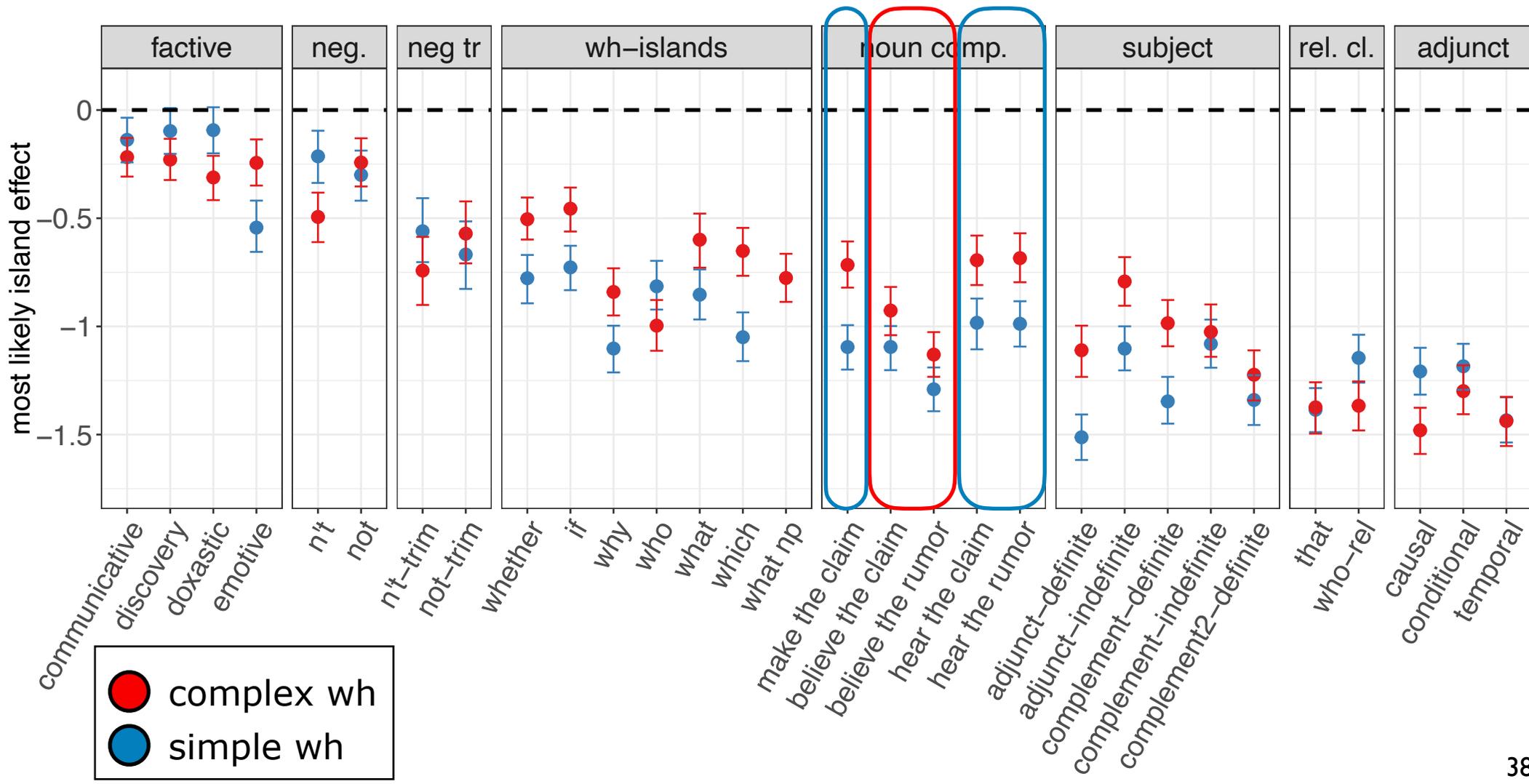
Question 2: Classifying islands

Wh-islands are uniformly *weak by our diagnostic* (setting aside the double-name penalty and the partial amelioration issue). These are all tensed, so we can also say that tensed wh-islands are not strong by our diagnostic. (But again, this does not preclude a further effect of wh-adjuncts.)



Question 2: Classifying islands

Noun complement islands are both **strong** and **weak**. It appears to be driven by the type of verb: **make** and **hear** are **weak**, while **believe** is **strong**. Semantically, this is **discovery** versus **doxastic**. But, crucially, neither of these are factive islands (and this is an NP, not CP, frame)!



Question 2: Classifying islands

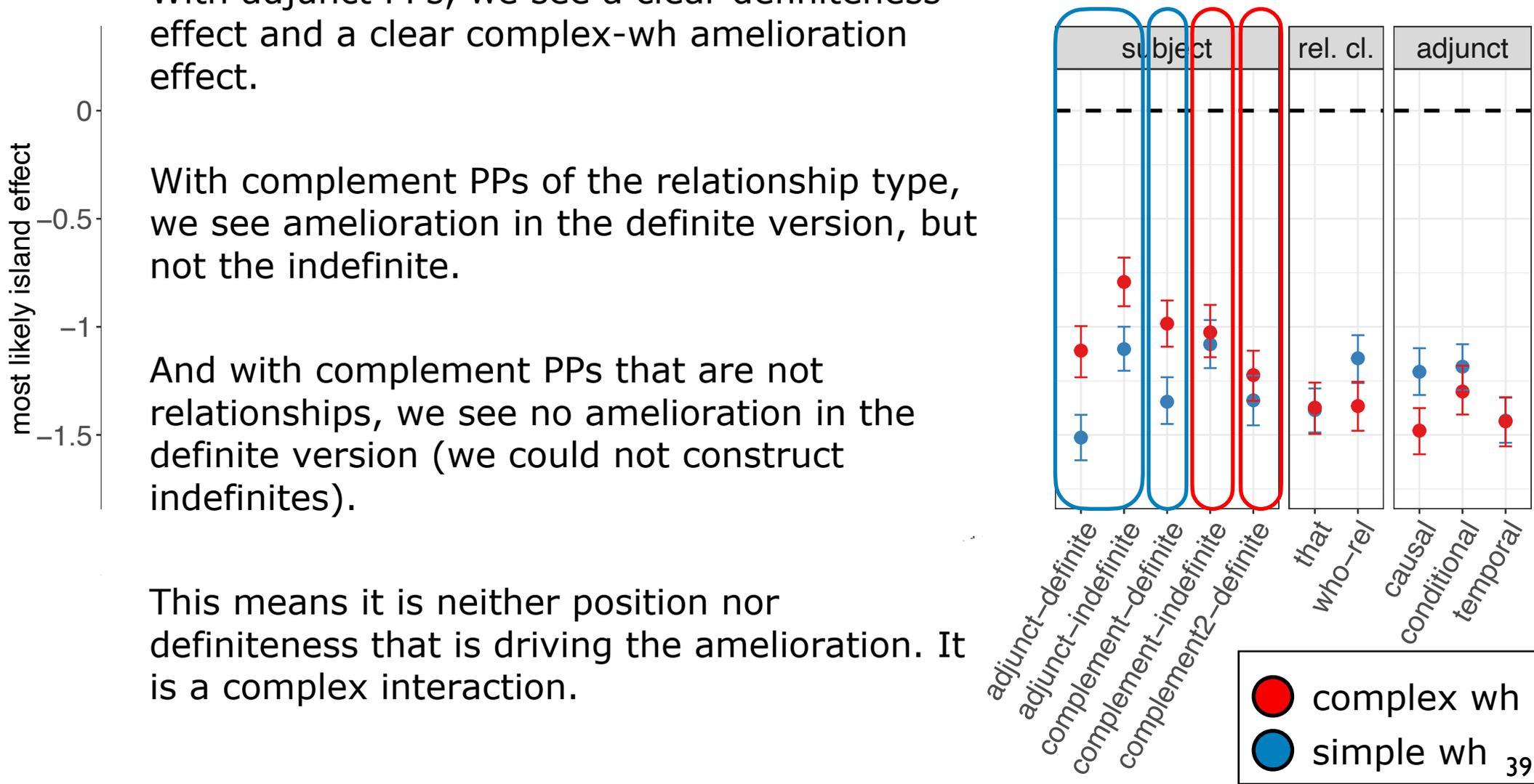
Subject islands are both **strong** and **weak**. The complication is that the pattern does not follow the properties we tested - it is an interaction.

With adjunct PPs, we see a clear definiteness effect and a clear complex-wh amelioration effect.

With complement PPs of the relationship type, we see amelioration in the definite version, but not the indefinite.

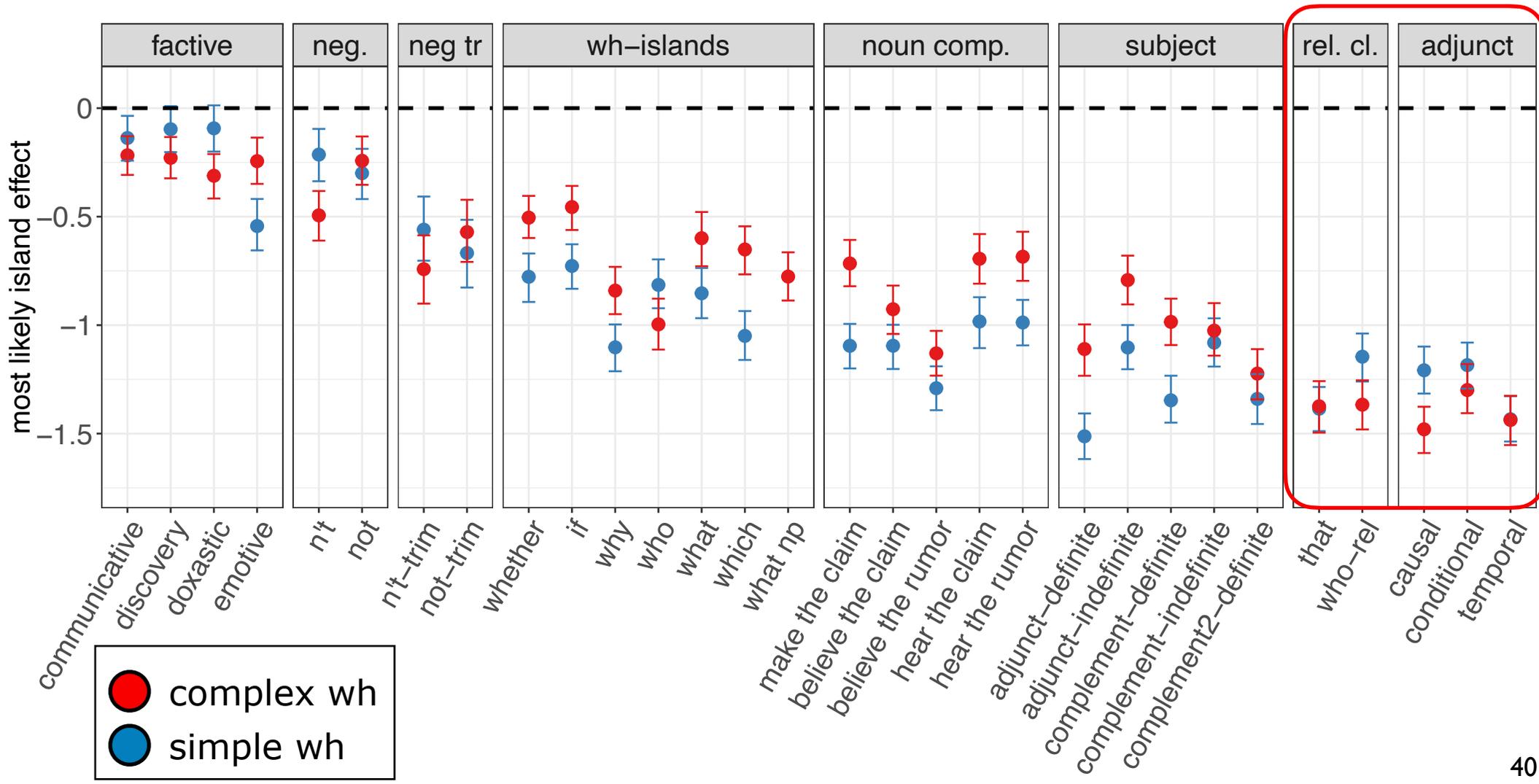
And with complement PPs that are not relationships, we see no amelioration in the definite version (we could not construct indefinites).

This means it is neither position nor definiteness that is driving the amelioration. It is a complex interaction.



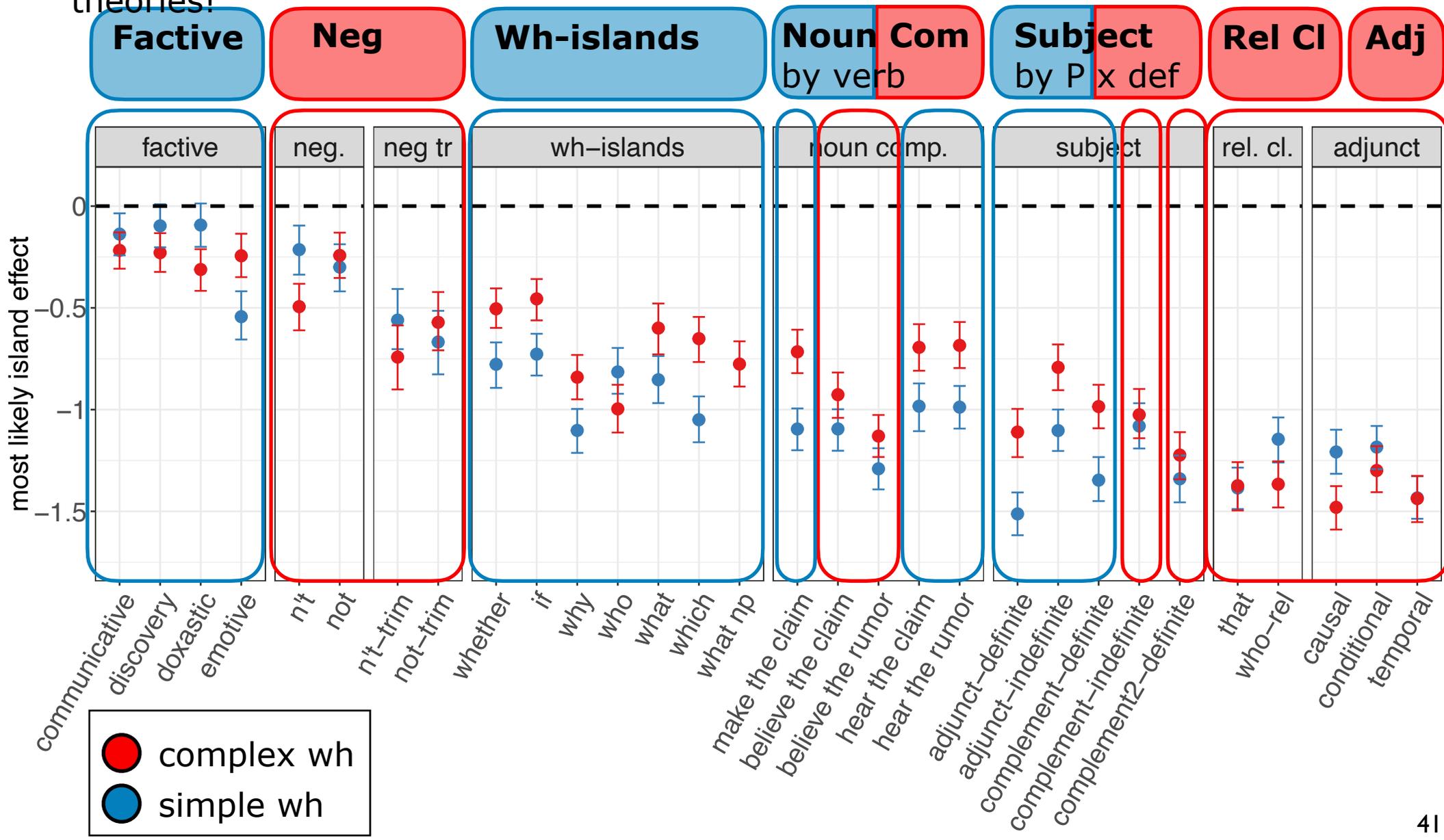
Question 2: Classifying islands

Relative clause and Adjunct islands are both **strong**. This seems like maybe adjuncts are strong islands in general. In fact, there is a bit of an inversion in the pattern for causal adjuncts such that the complex-wh shows a larger effect than simple. This is not predicted by any theory, so we set it aside for now.



An empirical classification of island types

I am sure you can already guess that the mismatch between what the literature has assumed and these results causes problems for all of the existing theories!



Question 3: Theories of islands

observed:

Fact

Neg

Wh

NC

Subj

RC

Adj

stacking:

Fact

Neg

Wh

NC

Subj

RC

Adj

input:

Fact

Neg

Wh

NC

Subj

RC

Adj

fRM:

Fact

Neg

Wh

NC

Subj

RC

Adj

IS default:

Fact

Neg

Wh

NC

Subj

RC

Adj

exception:

Fact

Neg

Wh

NC

Subj

RC

Adj

semantic:

Fact

Neg

Wh

NC

Subj

RC

Adj

Consequences for semantic theories

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
semantic:	Fact	Neg	Wh	NC	Subj	RC	Adj

We did not test wh-adjuncts, so we are not truly testing the full range of predictions of semantic theories.

But semantic theories make one crucial prediction - that wh-arguments (individuals) will not show factive or negative island effects.

We do find island effects for these. This suggests that the argument/adjunct (individual/non-individual) distinction is not the only source of island effects in these islands. There must be **some other source of island effects** also present.

This other source of island effects must also be (partially) sensitive to the complex/simple distinction.

Consequences for phase theories

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
stacking:	Fact	Neg	Wh	NC	Subj	RC	Adj
input:	Fact	Neg	Wh	NC	Subj	RC	Adj

Phases were never designed to account for the strong/weak distinction as far as we can tell.

One possibility is violation stacking, but this hinges on finding a second violation that is present in only some islands (factive, wh, some noun complement, and some subject) AND that is obviated by complex-wh.

Another possibility is to say that phases are extremely input-driven. This seems to run contrary to the fundamental idea of phases (that they derive from the architecture of the computational system), particularly once we are splitting types of noun complements or types of subjects.

Speaking for myself, I am more and more convinced that phases is not a particularly good theory of islands, though it may be a useful theory of cyclicity.

Consequences for IS theories

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
IS default:	Fact	Neg	Wh	NC	Subj	RC	Adj
exception:	Fact	Neg	Wh	NC	Subj	RC	Adj

Variation among the islands is a potential problem for the IS approach because all dependencies with the same IS properties should be sensitive to the same islands (i.e., those that create an IS clash).

Erteschik-Shir 1973 accounts for amelioration through complex-wh by creating an exception for complex-wh. But the partial amelioration is a problem for that - an exception must be total. If it is only partial, then some amount of clash must remain, and the variation in islands resurfaces as a problem.

The only other option we see is to say that IS clashes themselves can be gradient (to account for partial amelioration), and that the weak and strong sets each instantiate distinct IS properties. But neither set appears to form a unitary IS class (or a class that is distinct from the other), so this would take some clever theorizing by IS experts.

Consequences for relativized minimality

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
fRM:	Fact	Neg	Wh	NC	Subj	RC	Adj

Relativized Minimality does a fairly good job with **wh-islands** because it was created to account for those. The only issue is that [which...which...] should pattern like a “strong” island — no amelioration because the intervener has the same features as the moved element. But this island is weak like the others.

However, RM cannot explain our **negative islands** results. To capture the simple-wh island effect, RM must say that the neg feature is an intervener for a wh feature (the only feature in a simple wh). But that would then predict that complex-wh should partially ameliorate the intervention effect (a weak pattern), because complex-wh has an np feature, so only partial overlap. But we find negative islands to be strong.

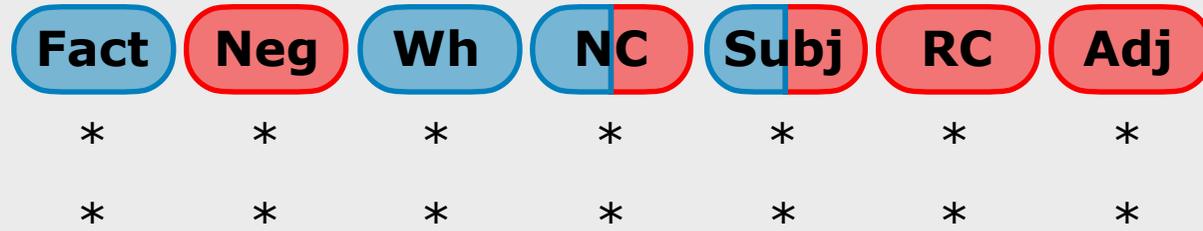
An even bigger challenge is explaining the **weak pattern** in **noun complements** and **subject islands**, because there are no interveners in these structures. One could say that these are caused by a different mechanism, but it is not ideal to have two mechanisms that lead to the very same amelioration effect.

How do we move forward with this?

previous:



observed:



Maybe reconsider a uniform approach?

previous:	Fact	Neg	Wh	NC	Subj	RC	Adj
	semantics		RM	IS theories or phases			
observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
simple:	*	*	*	*	*	*	*
complex:	*	*	*	*	*	*	*

The pattern that we have assumed until now lends itself to the idea that we can stitch together different approaches to islands into complete coverage.

But the results here resist our current theories - partly because some of the facts are wrong, partly because the members of the classes have changed, and partly because the primitives driving the classification are finer-grained.

And, we see that **even complex-wh shows an island effect**. So there is at least **one violation** in all of these for both wh-types.

This makes me think, perhaps for the first time since reading Barriers, that we should consider a **uniform approach again**. (The other option is to break things down into even smaller components.)

Maybe semantically-driven, syntactically encoded?

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
simple:	*	*	*	*	*	*	*
complex:	*	*	*	*	*	*	*

The insight of semantic and IS theories (and to some extent RM) is that island structures involve “spicy” semantics.

So, how about this fairly uninspired hybrid approach:

The meaning of (all?) island constituents creates a semantic/pragmatic clash with the semantics/pragmatics of various A' dependencies (that is potentially circumventable, given cross-linguistic variation in islands). This does not need to be the same in each island.

These meanings also require syntactic encoding. That syntactic encoding is something that the acquisition system tracks, perhaps in the form of sequences of nodes (similar to Barriers or Connectedness or Pearl's theory). So, if the semantic/pragmatic issues lead to low frequency in the input, the child learns **a grammaticalized (=syntactic) constraint**.

Maybe amelioration for semantics, but not syntax?

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
simple:	*	*	*	*	*	*	*
complex:	*	*	*	*	*	*	*

The **grammaticalized (=syntactic) constraint**, being not unlike Barriers or Connectedness, will be **insensitive to dependency types**.

Some (but not all) of the **semantic/pragmatic issues** that arise **de novo** in each island type can be **ameliorated by the meaning of certain dependencies**. That is the amelioration that we see with complex-wh (and possibly other dependency types).

I suspect this is ultimately about **presuppositions** of some sort given the empirical split we have between weak islands (factivity, embedded questions, the split within noun complements, the split within subjects based on noun types) and strong islands (adjuncts, RC adjuncts, and negative islands), and given the function of “d-linking”. But, like I said, I am not smart enough to do semantics/pragmatics, so I am unlikely to figure this out.

So, what should we do next?

observed:	Fact	Neg	Wh	NC	Subj	RC	Adj
simple:	*	*	*	*	*	*	*
complex:	*	*	*	*	*	*	*

We need to explore noun complement and subject islands further to determine what is driving the splits within those types.

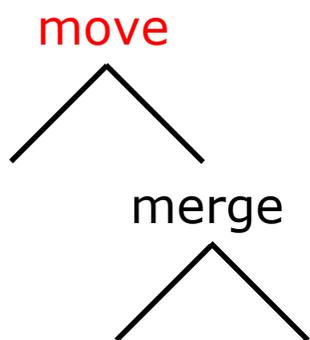
We need to figure out how to test wh-adjuncts (not easy).

We need to look at the frequency of the node sequences (probably relativized to dependency type) in child-directed input to see if this is plausibly learnable.

We need to look at this variation cross-linguistically (as many here are doing!) to further see what is driven by learning (vs potentially universal, as semantic or IS principles may be).

We need new collaborations between experimentalists and theorists — the patterns are more complex than we thought, requiring almost entirely new theories, at the levels of syntax, semantics, and maybe even acquisition!

A large-scale test of "d-linking" as a diagnostic of the strong/weak island distinction in English



Jon Sprouse & Sandra Villata
NYU Abu Dhabi

Penn Locality Workshop
04.01.23



I'd like to thank my wonderful collaborator Sandra for really being the heart of this project, the government of the UAE for the research funding, and most importantly all of you for the opportunity to share our results and hear your ideas about what might be going on!