Beyond Surprising: English Event Structure in the Maze

Lisa Levinson
University of Michigan

Experiments in Meaning (ELM) 2

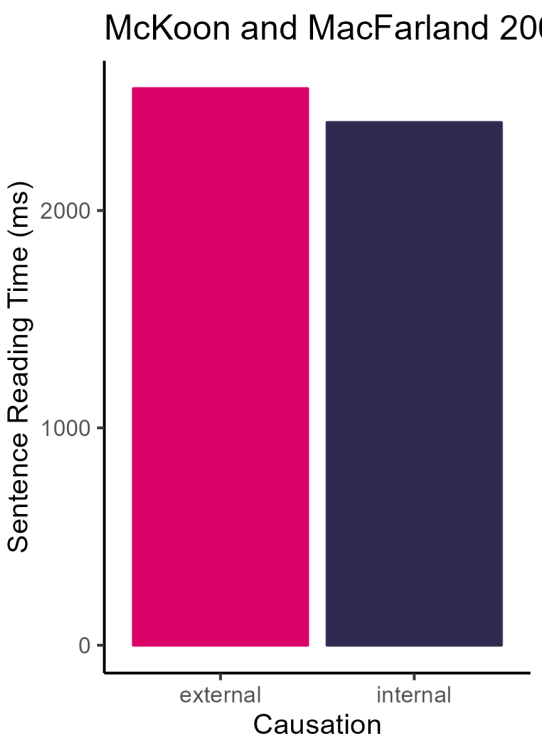
May 18, 2022

Rtistry by Danielle Navarro, CC-BY-SA

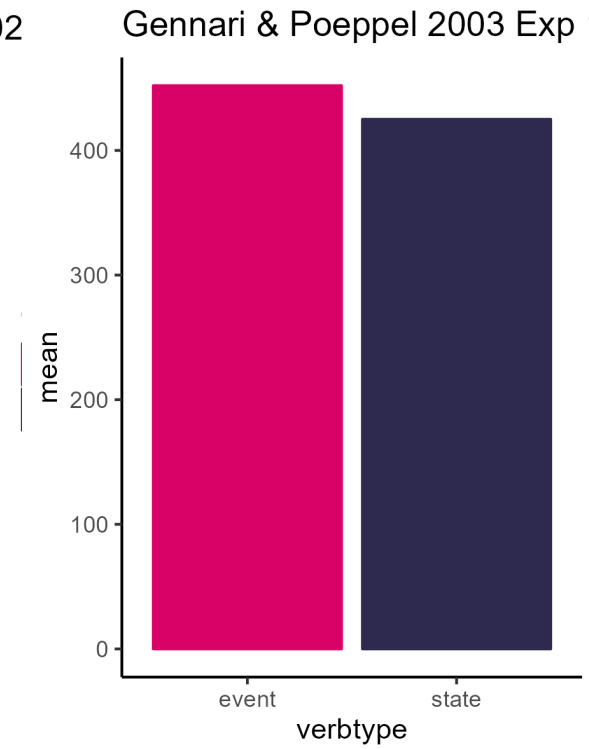
Various Event Structure Costs

Do they replicate?

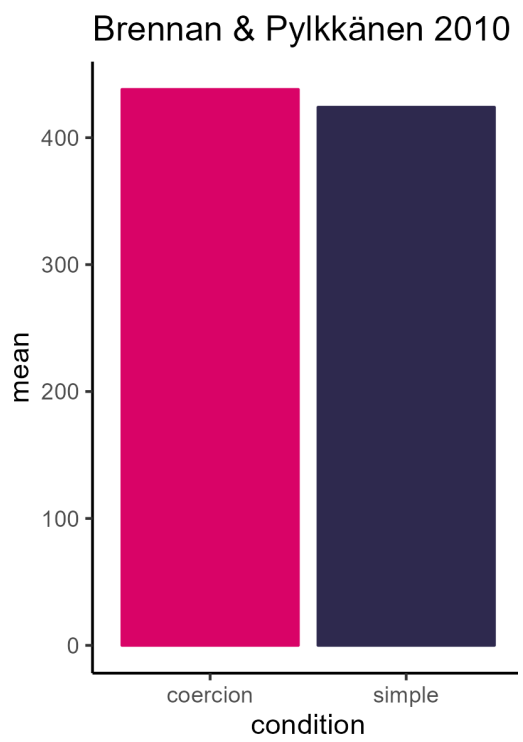
What is the link to semantic representations & processes?



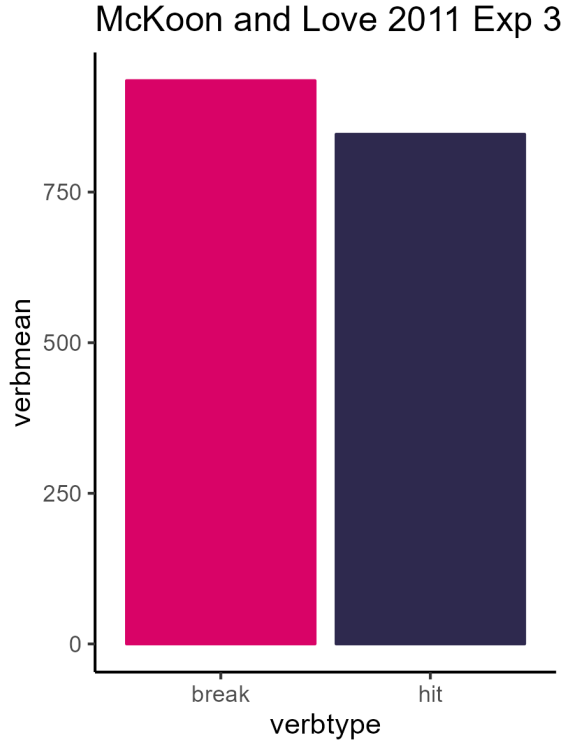
External vs. Internal Causation



Eventive vs. Stative



Aspectual Coercion



Result (Causative) vs. Manner

Agenda

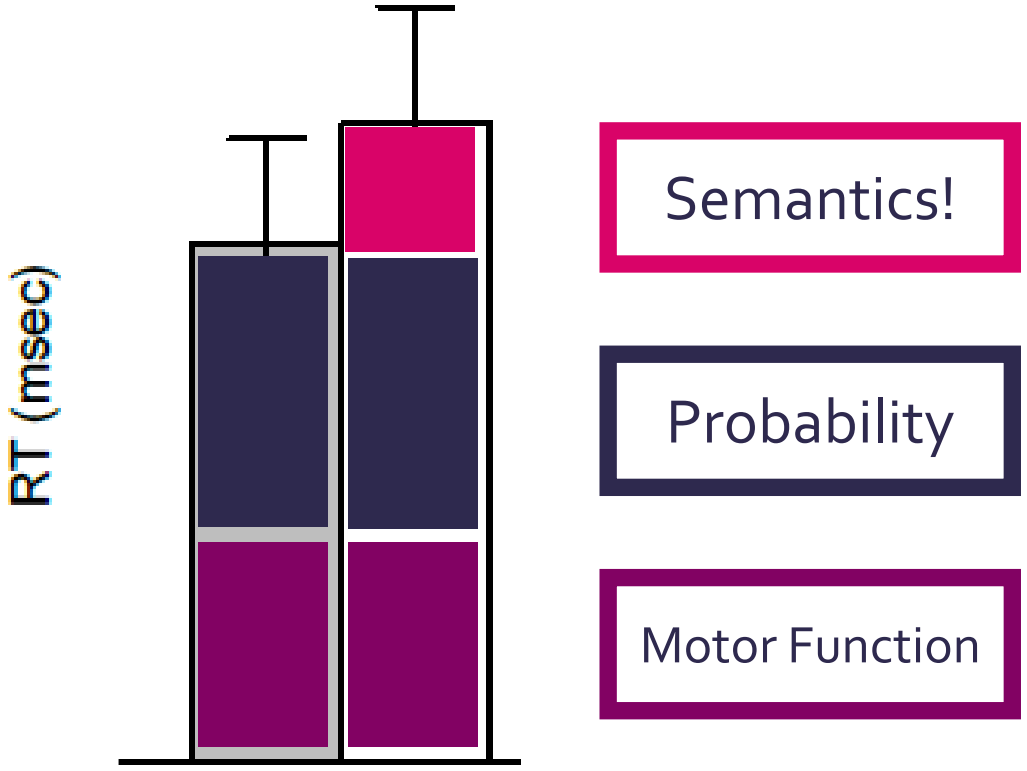
1. Semantics Beyond vs. Within Surprisal
2. Replicating Causativity Cost
3. Surprisals with Event Structure
4. Deeper exploration with the Maze Task

Agenda

1. **Semantics Beyond vs. Within Surprisal**
2. Replicating Causativity Cost
3. Surprisals with Event Structure
4. Deeper exploration with the Maze Task

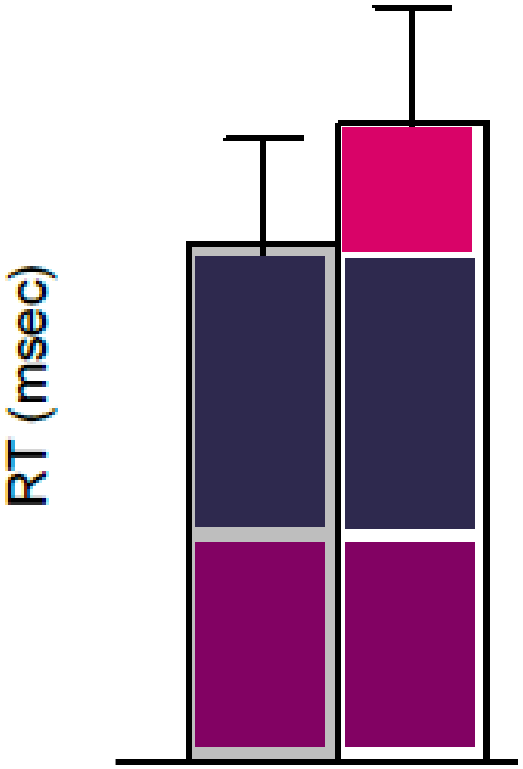
Can we tease apart semantics beyond (and within) probabilistic prediction?

Experiment Design Assumptions



Experiment Design Assumptions

Surprisal and Prediction (Hale 2001, Levy 2008)

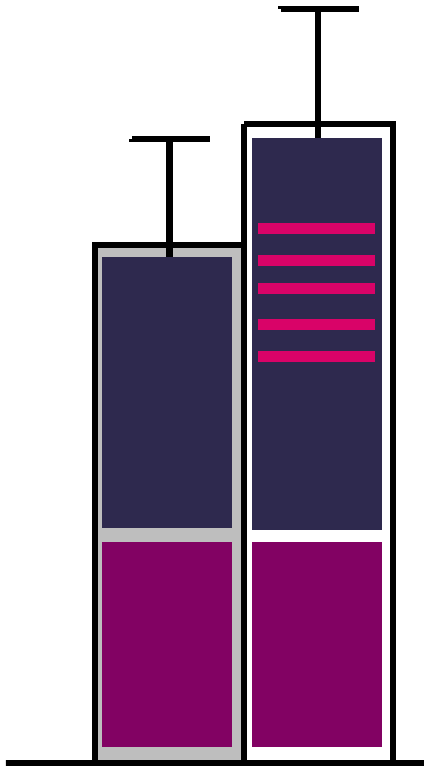


Semantics!

beyond surprisal!

Probabilistic Prediction / Surprisal

Motor Function



Semantics?

within surprisal!

Surprisal

Motor Function

What if…?

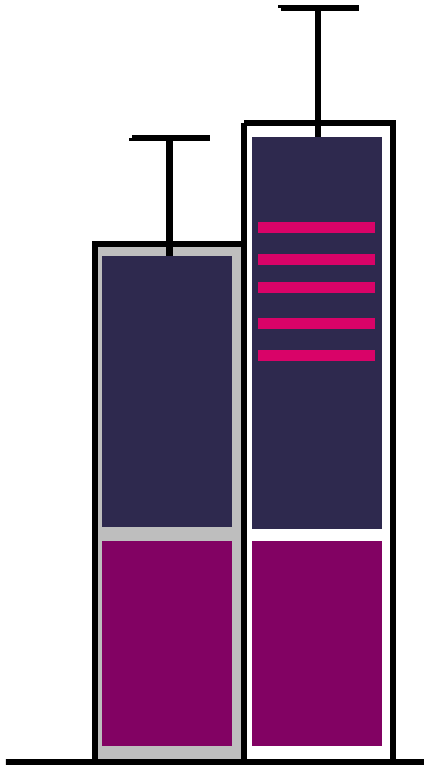
The effect is matched by probabilities?

How do the surprisals get to be what they are?

Human syntax, semantics…?

Is surprisal a confound, or a co-conspirator? A bottleneck?

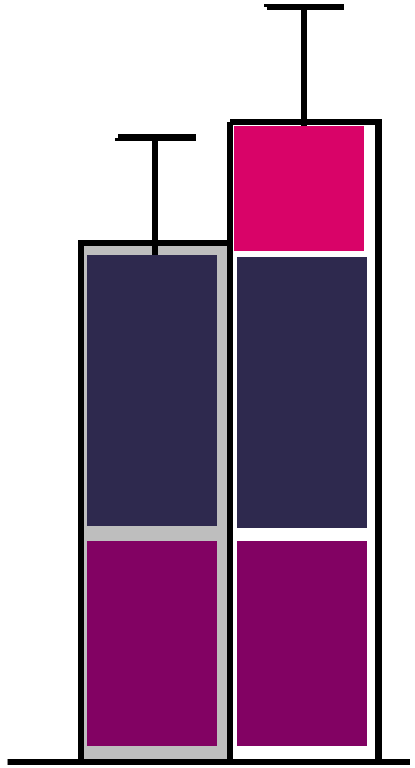
Semantics within surprisal



Surprisal

Motor Function

Semantics beyond surprisal

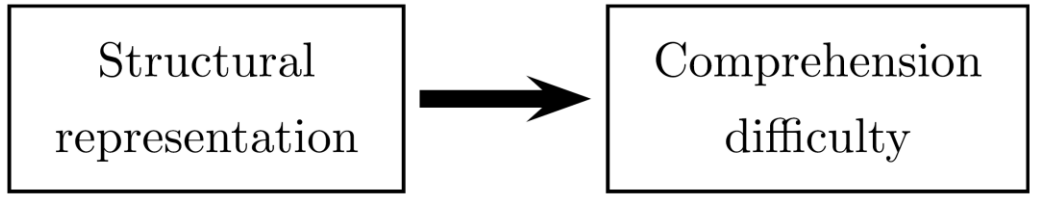


Surprisal

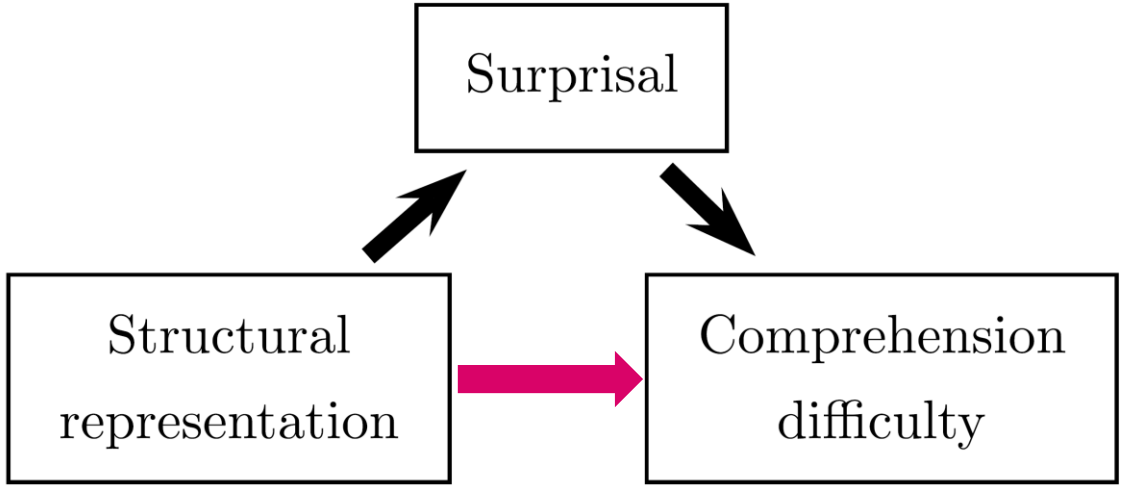
Motor Function

Could have both, or neither!

Causal Bottleneck (Levy 2008)



(a) Direct effect of representation on processing



(b) Surprisal as a causal bottleneck mediating effect of representation on processing

Why not both?

Agenda

1. Semantics Beyond vs. Within Surprisal
2. **Replicating Causativity Cost**
3. Surprisals with Event Structure
4. Deeper exploration with the Maze Task

A convenient replication: Levinson & Brennan (2016)

Hypothesis

Based on prior findings of lexical event structure complexity:

Processing cost for **causativity** within result verbs

vs. across manner/result, McKoon & Love 2011

Manner vs. Result Event Structure

Result/Change of State/Causative:

1) The door opened (inchoative/anticaus.)
2) The child opened the door. (causative)

Result

Manner/Implicit Theme:

3) The professor rea (implicit)
4) The professor rea (explicit)

Manner

Semantic Assumption (Simplified)

For English:

More causative in transitive (Dowty 1979, Bittner 1999)

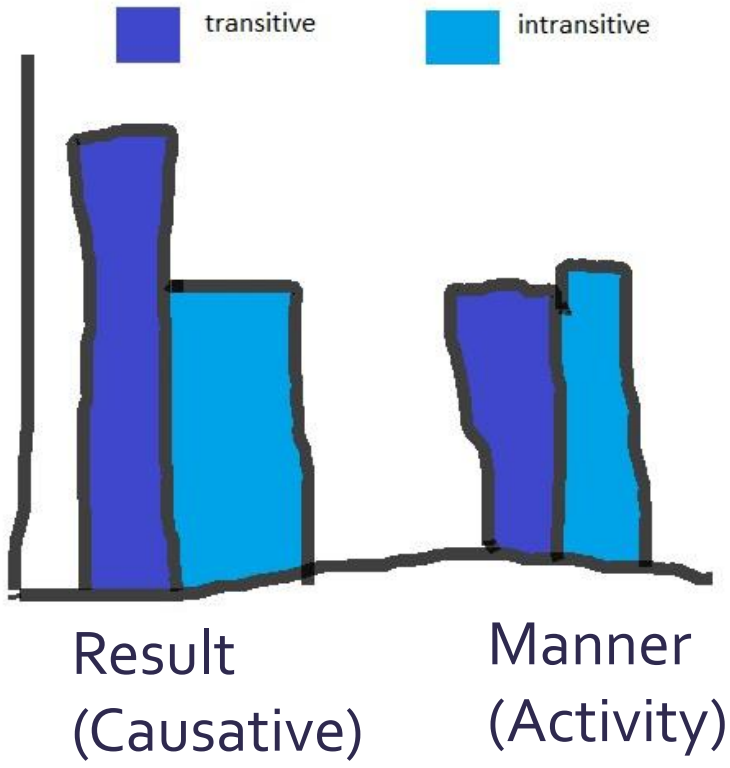
BECOME opened(the door)

the child CAUSE BECOME opened(the door)

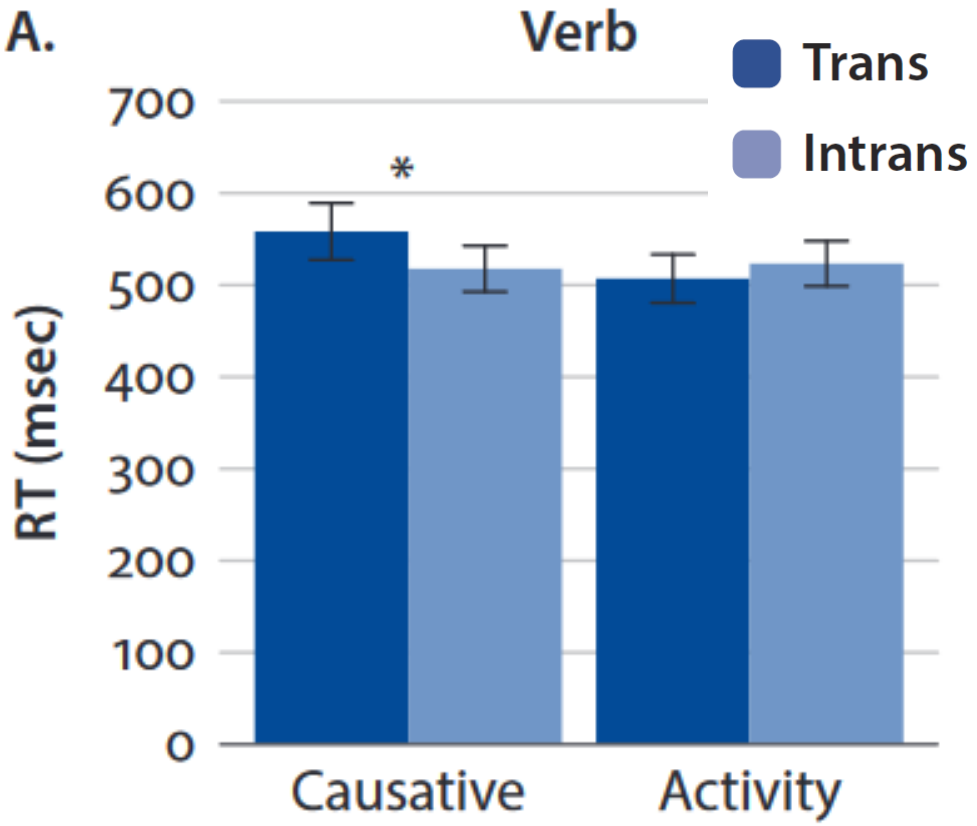
Transitive entails intransitive

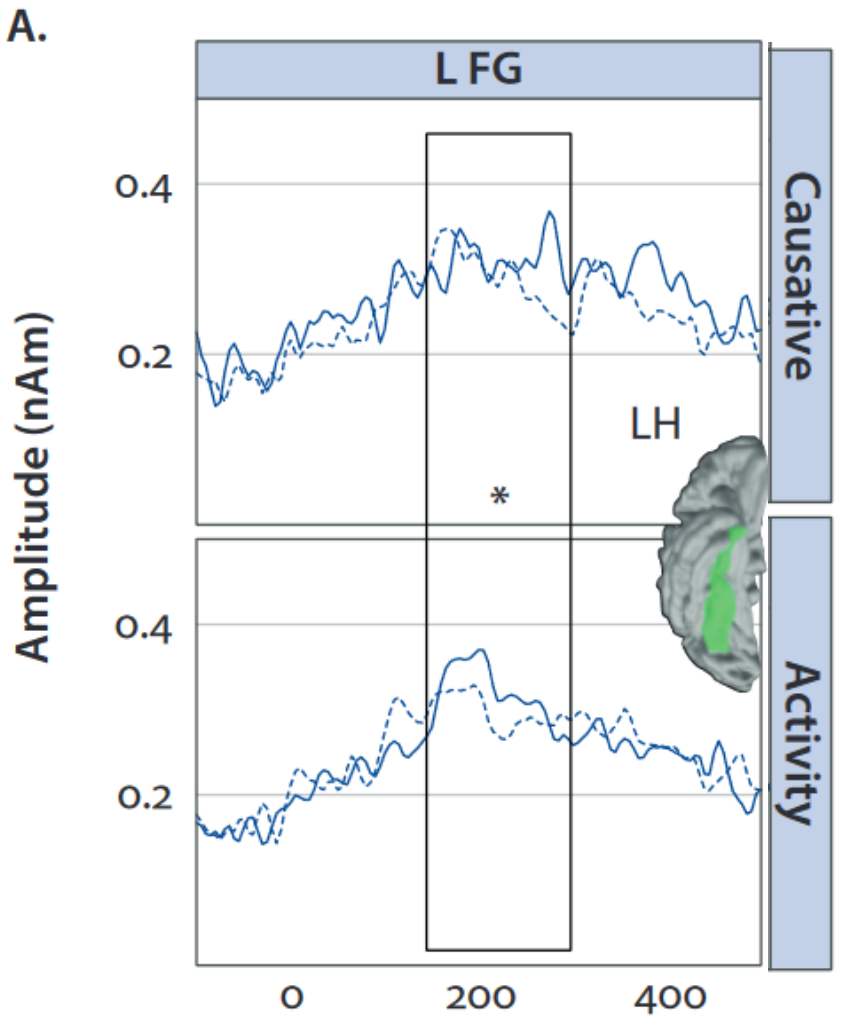
Processing Prediction: Interaction

Assuming a processing cost due to event structure complexity (causativity):



Levinson & Brennan 2016





Stimuli Design

Questions to indicate transitivity prior to verb:

1) What did the child open?
2) When did the door open?

Inanimate subject biases towards intransitive reading in (2)

Any "cost" of ambiguity works against hypothesized effect

Stimuli

Wh	did	D	N	V	P	D	N	Args	Verb Type
What	did	the	cook	thaw	in	the	cafeteria?	2	Result
When	did	the	popsicle	thaw	in	the	cafeteria?	1	Result
What	did	the	teacher	hum	for	the	students?	2	Manner
When	did	the	teacher	hum	for	the	students?	1	Manner

87 pairs (43 result, 44 manner), 110 fillers

Latin-square counter-balanced

Verb frequency and sentence acceptability matched

Replication 1: web-based self-paced reading time

Exp 1: Collection & Analysis

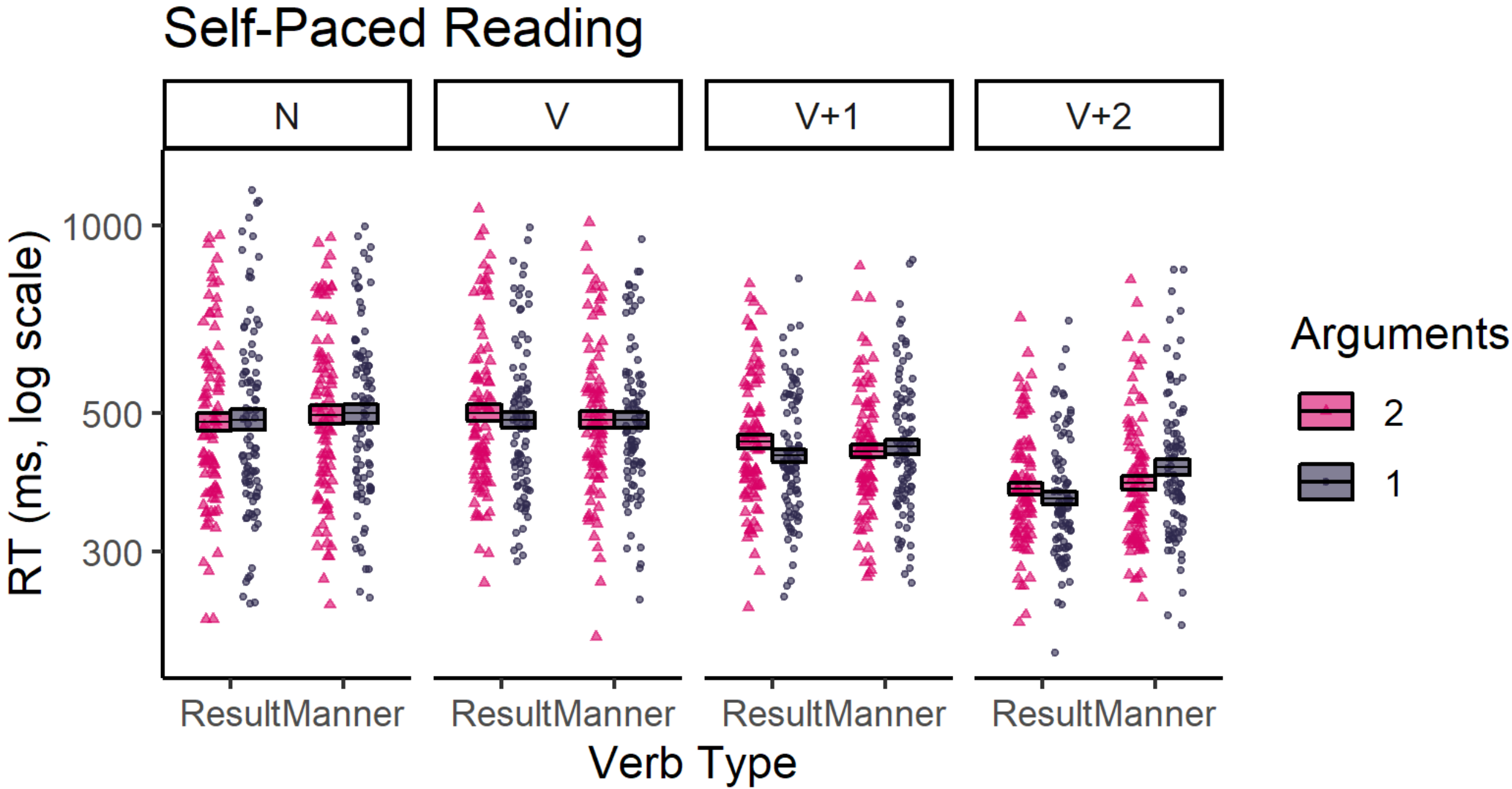
Data Collection

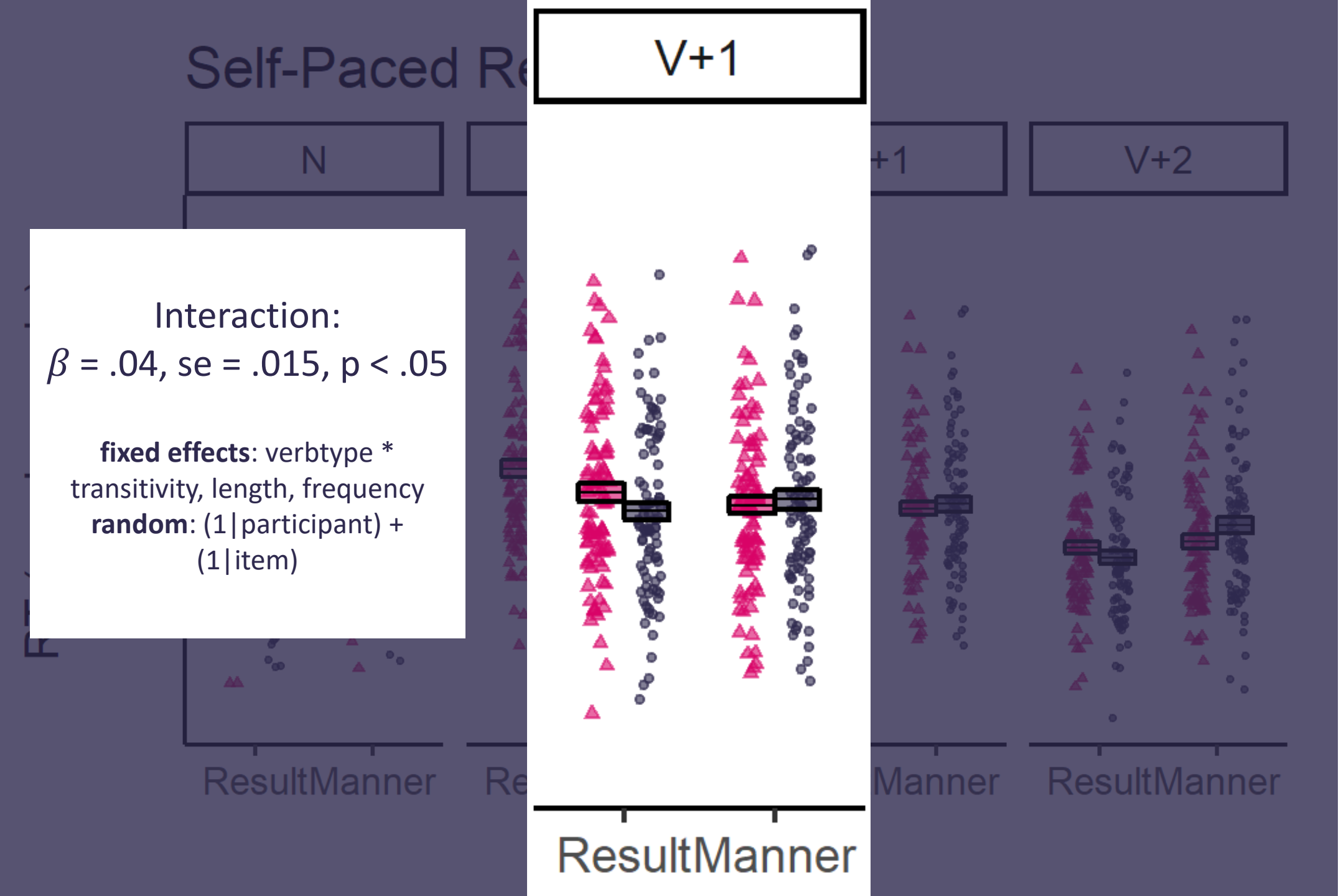
90 participants

Self-paced reading moving window via Ibexfarm

Statistical Analysis

Hierarchical regression (Gelman & Hill 2006; Baayen et al. 2008) with Participants and Items as random effects (more detail later)


Self-Paced Reading
N
V
V+1
V+2
RT (ms, log scale)
1000
500
300
ResultManner
ResultManner
ResultManner
ResultManner
Verb Type
Arguments
2
1

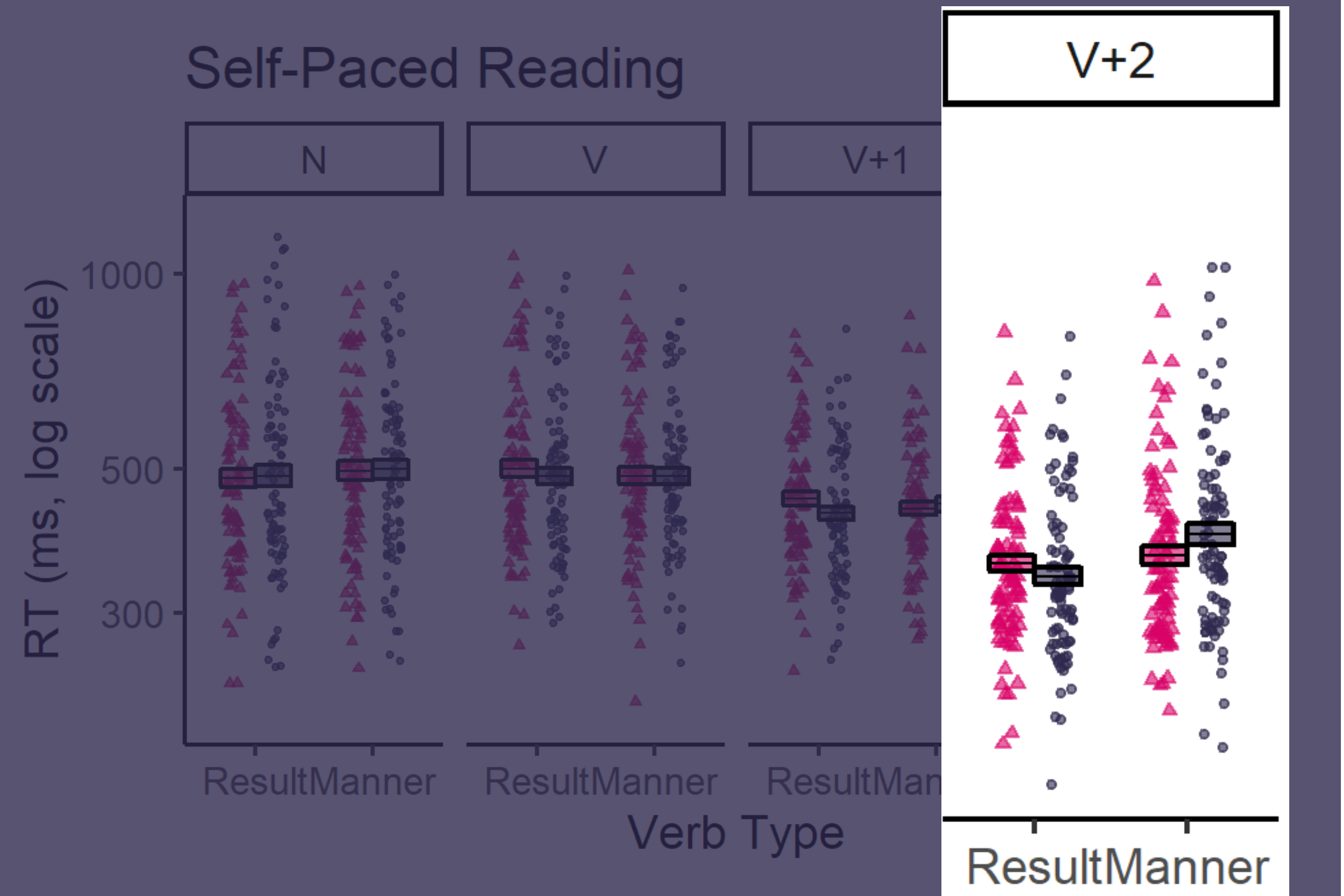


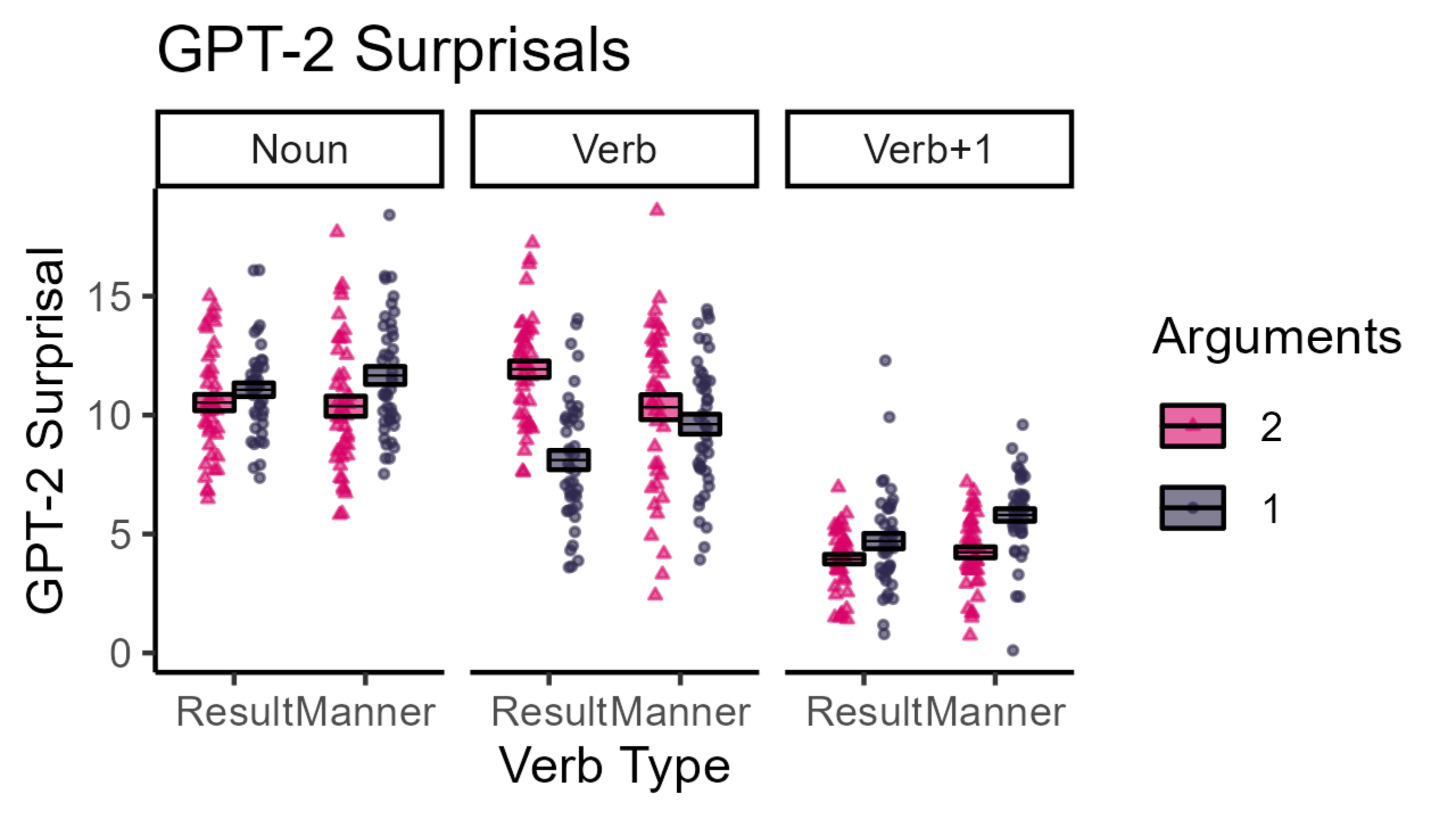
Interaction:

β = .04, se = .015, p < .05

fixed effects: verbtype * transitivity, length, frequency

random: (1|participant) + (1|item)


Self-Paced Reading
N
V
V+1
V+2
1000
500
300
RT (ms, log scale)
ResultManner
ResultManner
ResultMan
ResultManner
Verb Type
Arguments
2
1

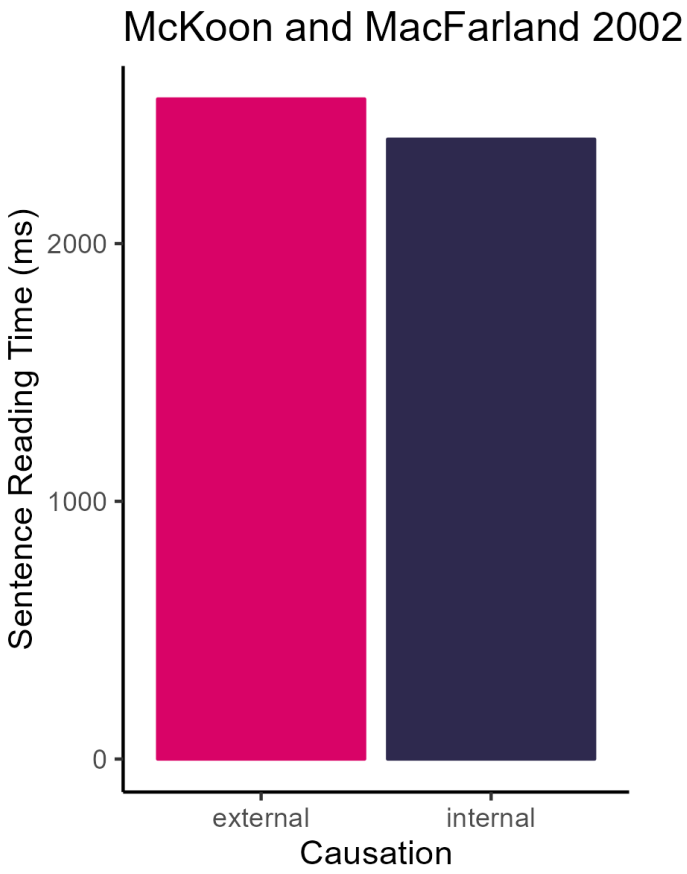

GPT-2 Surprisals
Noun
Verb
Verb+1
GPT-2 Surprisal
15
10
5
0
ResultManner
ResultManner
ResultManner
Verb Type
Arguments
2
1

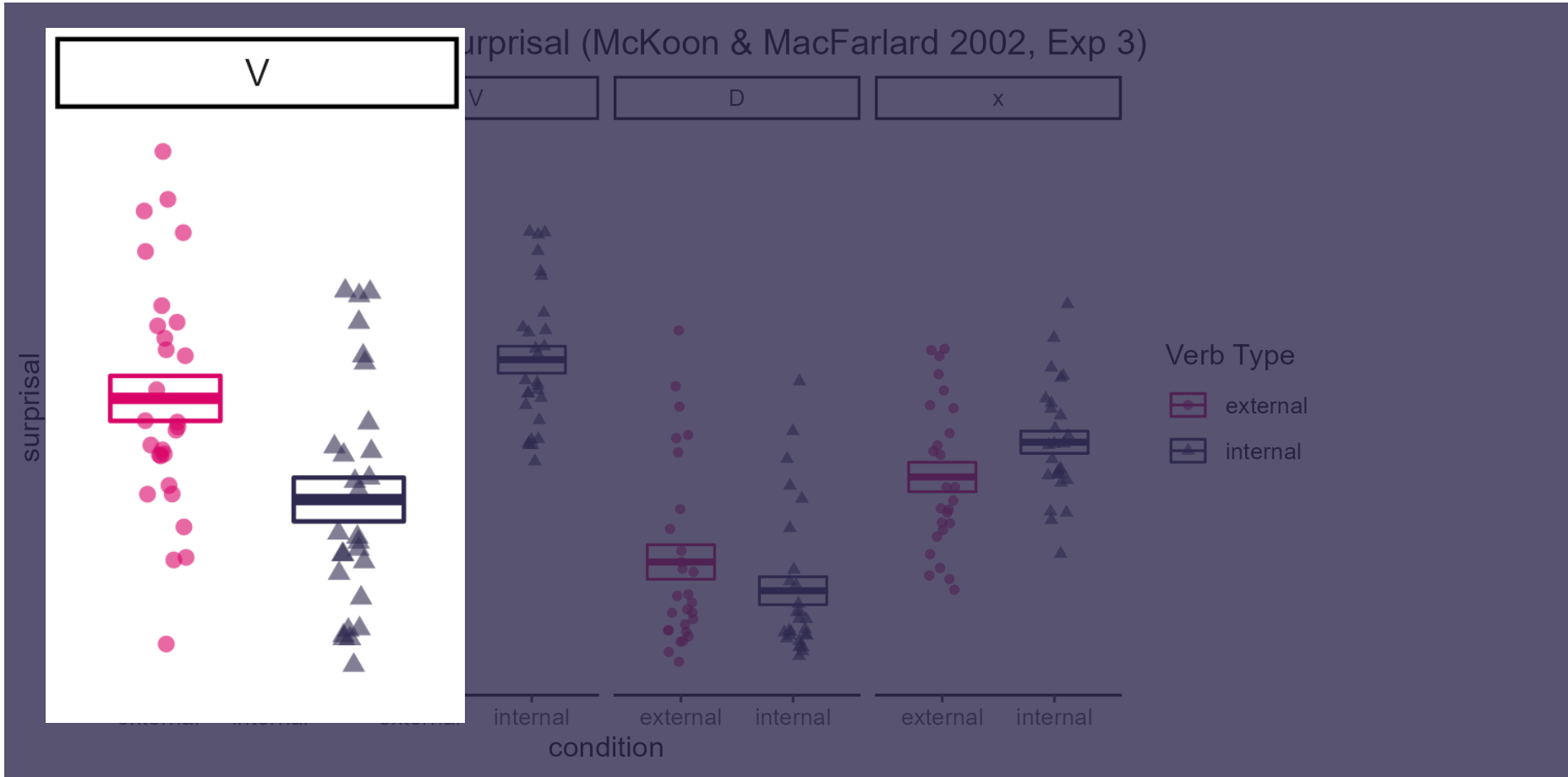
Agenda

1. Semantics Beyond vs. Within Surprisal
2. Replicating Causativity Cost
3. **Surprisals with Event Structure**
4. Deeper exploration with the Maze Task

McKoon & MacFarland 2002

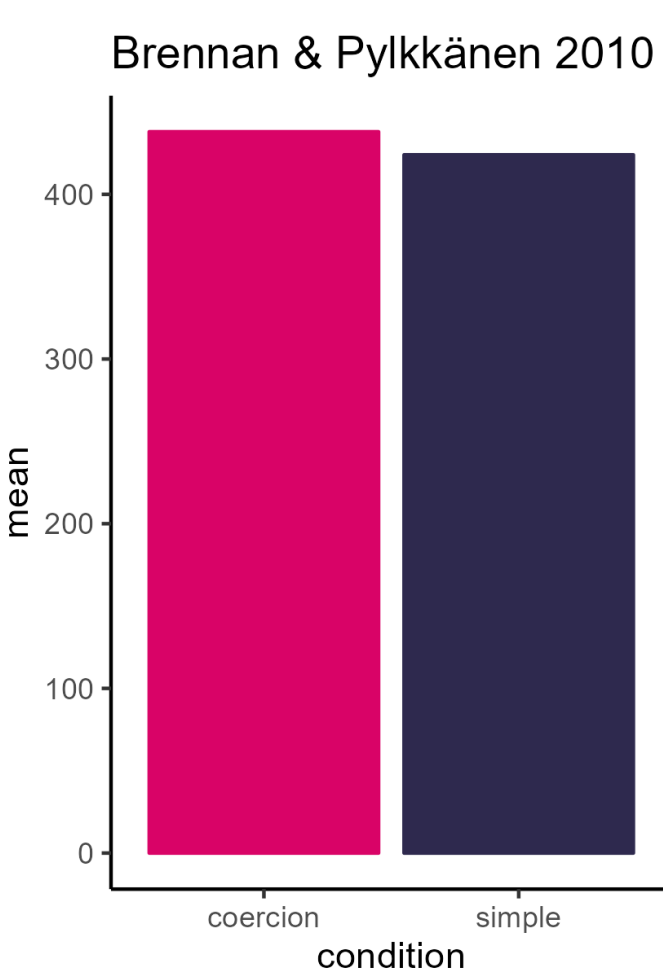
The illness atrophied the lower leg. (external)
The plants bloomed yellow blossoms. (internal)

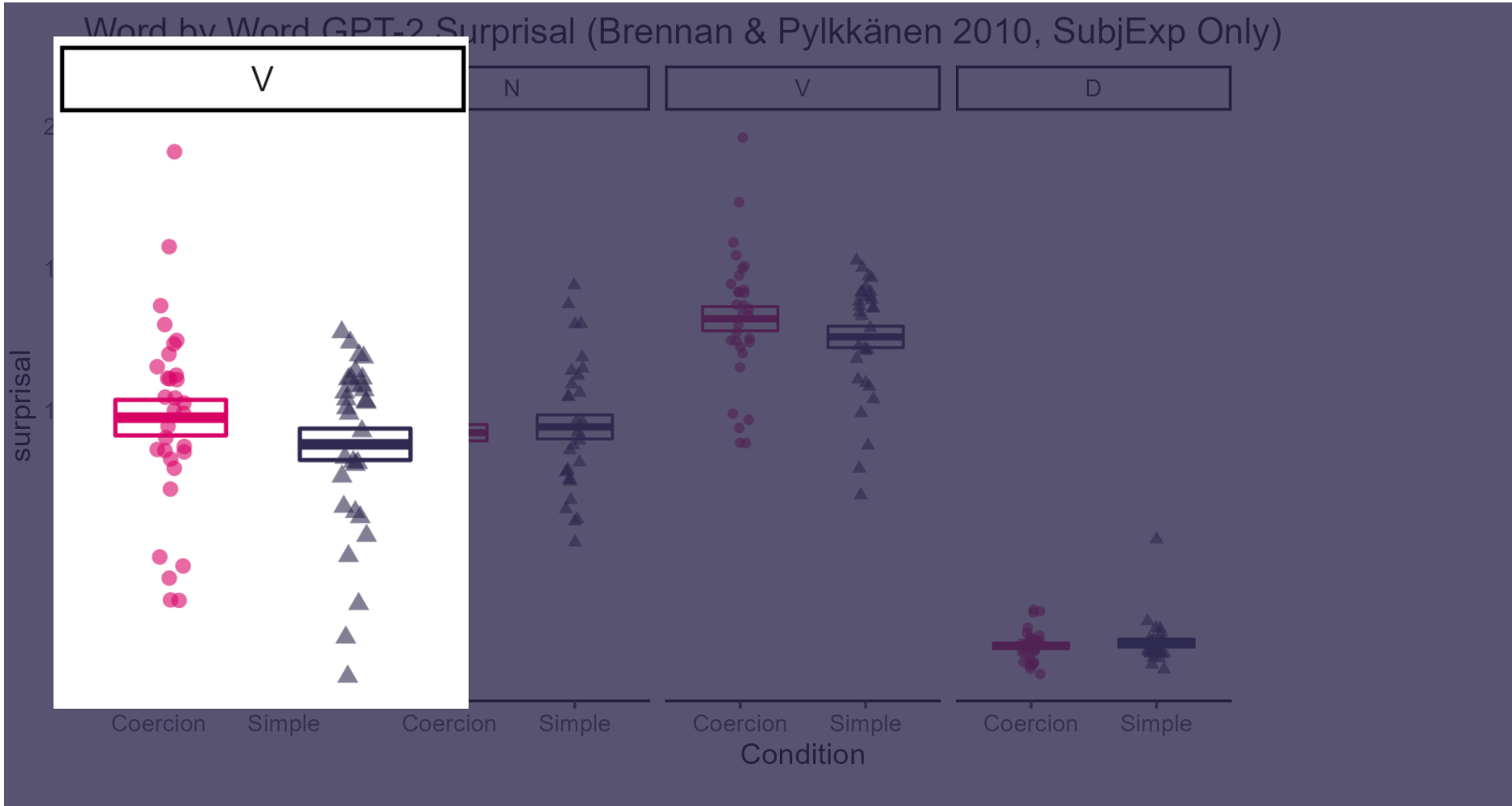




Brennan & Pylkkänen 2010

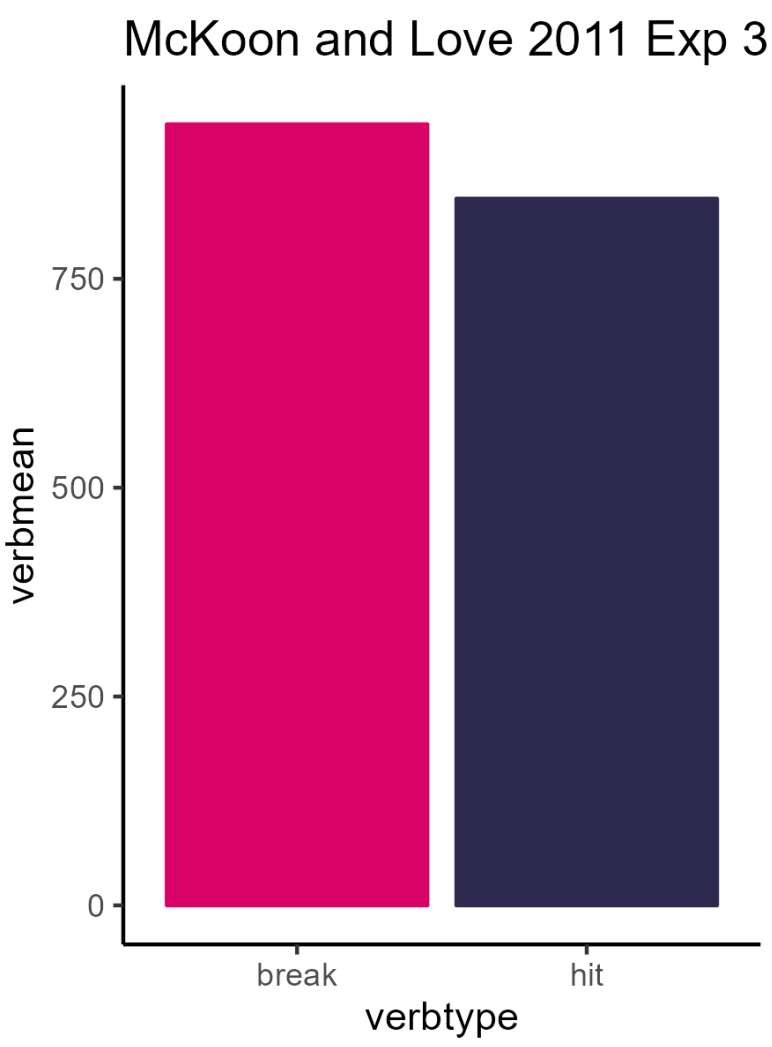
Within a few minutes, the child cherished the precious kitten. (coercion)
Without a doubt, the child cherished the precious kitten. (simple)

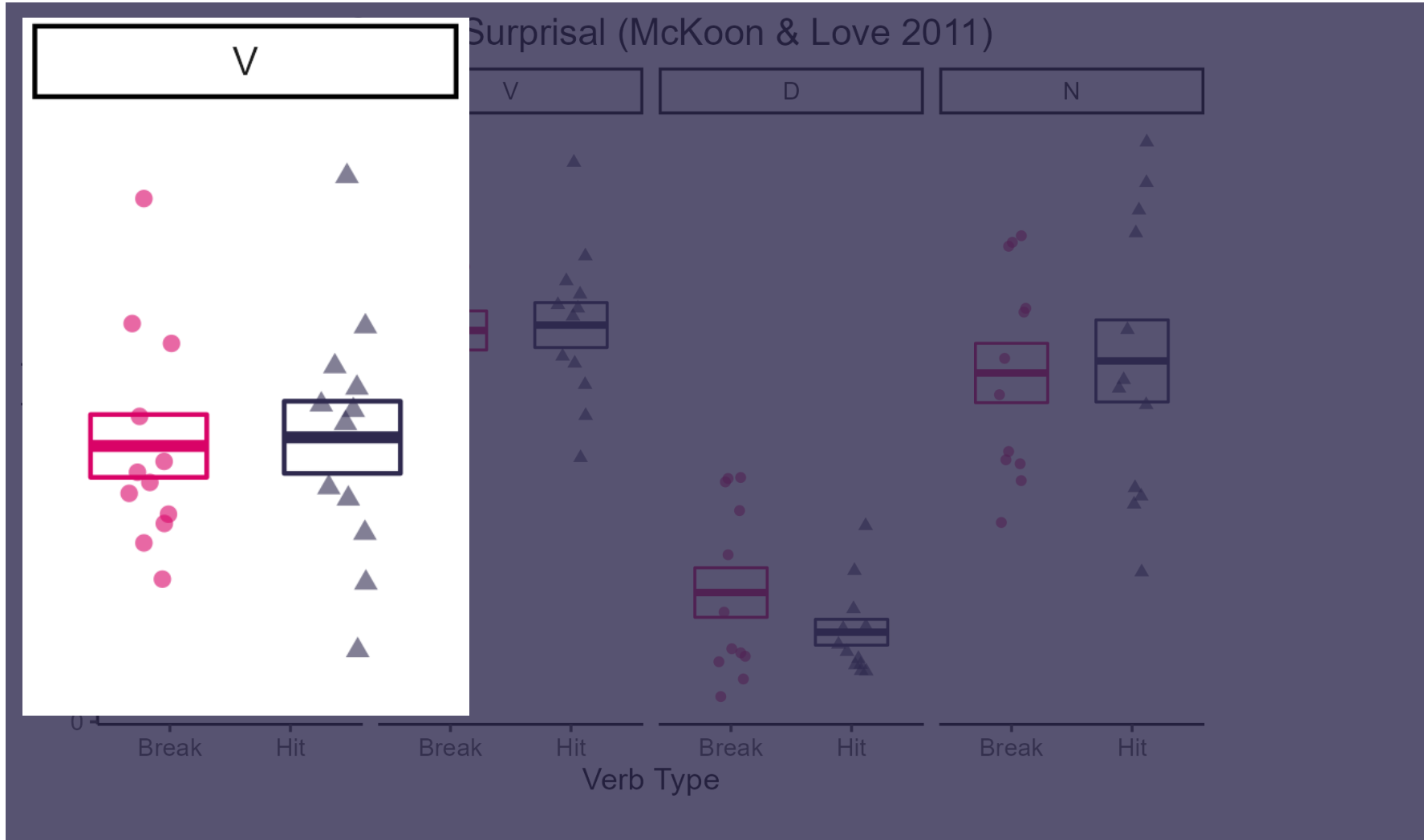




McKoon & Love 2011

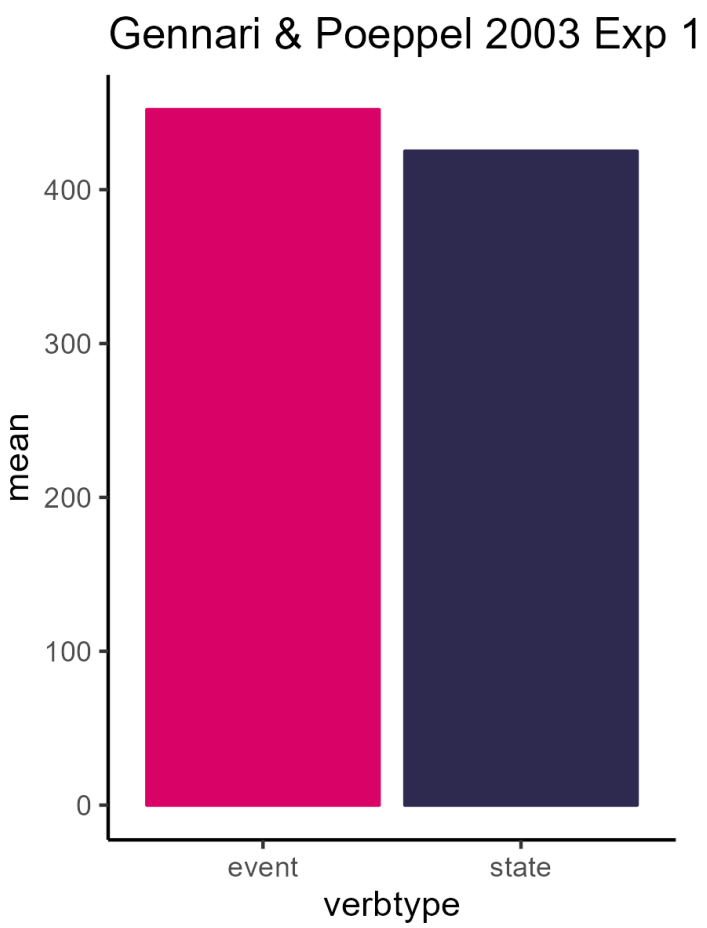
The stones cracked my windshield. (break)
The sheriff nudged my elbow. (hit)

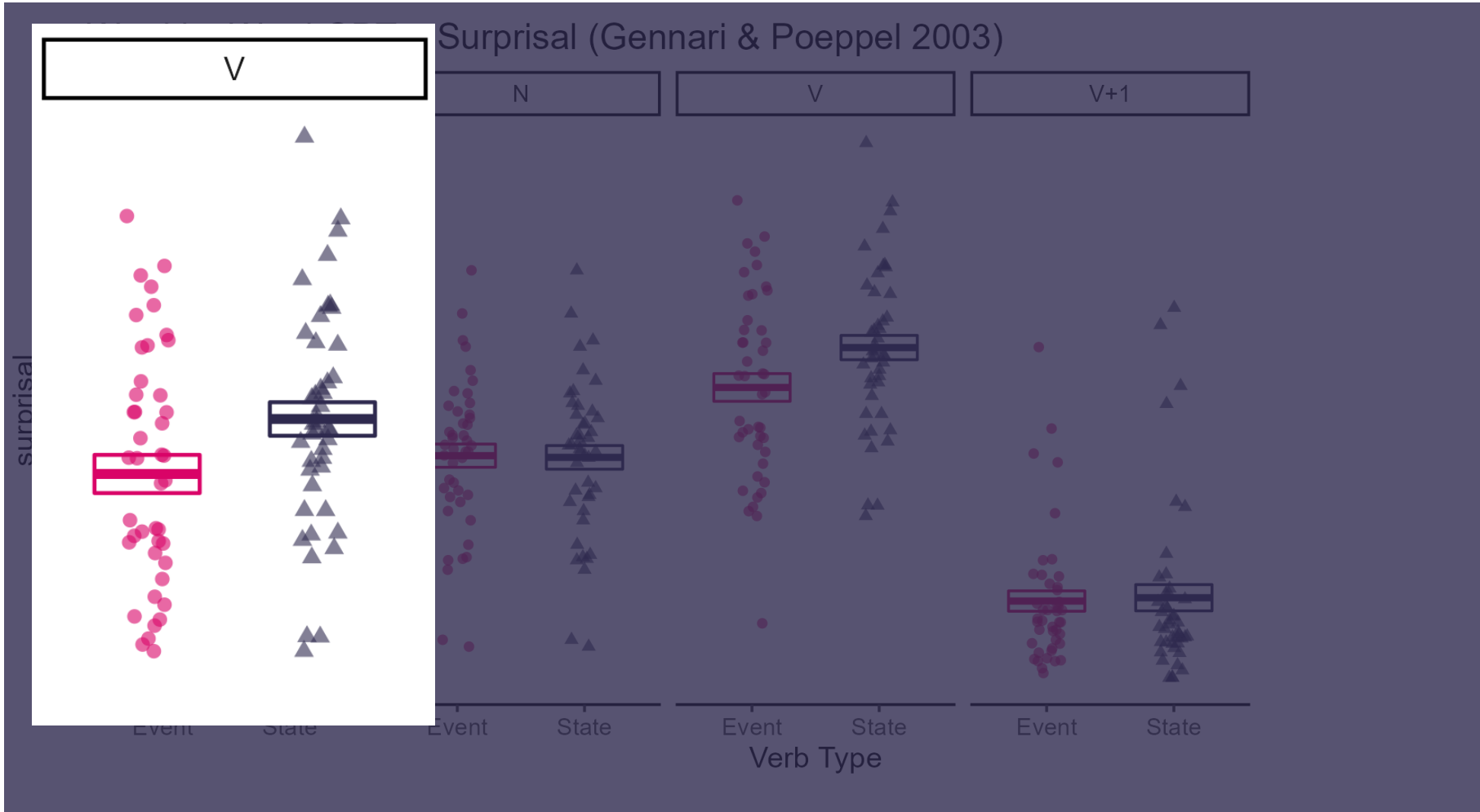




Gennari & Poeppel 2003

The head librarian arranged a new weekly meeting. (eventive)
The head librarian belonged to a union committee. (stative)





Why differences?

Some event structure contrasts **may** influence language use probabilities where there is "choice" (within surprisal):

transitive vs. intransitive result (agentive event)

Others **may** vary in complexity without influencing probabilities – no "choice" (beyond surprisal):

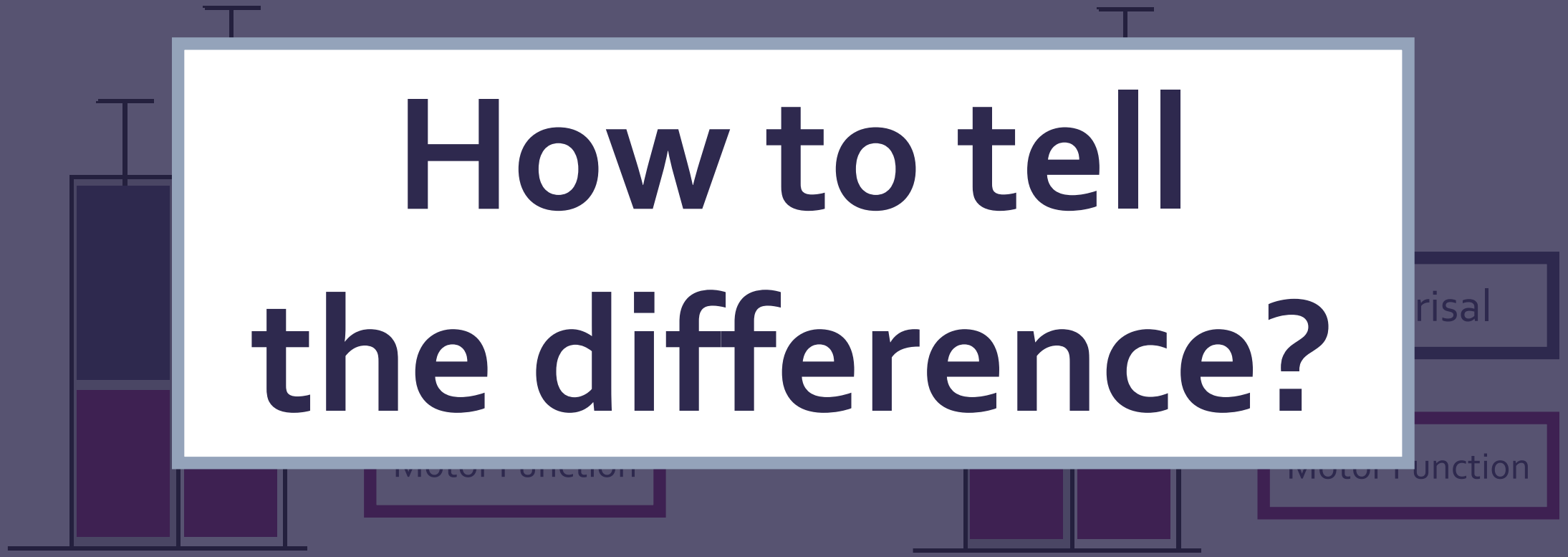
manner vs. result (causativity)

eventive vs. stative

Due to argument and event structure interactions, these are complex relationships!

Semantics within surprisal

Semantics beyond surprisal



How to tell the difference?

Beyond Surprisal

reading times

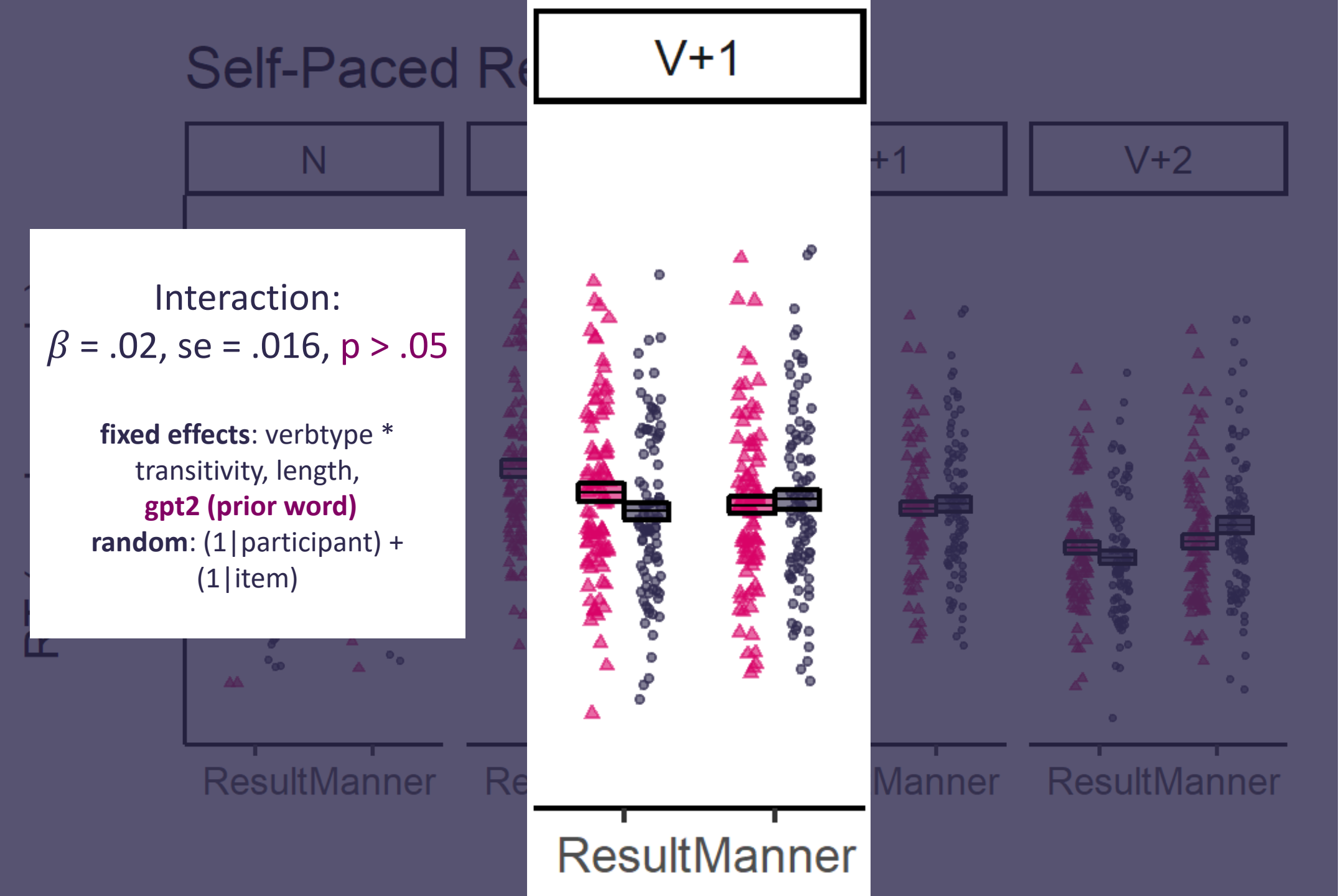
inaccurate language model?

Interaction:

β = .02, se = .016, p > .05

fixed effects: verbtype * transitivity, length, **gpt2 (prior word)**

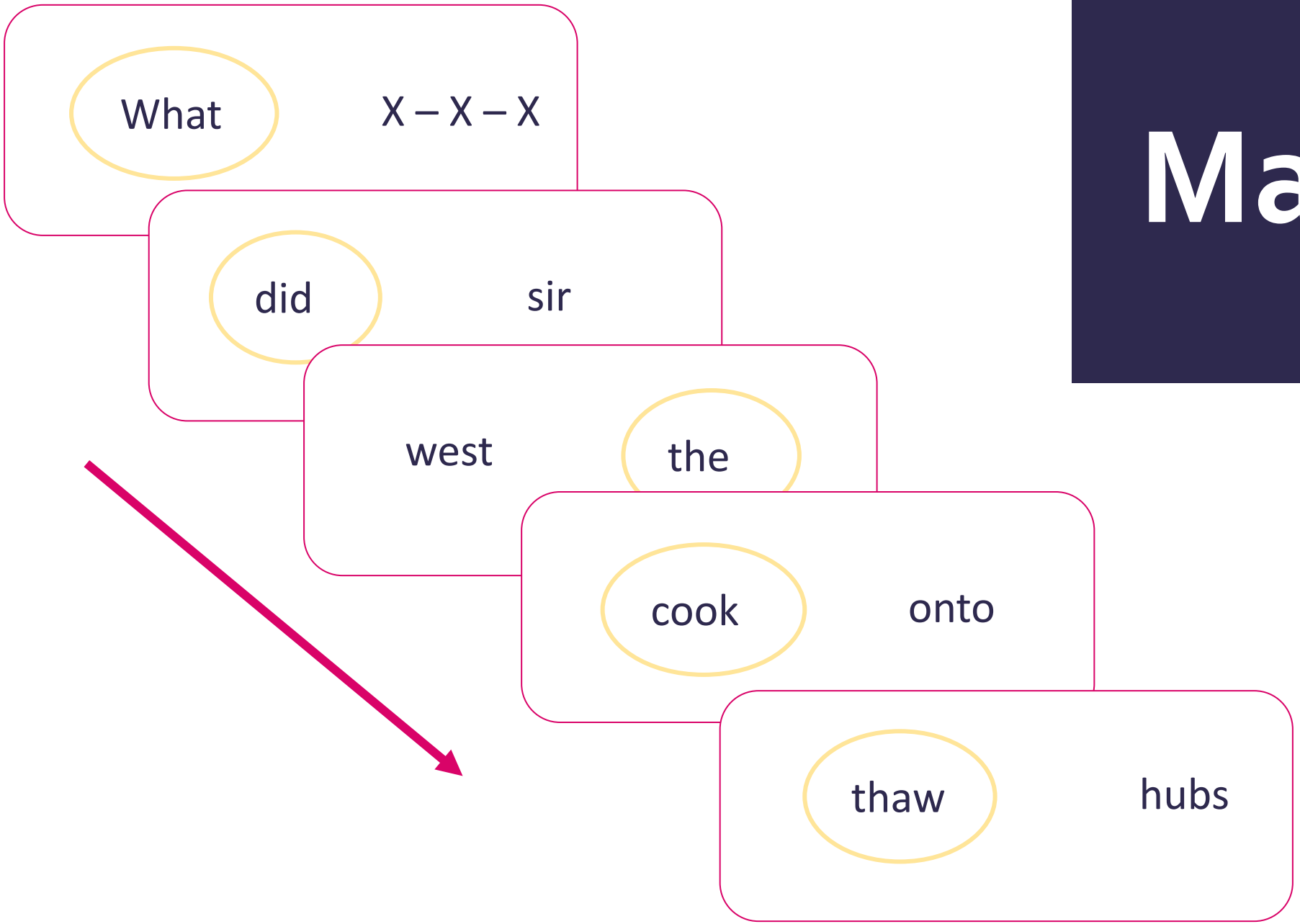
random: (1|participant) + (1|item)



Agenda

1. Semantics Beyond vs. Within Surprisal
2. Replicating Causativity Cost
3. Surprisals with Event Structure
4. **Deeper exploration with the Maze Task**

Maze Task



G-Maze Task (Forster et al 2009)

Less natural, but more incremental - no spillover "spread"

Greater power (Boyce et al 2020)

High surprisal "ungrammatical" alternatives generated with A-maze (Boyce et al 2020)

Exp 2: Collection & Analysis

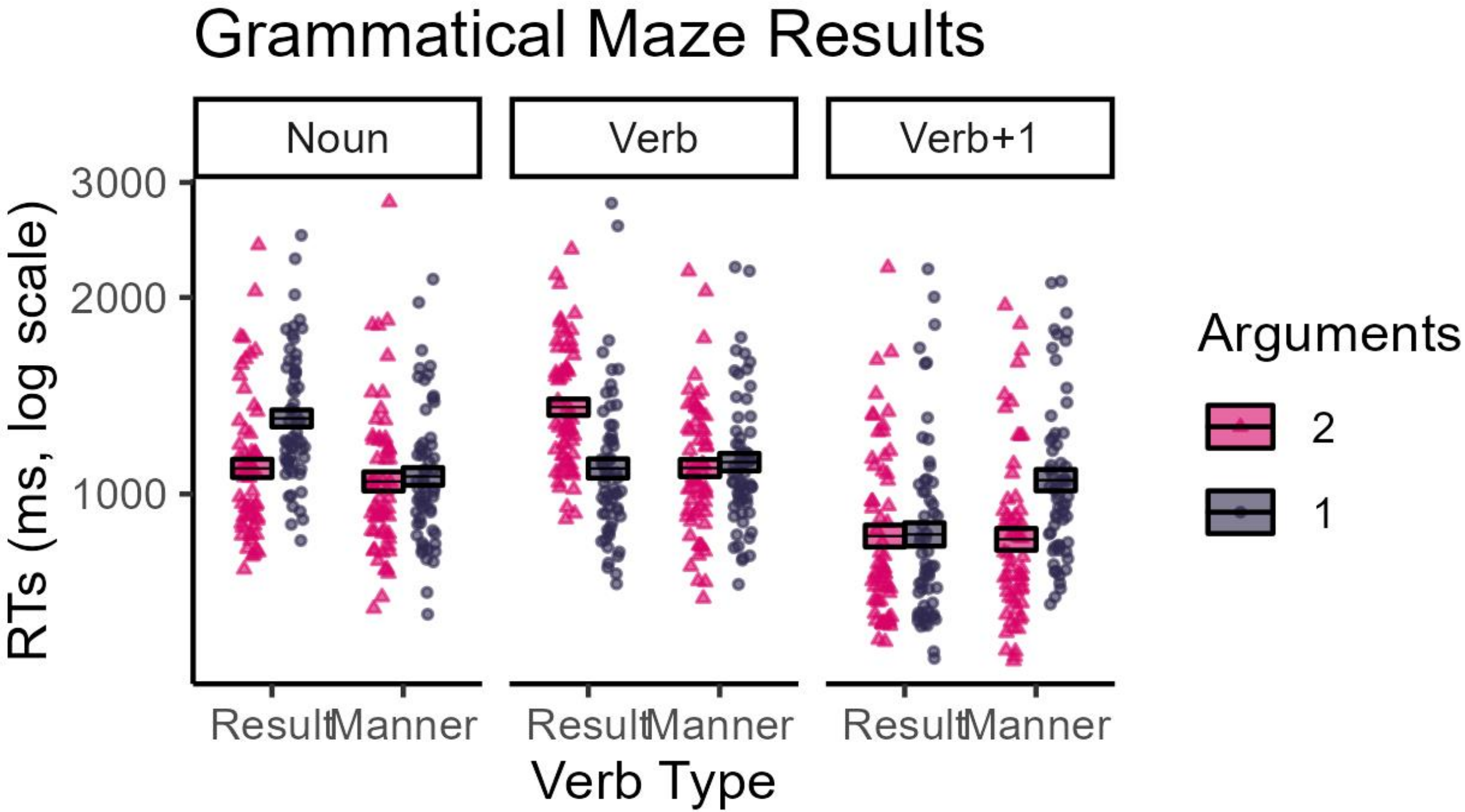
Data Collection

60 participants

Maze task via Ibexfarm

Statistical Analysis

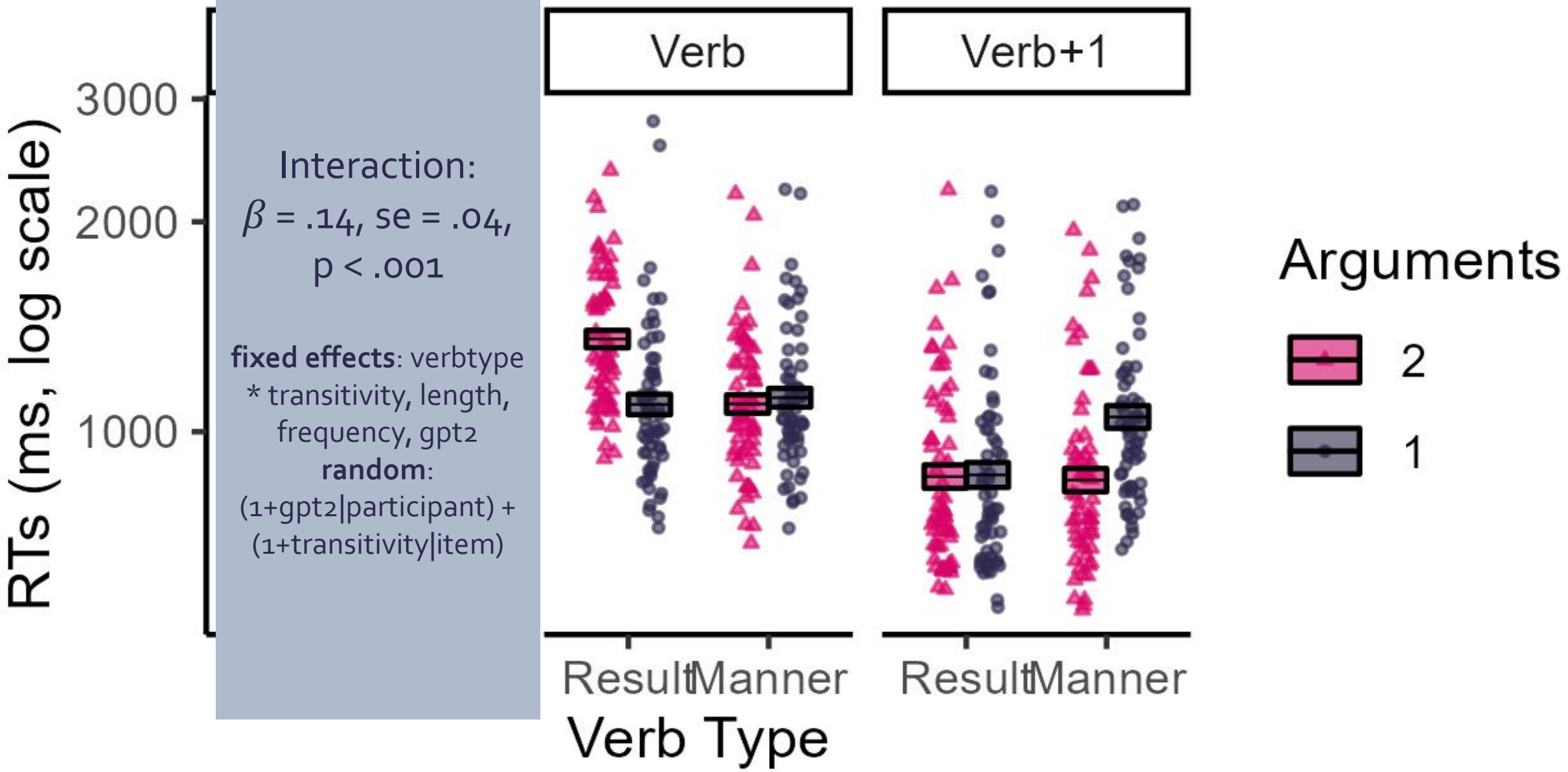
Hierarchical regression (Gelman & Hill 2006; Baayen et al. 2008) with Participants and Items as random effects


Grammatical Maze Results
Noun
Verb
Verb+1
3000
2000
1000
RTs (ms, log scale)
Result
Manner
Result
Manner
Result
Manner
Verb Type
Arguments
2
1

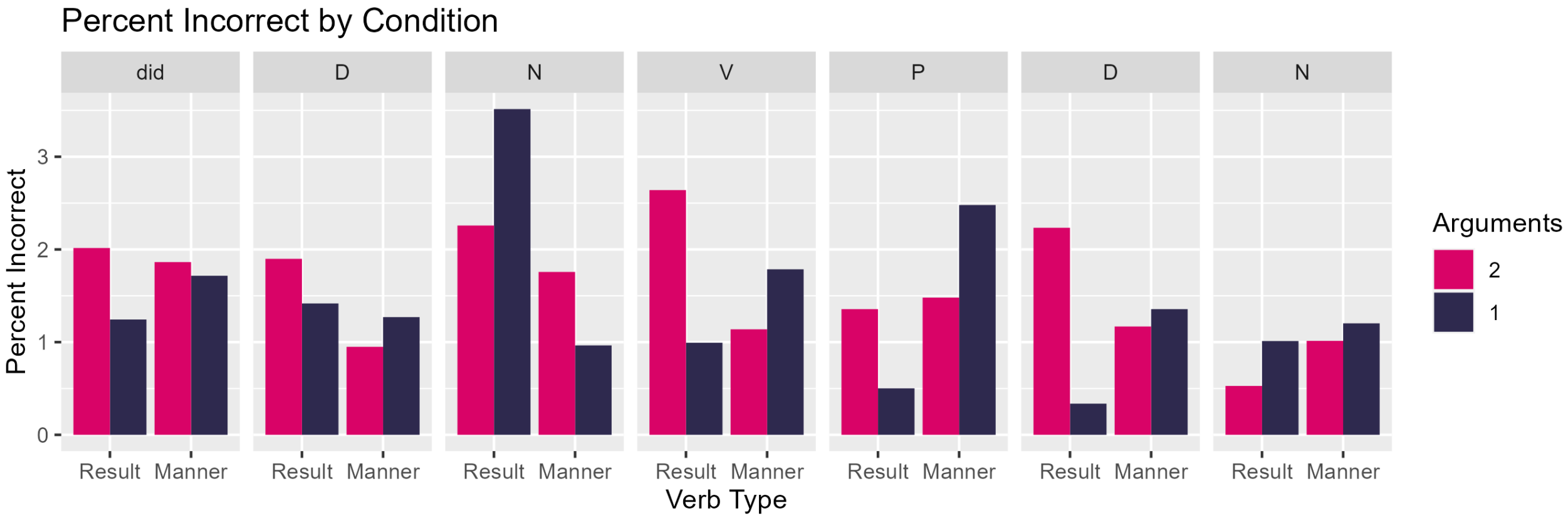
Grammatical Maze Results

Interaction:
β = .14, se = .04,
p < .001

fixed effects: verbtype * transitivity, length, frequency, gpt2
random: (1+gpt2|participant) + (1+transitivity|item)

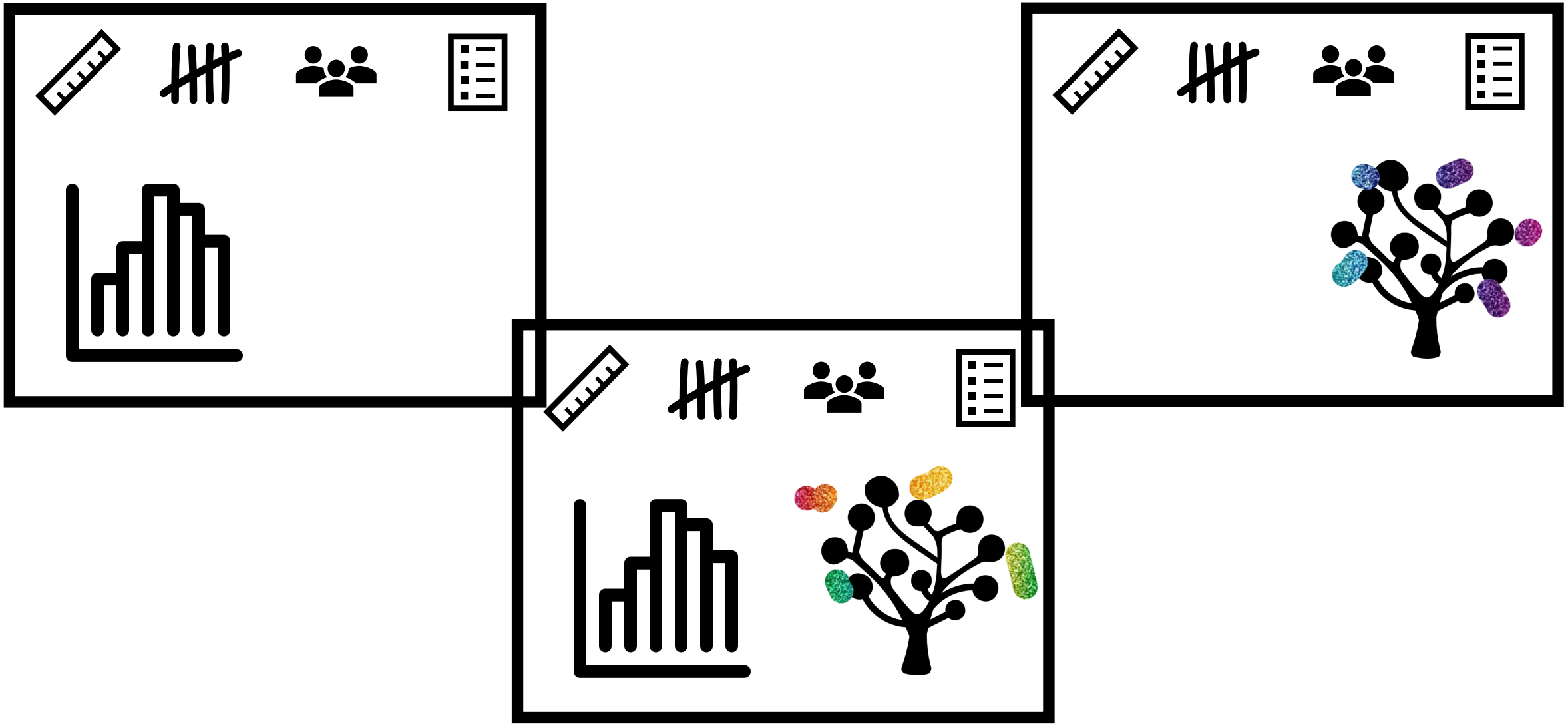


Percent Incorrect Responses

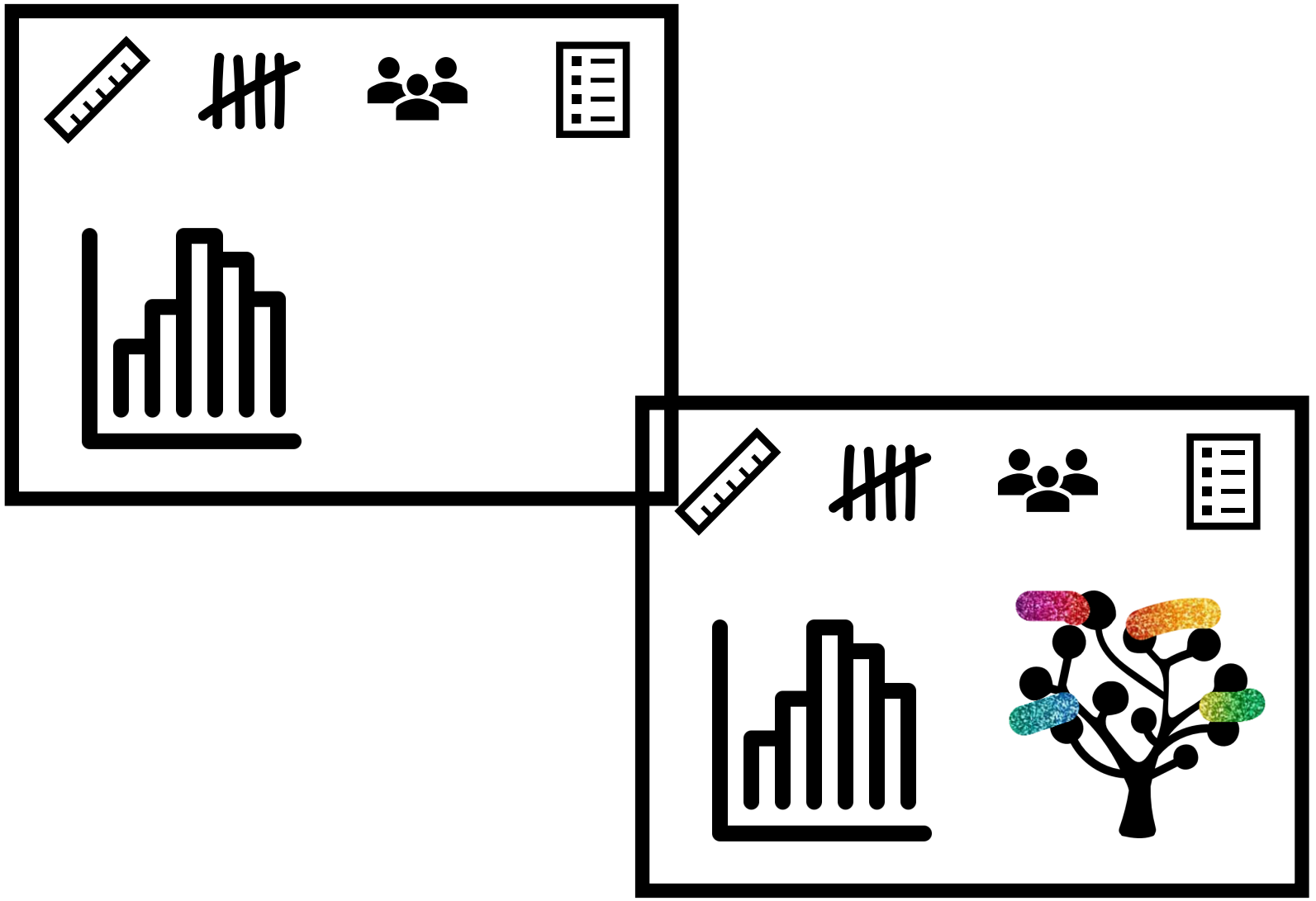


Replication 2: Statistical Model Comparison

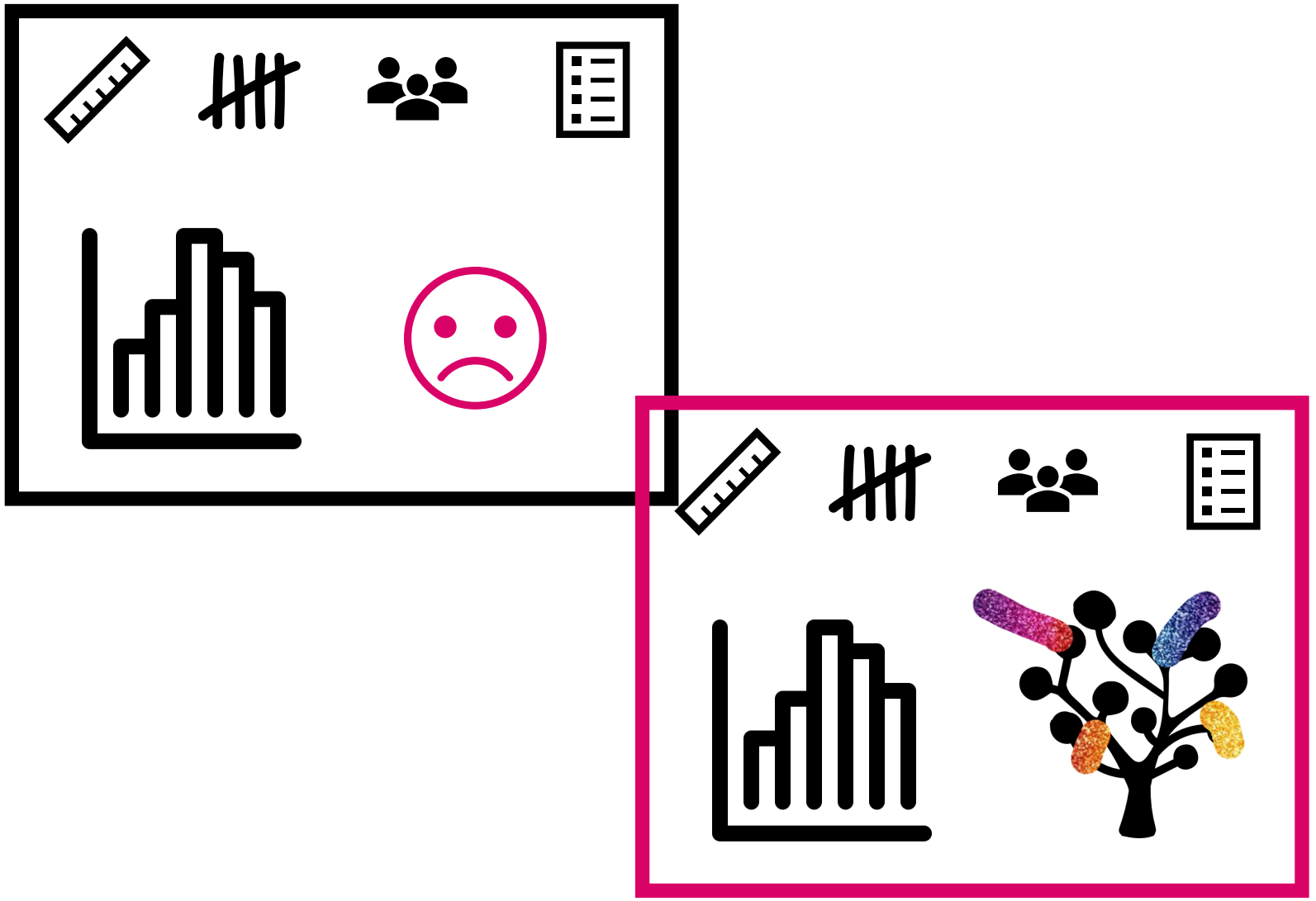
Model Comparison



Full vs. GPT-2 Only

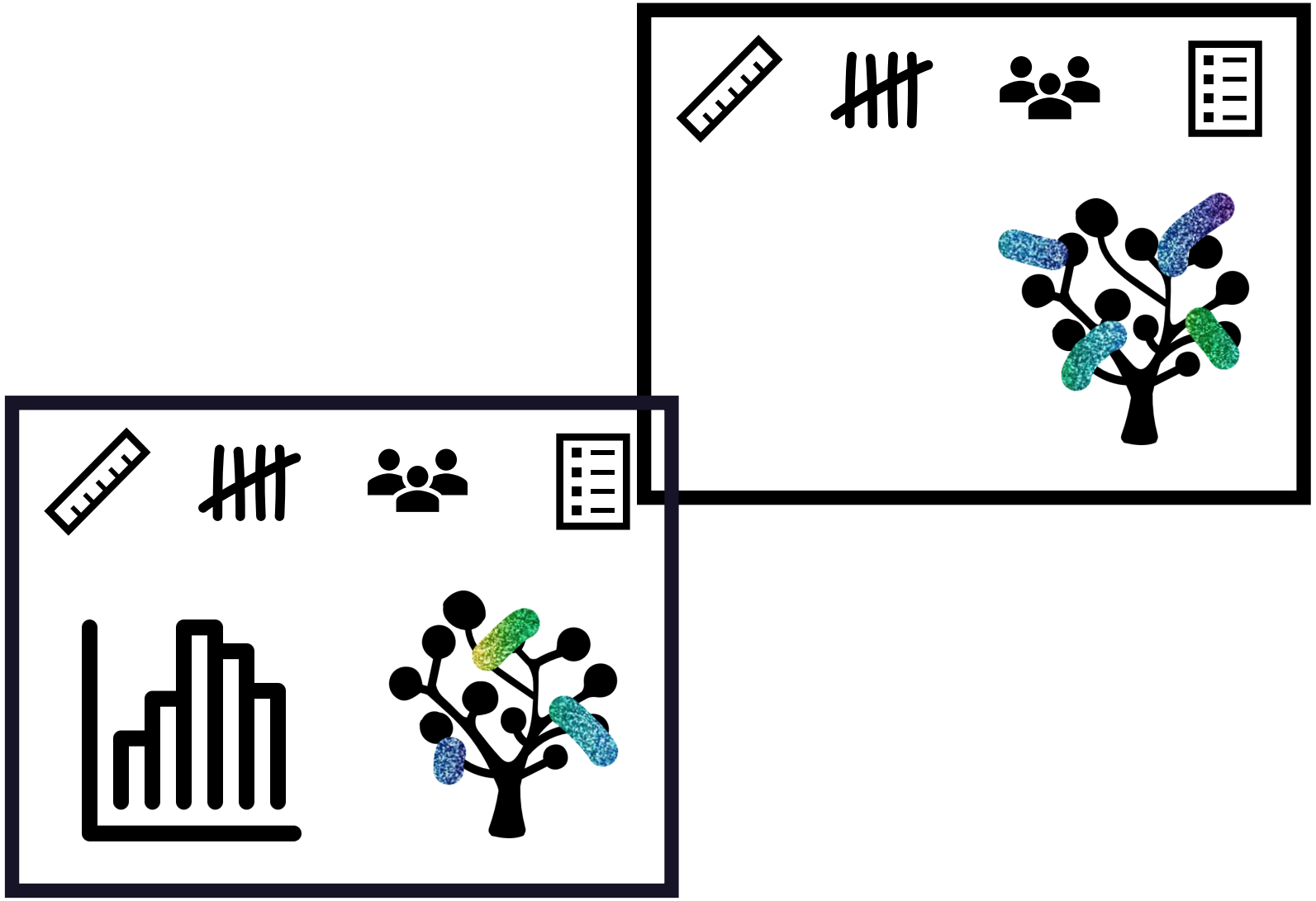


Full vs. GPT-2 Only

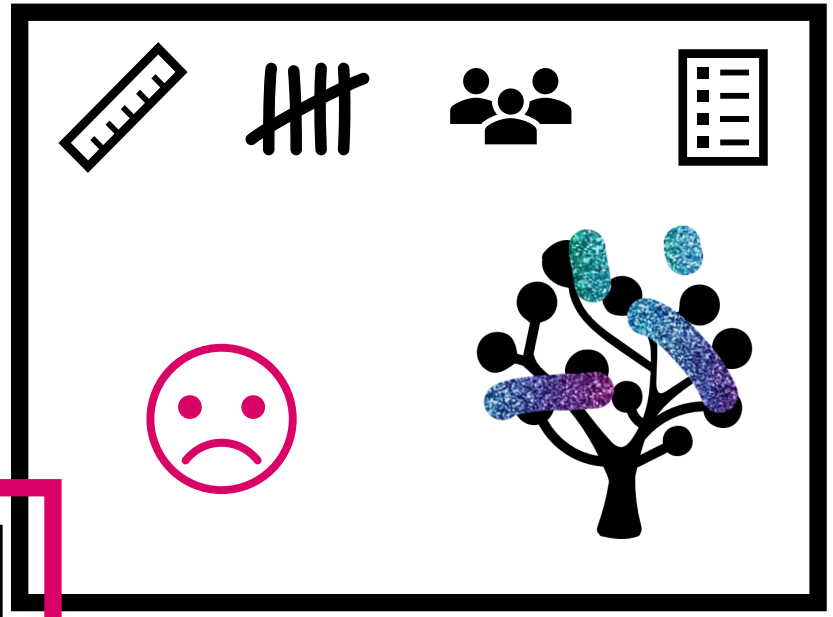


LRT $p < .0001$

Full vs. Event Structure Only



Full vs. Event Structure Only



LRT p < .0001

Model Comparison

Conclusions

Maze and other highly incremental measures may help us tease apart influences

Some event structure, such as causativity, “evades” the surprisal bottleneck

Need to continue exploring other contrasts and languages

Also can explore the relative timing of effects (e.g. via neural responses)

Thank you!

ELM organizers

RAs: Yizhi Tang, Lila Tappan, Brighton Pauli, Yasemin Gunal, Emma Thronson, and Thea Kendall-Green!

University of Michigan UROP.

Selected References

[1] G. McKoon and J. Love, Language and Cognition, 2011. [2] G. McKoon and T. MacFarland, Language, 2000. [3] G. McKoon and T. Macfarland, Cognitive Psychology, 2002. [4] S. Gennari and D. Poeppel, Cognition, 2003. [5] M. Rappaport Hovav, in Perspectives on Causation, 2020. [6] R. Levy, Cognition, 2008. [7] J. Hale, presented at the NAACL, 2001. [8] J. Hale, Language and Linguistics Compass, 2016. [9] K. I. Forster, C. Guerrera, and L. Elliot, Behavior Research Methods, 2009. [10] N. Witzel, J. Witzel, and K. Forster, J Psycholinguist Res, 2012. [11] L. Levinson and J. Brennan, in Morphological Metatheory, 2016. [12] L. Pylkkänen, Introducing Arguments. 2008. [13] A. Alexiadou, E. Anagnostopoulou, and F. Schäfer, in Phases of Interpretation, 2006. [14] M. Rappaport Hovav and B. Levin, in The Theta System, 2012. [15] A. Drummond, Ibex farm. Online server, 2013. [16] D. Bates and M. Maechler, 2009. [17] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, Journal of Statistical Software, 2017. [18] A. Korhonen, Y. Krymolowski, and T. Briscoe, presented at the LREC, 2006. [19] A. Radford et al., OpenAI blog, 2019. [20] V. Boyce, R. Futrell, and R. P. Levy, Journal of Memory and Language, 2020. [21] K. Gulordava, P. Bojanowski, E. Grave, T. Linzen, and M. Baroni, arXiv:1803.11138 [cs], 2018.