Which sentences do speakers favor?

ROC analysis of d-linking in filler-gap integration

Matt Wagers
Department of Linguistics - **UC Santa Cruz**
https://people.ucsc.edu/mwagers

Brian Dillon
Department of Linguistics - **UMass Amherst**
https://people.umass.edu/bwdillon

CEMS2020 (August 17-19)

Observation

BARE question words make sentences like (1)

1. **Who** do you believe the parrot could **imitate**?

harder to process & seem less acceptable than sentences like (2),

2. **Which diva** do you believe the parrot could **imitate**? [1,2,3]

which has a **D-LINKED** wh-phrase. (*d ~ discourse*)

Popular, reasonable hypothesis in psycholinguistics

***D*-linking** the **wh-phrase** creates a higher-quality encoding and increases likelihood of successful retrieval & integration at **verb**.

To test that idea, we adapted a **sentence acceptability judgment task** to measuring ROCs.

But our results lend it no support.

Anatomy of a question

To form a question, a **movement rule** [4] applies and **breaks** the **adjacency** between predicate and argument: (3) ➜ (4)

3. You believe the parrot could **imitate** them.

4. **Who** could the parrot **imitate** _ ?

↑ **filler** [moved argument]

↑ **gap** [site of interpretation]

An *unbounded* stretch of material can interrupt the **filler** and **gap**. E.g., (5).

5. **Who** does it seem least likely that the parrot could **imitate** _ ?

Many different expressions in language use this configuration: relative clauses, comparatives, topicalizations, etc. [4].

Anatomy of a question

To form a question, a **movement rule** [4] applies and **breaks** the **adjacency** between predicate and argument: (3) ➜ (4)

3. You believe the parrot could **imitate** them.

4. **Who** could the parrot **imitate** _ ?
   ↑ **filler** [moved argument] ↑ **gap** [site of interpretation]

An *unbounded* stretch of material can interrupt the **filler** and **gap**. E.g., (5).

5. **Who** does it seem least likely that the parrot could **imitate** _ ?

Many different expressions in language use this configuration: relative clauses, comparatives, topicalizations, etc. [4].

Comprehension challenge

- **Maintain** (enough of) the linguistic **context** to correctly parse the predicate.
- **Retrieve** the **filler** as required to interpret the predicate with the **gap**.

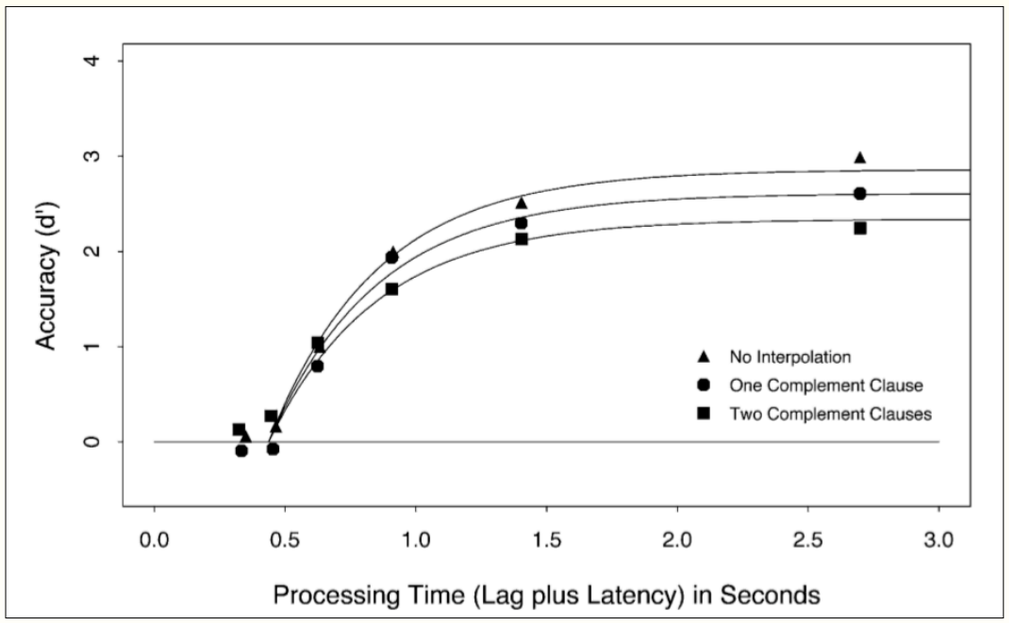
Comprehenders are usually accurate [5].

Evidence for retrieval (some of it)

McElree, Foraker & Dyer (2003) tested comprehenders' ability to discriminate plausible/implausible filler-gap dependencies in single-response SAT:

It was **the scandal** that ⌜...⌟ the celebrity **relished** / **panicked** _ .

⌜ ⌟ NO INTERPOLATION
⌜ the model believed that ⌟ ONE CLAUSE
⌜ the model believed that the journalist reported that ⌟ TWO CLAUSE



McElree, Foraker & Dyer (2003), Fig. 4

※ **Sensitivity** to plausibility falls off with greater distance

※ **Availability** of **filler** in memory contributes to success of filler-gap integration.

Our study

※ Measure sensitivity to wellformedness of filler-gap dependency to see if *d-linking* status impacts the filler's availability, as length/distance does.

Our design + experimental logic

Logic: If *d*-linking improves the ability to retrieve a filler, then *d*-linking should improve sentence acceptability when retrieval is necessary:

6. **Who** / **Which diva** do you believe that the parrot could imitate **imitate** _ ?

To construct a sensitivity measure, we scaled against ungrammatical baselines where no retrieval is necessary to detect ungrammaticality.

7. **Who** / **Which diva** _ believes that the parrot could **imitate** ? ← ⚠ obligatory intransitive ⚠ cues ungrammaticality

Key prediction of **popular, reasonable hypothesis**: Greater sensitivity to contrast in d-linked (**Which diva**) over bare wh (**Who**) conditions.

Empirical ROCs

We constructed **empirical, ratings-based ROCs** three ways:

E1: Binary Y/N + Secondary 3pt confidence ratings

E2: 6pt Likert confidence rating scale

E3: 6pt numerical Likert scale

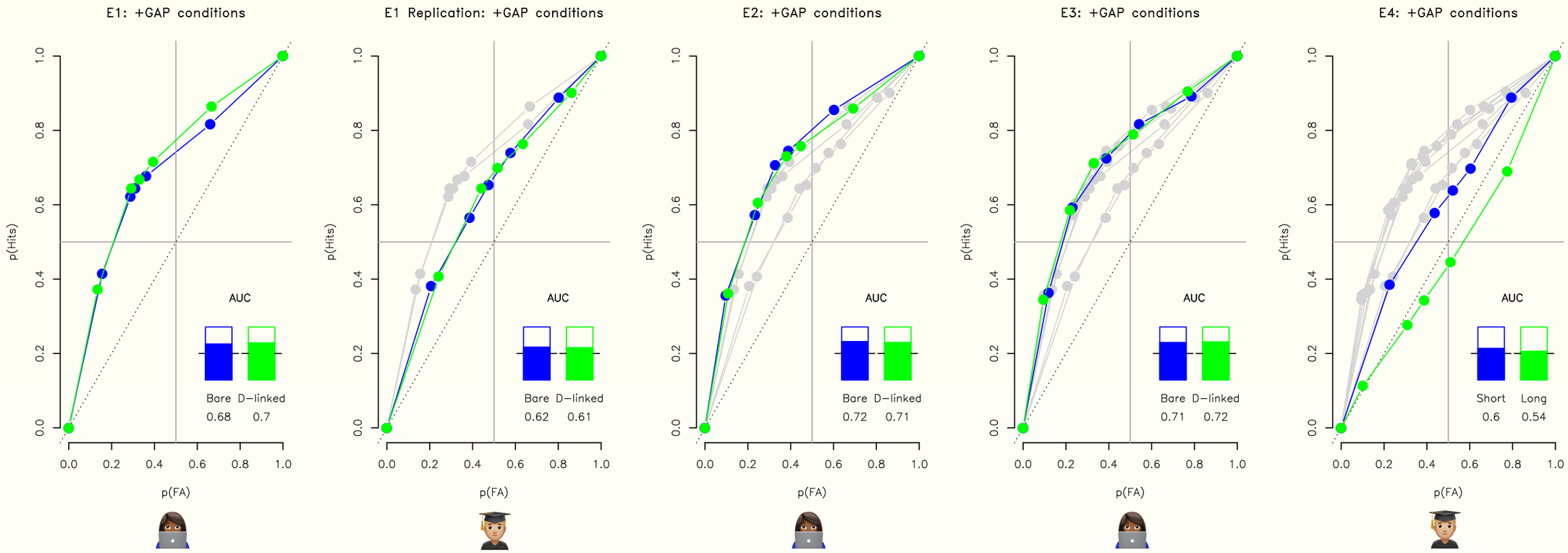
Empirical ROCs

E1: Binary Y/N + Secondary 3pt confidence ratings (N = 64 on Prolific 👩🏾‍💻; N = 86 in-lab at UCSC 👩🏽‍🎓)

E2: 6pt Likert confidence rating scale (N = 45 👩🏾‍💻)

E3: 6pt numerical Likert scale (N = 45 👩🏾‍💻)

E4: Conceptual replication of McElree et al (2003) with E1 technique (N = 47 👩🏽‍🎓)



Results

- Similar sensitivity to acceptability contrast across methods (E1-E3)
  - Lower sensitivity with in-lab participants than Prolific participants
- **No evidence that *d*-linking improves acceptability**
- Some evidence that length impacts acceptability in this paradigm (E4)

What of the popular, reasonable hypothesis?

Idea: **D-linking** the **wh-phrase** creates a higher-quality encoding and increases likelihood of successful retrieval & integration at **verb**.

We used **ratings-based ROCs** to test the idea, **but our results lend it no support.**

Broader result: sensitivity to sentence acceptability can be measured with empirical ROCs. Sensitivity isoclines were stable across ratings methods within a population.

Open for future investigation

- Does this conclusion generalize across different linguistic contexts?
  Islands for movement [1,4] might benefit from the *d*-linking boost more than syntactic structures tested here.

- Sensitivity to the grammaticality contrast in E4 (McElree et al (2003) conceptual replication) was near chance. Why?
  - Potentially, an indication of unintended difficulty with the experimental stimuli, or an effect of data collection during the COVID-19 pandemic. More work needed.

References

(1) Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology*, *5*, 1493

(2) Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, *5*, 1237.

(3) Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376-405.

(4) Phillips, C. (2003). Syntax. In Nadel, L. (Ed.): *Encyclopedia of Cognitive Science, (4),* pp. 319-329.

(5) Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, *67*(7), 1274-1304.

(6) McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67-91.