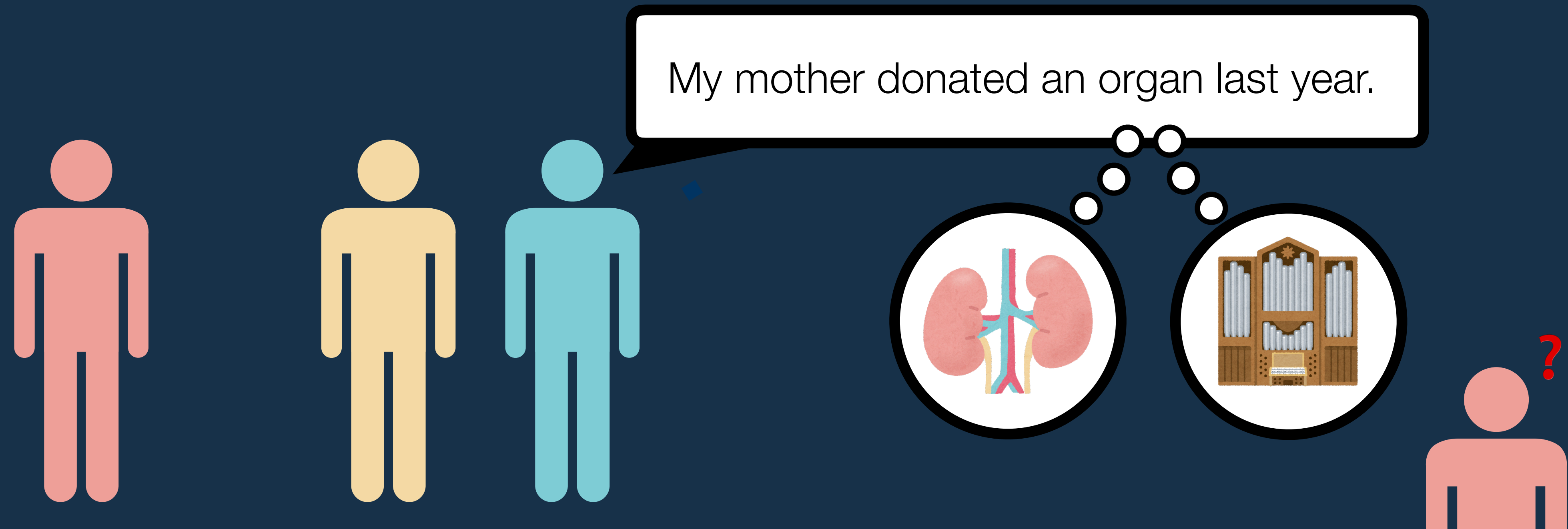# Semantic garden paths and the timing of comprehension decisions

John Duff ❖ Saarland University, Dept. of Language Science & Technology

6 May 2024

jduff@lst.uni-saarland.de
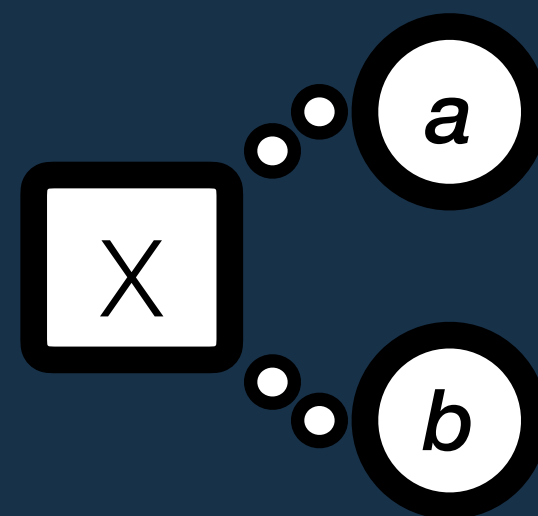
How and when do we move from awareness of multiple interpretations to something like a firm decision?

Different types of meaning

Different situations and pressures

We learn when to select a single meaning based on the expected utility and risk of a decision.

X → a
X → b

? ? ? ? ? ✅

Roadmap

1. **General background**

2. Case study #1: Polysemy and the strategic timing of decisions

3. Case study #2: Discourse inferences and the timecourse of delayed decision-making

4. Discussion and conclusions

Rapid decision-making

Though Hilda finally agreed to sing the German Christmas carols she chose <u>turned</u> out to be just awful.

(eyetracking; Frazier & Rayner 1982)

[to sing the German Christmas carols] ❌

[to sing], [the German Christmas carols… ✅

Garden-Path Effect

Comprehenders experience difficulty (= slow down) when a preferred analysis doesn't work out.

Garden-Path Effects

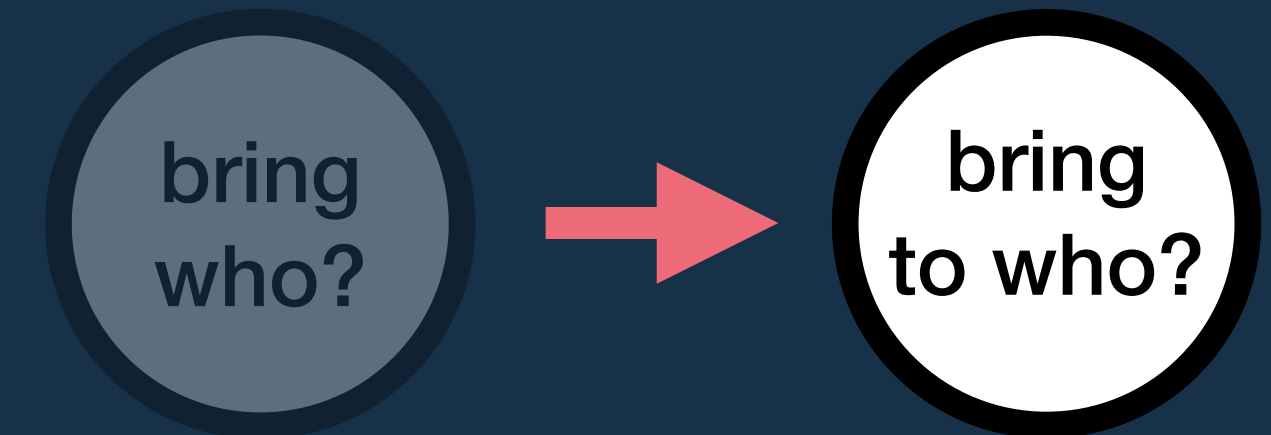**RAPID INCREMENTAL DECISION-MAKING:**
Comprehenders analyze the linguistic signal incrementally, making decisions before they have seen enough to be certain.

Many gardens, many garden paths

WH DEPENDENCIES

My brother wanted to know who Ruth will bring us home to at Christmas.

(eyetracking, Stowe 1986)

bring
who?

→

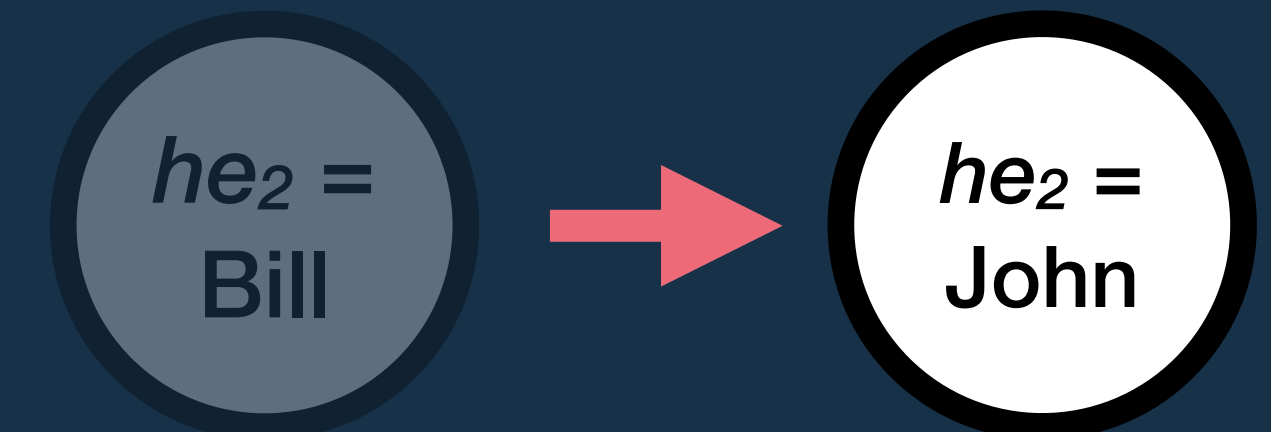bring
to who?

AMBIGUOUS PRONOUNS

Bill wanted John to look over some important documents. He had to mail him the documents. Unfortunately, he never received the papers.
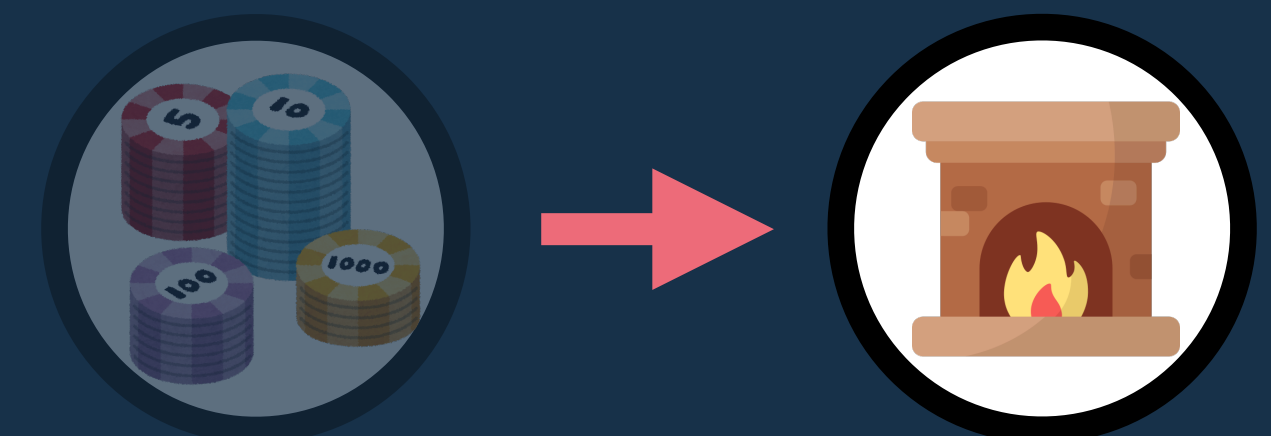
(self-paced reading, Gordon & Scearce 1995)

he_2 =
Bill

→

he_2 =
John

HOMONYMS

That type of poker has become rare due to changing fads in wood stoves.

(eyetracking, Frazier et al. 1990)

An exception!

HOMONYMY

poker

NOT related

POLYSEMY

newspaper

related!
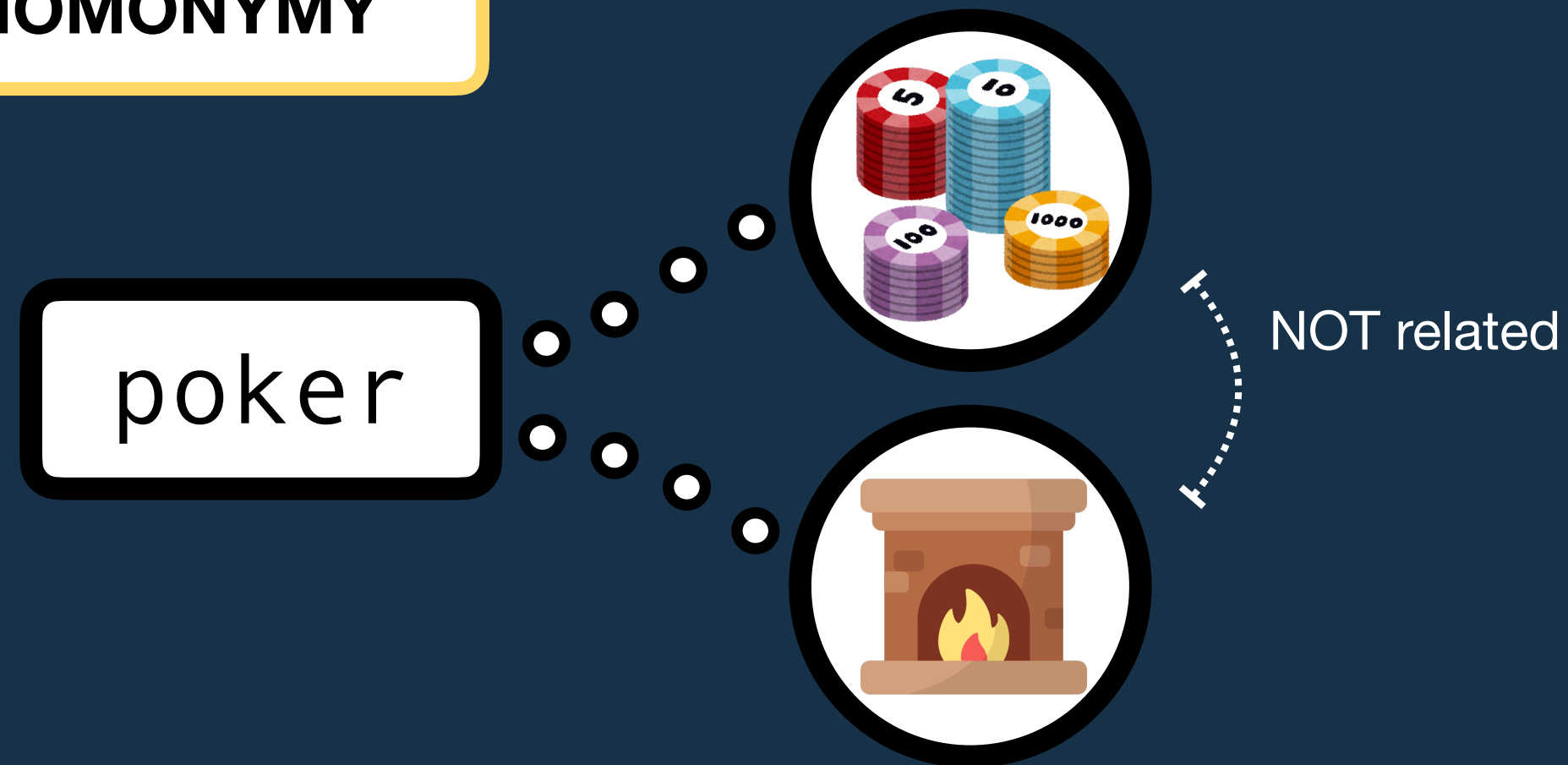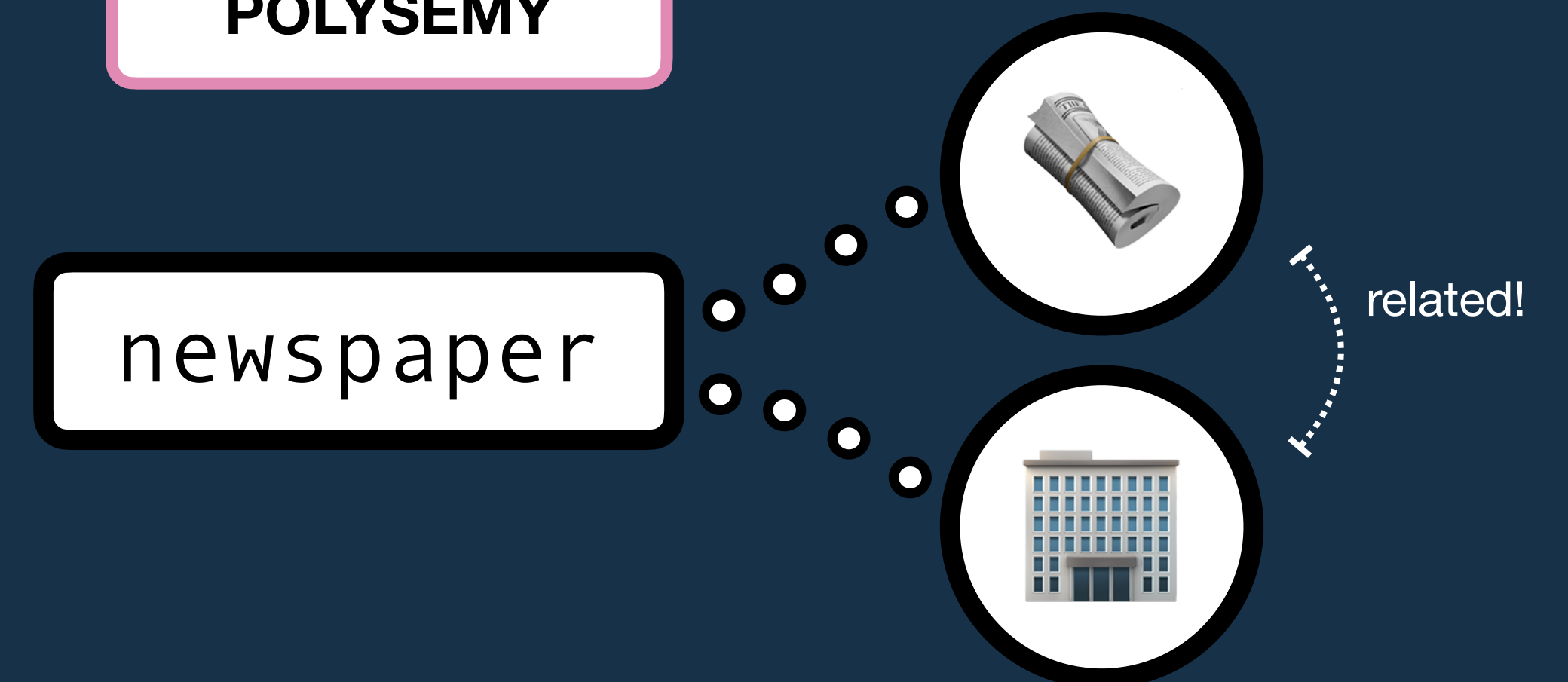
Unfortunately, the newspaper was destroyed, managing advertising so poorly. 👍

(eyetracking, Frazier et al. 1990)

Typical garden-path effects are not observed in the comprehension of polysemy.

(see also Frisson & Pickering 1999; Foraker & Murphy 2012; Brocher et al. 2016, 2018)

Absence of Garden-Path Effects

DEFERRED DECISIONS:
Comprehenders delay some decisions for longer than others.

Why wait for polysemy, in particular?

Frazier (1999): It's a matter of strategy.
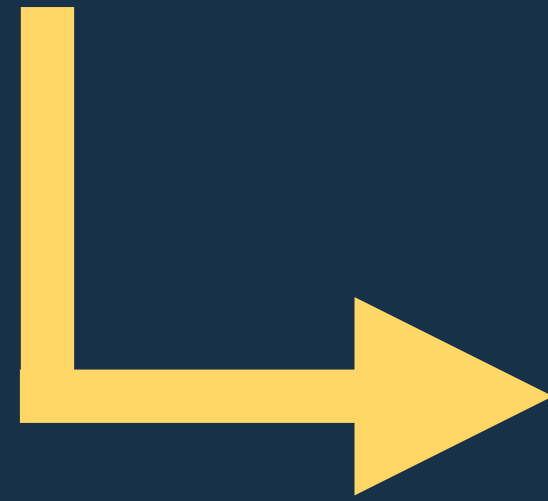
Why analyze rapidly?	Why delay analysis?
• Memory for linguistic representations is better than memory for raw sensory input. • Analysis of input $x… x_n$ is necessary to analyze input $x… x_{n+1}$	• Early analysis can lead to mistakes, requiring re-analysis, which is slow and difficult.

Make minimal decisions: just enough to generate a linguistic representation.

(Assumption: "Underspecified" representations are possible for polysemy, not homonymy.)

WHEN WE DEFER ANALYSIS (FIRST ATTEMPT):

We only select minimal linguistic representations during incremental comprehension.

Anything which is extra-grammatical will be left for later.

NEW OBSERVATION #1:

Comprehenders don't always defer the sense of a polyseme.
They select a sense rapidly when it would be useful for them.

STRATEGIC DEFERMENT IS FLEXIBLE

NEW OBSERVATION #2:

Comprehenders also postpone decisions about pragmatic inferences, even though they generate and exploit possible inferences rapidly.

RAPID CONSIDERATION WITHOUT SELECTION

Roadmap

1. General background

2. **Case study #1: Polysemy and the strategic timing of decisions**

3. Case study #2: Discourse inferences and the timecourse of delayed decision-making

4. Discussion and conclusions

The Maze task

- Participants must make word-by-word decisions between a plausible continuation and an implausible foil.

- Picking a foil results in an error message, ends the trial

- Here, foils generated by sampling low-probability continuations from a large language model, the "A-Maze" method of Boyce et al. (2020).

Surprising finding: Maze response latencies are highly correlated with measures of reading time in eyetracking experiments.

Why? It seems that participants comprehend and integrate the correct continuation as they pick it.

(Freedman & Forster 1985; Nicol et al. 1997; Forster et al. 2009; Witzel et al. 2012; Boyce et al. 2020; Boyce & Levy 2023; Levinson 2023; Hoeks et al. 2023)

The Maze as a probe for strategic performance

- **Forster et al. (2009):** the Maze encourages "strictly incremental processing"

  - Higher pressure: If you don't develop a complete interpretation of word *x*, you might make the wrong choice for word *x+1*.

- We can investigate the relationship between deferring decisions and strategy.

  - Comprehenders delay in selecting one analysis of a polyseme because…

    - *Option #1*: they follow a rigid heuristic: never commit to more than necessary for representation.

      PREDICTION: They will also delay in the Maze.

    - *Option #2*: they follow an active, flexible strategy based on the utility and risk of an early commitment.

      PREDICTION: We will see rapid selection in the Maze.

Two effects related to selection

Garden-Path Effects

EARLY M1 | due to changing fads in card games that type of poker has become rare

LATE M1 | that type of poker has become rare due to changing fads in card games

EARLY M2 | due to changing fads in wood stoves that type of poker has become rare

LATE M2 | that type of poker has become rare due to changing fads in wood stoves

Interaction: Difficulty associated with LATE **disambiguation is greater in** M2

Two effects related to selection

Subordinate Selection Costs

EARLY	M1	due to changing fads in card games that type of poker has become rare
LATE	M1	that type of poker has become rare due to changing fads in card games
EARLY	M2	due to changing fads in wood stoves that type of poker has become rare
LATE	M2	that type of poker has become rare due to changing fads in wood stoves

Interaction: Difficulty for targets with EARLY disambiguation is greater in M2

(Duffy et al. 1988; Rayner & Frazier 1989; Sereno et al. 1992; Rayner et al. 1994; Folk & Morris 1995; Binder & Rayner 1998; Kambe et al. 2001)

Why is subordinate selection costly?

- Lexical access literature: All meanings of a word are accessed initially.

  (cross-modal priming: Swinney 1979; Onifer & Swinney 1981… see Morris 2006 and Rodd 2018 for review)

- Exact mechanisms for making a choice proposed by psycholinguists differ.

  (Duffy et al. 1988; Rayner & Frazier 1989; Duffy et al. 2001)

- Psychology of decision-making offers a general explanation: Decisions are slower when sources of information are in conflict.

  (Luce 1986; Ratcliff 1985; Ratcliff & McKoon 2008)

- Bias based on context-free preference: M1
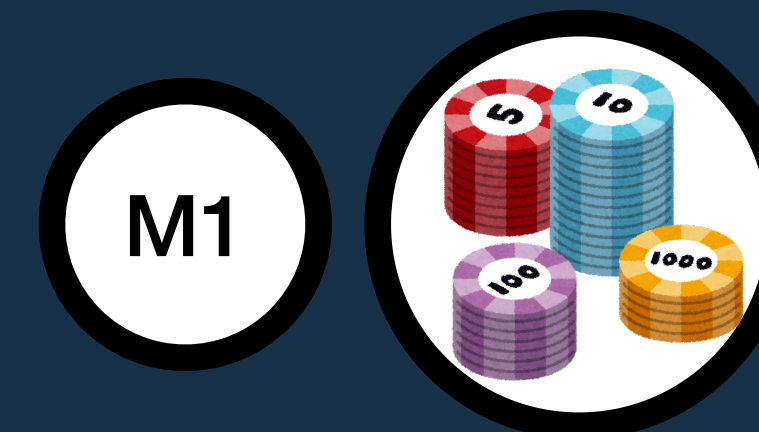
- Evidence from preceding context: M2

All incremental decisions might exhibit subordinate selection costs.

Extra source of evidence!

Experiment 2: Design

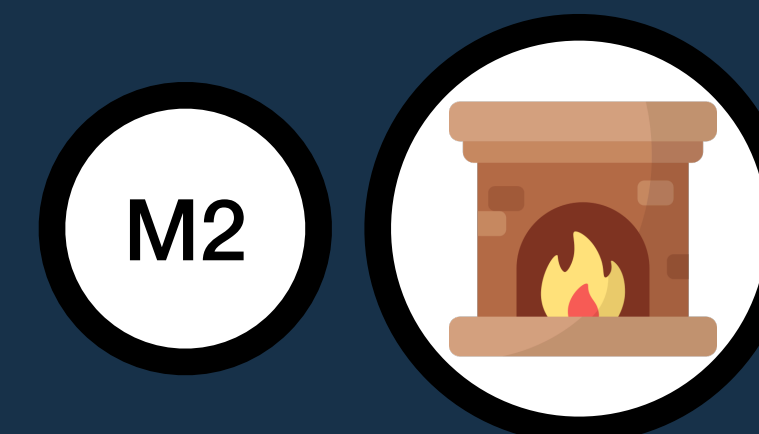- We're looking for whether Maze readers avoid typical delays in decisions about the meaning of a polysemous noun.

- Two effects in our toolbox: Garden Paths and Subordinate Selection Costs

- The present experiment: Frazier & Rayner (1990) in the Maze

    - Items expanded, with new norms for meaning dominance

    - 48 participants, 32 sentences with polysemes, 32 with homonyms

Experiment 2: Materials and predictions
(Polysemy)

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

Polysemes are still deferred   **Polysemes are interpreted immediately**

Disambiguation   M2 ≈ M1   extra **LATE** cost for M2   *(Garden Path)*

Target/Spillover   M2 ≈ M1   extra **EARLY** cost for M2   *(Subord. Sel.)*

Experiment 2: Results
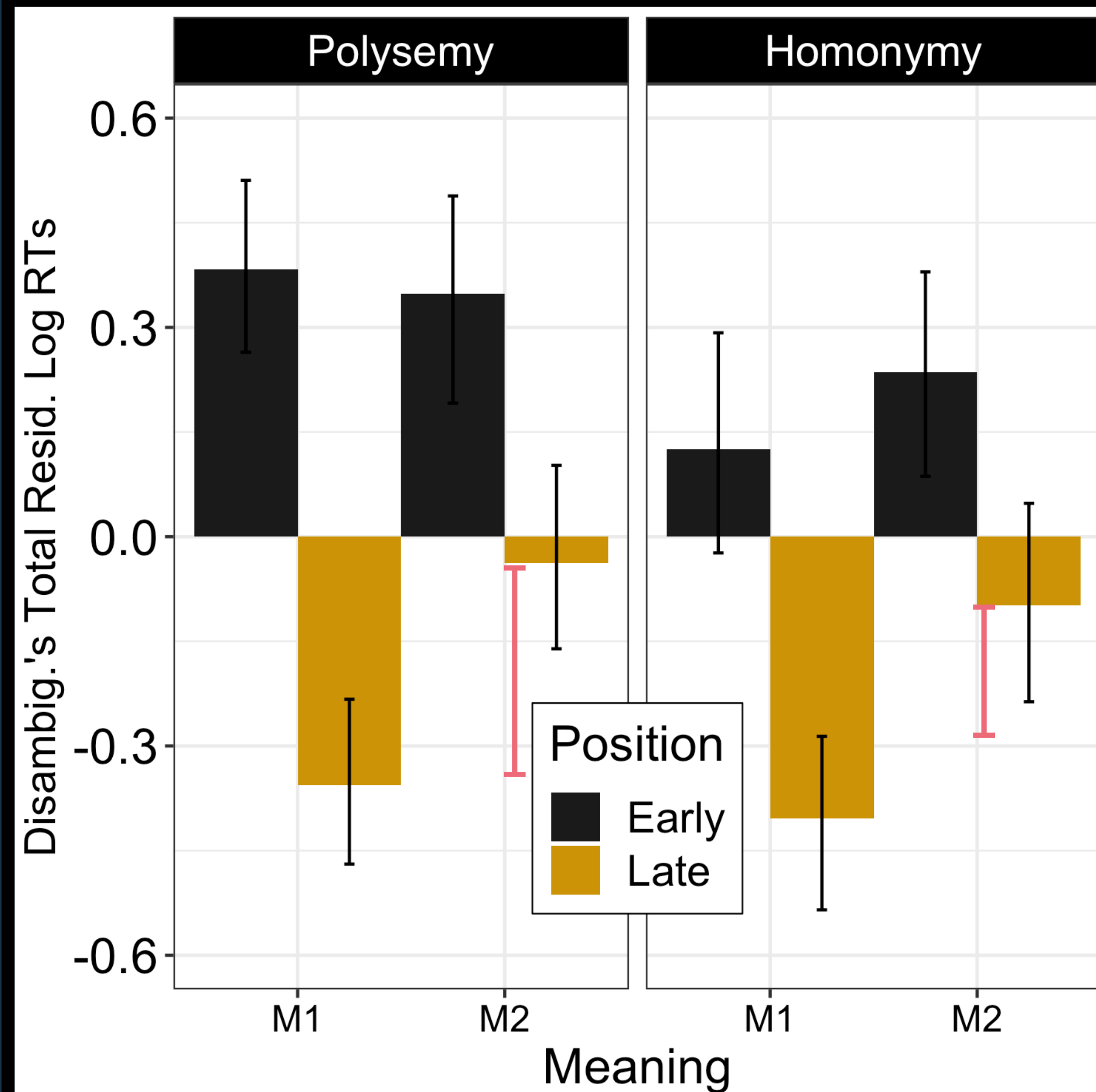
Disambiguation

Garden Path interaction for Polysemy:

extra LATE cost for M2

$$\hat{\delta} = 0.38, \; P(\delta > 0) = 0.99$$

Behavior matches Homonymy:

no credible differences across sub-experiments

$$\hat{\beta} = (-0.08, 0.04)_{.95}$$
$$BF_{10} = 0.55$$

Bayesian linear mixed-effects regression fit in `brms` using weakly-informed priors

β = regression coefficient
δ = marginal contrast

Experiment 2: Results

Spillover

Subordinate Selection interaction for Polysemy:

extra EARLY cost for M2

$$\hat{\delta} = 0.22, \; P(\delta > 0) \approx 1.00$$

Behavior matches Homonymy:

no credible differences across sub-experiments

$$\hat{\beta} = (-0.00, 0.06)_{.95}$$
$$BF_{10} = 0.06$$

Bayesian linear mixed-effects regression fit in `brms` using weakly-informed priors

β = regression coefficient
δ = marginal contrast

Experiment 2: Discussion

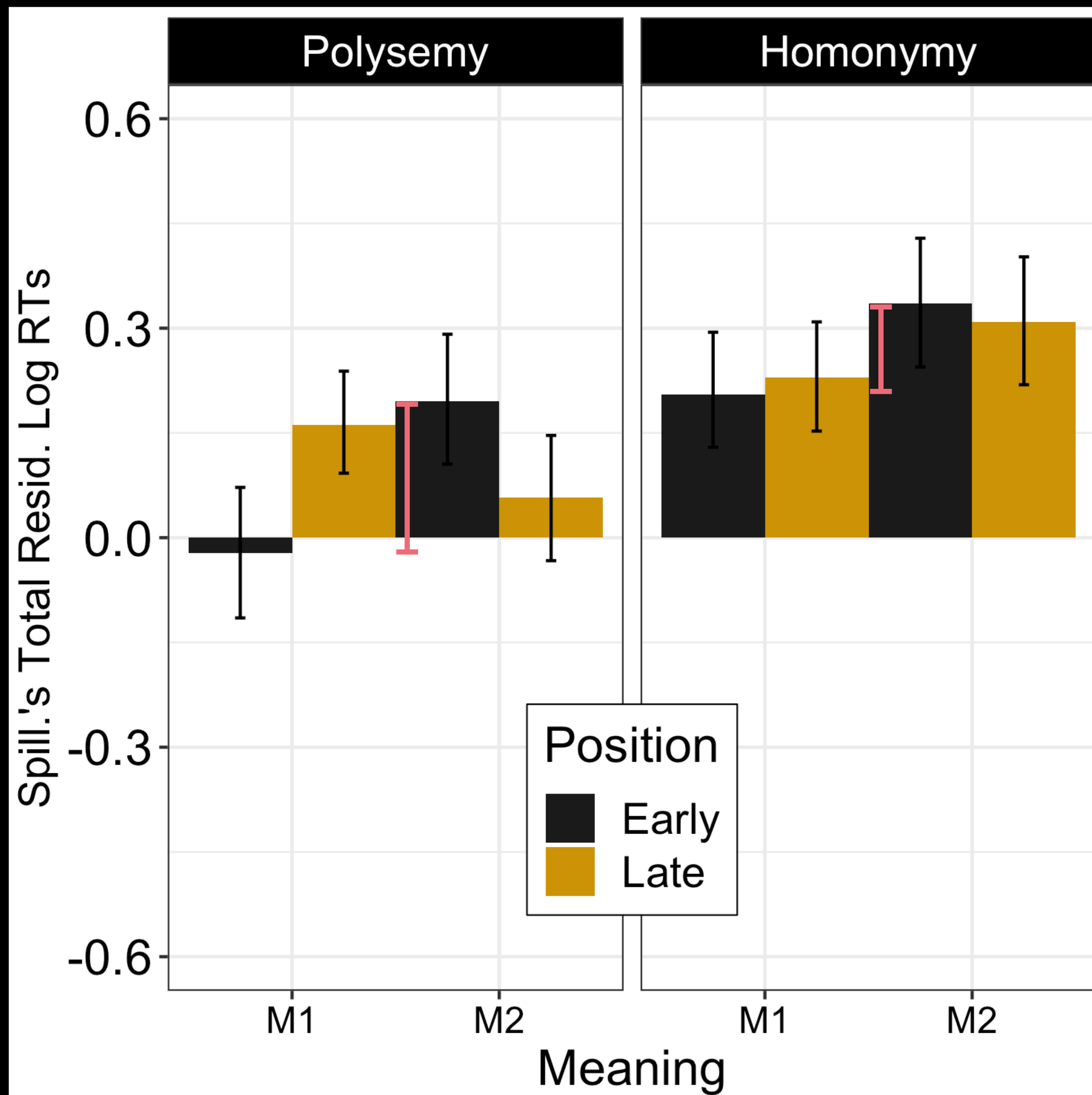|	Polysemes are still deferred	Polysemes are interpreted immediately
Disambiguation	M2 ≈ M1	extra **LATE** cost for M2 *(Garden Path)* ✅
Target/Spillover	M2 ≈ M1	extra **EARLY** cost for M2 *(Subord. Sel.)* ✅

In the Maze, **polysemous nouns are processed like homonyms**:
Comprehenders immediately select a complete interpretation.

(This really is about the Maze. In Experiment 1, I show that these early
commitment patterns don't show up in non-Maze self-paced reading.)

Experiment 2: Discussion

- Methodological upshot:

  - Maze comprehension is **not** a direct analogue for reading.

  - Forster et al. (2009) were right: Extra pressures for earlier decisions.

  - Experiments 3 and 4: similar task effects for distributivity ambiguities.

- Theoretical upshot:

  - Deferment behavior is not constant for all "post-grammatical" meaning.

  - We can better explain the data if deferment is based on a flexible calculus actively adapted to the comprehension task at hand.

STRATEGIC DEFERMENT IS FLEXIBLE

Experiment 2: Discussion

- Previous work has also observed selection effects for polysemy…

    - With rich preceding contexts.
      (eyetracking: Lowder & Gordon 2013; Brocher et al. 2016, 2018; even arguably Frazier & Rayner 1990; see also Bott et al. 2016)

    - Due to heuristic preferences for animate subjects.
      (eyetracking: Fishbein & Harris 2014)

    - Even in neutral contexts, Garden-Path Effects show up if disambiguation comes after the sentence boundary. *(eyetracking/SPR: Frisson & Frazier 2004; Foraker & Murphy 2012)*

- Even in "typical" reading of a polyseme, selection is only delayed in the absence of an informative context, and even then only for a little while.

Experiment 2: Discussion

- What specific factors encourage early selection?

  - Informative context, perhaps because confidence is high and therefore reanalysis is less likely.

  - The pressures of the Maze task, perhaps because…

    - A full analysis would be more useful for proceeding in the task.

    - Comprehenders lack the resources to maintain uncertainty.

Where to now?

- Proposed: Comprehenders defer selection due to a context-sensitive calculus over expected costs and benefits.

- A few reasonable questions:

  - What other kinds of interpretive decisions are often deferred?

  - Does the Maze encourage early decisions about **everything**?

  - What happens when we defer decisions?

    - Do we generate and consider the alternatives at hand?

    - Do we reach a certain level of analysis and stop?
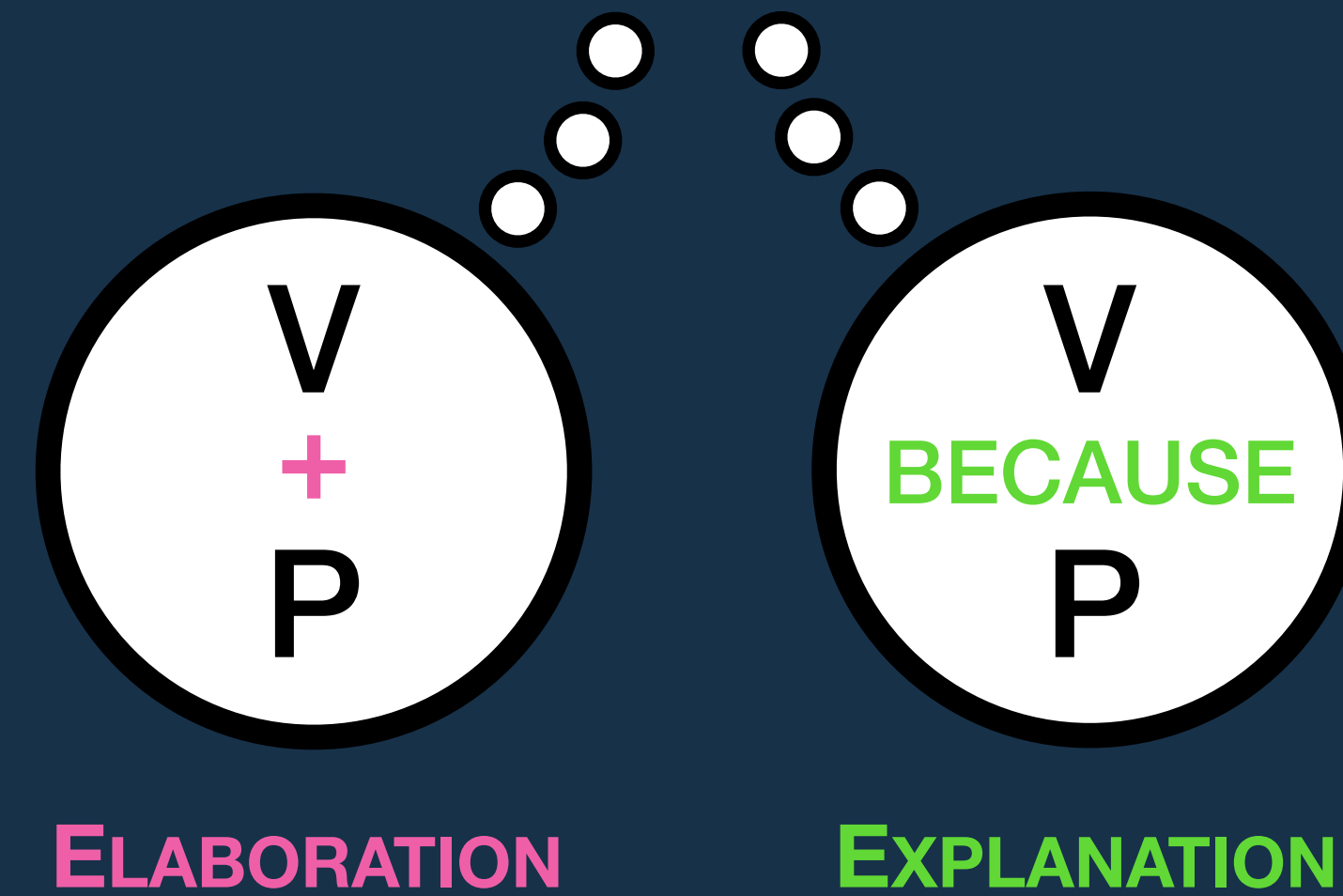
Roadmap

1. General background

2. Case study #1: Polysemy and the strategic timing of decisions

3. **Case study #2: Discourse inferences and the timecourse of delayed decision-making**

4. Discussion and conclusions

Uncertainties of discourse coherence

Where do these inferences come from?

head

V
+
P

ELABORATION

tail

V
BECAUSE
P

EXPLANATION

- Comprehenders reconstruct a communicative purpose for the current unit (e.g. P).

- To do so, they select from a small library of binary coherence relations, to categorize how P is related to some previous unit of the discourse (here, V).

  - Shallow inferencing: Coherence is a domain-general operation of communicative inferencing.

    (Hobbs 1979; Kehler 2002…)

  - Hierarchical structure-building: Coherence is a by-product of building hierarchical discourse structure, linked with other interpretational restrictions.

    (Reichman 1978; Polanyi 1985; Asher & Lascarides 2003…)

Where do these inferences come from?

head

V
+
P

ELABORATION

V
BECAUSE
P

EXPLANATION

tail

- Comprehenders reconstruct a communicative purpose for the current unit (e.g. P).

- To do so, they select from a small library of binary coherence relations, to categorize how P is related to some previous unit of the discourse (here, V).

- Shallow communicon ... of ... bs 1979; Kehler 2002...)

- Hierarchi... hierarchi... building ...nal restrictions. ...er & Lascarides 2003...)

This difference won't be important in this talk.

Either way, these inferences are derived as part of every-day interpretation.

(My conclusions won't depend on or argue for which approach is correct.)

When do we construct these inferences?
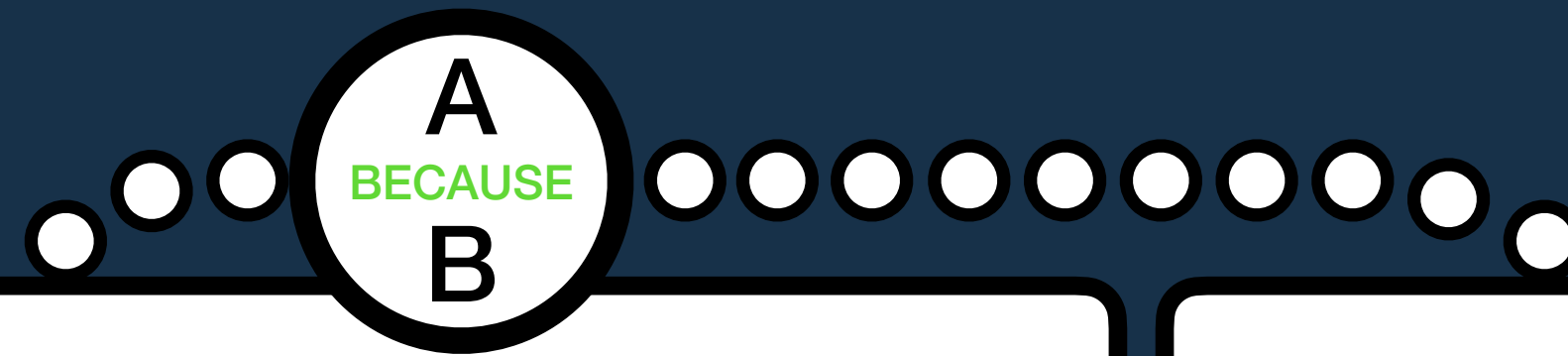
A
BECAUSE
B

| the protestor got a fine from the policemen | when he broke the rules during the demonstration 👍 |
| the protestor spoke with the policemen | when he broke the rules during the demonstration |

(eyetracking: Mak & Sanders 2013)

When contexts make an Explanation more likely, comprehenders processing the tail seem to anticipate Explanation-consistent content.

- Less difficulty on lexical processing for related vocabulary.

(eyetracking/SPR: Mak & Sanders 2013, Traxler et al. 1997, Cozijn 2000, Hoek et al. 2021)
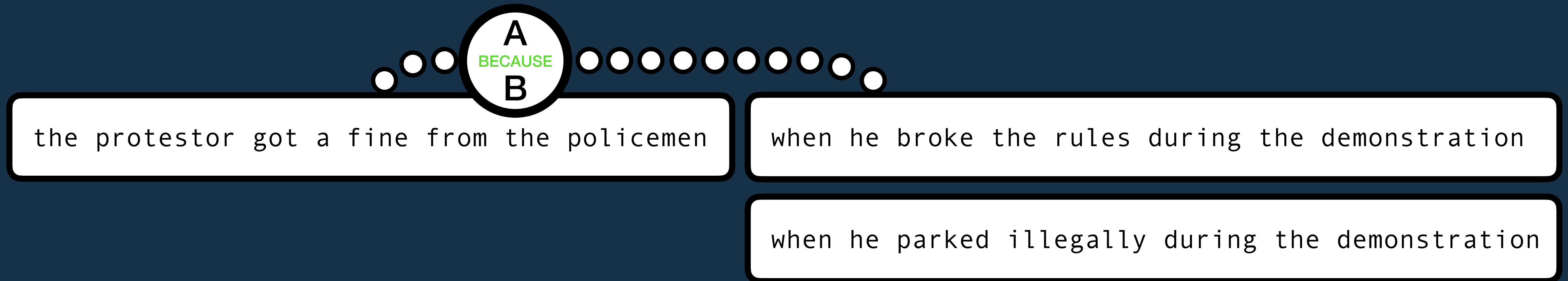
- Less difficulty processing anaphoric reference to causally-salient referent.

(eyetracking/SPR: Koornneef & Van Berkum 2006, Koornneef & Sanders 2013; priming: McDonald & MacWhinney 1995; offline: Kehler et al. 2008)

- Resolve structural ambiguities in favor of describing causally-salient referent.

(SPR: Rohde et al. 2011)

Is this really about coherence?

A
BECAUSE
B

the protestor got a fine from the policemen

when he broke the rules during the demonstration

when he parked illegally during the demonstration

- Facilitation effects are weaker when explanations are less prototypical.

(eyetracking/SPR: Traxler et al. 1997, Cozijn 2000)

Is this really about coherence?

A
BECAUSE
B

the protestor got a fine from the policemen

when he broke the rules during the demonstration

and he broke the rules during the demonstration
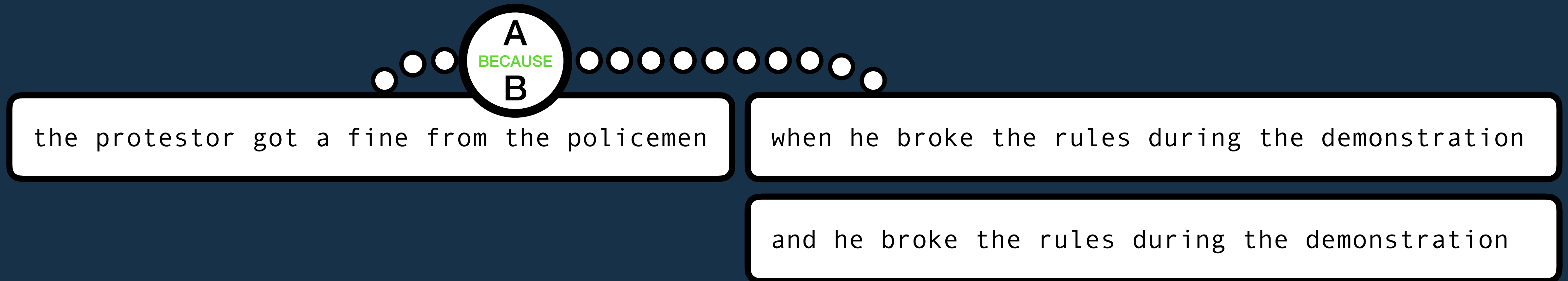
- Facilitation effects are eliminated when a connective prevents the relation.

(eyetracking: Koornneef & Sanders 2013)

So, rapid decision for an Explanation?
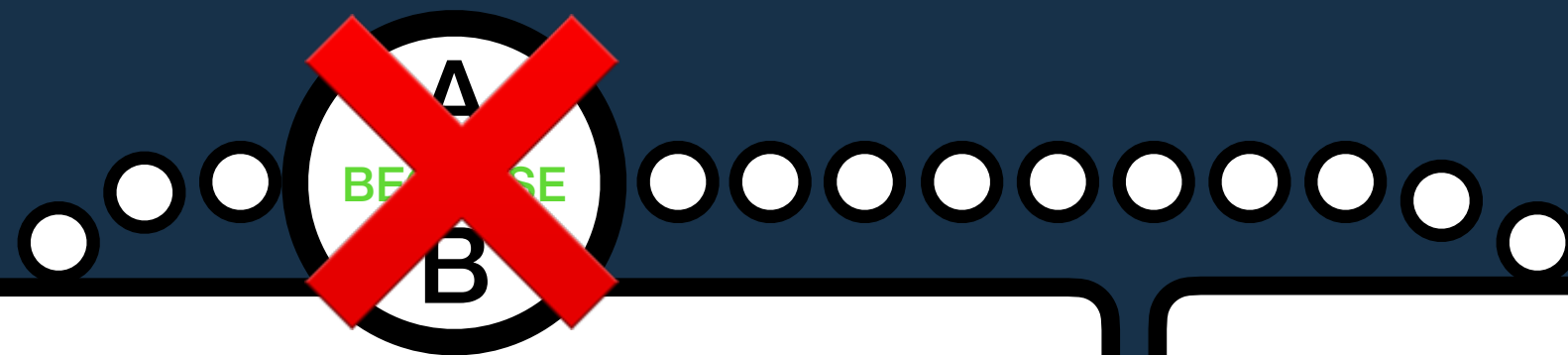
the protestor got a fine from the policemen

when he broke the rules during the demonstration

they fined him because he had an overdue library book

If so, we should also expect to see Garden-Path Effects when comprehenders are cued towards an Explanation, but then disambiguated away.

Experiment 9: Design

C1 Sally lives in a small city, where recently there was a citywide election for a new mayor with several candidates, and she had to decide among them on her mail-in ballot.

KNOWLEDGEABLE PROTAGONIST EX

C2 She spent some time reading everything she could about the candidates before mailing in her ballot.

She didn't have any time to read anything about the candidates before mailing in her ballot.

NEUTRAL PROTAGONIST EN

S1 In the end, she voted for Pat Mirabella.

S2 He has the most progressive platform in the race.

S3 He's from a very socio-economically diverse area... championed public programs.

S4 She voted for him because his name was first on the ballot. 🚩

Experiment 9: Results on *Because*

No Garden-Path Effect
for Knowl. Protagonist:

$$\hat{\beta} = (-0.02, 0.01)_{.95}$$

$$BF_{10} = 0.07$$

Bayesian linear mixed-effects regression fit to summed
log RTs in `brms` using weakly-informed priors

RTs by Segment within Because Clause
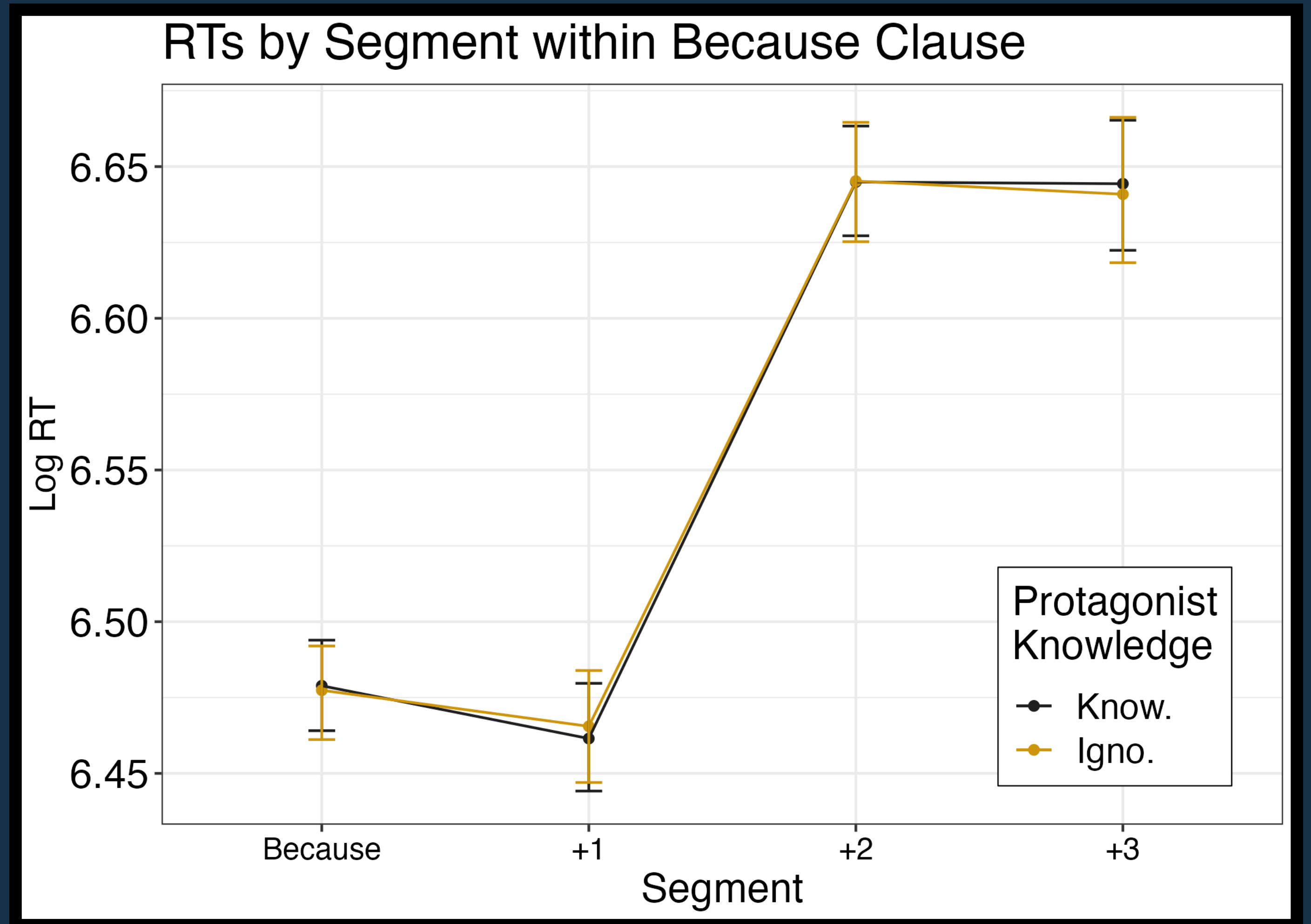
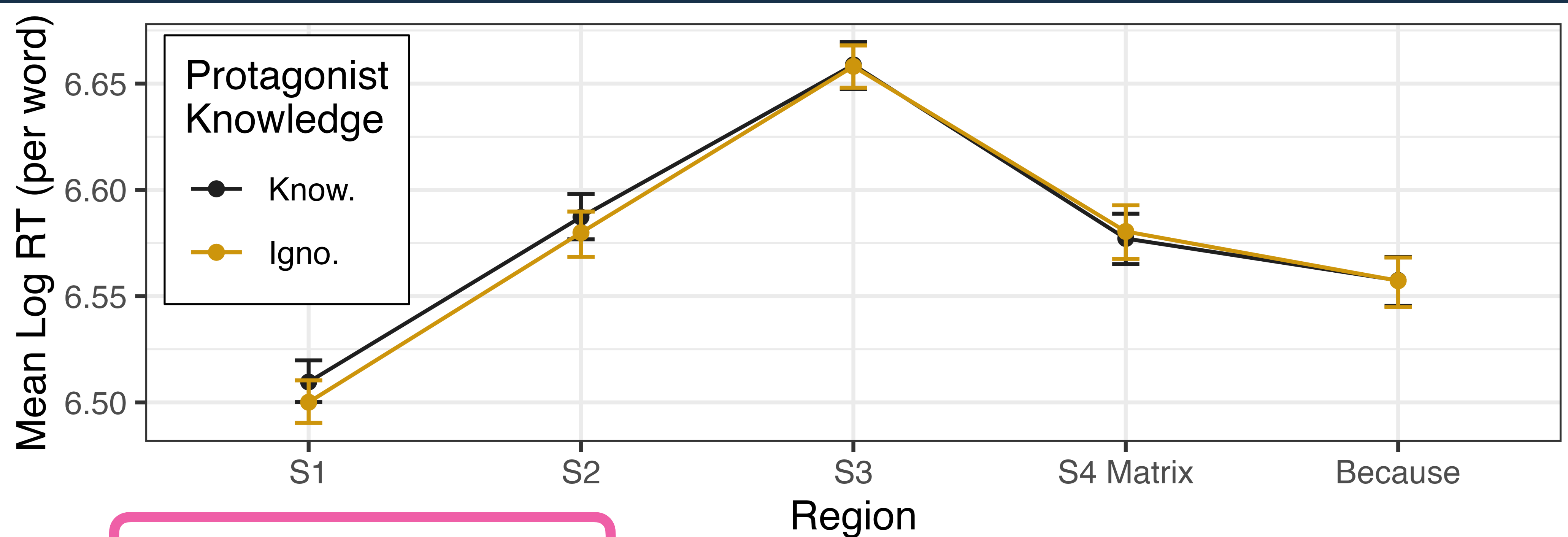**S4** She voted for him because his name was first on the ballot.

Experiment 9: Results across regions

No S2 facilitation effect for Knowl. Protagonist:

$$\hat{\beta} = (-0.02, 0.01)_{.95}$$

$$BF_{10} < 0.0001$$

Experiment 9: Discussion

- Immediate worry: Were comprehenders just insensitive to the manipulation?

  - Not a fact of the Maze: similar results in SPR (Expt. 8).

  - In general, readers are known to struggle to take protags.' knowledge into account when trying to explain their actions. *(Johnson & Keil 2014)*

  - Here, in belief attribution CQs, ~65% accuracy for ignorant protags.

- Still, converging evidence from other studies with similar manipulations.

  - Expt. 7: Explanation-linked facilitation effects, but still no Garden Paths.

  - Expts. 5 and 6: Likewise, evidence for rapid activation of scalar implicatures, but no evidence for Garden Paths.

Experiment 9: Discussion

- If comprehenders really don't reach a firm decision to infer an Explanation, how can we explain the facilitatory effects noted in the literature?

  - One tactic: Admit the possibility of graded expectations without selection.

  - Stage 1: Comprehenders are aware of multiple competing meanings.

    - Some meanings may be favored over other meanings.

    - Decision-contingent expectations may be generated proportionally.

  - Stage 2: Comprehenders select a single meaning.

    - Any revision will provoke a Garden Path.

RAPID CONSIDERATION WITHOUT SELECTION   **(NO WORRIES IF NOT)**

Experiment 9: Discussion

Homonym resolution

RAPID SELECTION

Polyseme specification

FLEXIBLE DEFERMENT OF SELECTION

Discourse coherence

RAPID CONSIDERATION WITHOUT SELECTION

For higher-level pragmatic inferences, selection may be deferred indefinitely, even in the Maze. Why?

- Task-motivated strategic selection is limited to certain domains?

- There is less utility (higher risk/lower reward) for incremental pragmatic decisions?

Roadmap

1. General background

2. Case study #1: Polysemy and the strategic timing of decisions

3. Case study #2: Discourse inferences and the timecourse of delayed decision-making

4. **Discussion and conclusions**

Core hypothesis

The selection of a single interpretation:

• Can be understood in the same way across different ambiguities

• Is variably-timed based on a flexible calculus of expected value

• Often follows an earlier stage where all candidates are generated

A (very) generalized model of comprehension decisions

ENCOUNTER INPUT WITH UNCERTAIN MEANING

GENERATE POSSIBLE INTERPRETATIONS

May develop preferred interpretation, and use to drive expectations.

Length of delay is subject to strategic variation.

Dispreferred disambiguation is not costly.

input

M1

M2 ✅

SELECT ONE INTERPRETATION

Costly process when faced with conflicting evidence/biases.

Dispreferred disambiguation becomes costly (Garden-Path Effects).

Contrasting with existing proposals

Minimal Commitment/Underspecification/"Good Enough" Models

(e.g. Frazier 1999; Frisson 2009; Swets et al. 2008)

Similar: Two discrete stages, full interpretation can be deferred.

Different: Full analyses are rapidly generated even when selection is deferred.

Parallel/"Constraint-Satisfaction"/Probabilistic Expectation Models

(e.g. MacDonald et al. 1994; Levy 2008)

Similar: Rapid parallel consideration with sensitivity to context, meaning.

Different: Two discrete stages, selection is real, and Garden Paths are true reanalysis.

A typology of selection timing

Immediately   Phrase boundary   Sentence boundary   Beyond

pronouns

homonymy

distributivity

modifier attachment

polysemy

aspect

quantifier scope

implicature

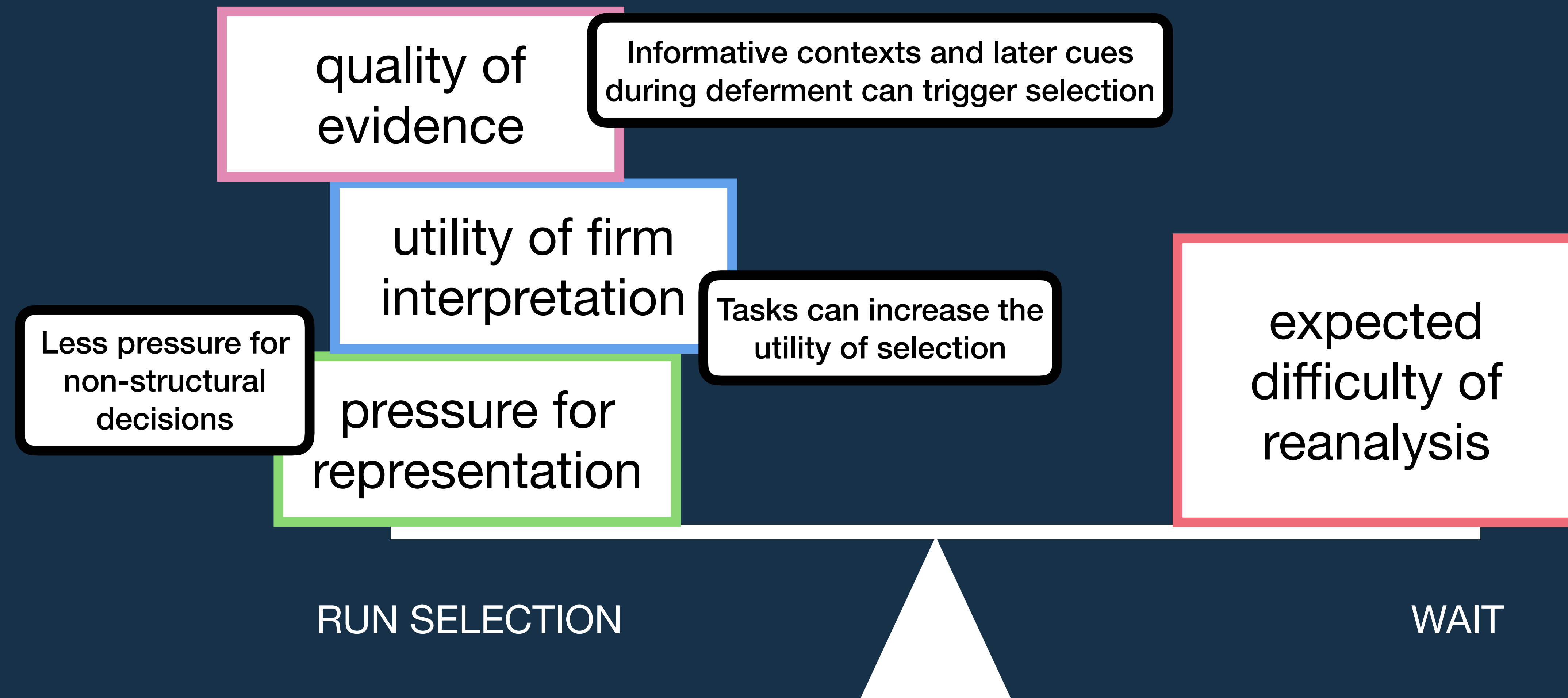**causal inference**

temporal order

A goal: Concrete theory for the variability in timing

Can we theorize a calculus for deferment that can predict both
(a) task effects and (b) differences among types of ambiguity?

quality of evidence

Informative contexts and later cues during deferment can trigger selection

utility of firm interpretation

Less pressure for non-structural decisions

Tasks can increase the utility of selection

pressure for representation

expected difficulty of reanalysis

RUN SELECTION

WAIT

A goal: Concrete theory for the variability in timing

How to independently estimate these parameters?
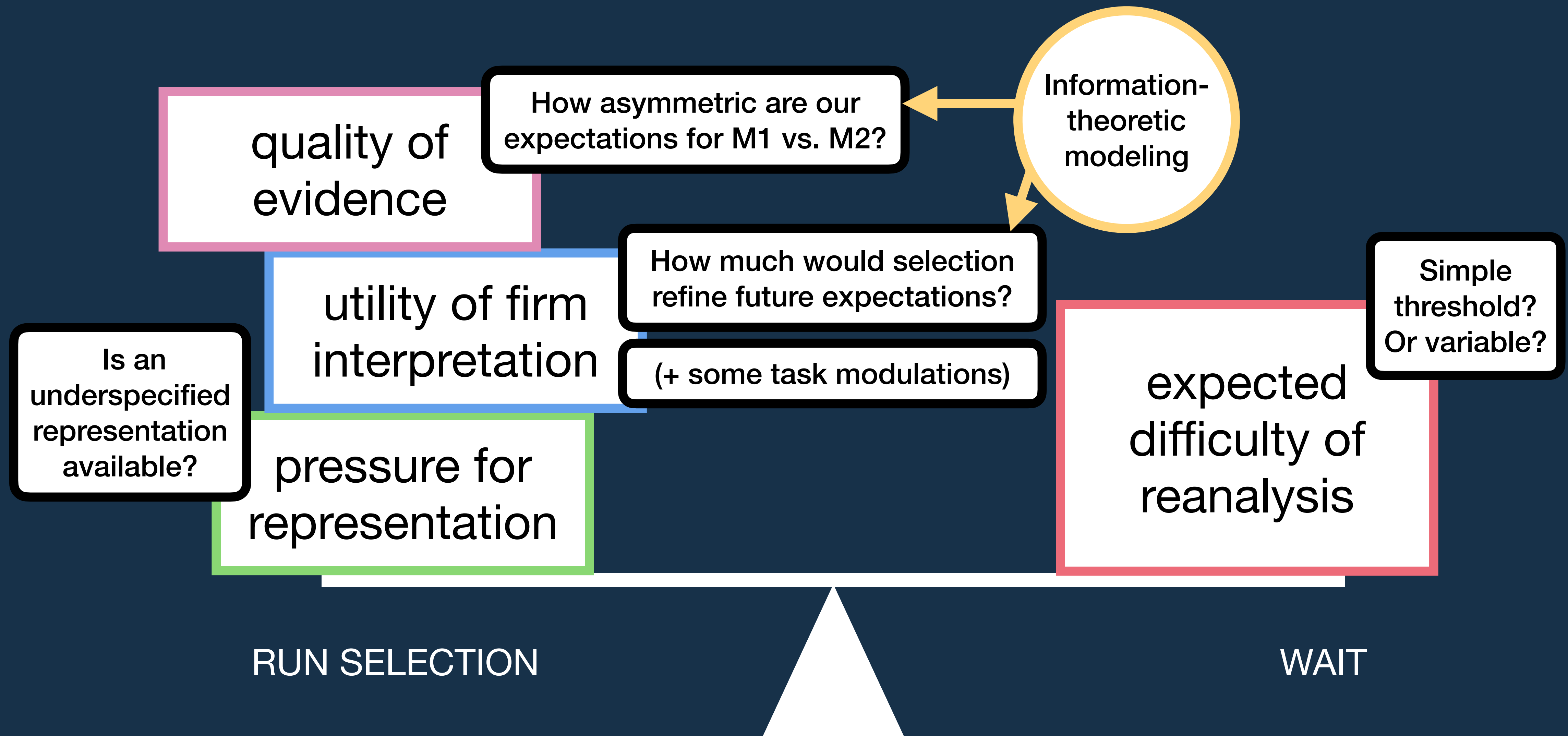
What about the children?

- If selection timing is a experience-based value-maximizing calculus…

  - Modelable using domain-general computational theories of on-the-job process optimization (e.g. Reinforcement Learning).

  - Could be acquired through trial and error, exploration of the costs and benefits to different strategies for uncertain meaning.

    (Sutton & Barto 2018; see e.g. Brasoveanu & Dotlačil 2021)

- How do developing comprehenders (children) handle ambiguity?

  - Like adults, rapid consideration of possible meanings.

  - Unlike adults, not sensitive to all cues, and struggle with reanalysis.

  - Not so much work separating expectations from selection (…no reading!), room for other kinds of incremental interpretation tasks to be recruited.

    (Trueswell et al. 1991; reviews in Snedeker 2013; Omaki et al. 2015; Lidz 2018)

Thanks!

Ask me about:

Task effects in the processing of distributivity

Evidence for Rapid Consideration Without Selection in the processing of scalar implicatures

Literature snapshots about pronouns, modifier attachment, aspect, quantifier scope, and temporal order

SPR comparisons for this Maze data

Garden Paths vs. effects of unmet expectations

More plots

The possibility of multiple explanations

More sample items

Within-trial correlation data

Parallels for syntactic parsing

V P BECAUSE

Amanda Rysling

Pranav Anand

Adrian Brasoveanu

RAs: Kasey La Sebastian Bissiri