Task effects on the processing of predicate ambiguity: Distributivity in the Maze

John Duff, Adrian Brasoveanu, and Amanda Rysling
UC Santa Cruz

ELM 2 @ UPenn ❖ 20 May 2022

jduff@ucsc.edu

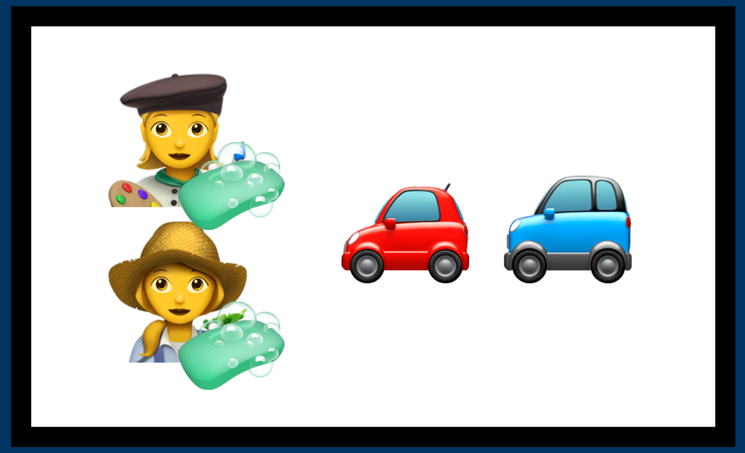
Distributive ambiguities

Bernadette and Jackie washed two cars.

Distributive

washed two cars each

Collective



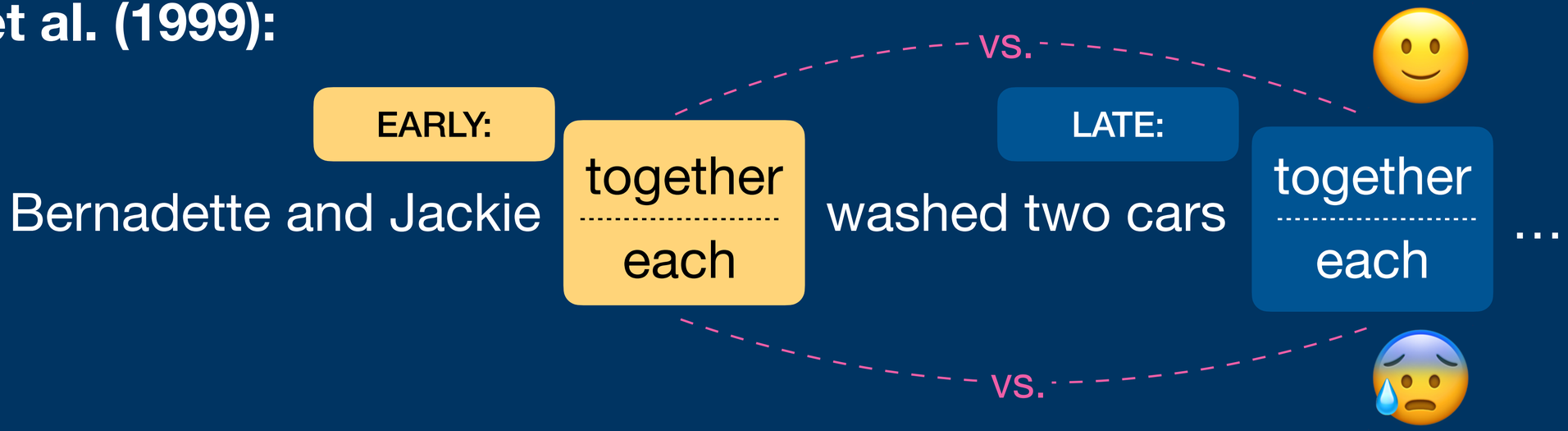
washed two cars together

Cumulative

Roberts (1987), Landman (2000), Nouwen (2012), Brasoveanu (2013), Champollion (2020)

Comprehending distributivity online

Frazier et al. (1999):



- Collective vs. distributive: systematic ambiguity of verbal structure
- Ambiguous content must be settled as encountered (Frazier & Rayner, 1990)
- By default the parser doesn't distribute, so LATE *each* triggers reanalysis 😰

Eyetracking: Frazier, Pacht & Rayner (1999); SPR: Dotlačil & Brasoveanu (2021)

Comprehension in the Maze

The A-Maze task (Boyce et al. 2020):

localizes	brown	ten	may	introduce	pear	effects
potatoes	costs	but	hip	riverbeds	task	closest

Duff et al. 2020/21:

Eager lexical commitments in the Maze: even polysemes are specified early.

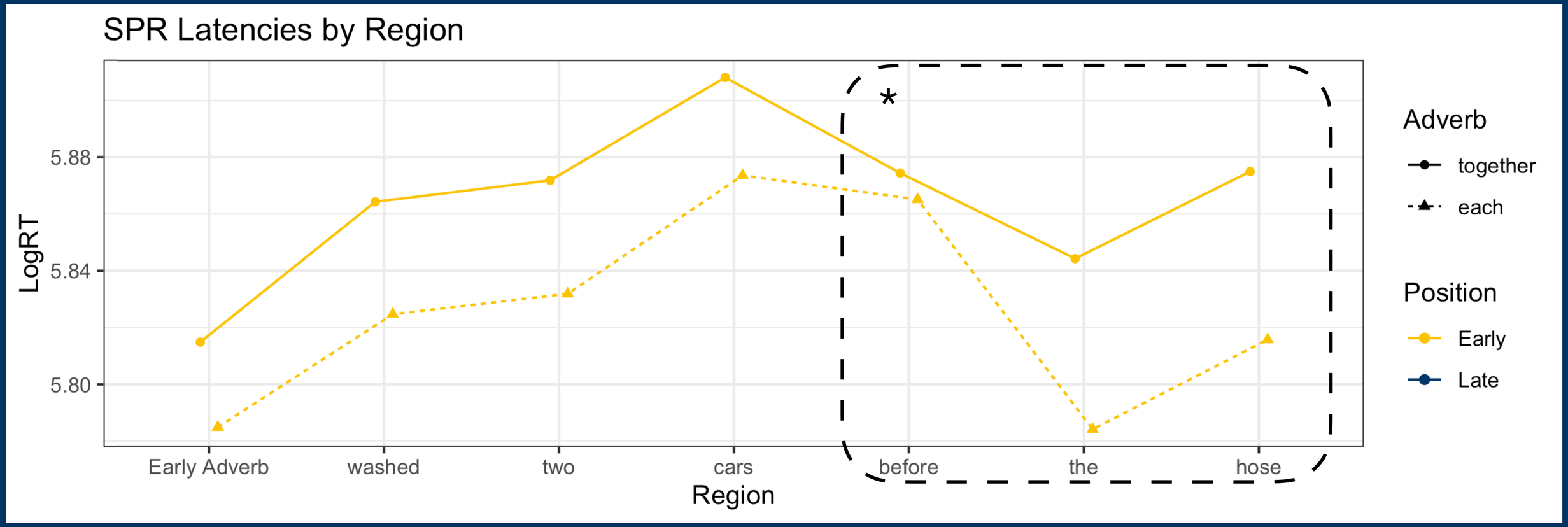
- Can we replicate predicate ambiguity effects in the Maze as well?
- Can we learn anything more about strategic comprehension in the Maze?

Forster, Guerrera & Elliot (2009), Boyce, Futrell & Levy (2020), Duff, Brasoveanu & Rysling (Posters @ AMLaP 2020, CUNY 2021)

32 Latin-squared items in 4 conditions; 130 fillers; *n* = 48 (24 UCSC + 24 Prolific).
Summed residualized log RTs analyzed via LMER fit in brms, fixed effects treatment-coded.

Experiment 1: SPR

In spillovers after an EARLY adverb, *together* >RT *each*

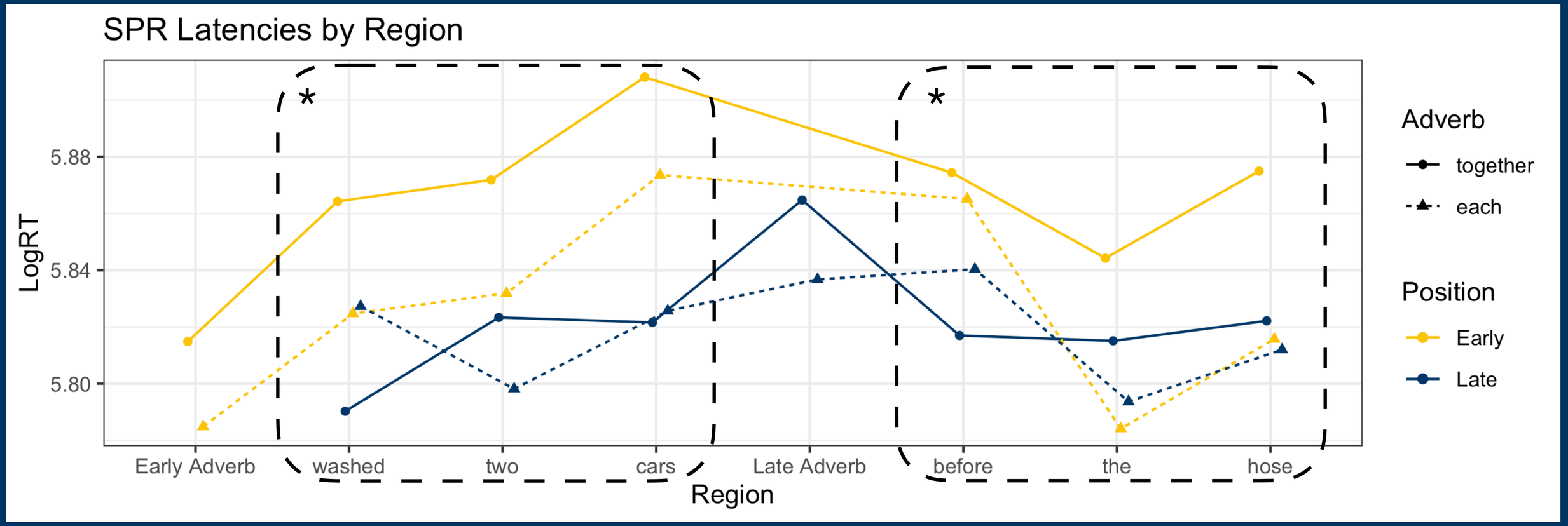


32 Latin-squared items in 4 conditions; 130 fillers; *n* = 48 (24 UCSC + 24 Prolific).
Summed residualized log RTs analyzed via LMER fit in brms, fixed effects treatment-coded.

Experiment 1: SPR

In spillovers after an EARLY adverb, *together* >RT *each*

Predicates with EARLY *together* >RT LATE *together*, continuing into spillover



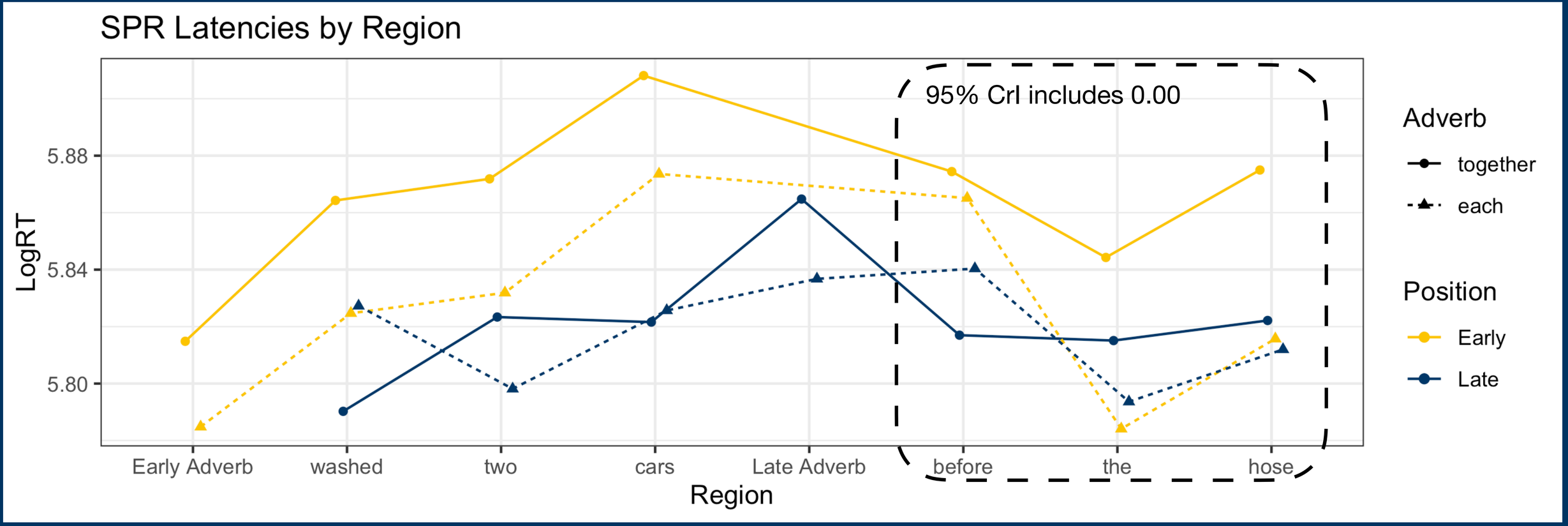
32 Latin-squared items in 4 conditions; 130 fillers; *n* = 48 (24 UCSC + 24 Prolific).
Summed residualized log RTs analyzed via LMER fit in brms, fixed effects treatment-coded.

Experiment 1: SPR

In spillovers after an EARLY adverb, *together* >RT *each*

Predicates with EARLY *together* >RT LATE *together*, continuing into spillover

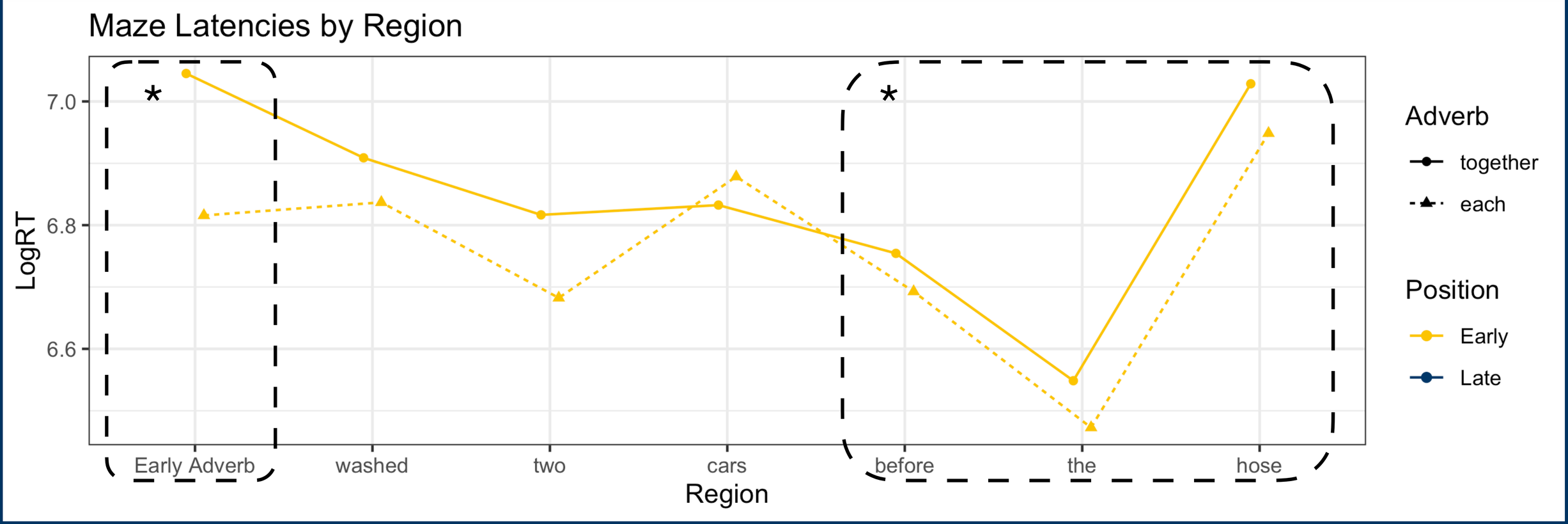
LATE *each* does not trigger a convincing penalty in spillover



32 Latin-squared items in 4 conditions; 130 fillers; *n* = 48 (24 UCSC + 24 Prolific).
Summed residualized log RTs analyzed via LMER fit in brms, fixed effects treatment-coded.

Experiment 2: Maze

together >RT *each* on EARLY adverb itself, plus in the spillover as in SPR

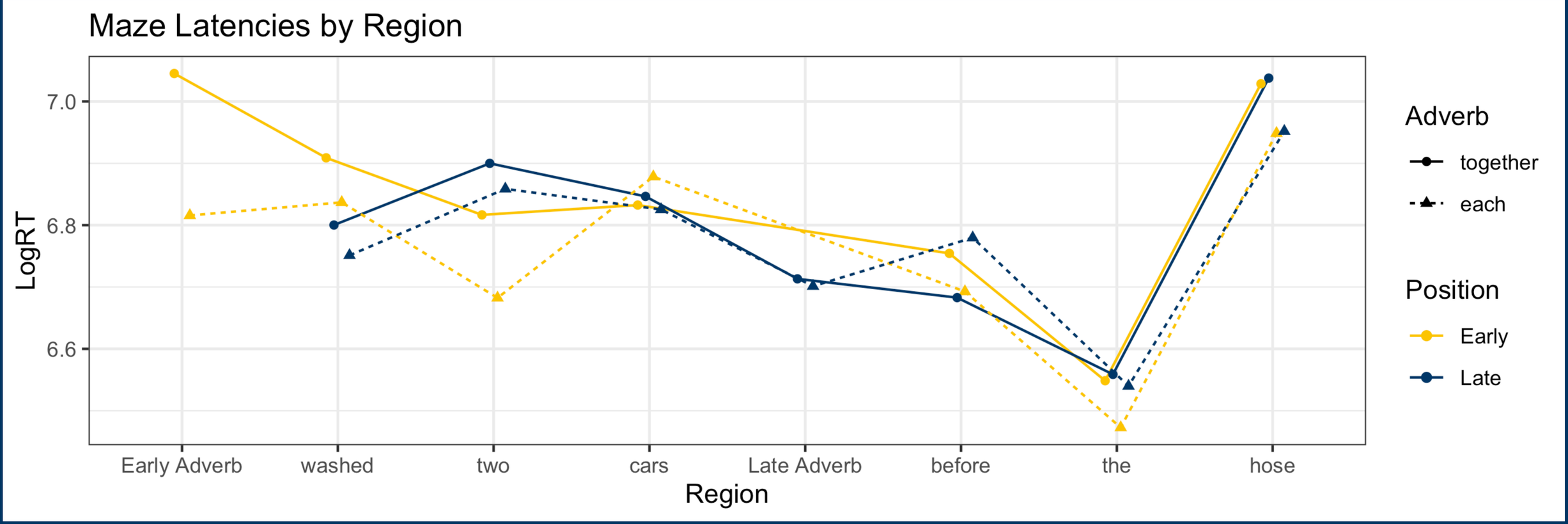


32 Latin-squared items in 4 conditions; 130 fillers; n = 48 (24 UCSC + 24 Prolific).
Summed residualized log RTs analyzed via LMER fit in brms, fixed effects treatment-coded.

Experiment 2: Maze

together >RT *each* on EARLY adverb itself, plus in the spillover as in SPR

Unlike SPR, LATE -disambiguated predicates are no faster



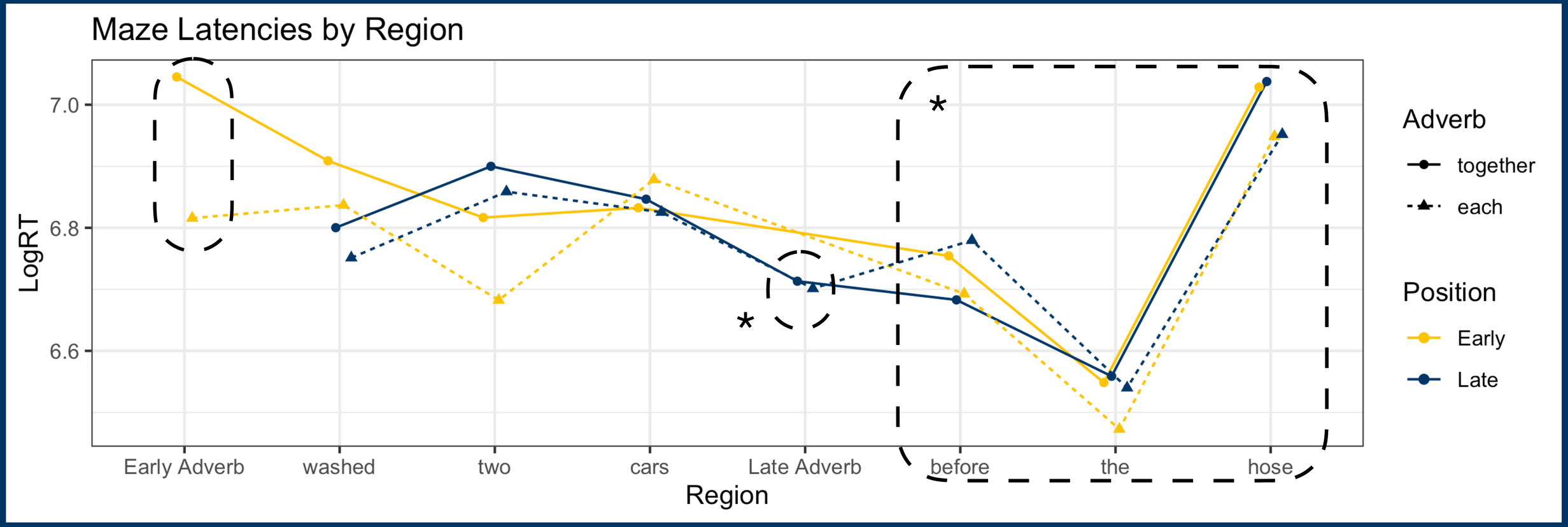
32 Latin-squared items in 4 conditions; 130 fillers; *n* = 48 (24 UCSC + 24 Prolific).
Summed residualized log RTs analyzed via LMER fit in brms, fixed effects treatment-coded.

Experiment 2: Maze

together >RT *each* on EARLY adverb itself, plus in the spillover as in SPR

Unlike SPR, LATE -disambiguated predicates are no faster

LATE *each* shows predicted penalties on the adverb and spillover



Discussion: Why is *together* so costly?

- *Together* → Potential difficulty imagining collaborative participation.
- Why do we see *together* $>_{RT}$ default reading in SPR?
    - Prior assumption: default ≠ collective ≠ *together*
    - Option #1: The collective default lacks the detail of *together*
    - Option #2: Not a collective default, but a heuristic avoidance of extra object entities (see Fodor 1982, Brasoveanu & Dotlačil 2015 on scope)
- Both predict special costs for late *each* overriding a default, vs. constant costs for *together*: it's always compatible with the default but still extra work.

Discussion: Apparent task effects

- As expected, the Maze localized costs more tightly to the adverb.
- A more puzzling contrast: *together* and default readings are equally difficult, unlike SPR.
    - Evidence for an enriched default reading in the Maze.
    - Fits with polysemy evidence: The task encourages earlier commitments.

Thanks!

Happy to answer questions about:

- Details of norming, design, and analysis for these experiments
- Close comparisons with Frazier et al. (1999) and Dotlačil & Brasoveanu (2021)
- The (A-)Maze, including musings on the reasons for eager commitment



(Jack, Adrian, and Amanda stood in front of a tree.)

Appendix A: Items & Norming Data

Items: Norming Task

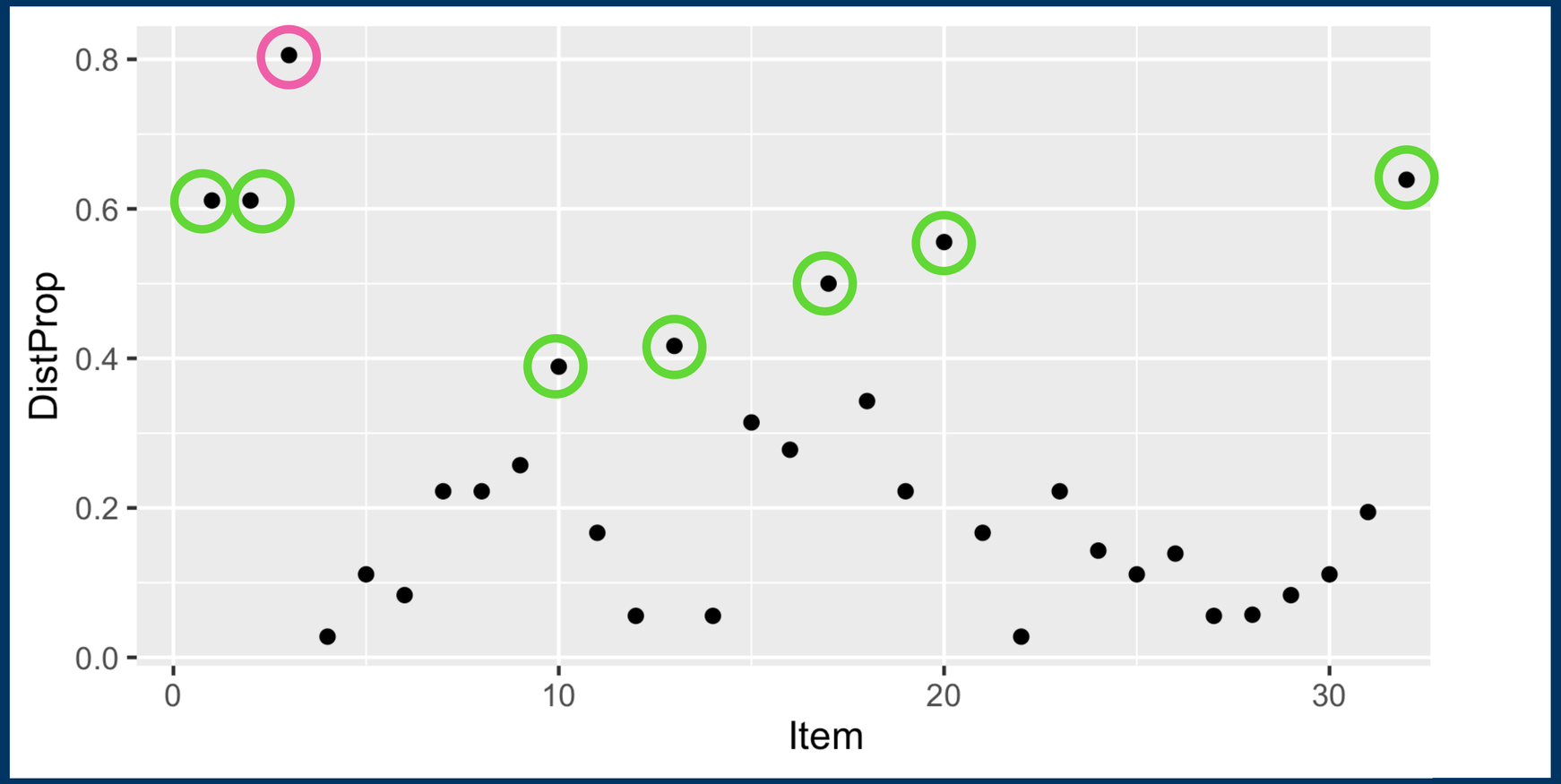
Luckily, Bernadette and Jackie washed two cars before the hose broke.

Which meaning is most likely?

They washed two cars together

They washed two cars each

Items: Norming results



- **1 item** with reliable distributive preference (bootstrapped 95% CI > 0.5)
- **7 items** with no reliable preference (bootstrapped 95% CI includes 0.5)
- 24 items with reliable collective preference (bootstrapped 95% CI < 0.5)

Collective-biased items: Singular objects

- Yesterday, Fred and Susan sang a Bob Dylan song after the bar closed.
- Reportedly, Lou and Deborah danced one tango when the ballroom reopened.
- Thankfully, Felix and Sarah wrote one song after the director asked.
- Hilariously, Emmon and Barbara made up one ghost story when the power died.
- Recently, Jackson and Beverly painted a room before the temperature rose.
- Amazingly, Edna and Milton mailed one care package when the crisis began.
- Typically, Martin and Kimberly play one video game before the counselor notices.
- Apparently, Matt and Carrie saw one concert before the semester started.
- Luckily, Maura and Jerry hosted one party before the landlord complained.

Collective-biased items: Plural objects

- Luckily, Bernadette and Jackie washed two cars before the hose broke.
- Today, Julie and Jason baked several cakes before the alarm sounded.
- Fortunately, Lyn and Patrick saved 1000 dollars after the recession ended.
- Clearly, Ted and Jim destroyed 50 dollars worth of posters after the party began.
- Incredibly, Molly and Jack ate four burritos before the weekend ended.
- Supposedly, Rick and Angie sold five paintings when the artist died.
- Unsurprisingly, Mandy and Paul talked for twenty minutes after the conference ended.
- Clearly, Mitch and Rebecca pruned six rosebushes before the sun set.

Distributive-biased and equibiased items

- Incredibly, Jane and Martha weighed 220 pounds when the contest began.
- Supposedly, Sam and Maria carried one suitcase when the family traveled.
- Apparently, Billy and Francis got one ice cream cone when the vendor came.
- Worryingly, Paul and Marcie drew three nasty political cartoons when the mayor resigned.
- Charmingly, Tom and Laura sent one letter when the ship docked.
- Usually, Tom and Allison review four papers when the seminar meets.
- Reportedly, Burt and Sierra taught two classes when the school opened.
- Charmingly, Jed and Harriet read one book when the television broke.

Appendix B: Analysis Details

Analysis: Model specifications

- Automatic weakly informative priors set in brms
    - E.g. for the spillover model in Experiment 1:
        - Intercept: Normal($\mu = -0.16, \sigma = 1.5$)
        - Fixed effects: Normal($\mu = 0, \sigma = 3.04 - 3.29$)
        - Standard deviations for random effects: Exponential($\lambda = 1.6$)
- STAN parameters:
    - 6 chains of 10,000 iterations each, including 2,000 iterations of warmup

Analysis: Expt. 1, Predicate

Effect	$\hat{\beta}$	$\hat{\sigma}$	**95% HPDI**		$\hat{R}$
Intercept	0.0028	0.052	-0.101	0.105	1
DisambigLate	-0.2042	0.067	-0.336	-0.071	1
MeaningDist	-0.0947	0.061	-0.215	0.026	1
Disambig:Meaning	0.1076	0.083	-0.054	0.269	1

Analysis: Expt. 1, Critical Adverb

Effect	$\hat{\beta}$	$\hat{\sigma}$	**95% HPDI**		$\hat{R}$
Intercept	-0.0988	0.021	-0.139	-0.058	1
DisambigLate	0.0427	0.027	-0.010	0.096	1
MeaningDist	0.0250	0.024	-0.022	0.072	1
Disambig:Meaning	-0.0053	0.033	-0.071	0.060	1

Analysis: Expt. 1, Spillover

Effect	$\hat{\beta}$	$\hat{\sigma}$	**95% HPDI**		$\hat{R}$
Intercept	-0.044	0.054	-0.15	0.063	1
DisambigLate	-0.137	0.061	-0.26	-0.018	1
MeaningDist	-0.135	0.058	-0.25	-0.021	1
Disambig:Meaning	0.097	0.075	-0.05	0.243	1

Analysis: Expt. 2, Predicate

Effect	$\hat{\beta}$	$\hat{\sigma}$	**95% HPDI**		$\hat{R}$
Intercept	0.150	0.069	0.014	0.287	1
DisambigLate	0.121	0.067	-0.010	0.254	1
MeaningDist	-0.092	0.064	-0.216	0.033	1
Disambig:Meaning	0.034	0.085	-0.132	0.201	1

Analysis: Expt. 2, Critical Adverb

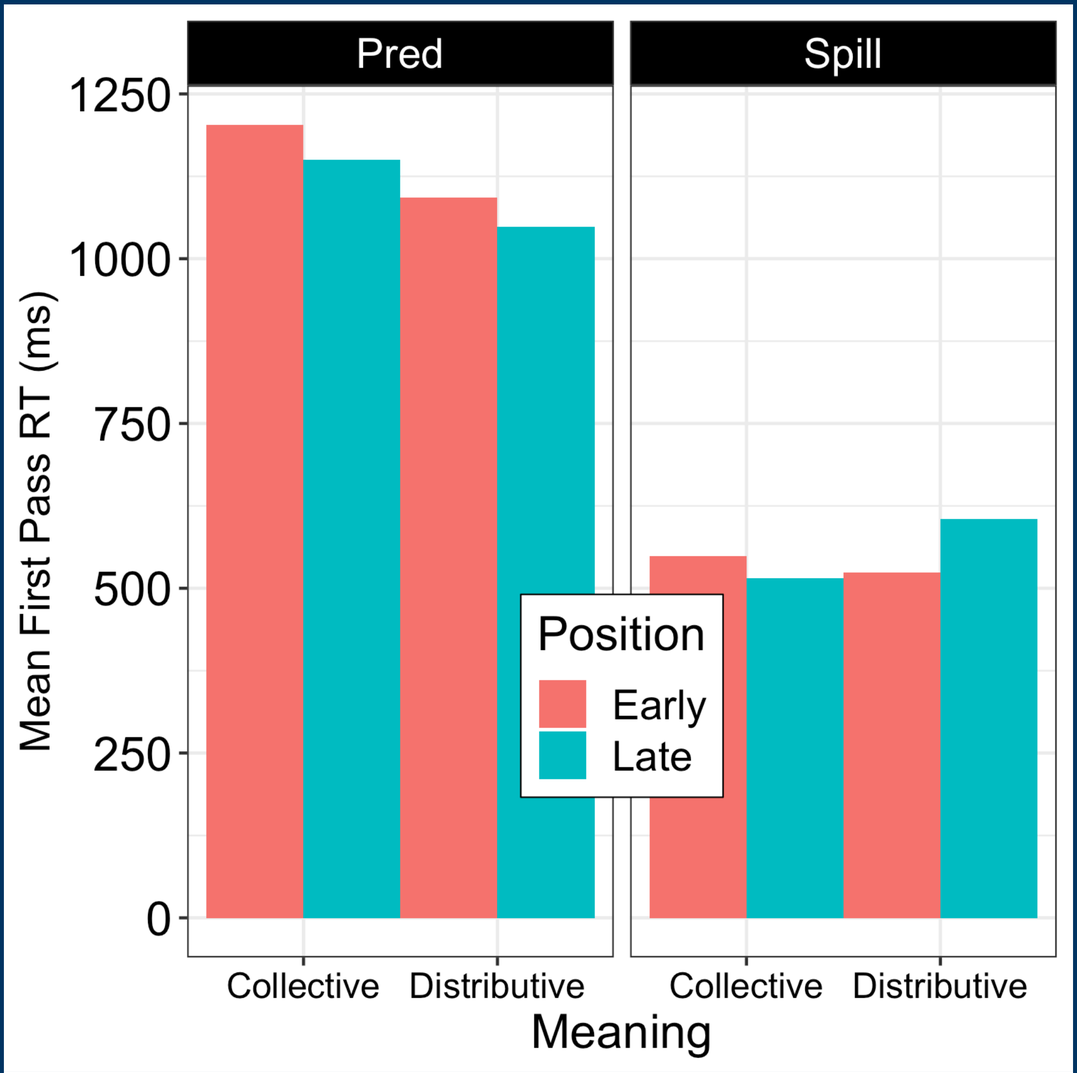
Effect	$\hat{\beta}$	$\hat{\sigma}$	**95% HPDI**		$\hat{R}$
Intercept	0.13	0.026	0.079	0.180	1
DisambigLate	-0.331	0.035	-0.400	-0.263	1
MeaningDist	-0.063	0.032	-0.125	-0.000	1
Disambig:Meaning	0.206	0.044	0.120	0.292	1

Analysis: Expt. 2, Spillover

Effect	$\hat{\beta}$	$\hat{\sigma}$	**95% HPDI**		$\hat{R}$
Intercept	-0.133	0.064	-0.260	-0.01	1
DisambigLate	-0.027	0.069	-0.163	0.11	1
MeaningDist	-0.139	0.065	-0.266	-0.01	1
Disambig:Meaning	0.180	0.083	0.017	0.34	1

Appendix C: Comparisons with previous work

Frazier, Pacht & Rayner (1990)



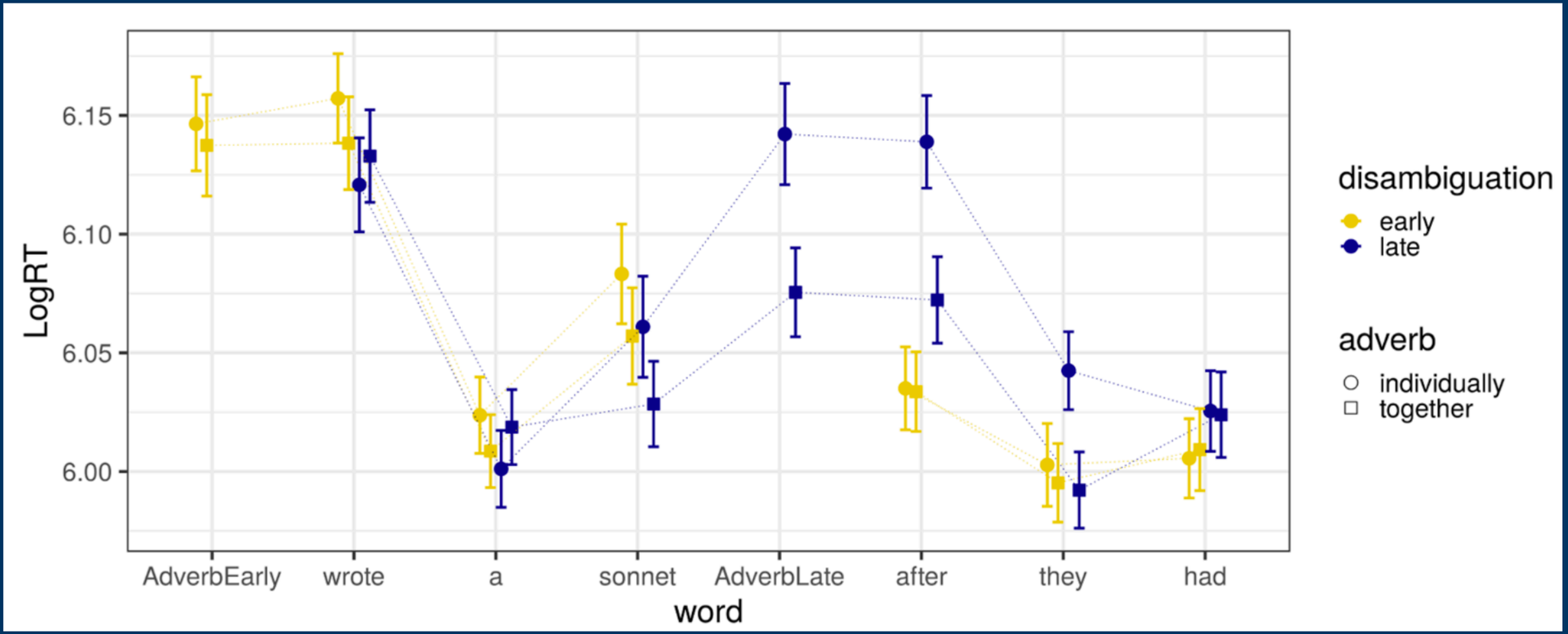
- 16 (-1) items, *n* = 60 in-lab at UMass
- First pass RTs:
  - Main effect of LATE <$_{RT}$ EARLY in the predicate region
  - Marginal main effect of *each* <$_{RT}$ *together*
  - Critical interaction in spillover: LATE *each* prompts particular delays
- LATE had more regressions from spillover

Frazier et al. (1999) vs. This Study

- Merely trending cost for *together*, which is borne out as significant here in SPR and in the Maze.
- Temporarily ambiguous predicates are read faster, which is borne out here in SPR as well, though not the Maze.
  - Their explanation: Early adverbs prompt difficulty because that position is odd, post-verbal positions are more natural.
  - Given the strength and persistence of the effect in our SPR study, we are not convinced this is solely about a dispreferred word order.
- Critical interaction here is borne out in the Maze as well, but not SPR.

Dotlačil & Brasoveanu (2021): Expt. 1

- 28 items, *n* = 87 at UCSC (NB: *individually* instead of *each*)
- *Together* associated with slower residuals on the adverb
- No credible effects in predicate region, but LATE >RT EARLY on spillover
- Classic interaction on adverb and spillover

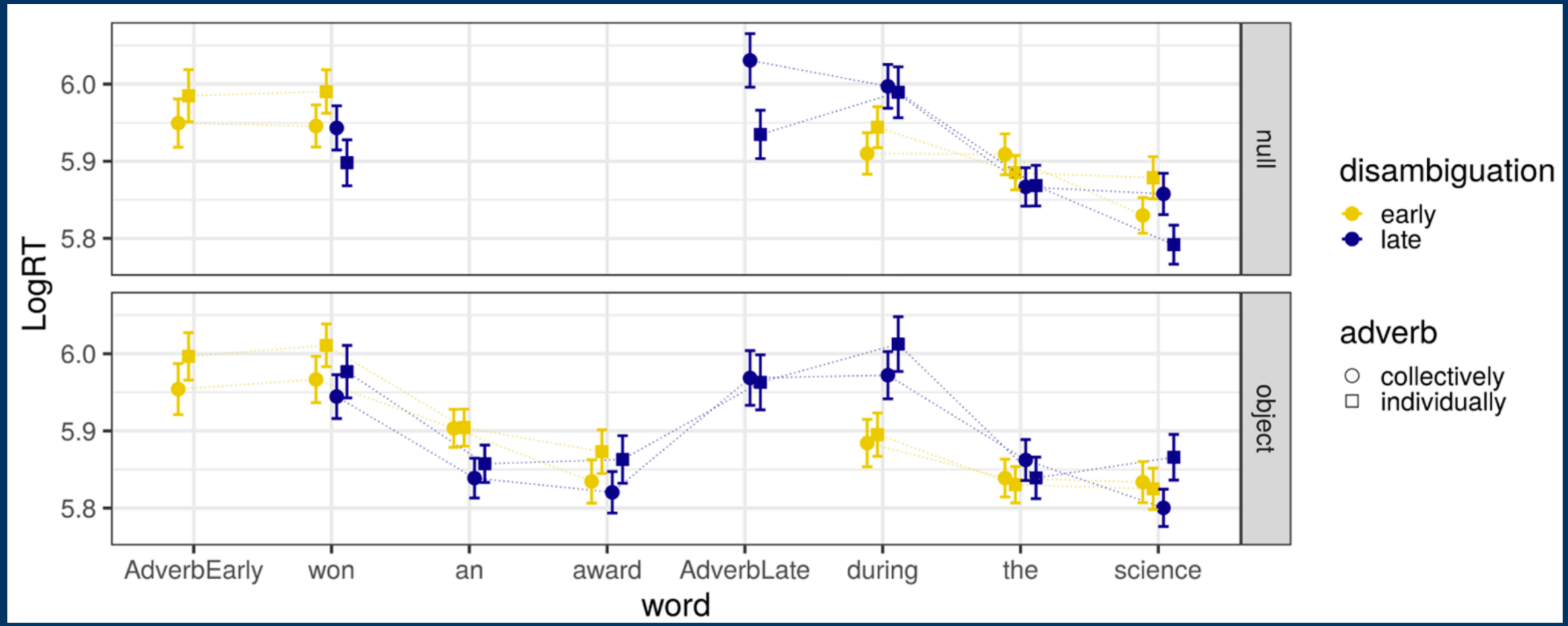


Dotlačil & Brasoveanu (2021) vs. This Study

- Small local cost for *together*, which we find more broadly in SPR & Maze.
- Temporarily ambiguous predicates are not read credibly faster for them.
  - In fact, late resolution is associated with a cost in the spillover.
  - Unclear why the findings are so different in this regard.
- Critical interaction here is borne out in the Maze as well, but not SPR
  - Their estimates:
    - Adverb: $\hat{\beta}$ = 0.05 $(0.01, 0.12)_{89\%}$
    - Spillover: $\hat{\beta}$ = 0.06 $(0.02, 0.11)_{89\%}$
  - Our estimates:
    - Adverb: $\hat{\beta}$ = -0.01 $(-0.07, 0.06)_{95\%}$
    - Spillover: $\hat{\beta}$ = 0.10 $(-0.05, 0.24)_{95\%}$

Dotlačil & Brasoveanu (2021): Expt. 2

- LATE *individually* speeds up reading for intransitives, slows for transitives
  - Compatible with an entity-cost explanation for the default?



Defaults and reanalysis with quantifier scope

A child saw every squirrel.

- Fodor (1982):
  - Inverse scope here is hard becausethe comprehender reads *a child*, and first prepares a mental model involving 1 child.
  - To get inverse scope, they would have to go back and imagine a (possibly) much larger set of children.
  - See also the *Principle of Parsimony* of Crain & Steedman (1985).
- Brasoveanu & Dotlačil (2015): Indeed, that cost is eliminated for inverse scope reanalysis that doesn’t affect the number of entities.

The same child saw every squirrel.

Appendix D: Eager commitment in the Maze

(Duff, Brasoveanu & Rysling @ CUNY 2021)

Underspecification

Unfortunately, the newspaper was destroyed…

after it lost its advertising profits.

(M1) (M2)

Reportedly, the jam displeased Tom…

after it doubled his morning commute.

(M1) (M2)

Claim: Full commitment to a particular meaning of a polyseme is delayed.

Why?

Utility: Because it's efficient when possible: prevents costly reanalysis.

Necessity: Because the processor cannot resolve polysemes without context.

What happens when underspecification wouldn't be useful?

Eye-tracking-while-reading: Frazier & Rayner (1990), Frisson & Pickering (1999), Pickering & Frisson (2001), Foraker & Murphy (2012), Brocher, Koenig, Mauner, & Foraker (2018)

Enter the Maze

(Obviously, the referee had…)

WELFARE DROPPED

(~40%)

The A-Maze (Boyce et al. 2020) encourages eager interpretation.

- Representing semantic context necessary to pick the correct target

↳ **Underspecification is no longer useful.**

If underspecification is utility-based / necessary then we won't / will see it in the Maze.

Reanalysis costs for homonymy and polysemy.

More reanalysis costs for homonymy.

Forster, Guerrera, & Elliot (2009), Boyce, Futrell, & Levy (2020)

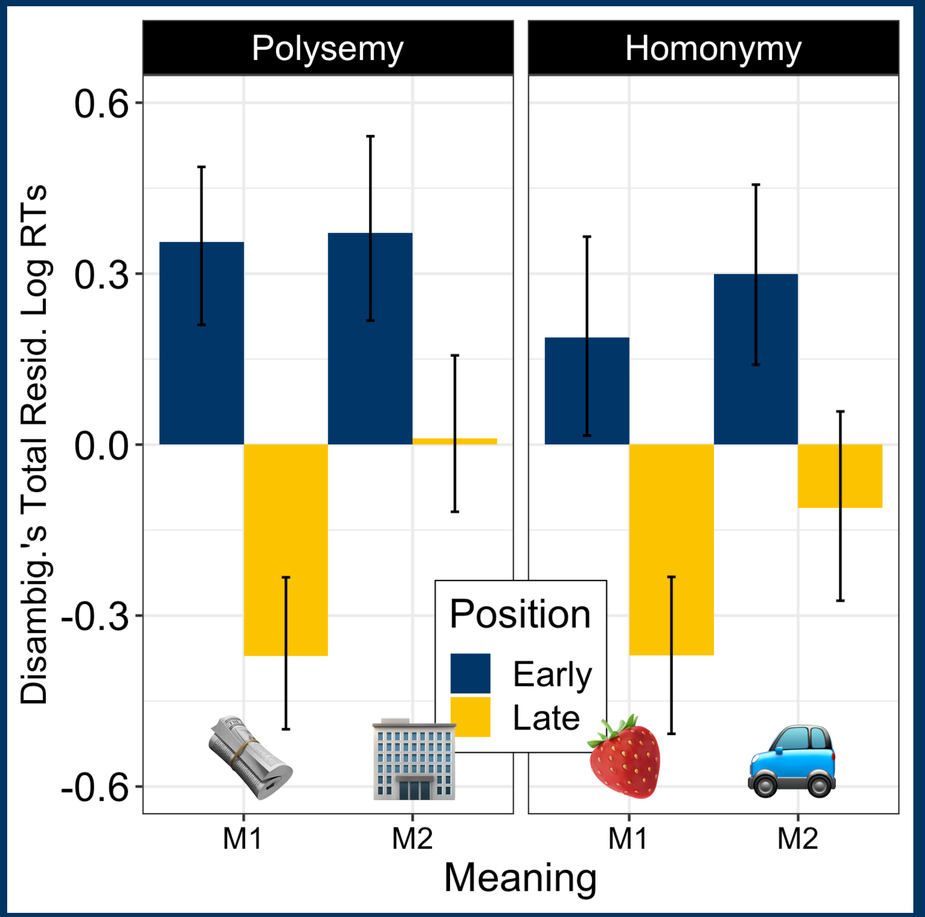
64 Latin-squared items (32 POL, 32 HOM); 128 fillers; *n* = 24 UCSC + 24 Prolific

E1: No underspecification in the Maze

M1, EARLY	Unfortunately, after it was soaked with rain the **newspaper** was destroyed.
M1, LATE	Unfortunately, the **newspaper** was destroyed after it was soaked with rain.
M2, EARLY	Unfortunately, after it lost its advertising profits the **newspaper** was destroyed.
M2, LATE	Unfortunately, the **newspaper** was destroyed after it lost its advertising profits.

- POSITION: **LATE** read faster, presumably due to cataphora in **EARLY**
- POSITION X MEANING: Reduced for M2, apparent reanalysis costs
- No POS X POL/HOM (X M): no difference in reanalysis for POL V. HOM
- Replicated in error rates (not shown): No POL/HOM difference

↳ **No evidence for necessary underspecification in the Maze.**



Log RTs residualized over position and length, summed, analyzed via LMER fit in STAN, fixed effects treatment-coded. Effects reported if 95% credible interval excludes 0.

64 Latin-squared items (32 POL, 32 HOM); 128 fillers; *n* = 24 UCSC + 24 Prolific

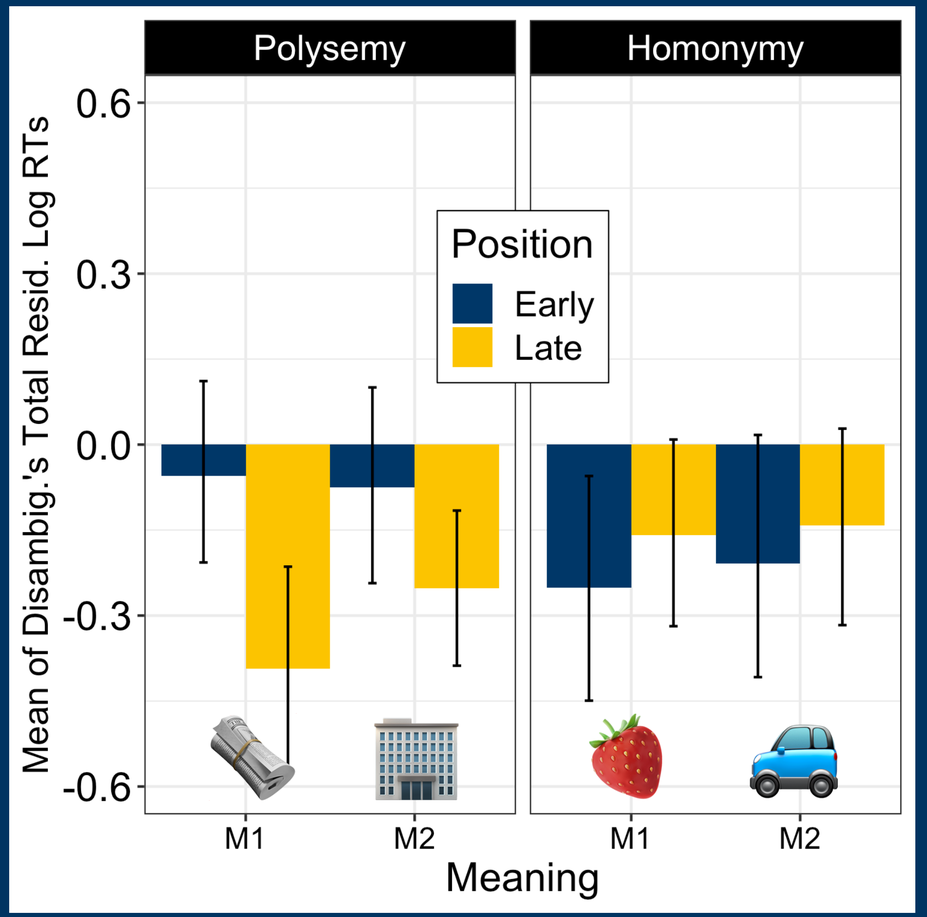
E2: Underspecification in SPR

M1, EARLY	Unfortunately, after it was soaked with rain the **newspaper** was destroyed.
M1, LATE	Unfortunately, the **newspaper** was destroyed after it was soaked with rain.
M2, EARLY	Unfortunately, after it lost its advertising profits the **newspaper** was destroyed.
M2, LATE	Unfortunately, the **newspaper** was destroyed after it lost its advertising profits.

SPR replication to ensure the Maze results are due to the task.

- POSITION: **LATE** read faster, again due to cataphora in **EARLY**
- POSITION X POL/HOM: Crossover for HOM, **extra reanalysis costs**

↳ **E1 results can be attributed to a Maze-specific task effect.**



Log RTs residualized over position and length, summed, analyzed via LMER fit in STAN, fixed effects treatment-coded. Effects reported if 95% credible interval excludes 0.

Upshots

Underspecification effects in polysemy are mediated by task demands.

- ↳ **Underspecification is optional and apparently strategic.**
- ↳ **Open questions remain: what makes it possible?**

The Maze task modulates strategies of incremental interpretation.

- ↳ **Shouldn’t be used as a 1:1 replacement for eyetracking or SPR.**
- ↳ **BUT: a powerful tool for clarifying the source of behavior.***

* e.g. Sloggett, Van Handel, Sasaki, Duff, Rich, Orth, Anand, & Rysling (2020 CUNY Poster)