**Multiverse analysis of eye-tracking data: Reexamining the ambiguity advantage effect**

Caren Rotello & Brian Dillon (UMass Amherst), Caroline Andrews (University of Zürich)

Statistical analysis of eye-tracking-while-reading data involves many decisions. For example, researchers may analyze different dependent measures (e.g. regression path duration, or total reading time). These choices create multiple-comparisons issues in eye-tracking research, leading to unacceptably high Type I error rates [1]. However, similar multiple comparisons issues implicitly arise whenever a researcher faces choice points in constructing her dataset, such as when semi-arbitrary subject exclusion criteria are set [2]. Counterintuitively, the existence of these alternative datasets (Gelman's *garden of forking paths*) can create a multiple comparisons problem <u>even if only a single dataset is analyzed</u>, <u>and only a single statistical test is ever performed</u> [2]. This reality cannot be remedied using familiar corrections. A new strategy to manage this is *multiverse analysis,* which involves enumerating all plausible alternative datasets that could be used in statistical analysis (i.e., all reasonable choices a researcher might make), analyzing all possible datasets at once, and evaluating how robust the results are to different choice points in the analysis [3].

We apply the multiverse approach to eye-tracking-while-reading data. Eye-tracking data are a good candidate for multiverse analysis because there is often uncertainty about where (target or spillover) or in what measure post-lexical effects will be seen [4]. We investigate the *ambiguity advantage effect*, the finding that some globally ambiguous sentences are read more quickly than unambiguous counterparts [e.g., 5]. Participants ($N_{subj}$ = 84) read sentences like (1) in either ambiguous (**AMBIG**) or unambiguous (**HIGH ATTACH, LOW ATTACH**) variants ($N_{item}$ = 27, verb number counterbalanced across items). We identified 7 different choice points, such as the measure and region of interest (ROI) to use; see (2) for a summary. ROIs were generated by considering 1-4 word spans centered on the critical disambiguating word <u>*was*</u> and two spillover words. Taking all possible combinations of the decision points yielded 2,880 possible datasets. For each dataset, we fit a linear mixed-effects regression model to the data to estimate the effect of *AMBIGUITY* on RT; Random-effects structure for each model was determined using the parsimonious approach of [6]. We obtained the *p*-value for *AMBIGUITY* using the Satterthwaite approximation [7].

Figure 1 plots the distribution of *p*-values for the effect of *AMBIGUITY* across datasets. Overall, two analysis choice points substantially shift the distribution of *p*-values across datasets: the eye-tracking measure used, and the ROI. Figure 1 suggests limited evidence for the ambiguity advantage effect in first fixation measures, with a somewhat uniform distribution of *p*-values across datasets. In contrast the distribution of *p*-values is more concentrated below 0.05 in first pass, go-past, and especially total time measures. These trends interact with ROI: 'spillover only' ROIs that did not include the critical disambiguating word (<u>*was*</u>) were overall less likely to have *p*-values less than 0.05, but multiword ROIs including the disambiguating word revealed the opposite tendency. The other choice points considered largely did not systematically shift the distribution of *p*-values across datasets. Overall, our analysis yields evidence for the ambiguity advantage effect, but it does not appear in all combinations of ROI and eye-tracking measure. We suggest that multiverse analyses may profitably serve as guides for strong pre-registered studies on eye-tracking while reading.

**(1)** Edwin has been reading about...

**AMBIG:** the sister of the actor who **was** visiting the resort…

**HIGH ATTACH:** the sister of the actors who **was** visiting the resort…

**LOW ATTACH:** the sisters of the actor who **was** visiting the resort…

**(2)** Analysis choice points:

**Eye-tracking measure:** {first fixation, first pass, go-past time, total times}

**Duration scale:** {raw RT (ms), log-transformed RT (log ms)}

**ROI:** {*was, was visiting, was visiting the, who was*, *who was visiting, who was visiting the, the, visiting, visiting the*}

**Exclude subjects by accuracy:** {No cut off, >50% accuracy, > 60%, > 70%, > 80%}

**Exclude subjects with excessive track loss:** {Yes, No}

**Exclude trials with first pass regression:** {Yes, No}

**Exclusion of fixations < 80ms or > 1000ms:** {Yes, No}

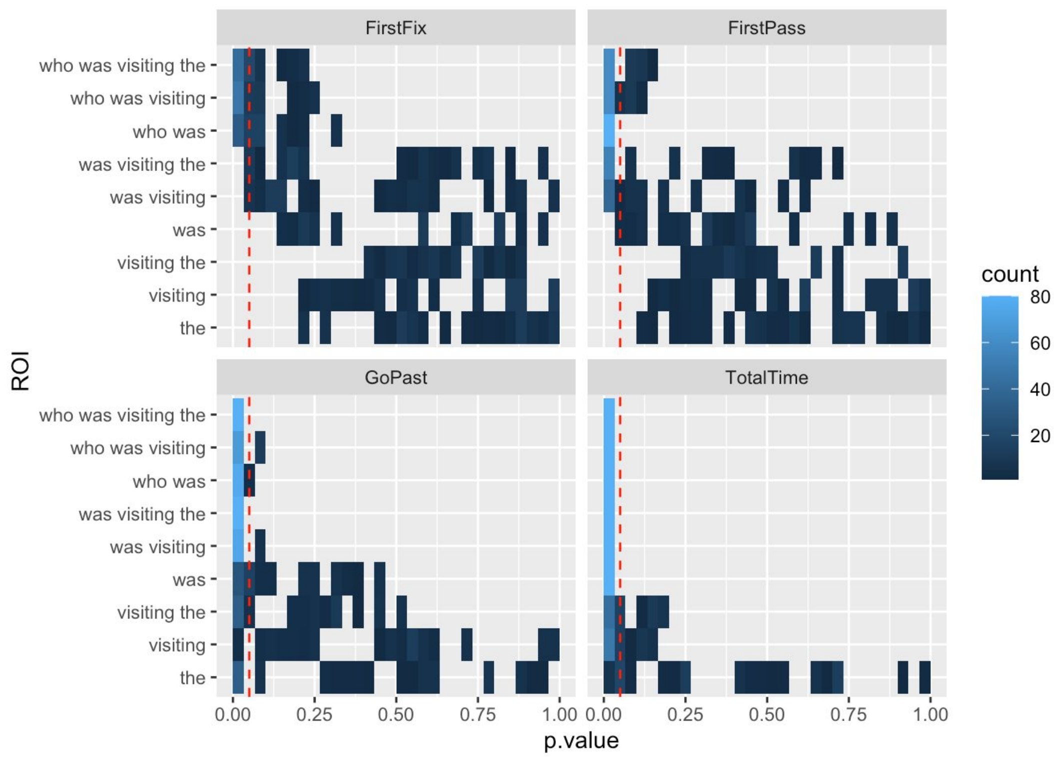


**Figure 1:** Distribution of *p*-values for effect of *AMBIGUITY* across datasets. Panels represent different eye-tracking measures. Different ROI are represented in the rows; each row has 80 total datasets.


[1] von der Malsburg & Angele (2017). *JML.* [2] Gelman & Loken (2014). *American Scientist.* [3]. Steegen et al. (2016). *Persp on Psych Science* [4] Clifton, Staub & Rayner. (2007) *Eye movements and reading.* [5] van Gompel et al. (2005). *JML.* [6] Matuschek et al. (2017). *JML.* [7] Kuznetsova et al. (2017). *J of Stat Software.*