CRISSP Lecture Series

A program for experimental syntax: data, theory, and biology

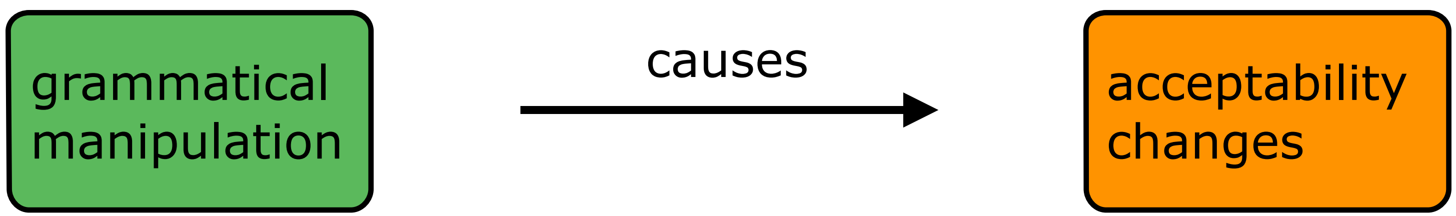
Jon Sprouse
University of Connecticut

Data 2: Experimental Syntax and the source of acceptability judgments

KU Leuven - Brussels 03.16.15

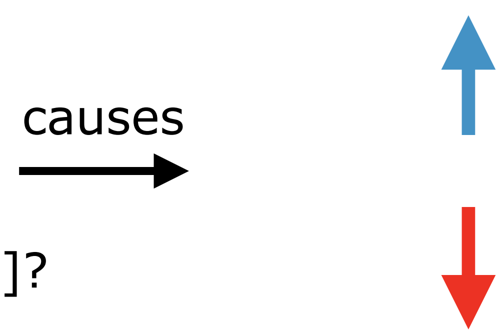
Evidence in syntax

The primary evidence in syntactic theory comes from **differences in acceptability**.



What do you [think that John bought]?

*What do you [wonder whether John bought]?



Syntax assumes a **linking hypothesis** between grammaticality and acceptability: changes in grammaticality will lead to changes in acceptability.

There are tons of linking hypotheses

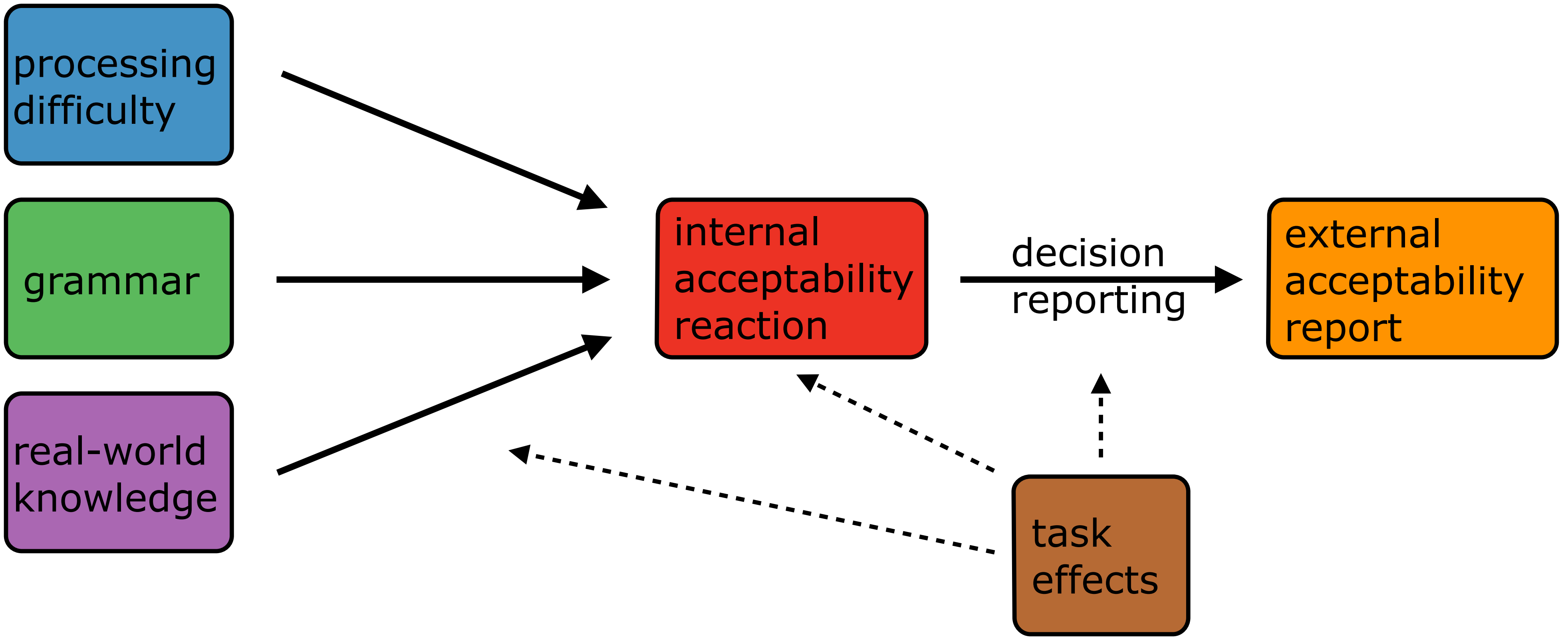
Here are some common linking hypotheses in experimental work in linguistics.

	unobservable cognitive construct		observable response
syntax:	grammar	→	acceptability
psycholinguistics:	processes	→	reading times
neurolinguistics:	processes	→	BOLD signal

In general, linking hypotheses are not independently testable because one half of the relationship (the cognitive construct) is not independently observable (or measurable). **Instead of testing them, we simply gain confidence in them by seeing how well they lead to usable theories.**

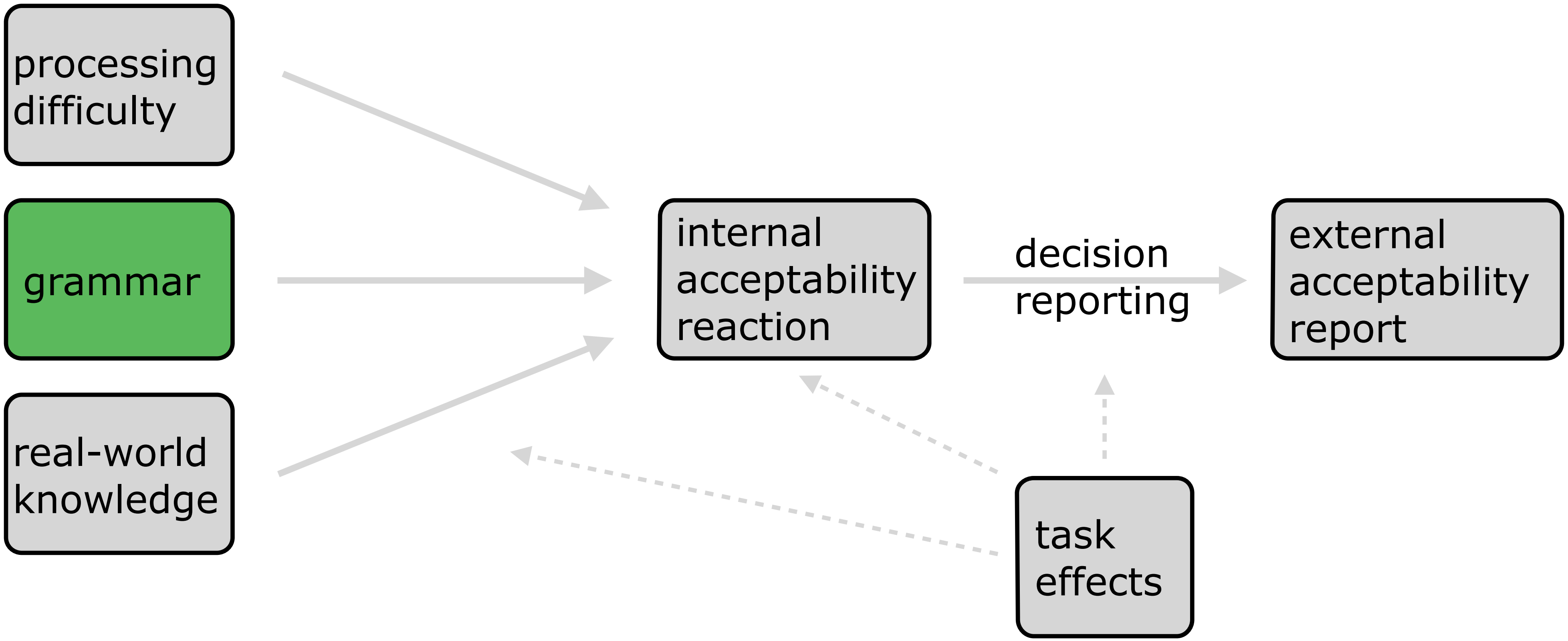
There could be other explanations

Because acceptability is the result of several cognitive systems, there could be other explanations (other linking hypotheses) for any observed acceptability difference.



There could be other explanations

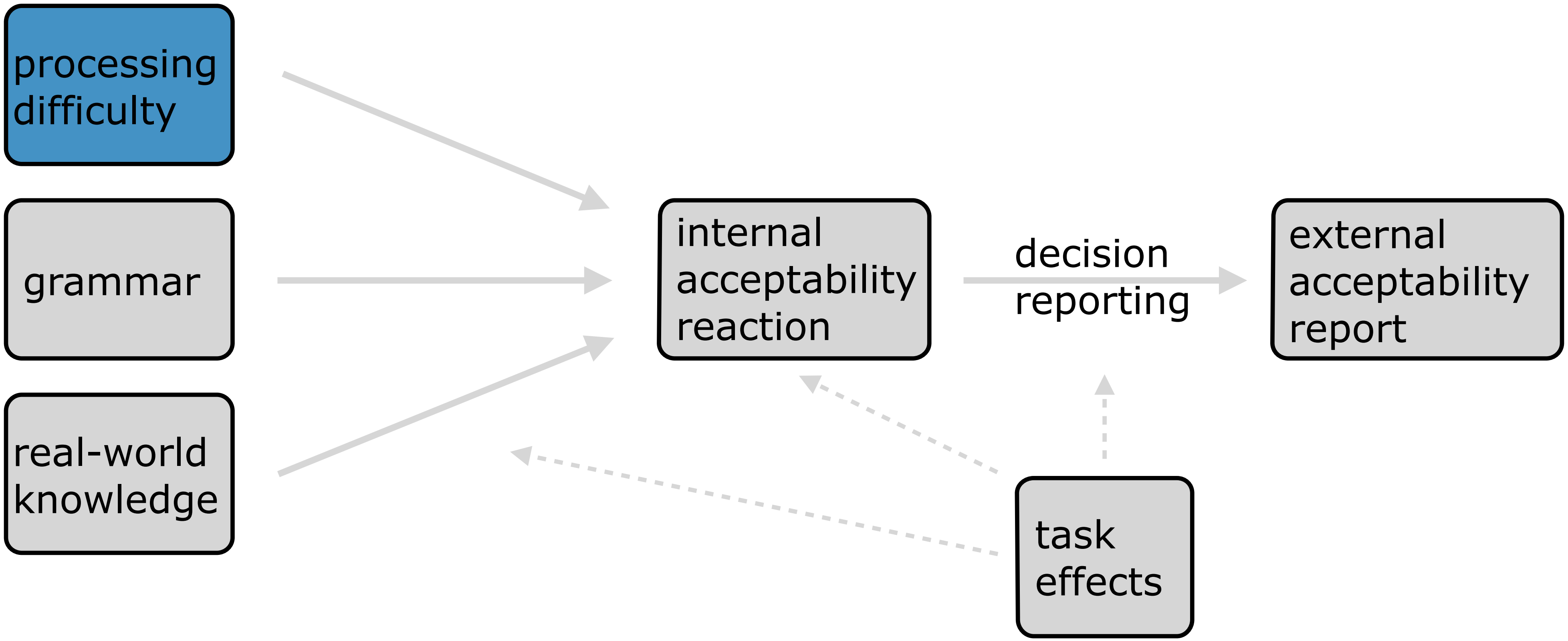
In fact, this is very familiar to every linguist. Many papers, especially in syntax, devote time to establishing that a given phenomenon is the result of syntax, as opposed to phonology, morphology, semantics, pragmatics, etc.



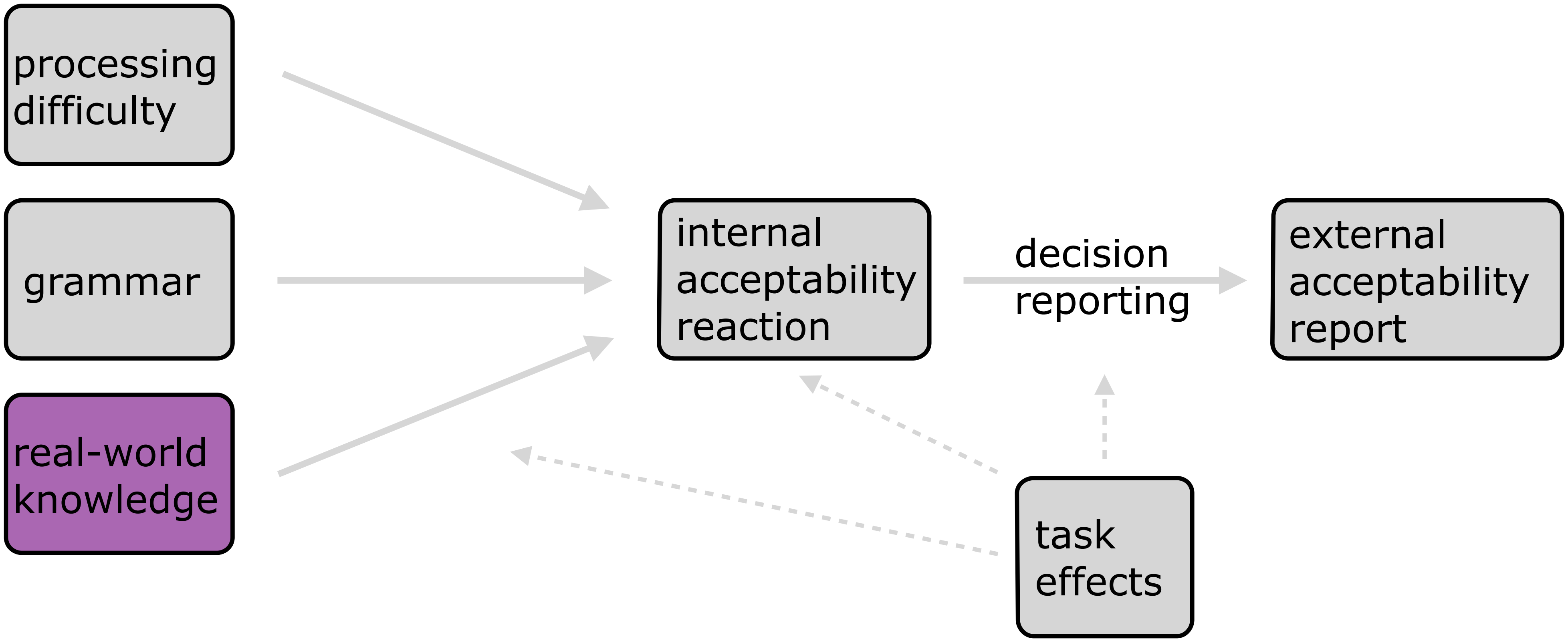
Linguists sometimes call this a **boundary problem**, but it is a general problem in all of cognitive science.

There could be other explanations

Structures that are difficult to process can lead to lower acceptability judgments.

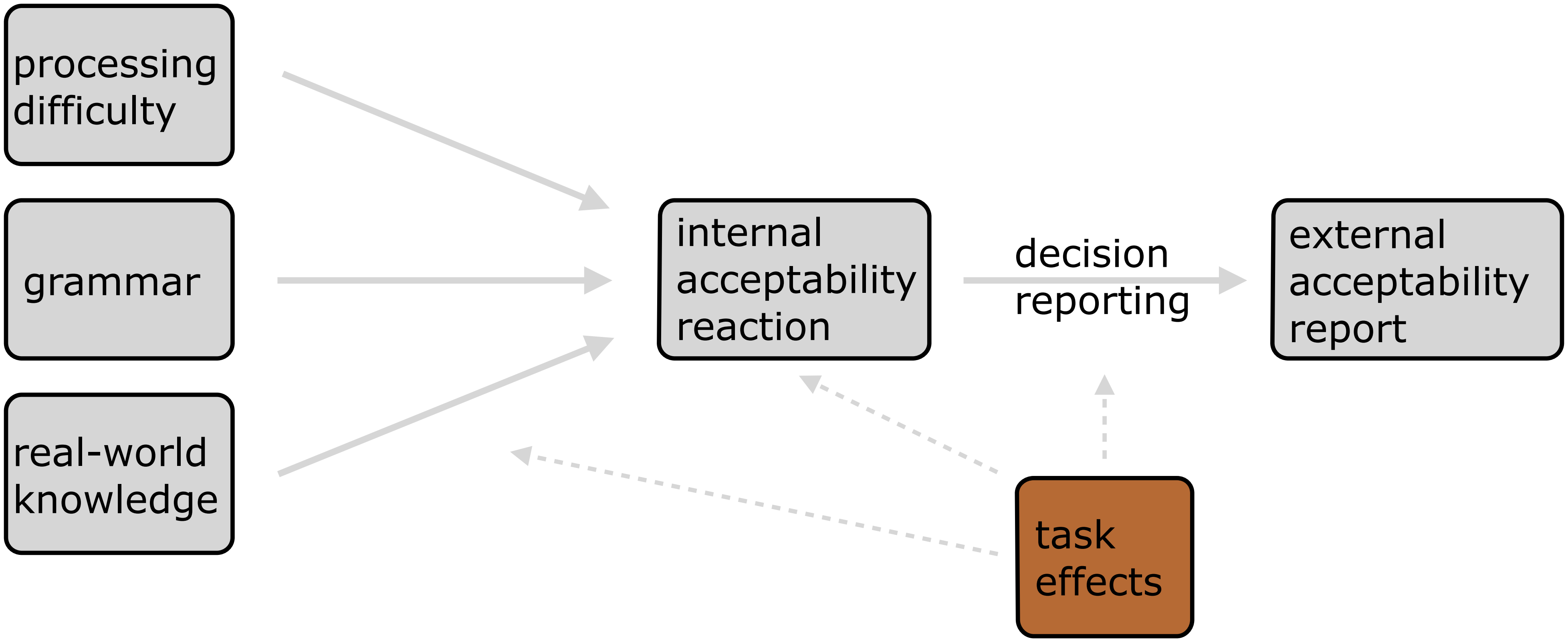


There could be other explanations



The meaning of a sentence can influence the acceptability; for example, implausible sentences tend to be rated lower than plausible sentences.

There could be other explanations



The task that we ask participants to complete can affect their judgments; for example, long experiments or repetitions of constructions can lead to different ratings.

There could be other explanations

processing difficulty	Structures that are difficult to process can lead to lower acceptability judgments.
other grammar	Even if we can be sure that the source of the effect is the grammar, it could be a **different** component of the grammar than the one we think.
knowledge/ meaning	The meaning of a sentence can influence the acceptability; for example, implausible sentences tend to be rated lower than plausible sentences.
task effects	The task that we ask participants to complete can affect their judgments; for example, long experiments or repetitions of constructions can lead to different ratings.

This is probably not an exhaustive list, but it is a good first approximation of the types of other explanations that are possible for acceptability judgment effects.

Processing difficulty is trendy

processing difficulty	Structures that are difficult to process can lead to lower acceptability judgments.
other grammar	Even if we can be sure that the source of the effect is the grammar, it could be a **different** component of the grammar than the one we think.
knowledge/ meaning	The meaning of a sentence can influence the acceptability; for example, implausible sentences tend to be rated lower than plausible sentences.
task effects	The task that we ask participants to complete can affect their judgments; for example, long experiments or repetitions of constructions can lead to different ratings.

Boundary problems are inherently interesting (and difficult), but in syntax there seems to be a particularly strong interest in so-called "processing explanations". There are probably many reasons for this.

Island Effects as a case study

Island effects are typically defined as extreme unacceptability arising from movement out of certain phrases (islands).

	You wonder [whether Jack stole a necklace]
Whether:	* What do you wonder [whether Jack stole __]?
	You make [the claim that Jack stole a necklace]
Complex NP:	* What did you make [the claim that Jack stole __]?
	You worry [if Jack forgets the necklace]
Adjunct:	* What do you worry [if Jack forgets __]?
	You think [the necklace for Jack] is pretty
Subject:	* What do you think [the necklace for __] is pretty?

This unacceptability, relative to a grammatical control, is the fact that needs to be explained (explanandum).

The question for linguists is what cognitive mechanisms lead to this unacceptability (explanans).

All island sentences involve processing difficulty

All sentences that contain island effects also contain at least two properties that are independently known to lower acceptability: long-distance dependencies and syntactically complex phrases.

Whether island: * What do you wonder [whether Jack stole ___]?

1. Dependency: What do you wonder [whether Jack stole ___]?

2. Complex phrase: What do you wonder [whether Jack stole ___]?

The empirical question is whether the unacceptability of these sentences can be **completely explained** by the dependency and complexity difficulties alone. If so, then there is no need for a constraint. If not, then we need the constraint to fully explain the effect!

3. Constraint: What do you wonder [whether Jack stole ___]?

Grammatical Approach

The unacceptability is due to processing difficulties AND a constraint in the grammar

wh .... __ .... **Dependency cost**

+

.... [ISL ....] **Structure cost**

+

*

wh .... [ISL ...... __] **Constraint**

process effect 1

process effect 2

…

\+ grammatical constraint

unacceptability/island effect

VS

Reductionist Approach

The unacceptability is due to processing difficulties alone.

wh .... __ .... **Dependency cost**

+

.... [ISL ....] **Structure cost**

process effect 1

process effect 2

…

\+

unacceptability/island effect

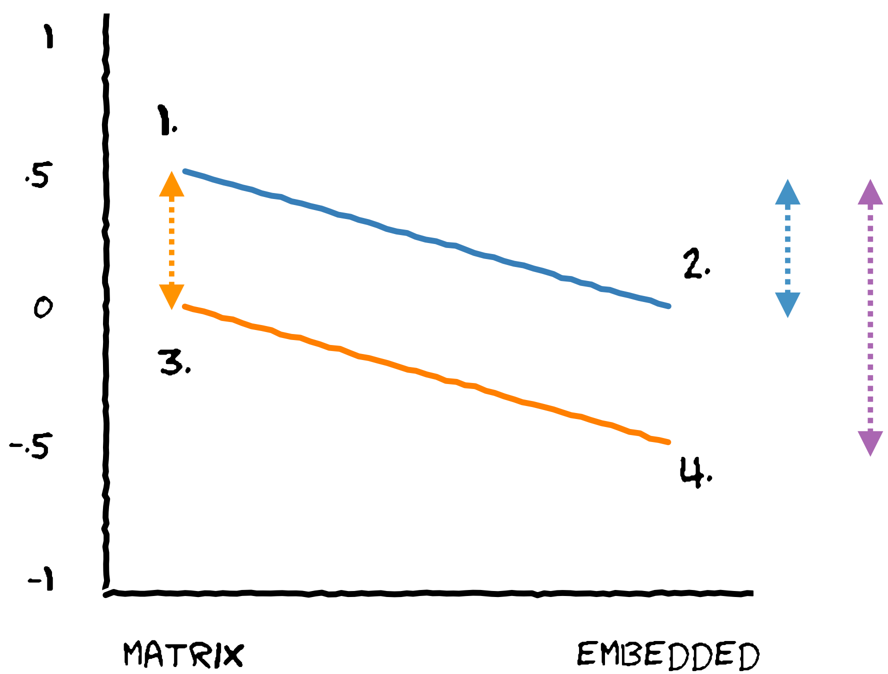
Using experimental syntax to isolate each of these components

complex phrase effect

global effect

1. Who __ thinks that Jack stole the necklace?
2. What do you think that Jack stole __?
3. Who __ wonders whether Jack stole the necklace?
4. * What do you wonder whether Jack stole __?

dependency effect



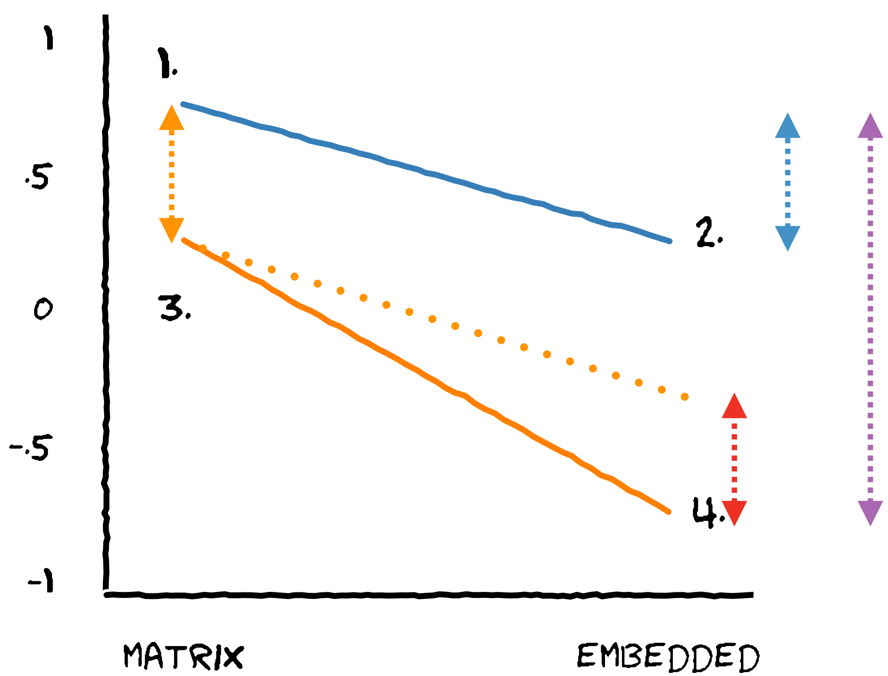
	dependency effect		(1-2)
+	complexity effect	+	(1-3)
	global effect		(1-4)

If the two processing costs can completely explain the total unacceptability, then there is no work left for the grammatical constraint to do.

Using experimental syntax to isolate each of these components

complex phrase effect | **dependency effect** | global effect

1. Who __ thinks that Jack stole the necklace?
2. What do you think that Jack stole __?
3. Who __ wonders whether Jack stole the necklace?
4. * What do you wonder whether Jack stole __?



dependency effect	(1-2)
complexity effect	(1-3)
+ constraint	+ X
global effect	(1-4)

If there is a grammatical constraint, the two independent effects won't be enough to explain the total global effect. We'll need to add the constraint's effect in.

The framework: factorial logic

What we are doing here is applying factorial logic to the reductionist claim that dependency cost and complexity cost can explain island effects.

Factor: A property that you can manipulate

Level: The values that a factor can take

We are converting each of these costs into a factor with two levels (basically, high cost and low cost), like so:

Factor 1:	the cost of long distance dependencies	**Levels**
	Who __ thinks that Jack stole the necklace?	short
	What do you think that Jack stole __?	long
Factor 2:	the cost of island structures	**Levels**
	Who __ thinks that Jack stole the necklace?	non-island
	Who __ wonders whether Jack stole the necklace?	island

Crossing the factors

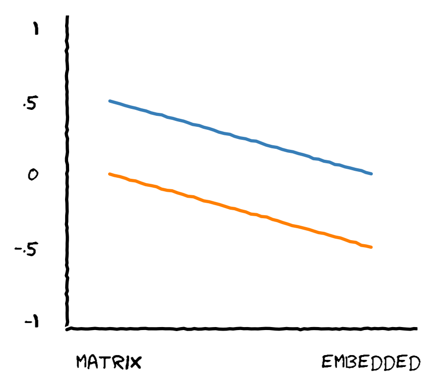
Since each factor has 2 levels, there are actually 4 combinations of the levels.

What we've done is create all four combinations. This is called **crossing the factors**. We now have a fully crossed design, which we call a 2x2 design because there are two factors, each with two levels.

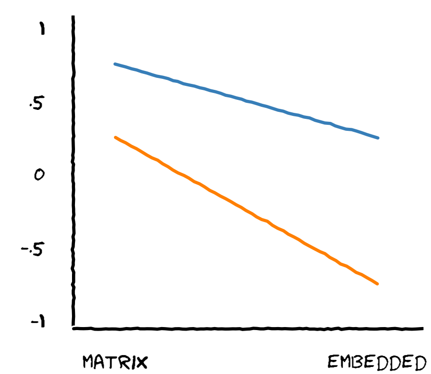
		Dependency	Structure
1.	Who __ thinks that Jack stole the necklace?	short	non-island
2.	What do you think that Jack stole __?	long	non-island
3.	Who __ wonders whether Jack stole the necklace?	short	island
4.	* What do you wonder whether Jack stole __?	long	island

There are a number of patterns that can result from a 2x2 design, but two are important to us: linearly additive (parallel lines) and super-additive (non-crossing non-parallel lines).

linearly additive



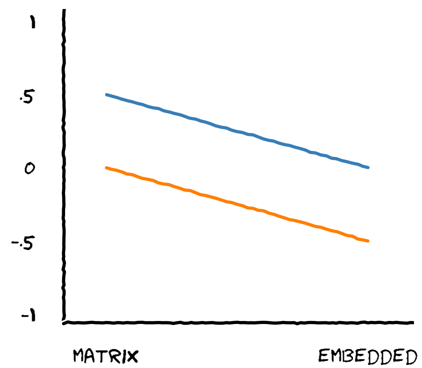
super-additive



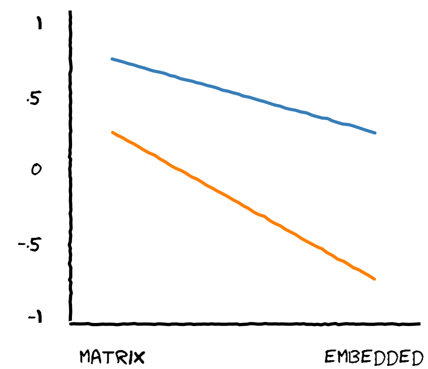
The logic of the 2x2 design

So what can we conclude from the 2x2 design?

linearly additive



super-additive



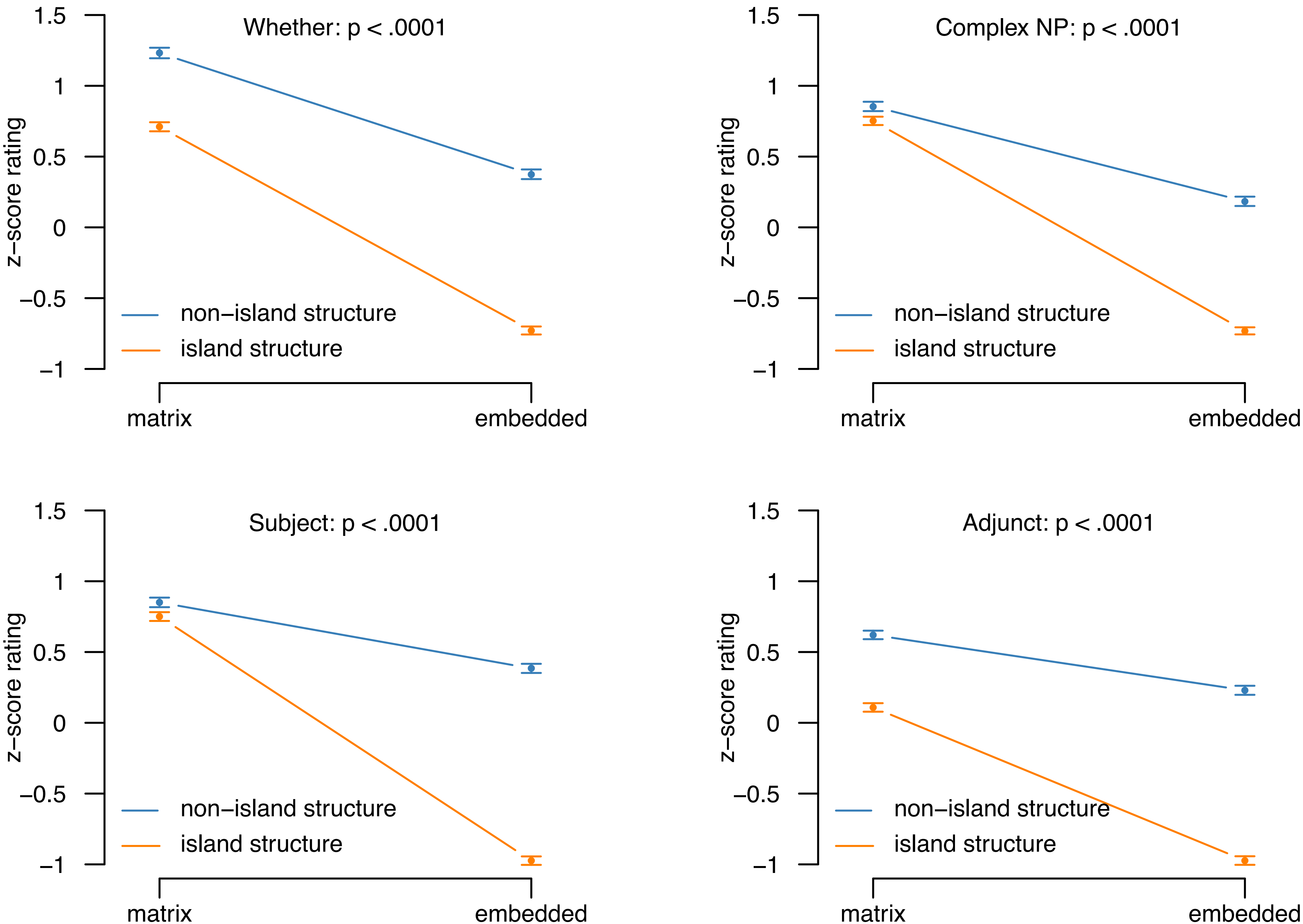
Reductionism holds

There is a mystery component

Up until now, I've been discussing the super-additive result as if it is evidence for a grammatical constraint. But it is not evidence for the constraint. It is a necessary, but not sufficient, condition for the existence of a constraint.

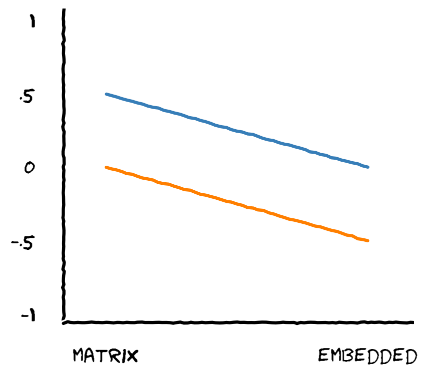
The super-additivity is actually just evidence for an unnamed component lowering acceptability. This component could be a constraint, or it could be a more complex theory of reductionism that explains the super-additivity.

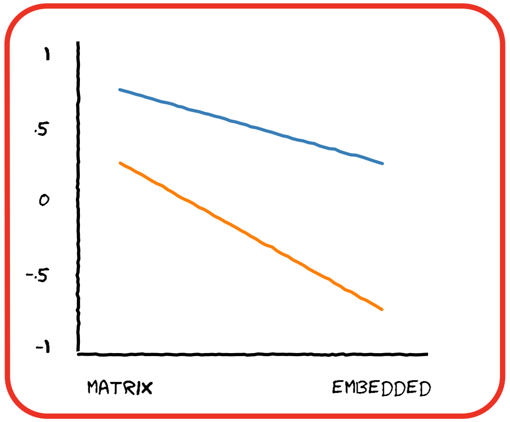
The results of a real experiment using the **factorial design**:



The logic of the 2x2 design

Step 1: Use factorial logic to test linear versus super-additivity.





Reductionism holds

Grammar or
More Complex Reductionism

Step 2: Postulate a mechanism that explains the super-additivity.

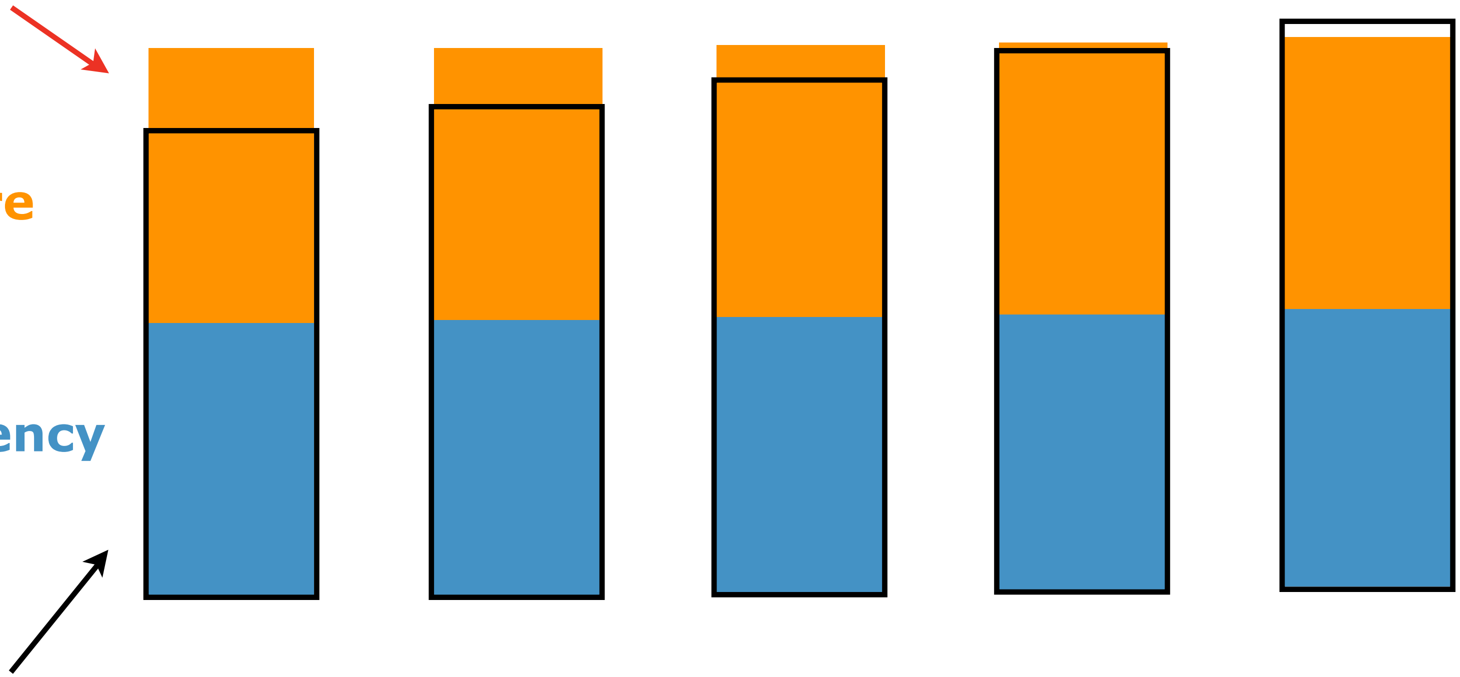
Step 3: Test the mechanism

The **super-additivity** arises from exceeding the pool of resources.

Kluender and Kutas 1993

Structure Cost

Dependency Cost



The **limited** pool of resources.

Reductionist Prediction: Variability in the pool of resources will lead to variability in the size of the super-additive component.

Testing the Prediction: Look for correlations between the size of the super-additive component and measures of working memory capacity.

Grammatical Approach

The unacceptability is due to processing difficulties AND a constraint in the grammar

wh .... __ ....

Dependency cost

+

.... [ISL ....]

Structure cost

+

*

wh .... [ISL ...... __]

Constraint

process effect 1

process effect 2

…

\+ grammatical constraint

unacceptability/island effect

VS

Complex Reductionism

The unacceptability is due to processing difficulties alone.

wh .... __ ....

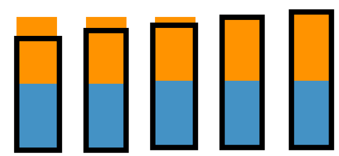
Dependency cost

+

.... [ISL ....]

Structure cost

+



WM interaction

process effect 1

process effect 2

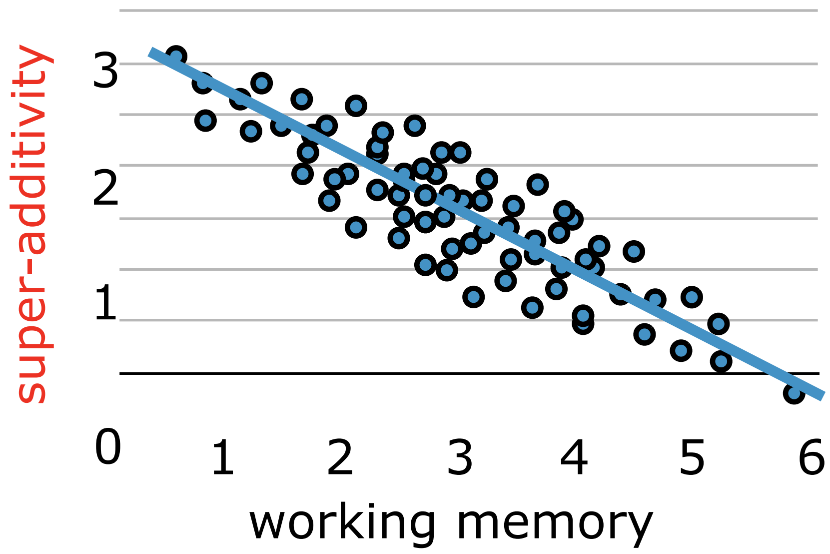
…

\+ WM interaction

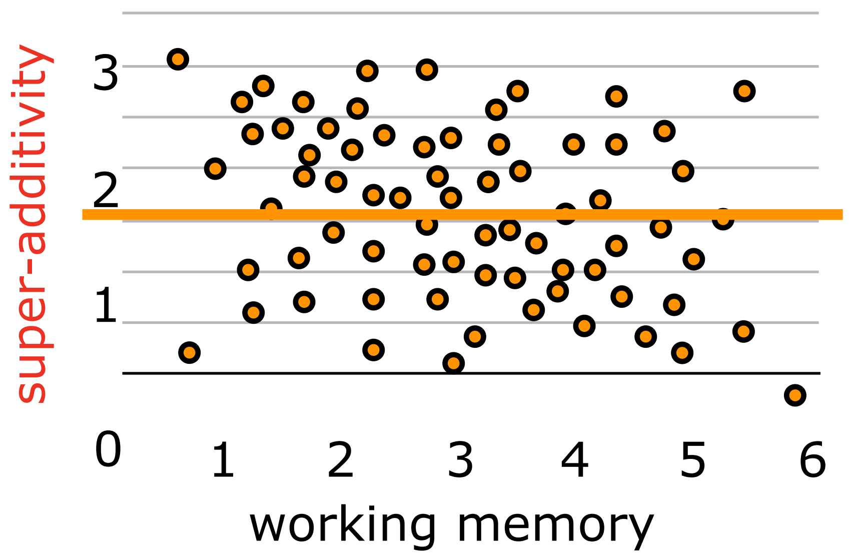
unacceptability/island effect

Potential predictions of this Reductionist Approach:

One possible prediction of the WM interaction mechanism is that as memory capacity changes so will the size of the island effect.



One prediction of a grammatical constraint is that there will be no relationship between WM and superadditivity.



This is a null result, but failure to find a correlation is **some amount of evidence** against the K&K93 theory (we need Bayes theorem to figure out how much).

Serial Recall

1. ________________
2. ________________
3. ________________
4. ________________
5. ________________
6. ________________

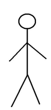
1. Listen to 6 words

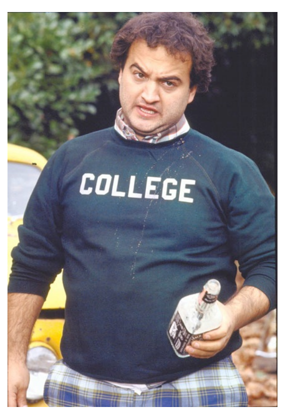
2. Write them down in the correct order

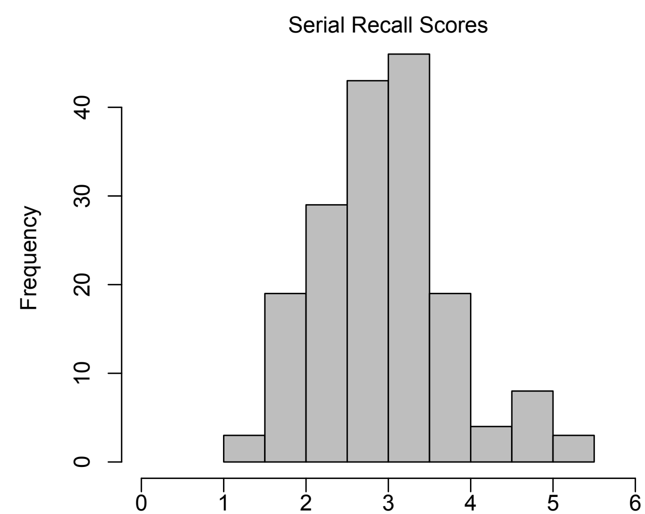
To avoid mnemonics, the 6 words were always chosen from a pool of 8 words

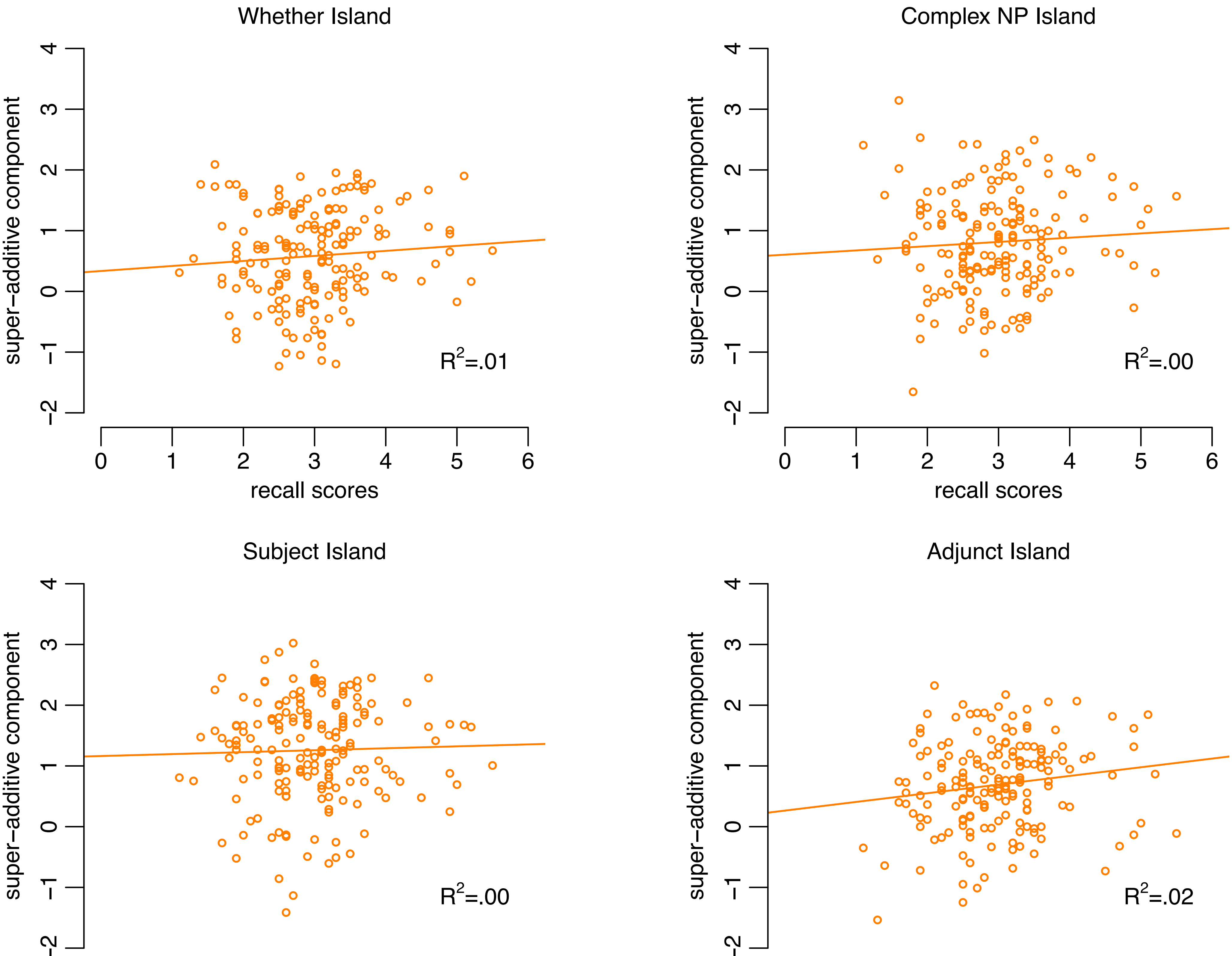
To avoid repetition, subjects were told to whisper the repeatedly during presentation

The words were matched for length (CVCVC), frequency, and neighborhood density

173 X 







Whether Island
super–additive component
recall scores
R²=.01
Complex NP Island
super–additive component
recall scores
R²=.00
Subject Island
super–additive component
R²=.00
Adjunct Island
super–additive component
R²=.02

A different correlation measure?

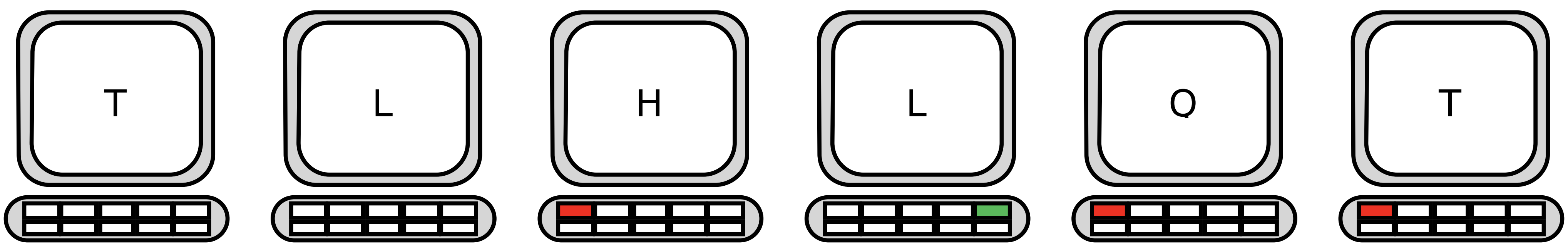
When looking for a correlation between data types, perhaps the most important decision is to choose the right data types to compare.

In the case of working memory, there are a number of tasks to choose from. It turns out that serial recall correlates well with most other WM tasks. But there is at least one task, the n-back task, that it does not correlate with. So we tested that too!

Example 1: the 2-back

Subjects are shown a series of letters, one at a time:

The task is to press a button if the letter shown also appeared 2 letters before.



A different correlation measure?

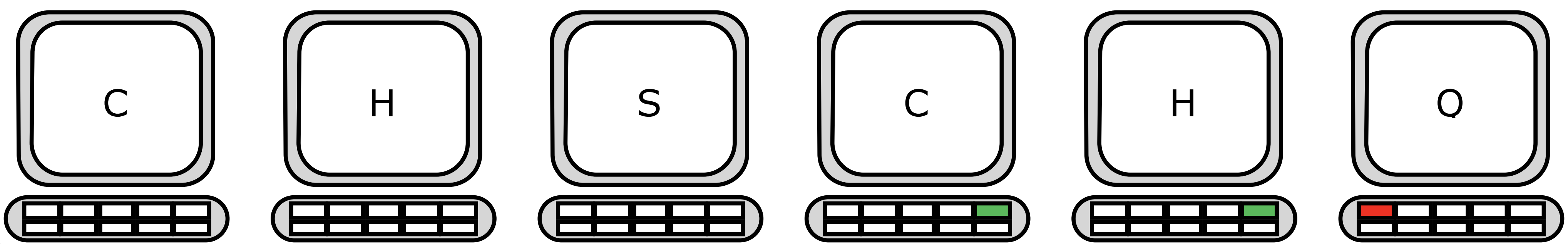
When looking for a correlation between data types, perhaps the most important decision is to choose the right data types to compare.

In the case of working memory, there are a number of tasks to choose from. It turns out that serial recall correlates well with most other WM tasks. But there is at least one task, the n-back task, that it does not correlate with. So we tested that too!

Example 2: the 3-back

Subjects are shown a series of letters, one at a time:

The task is to press a button if the letter shown also appeared 2 letters before.

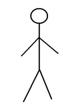


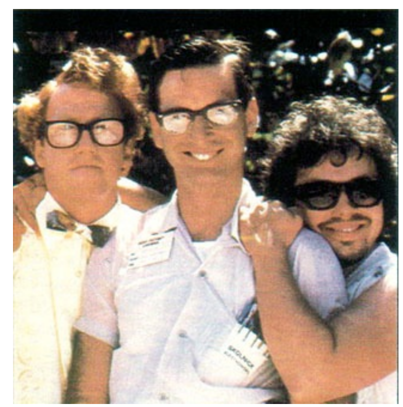
A different correlation measure?

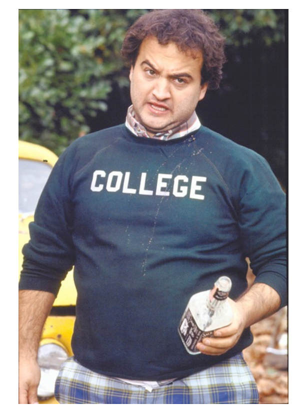
We tested serial recall, 2-back, 3-back, and 4-back during the testing session.

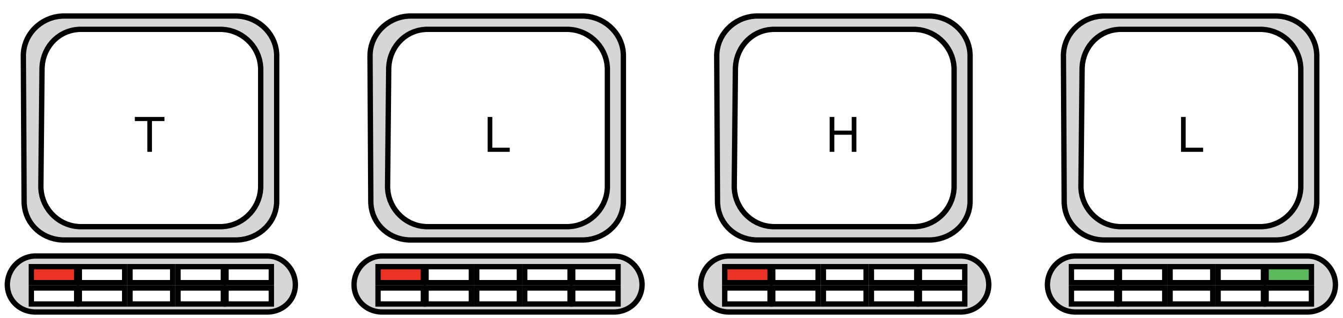
Serial Recall

1. ______________
2. ______________
3. ______________
4. ______________
5. ______________
6. ______________

173 X 





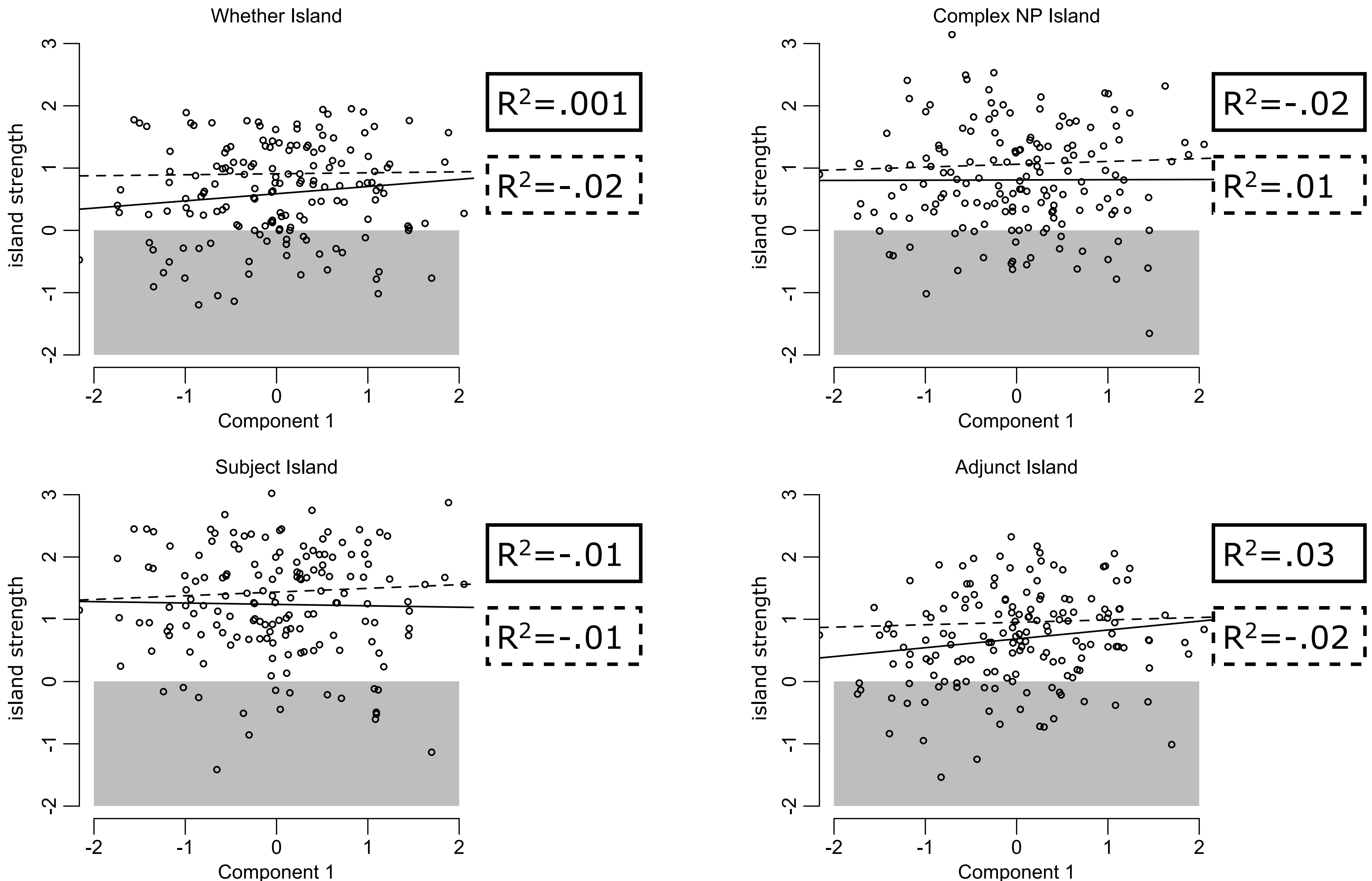


$n = 2$

$n = 3$

$n = 4$

PCA for 2,3,4-back

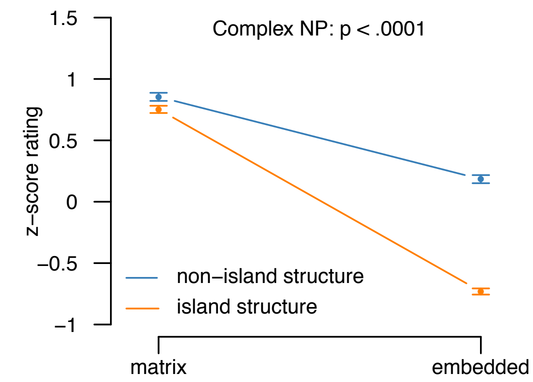


Simple Reductionism:

dependency cost

\+ structure cost

unacceptability/island effect



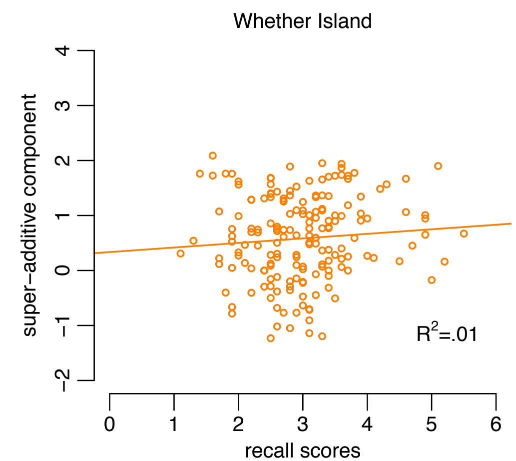
Complex Reductionism:

dependency cost

structure cost

\+ WM interaction

unacceptability/island effect



Grammatical Approach:

dependency cost

structure cost

\+ grammatical constraint

unacceptability/island effect

I am not sure that there can be direct evidence for this. Grammar is a hypothesis of exclusion.

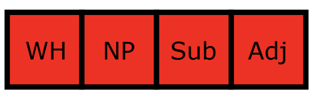
Additional evidence against reductionist approaches to island effects

This WM study is just one piece of evidence among many that island effects cannot be easily reduced to independent constraints on processing. Here are some others:

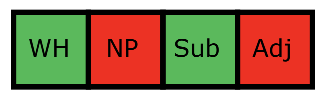
1. Cross linguistic variation.

   If there are languages that don't show all of the island effects that English shows, then it is very difficult to make the case that island effects are due to something like WM.

English

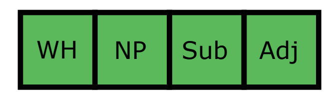


Italian



Rizzi 1982

Swedish



Engdahl 1980

   I don't think anybody wants to say that WM capacity varies as a function of country, or even as a function fo language.

   So this variation would have to be tied to some morpho-syntactic differences between the languages that interacts with WM. But there are no obvious candidates in these languages.

Additional evidence against reductionist approaches to island effects

This WM study is just one piece of evidence among many that island effects cannot be easily reduced to independent constraints on processing. Here are some others:

2. Binding-dependencies versus wh-dependencies (e.g., Yoshida et al. 2014)

   Binding-dependencies share many of the processing properties of wh-dependencies, especially in so-called "backward anaphora" constructions:

   search for an antecedent →

   He revealed that the studio that notified John Stewart about the new film selected a novel script.

   Despite these processing similarities, binding-dependencies don't show island effects. Binding dependencies show Binding Constraint effects (e.g., Principle C). And wh-dependencies don't show Binding Constraint effects, they show island effects.

   This difference is hard to explain under an approach that ties island effects to processing difficulty.

Additional evidence against reductionist approaches to island effects

This WM study is just one piece of evidence among many that island effects cannot be easily reduced to independent constraints on processing. Here are some others:

3. Subject parasitic gaps (Phillips 2006)

   All parasitic gap constructions are interesting because the parasitic gap appears inside an island, yet the sentence is acceptable.

   Which school did [the proposal to expand __] ultimately overburden __?

   ↑ typically bad ↑ licensor

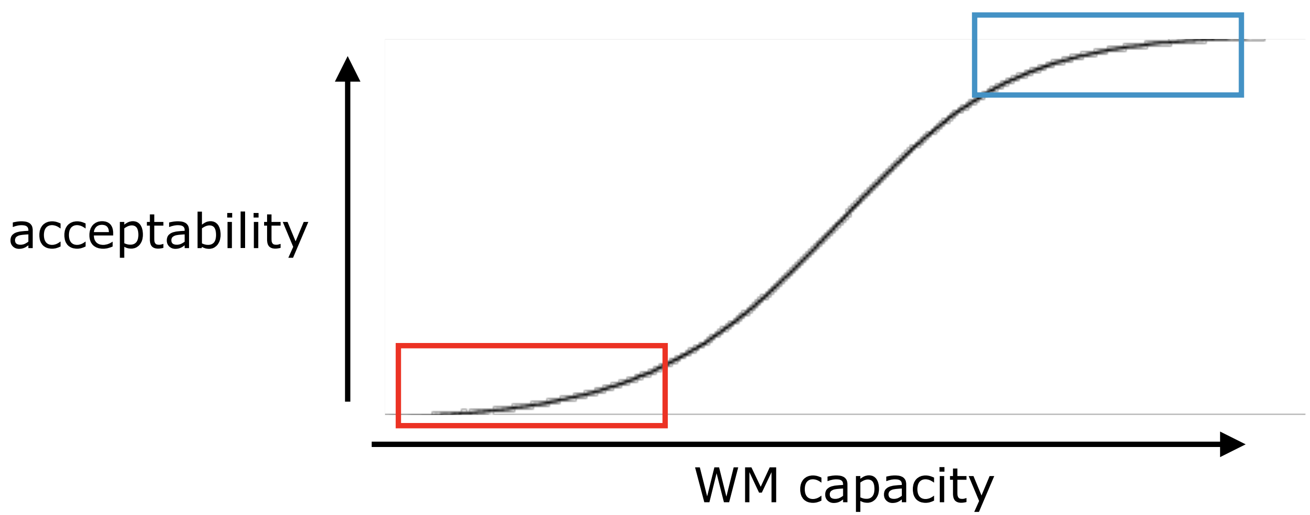
   Subject parasitic gaps are particularly interesting because the typically-bad gap appears first in the sentence, before the gap that licenses it.

   Phillips 2006 shows that the parser attempts to fill the typically-bad gap immediately (before it sees the licensor). This means that the parser can fill gaps inside of subject island structures. This means that whatever is causing the unacceptability in subject islands is not due to a failure in parsing. It must be something else, such as a grammatical constraint).

What do reductionists think about all of this?

This is difficult to answer because there is currently no explicit reductionist theory that can account for all of these facts.

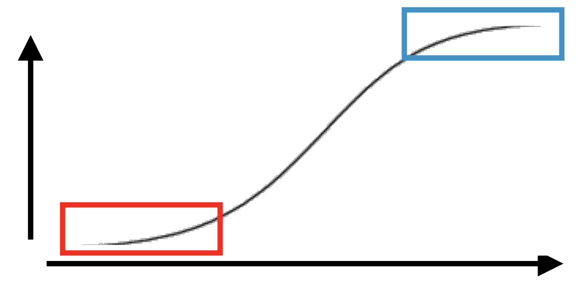
Hofmeister et al. 2012 is a series of two replies to the WM study presented earlier. In it, they suggest that the mapping between WM capacity and acceptability judgments could potentially be non-linear (sigmoidal or step), and that (English) speakers might never have enough WM to move into the acceptable section:



To be fair, the previous study only tested a linear mapping between WM and acceptability, and assumed that variation would be large enough to find a relationship.

What do reductionists think about all of this?

The first problem with the opaque-mapping hypothesis is that, while it is logically possible, it is not directly testable. It stipulates that there is not enough variation in WM to actually see the function.



The second problem is that it begins to betray the spirit of the reductionist approach.

The value of the reductionist approach is that it simplifies the explanation of island effects (by eliminating complex grammatical constraints).

But the opaque-mapping hypothesis doesn't reduce the complexity of the explanation. It simply shifts the complexity into a different component (WM). This is just a rehash of the classic boundary problem!

Unfortunately, this shift is untestable (at least directly) so it leaves little room for an empirical continuation of the conversation.

THANK YOU!
and thank you to my generous collaborators!

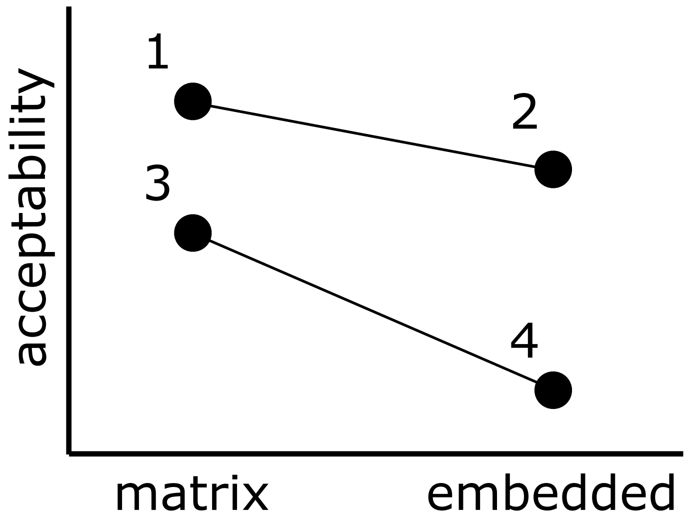
Colin Phillips
UMD

Matt Wagers
UC Santa Cruz

Extra slides about factorial designs

Some nice properties of factorial designs

1. They allow us to quantify two confounds simultaneously (to control for more would require additional factors, e.g. 2x2x2, or 2x2x2x2)



confound captured by factor 1: 1 - 2

confound captured by factor 2: 1 - 3

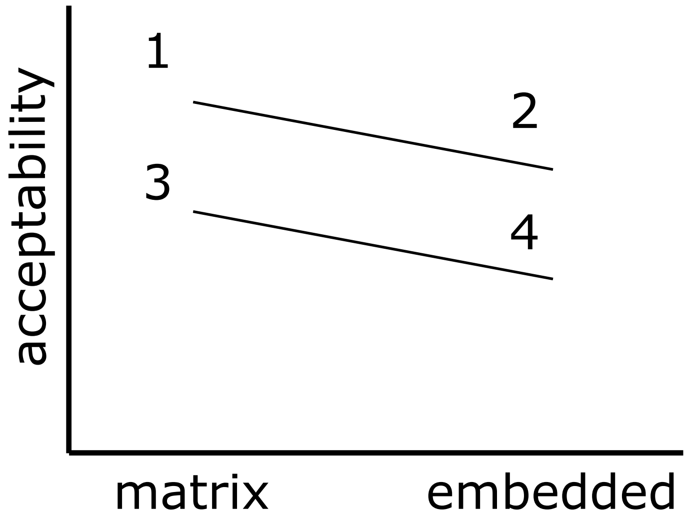
full effect: 1 - 4

target effect: (1-4) - (1-2) - (1-3)

The trick is to capture each confound in a factor based on the sentence of interest.

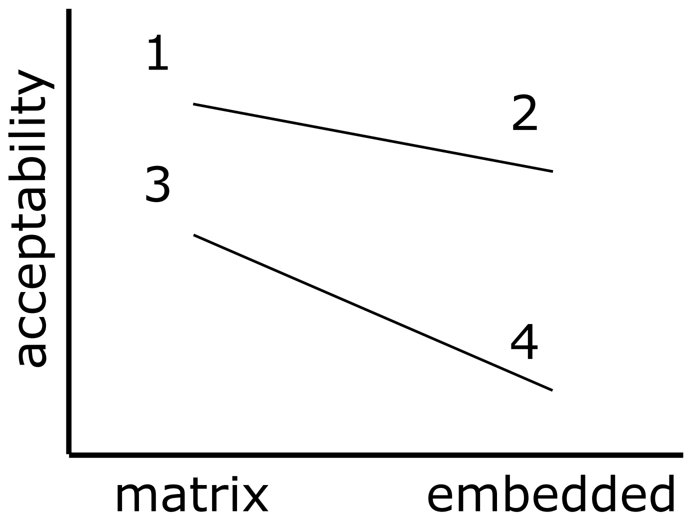
Some nice properties of factorial designs

2. The results can be interpreted visually using an interaction plot.



Parallel lines indicate (linear) additivity of the two factors. This means there is no target effect. There is just the two effects of the two factors (the two confounds).

target effect: (1-4) - (1-2) - (1-3) = 0

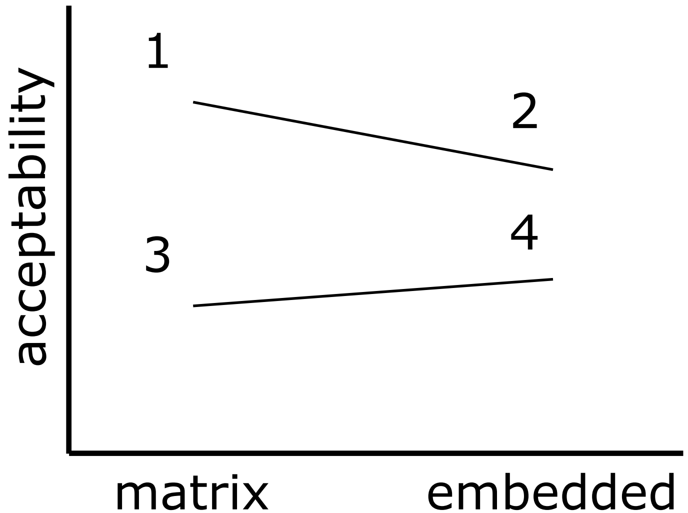


Non-parallel lines that open to the right indicate superadditivity of the two factors. This means the two factors don't account for the entire effect, therefore there is a target effect that needs to be explained.

target effect: (1-4) - (1-2) - (1-3) = >0

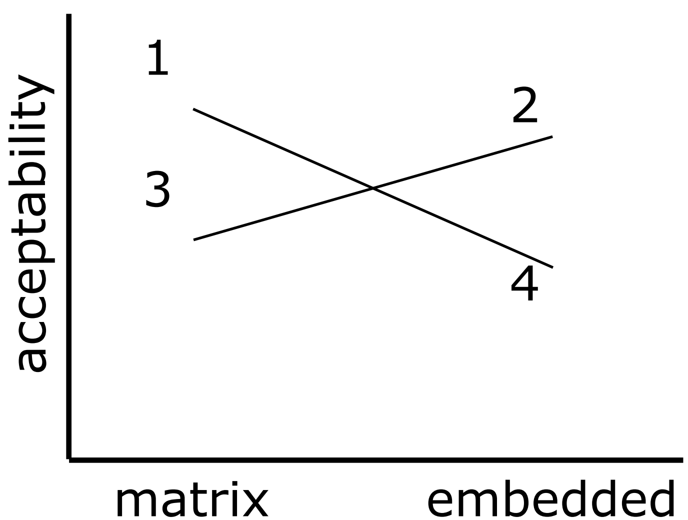
Some nice properties of factorial designs

There are (at least) two other patterns that could arise. But these will be rarer in experimental syntax.



Non-parallel lines that open to the left indicate sub-additivity of the two factors. This means the target effect is less than what we'd expect from the two factors.

target effect: (1-4) - (1-2) - (1-3) = <0



Crossing lines indicate a non-monotonic interaction. This is difficult to interpret, because it isn't the case that the factors are confounds to be separated from the target effect.

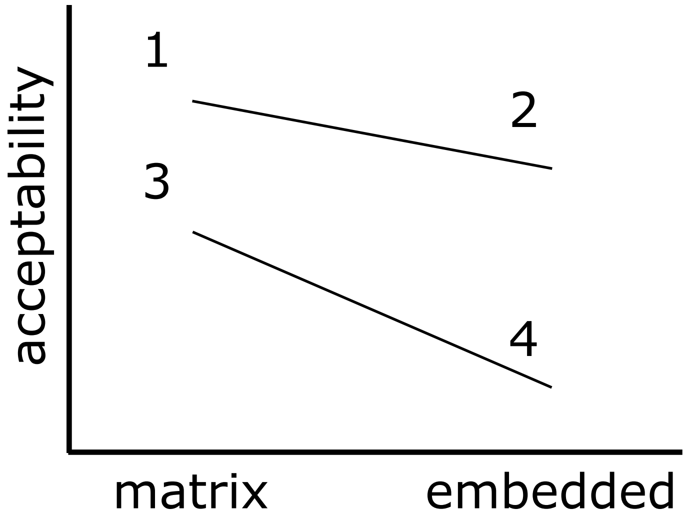
target effect: (1-4) - (1-2) - (1-3) = ??

Some nice properties of factorial designs

3. They allow us to calculate the size of the target effect. Here is the full calculation:

target effect = the full effect — confound 1 — confound 2.

(1-4) (1-2) (1-3)



But there is also a short version called a **differences-in-differences score**:

DD = (2-4) - (1-3)

or

DD = (3-4) - (1-2)

Basically, you subtract either both columns or both rows from each other. You can use algebra to see it is identical to the full equation above.

Some nice properties of factorial designs

4. They allow us to subtract out additional potential confounds without adding additional factors.

Condition 1:	Who __ thinks that John bought a car	matrix | that
Condition 2:	What do you think that John bought __	embedded | that
Condition 3:	Who __ wonders whether John bought a car	matrix | whether
Condition 4:	What do you wonder whether John bought __	embedded | whether

One potential confound in these conditions that could contribute to our target effect is the type of wh-word: who versus what.

In order to see the effect of wh-word, we would need to add another factor, leading to a 2x2x2 design.

That doubles the size of our experiment, and creates extra work to see an effect that we aren't really interested in. So what can we do?

Some nice properties of factorial designs

If we look again at the calculation of the target effect using the DD score equation, we can see that we can subtract out this effect if we put the two levels of the confound in our conditions in one of two specific ways.

Option 1: put the levels of the confound in 1 & 3, and 2 & 4

This is what we did in this experiment. The who wh-words are in 1&3, and the what wh-words are in 2&4.

Condition 1:	Who __ thinks that John bought a car	matrix | that
Condition 2:	What do you think that John bought __	embedded | that
Condition 3:	Who __ wonders whether John bought a car	matrix | whether
Condition 4:	What do you wonder whether John bought __	embedded | whether

When you put the levels of the confound in the DD equation, you can see that they subtract out, leaving nothing behind!

DD = (2-4) - (1-3)

DD = (what-what) - (who-who) = no effect of wh-word!

Some nice properties of factorial designs

If we look again at the calculation of the target effect using the DD score equation, we can see that we can subtract out this effect if we put the two levels of the confound in our conditions in one of two specific ways.

Option 2: put the levels of the confound in 1 & 2, and 3 & 4

This is not what we did in this design, because it doesn't make much sense in this construction. But we could have. Here is what it would look like:

Condition 1:	Who __ thinks that John likes a car	matrix | that
Condition 2:	Who do you think that John likes __	embedded | that
Condition 3:	What __ wonders whether John likes a car	matrix | whether
Condition 4:	What do you wonder whether John likes __	embedded | whether

When you put the levels of the confound in the DD equation, you can see that they subtract out, again leaving nothing behind!

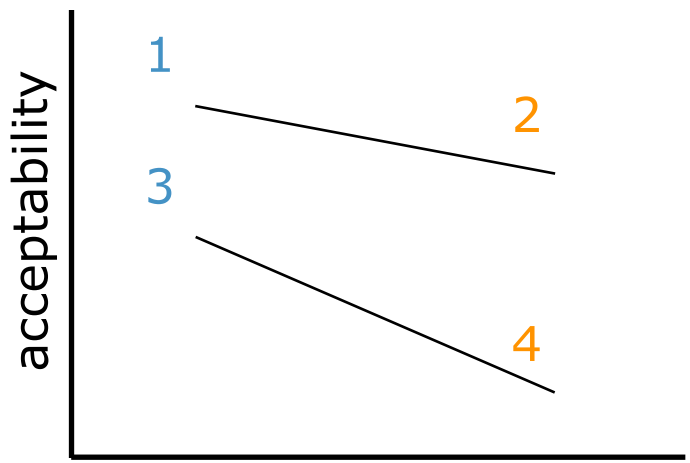
DD = (2-4) - (1-3)

DD = (who-what) - (who-what) = no effect of wh-word!

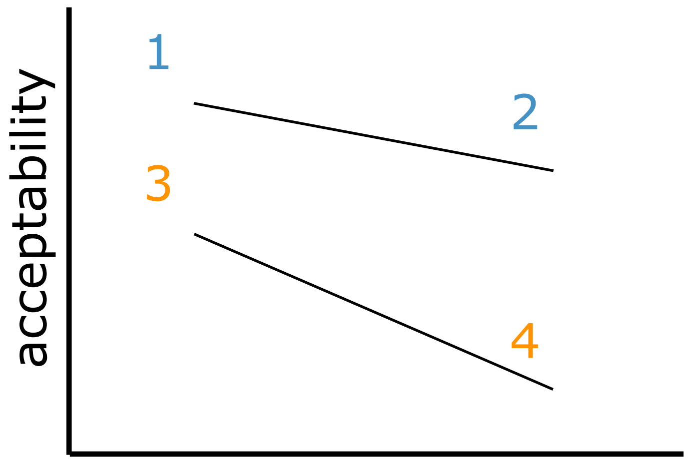
Some nice properties of factorial designs

4. They allow us to subtract out additional potential confounds without adding additional factors.

Basically, as long as the two levels of the potential confound are split along a row or column in the interaction plot, they will subtract out in the calculation of the target effect:



or

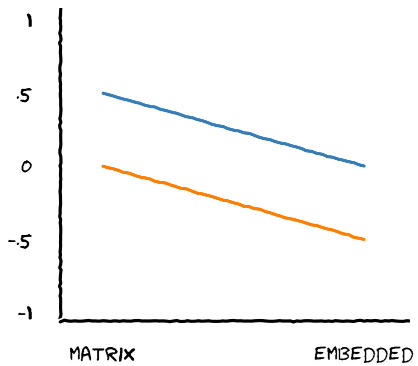


This is a powerful tactic for controlling potential confounds.

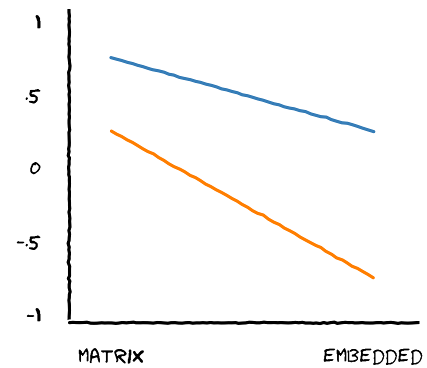
However, if the levels are split along the diagonals (1/4 or 2/3) the subtraction will not succeed and the confound will contaminate the results.

Some nice properties of factorial designs

5. Finally, they provide a necessary (but not sufficient) condition for the postulation of a syntactic constraint.



Reductionism holds



Grammar or
More Complex Reductionism

If the factors are designed using reductionist components, then the superadditive component is the effect that needs to be explained.

In order for a syntactic constraint to have something to do, there must be an effect that needs to be explained!

Some nice properties of factorial designs

1. They allow us to control for two confounds simultaneously (to control for more would require additional factors, e.g. 2x2x2, or 2x2x2x2)

2. The results can be interpreted visually using an interaction plot.

3. They allow us to calculate the size of the target effect using differences-in-differences scores

4. They allow us to subtract out additional potential confounds without adding additional factors, as long as the confound’s levels are split appropriately

5. Finally, they provide a necessary (but not sufficient) condition for the postulation of a syntactic constraint.

These properties hold generally for all crossed factorial designs. However, general practice is to start with 2x2 designs first, and only include additional factors if there are additional properties that must be quantified to test an hypothesis.