

Reassessing the grammaticality asymmetry in agreement attraction: An ROC analysis

Christopher Hammerly (UMass), Brian Dillon (UMass), and Matt Wagers (UCSC)

Agreement attraction occurs when a distractor noun that mismatches the number of the subject affects the computation of verb agreement. For example, in the PP-modifier construction *The critic of the movie(s) is/*are crazy*, the presence of a mismatching plural distractor (*movies*) in ungrammatical sentences increases the proportion of grammatical responses in judgments, relative to the ungrammatical distractor match baseline (e.g. [7])

One theoretically central, but empirically contested, question is whether mismatch effects arise in grammatical sentences. The **representational** Marking & Morphing account [2] predicts a *symmetric* effect of attraction from both grammatical and ungrammatical strings via spreading activation of number features to the subject root phrase. By contrast, the **operational** cue-based retrieval model predicts an asymmetric effect, as the features of the verb in grammatical strings completely match the subject head, while the verb in ungrammatical strings partially matches with both the subject head and distractor. Recent implemented models strengthen this prediction: under common assumptions about the cues and their combinations (e.g. [7]), the asymmetry is predicted to be total, with no attraction in grammatical strings. While many studies have found an asymmetry consistent with current cue-based models, recent evidence from [3] suggests that the asymmetry is the result of response bias, calling into question the conclusion that the asymmetry supports the cue-based account. [3] argue that the presence of symmetrical attraction in the absence of bias supports a representational account.

The present study used signal detection theory (SDT) to partial out the effects of bias in judgments to test whether there is a grammaticality asymmetry. We examined two attractor positions: embedded in PP, and the head of an object relative clause (ORC). [6] argued that attraction only arises due to retrieval with ORCs, but has a representational source with PPs. If this is the case, **we expect an asymmetry to arise with ORC but not PP attractors**. To test this claim directly, we maintained a constant syntactic structure across conditions. We manipulated factors GRAMMATICALITY and ATTRACTOR using the sentence structure illustrated in (1) ($N_{\text{items}} = 120$ experimental + 70 fillers). We used the ratings method design [1,4]: sentences were presented word-by-word (325ms/word), and participants ($N = 84$) made a binary grammaticality judgment followed by a three-point confidence rating.

- (1) a. Alex lost the phonebook that the *lawyer* for the company often *use(s)* **Match**
b. Alex lost the phonebooks that the *lawyer* for the company often *use(s)* **ORC Mismatch**
c. Alex lost the phonebook that the *lawyer* for the companies often *use(s)* **PP Mismatch**

The results show a similar picture in both PPs and ORCs. For reference, the raw data for judgments, judgment RT, and confidence are shown in Table 1. We constructed empirical Receiver Operating Characteristic (ROC) curves with the ratings, shown in Figure 1, to determine the discriminability of the mismatch condition against its match baseline for each level of grammaticality in both the PP and ORC conditions. Using the *pROC* package in *R* [5], we computed the d_A value from the best-fitting unequal-variance SDT model for each of these contrasts, and conducted a bootstrap test for on the difference in mismatch effect for grammatical vs. ungrammatical sentences for each construction. While **there is a significant effect of attractor mismatch in both grammatical and ungrammatical sentences**, the effect is not symmetrical. The comparison of grammatical/ungrammatical sentence ROCs reveals the **mismatch effect is larger in ungrammatical sentences** ($p < 0.001$ for both PPs and ORCs).

The results do not wholly support either existing account. While current implementations of retrieval capture the grammaticality asymmetry, they do not capture the mismatch effect in grammatical sentences. Meanwhile, Marking & Morphing captures the grammatical mismatch effect, but not the asymmetry. The results are consistent with two possible explanations: (i) a revision to cue specification in the retrieval model; (ii) the adoption of a hybrid model where both retrieval-based and representational effects of distractor mismatch can arise.

<i>Condition</i>	%	Grammatical responses		Ungrammatical responses	
	Gram	RT	Confidence	RT	Confidence
Gram. Match	0.75	924 (24)	2.44 (0.03)	1088 (30)	1.95 (0.04)
Gram. PP Mismatch	0.68	979 (22)	2.31 (0.03)	1048 (28)	1.89 (0.05)
Gram. ORC Mismatch	0.66	989 (25)	2.27 (0.04)	1100 (27)	1.88 (0.05)
Ungram. PP Mismatch	0.37	1057 (30)	1.98 (0.05)	975 (22)	2.31 (0.04)
Ungram ORC. Mismatch	0.40	1041 (26)	2.13 (0.05)	1014 (23)	2.27 (0.04)
Ungram. Match	0.18	1091 (39)	1.94 (0.08)	914 (21)	2.46 (0.03)

Table 1: Summary of percent grammatical responses and RT in ms for binary judgment task, as well as mean confidence ratings. Confidence ranged from 1 (not at all confident) to 3 (very confident). By-participant standard error is in parentheses.

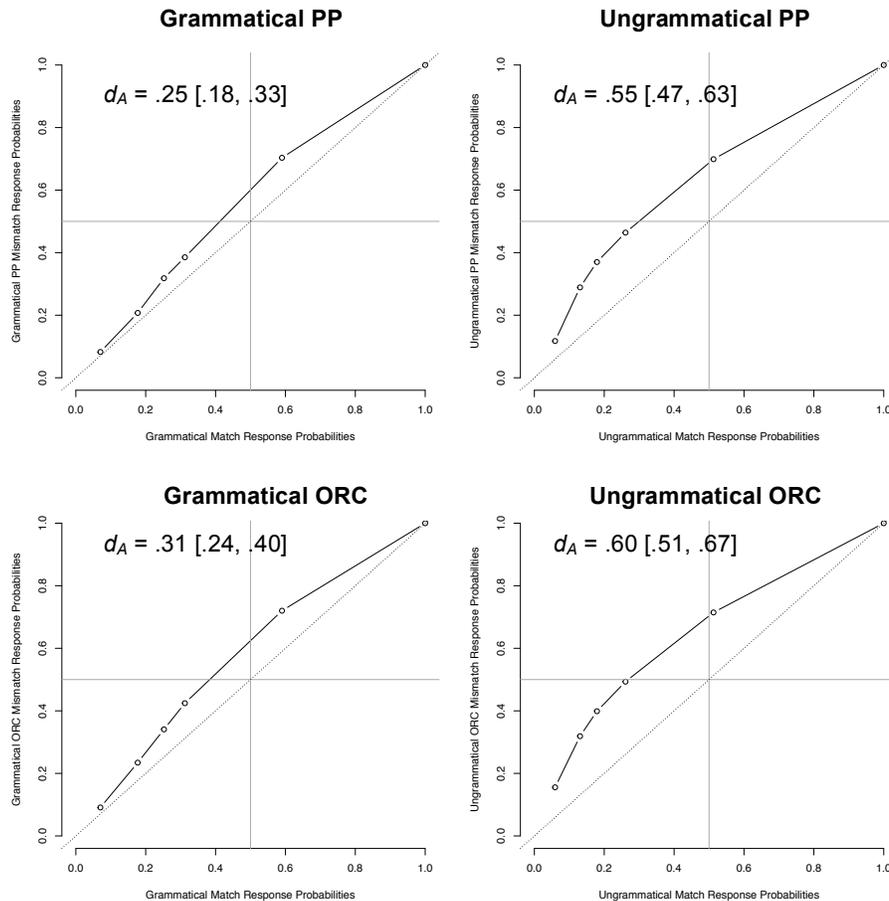


Figure 1: Empirical ROC curves for match vs. mismatch contrast by grammaticality and construction. Greater area under the curve corresponds to higher discrimination between match and mismatch conditions, and is related to the d_A value derived from the best-fitting unequal variance SDT model. Corresponding d_A value and 95% CI are given next to each curve.

- [1] Dillon et al. (2018) A new argument for co-active parses during language comprehension
[2] Eberhard et al. (2005) Making syntax of sense: Number agreement in sentence production.
[3] Hammerly et al. (2018) The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence.
[4] Macmillan & Creelman (1991) Detection Theory: A User's Guide.
[5] Robin et al. (2011) *pROC* package in R.
[6] Staub (2010) Response time distributional evidence for distinct varieties of number attraction.
[7] Wagers et al. (2009) Agreement attraction in comprehension: Representations and processes.