Rational inference does not predict agreement errors: Gender vs number attraction in Hebrew comprehension

Cartner M.[1], Dillon B.[2], Meltzer-Asscher A.[1] & Keshev M.[3]   [1]TAU, [2]UMass Amherst, [3]HUJI

Agreement attraction is the phenomenon whereby ungrammatically inflected verbs may be erroneously licensed when a noun in the sentence matches their features, in production [1] and comprehension [2-3]. Ryskin et al. (2021) [4] suggest that agreement attraction, as in (1), is the result of a rational inference whereby the target DP (*the key to the cabinets*) is interpreted as a noisy form of a higher probability alternative (*the keys to the cabinets*).

(1) *The key to the cabinets were rusty.

According to this account, because a singular-plural DP (as in 1) is arguably less likely than a plural-plural DP, and producers are likely to erroneously omit a plurality marker (*-s*), comprehenders may infer that "the key to the cabinets" is a corruption of "the keys to the cabinets". This rational inference results in acceptance of plural agreement.

Gender features also give rise to agreement attraction, such that a masculine-feminine DP may erroneously license a feminine verb [5]. Does prior probability motivate misrepresentation of a masculine-feminine DP as feminine-feminine?

Ryskin et al. predict that agreement error rates should correlate with the frequency of the relevant alternative agreement target. Thus, if a feminine-feminine DP is not more likely than a masculine-feminine DP, gender should display little attraction effects compared to number. On the other hand, accounts that attribute agreement attraction to retrieval interference [2] or representational interference [6-7] controlled by feature markedness, predict similar rates of attraction for both features. We test these predictions by examining number vs gender attraction in Hebrew comprehension, and comparing them against the relative frequency of the relevant misinterpretation in a corpus.

Corpus data. We extracted 10K [DP_1 of DP_2] tokens from the Hebrew web corpus HeTenTen21 (Table 1). We calculated the conditional probability of the features of DP_1 based on the features of DP_2. Number: As in English, DP_1 is more likely to be plural when DP_2 is also plural (26.8%), compared to when DP_2 is singular (18.1%). Gender: DP_1 is **not** more likely to be feminine when DP_2 is feminine (40.5%), compared to when DP_2 is masculine (42.5%). Thus, the noisy-channel account predicts no attraction for Hebrew gender, as prior probability does not support greater misinterpretation when DP_2 is feminine.

Speeded acceptability. 94 Hebrew speakers read sentences in RSVP and provided speeded, binary acceptability judgments. The task included 36 sets, crossing: *Feature* (*Gender, Number*) and *Agreement match* (*Target*, *None*, *Distractor*) (Table 2). Acceptability accuracy was analyzed in a GLMM using *brm* in R. The model revealed evidence for attraction in both *Number* (Est. 0.46, SE 0.11, CRI 95% [0.24, 0.7]) and *Gender* (Est. 0.31, SE 0.1, 95% CRI [0.11, 0.52]), such that accuracy in the *No match* condition was higher than the *Distractor match* condition. The interaction between this comparison and feature type was not reliable (Est. 0.21, SE 0.17, 95% CRI [-0.13, 0.54]), meaning that the rates of attraction are likely similar across number and gender in Hebrew (Figure 1). A GLMM fit to RTs in accurate trials (Figure 2) revealed slower RTs in the *Distractor* compared to the *No match* condition (Est. -0.07, SE 0.02, CRI 95% [-0.11, -0.03]), but no reliable interaction with *Feature* (Est. -0.02, SE 0.02, CRI 95% [-0.06, 0.01]).

Discussion. Our Hebrew corpus analysis provides some support for the claim that [DP_{SG} of DP_{PL}] is likely to be misinterpreted as [DP_{PL} of DP_{PL}]. However, we do not find evidence suggesting that prior knowledge will facilitate misrepresentation of [DP_{MASC} of DP_{FEM}] as [DP_{FEM} of DP_{FEM}]. If agreement attraction were the result of a rational integration of prior knowledge and awareness of potential noise, a grammaticality illusion should be more likely in number compared to gender attraction. However, we found similar rates of number and gender attraction in Hebrew comprehension. This is compatible with previous findings from Hebrew production [5], Modern Standard Arabic reading [8], and Spanish production [9].

The similar rates of attraction across feature types provide evidence that abstract feature markedness, rather than form frequency, is an important determining factor in agreement errors. It also does not lend support to the claim that number features are more cognitively salient, or have a privileged representation compared to gender [10].

Table 1. Number of hits (out of 10K) per each combination of features in [DP$_1$ of DP$_2$] tokens from HeTenTen21 (log 10).

| DP1 (target)	DP2 (distractor)			
	M.SG	F.SG	M.PL	F.PL
M.SG	2263 (3.35)	1679 (3.23)	503 (2.7)	297 (2.47)
F.SG	1389 (3.14)	946 (2.98)	390 (2.59)	191 (2.28)
M.PL	582 (2.76)	507 (2.71)	214 (2.33)	131 (2.12)
F.PL	192 (2.28)	119 (2.08)	56 (1.75)	35 (1.54)

Table 2. Example set from the speeded acceptability judgement task.

Feature: Gender

Target match	zaxarti še-ha-xayetet šel ha-me'acev tamid me'alteret likrat ha-tecuga
	I.remember that-the-**tailor.F** of the-**designer.M** always **improvises.F** before the show
No match	* zaxarti še-ha-xayat šel ha-me'acev tamid me'alteret likrat ha-tecuga
	I.remember that-the-**tailor.M** of the-**designer.M** always **improvises.F** before the show
Distractor match	* zaxarti še-ha-xayat šel ha-me'acevet tamid me'alteret likrat ha-tecuga
	I.remember that-the-**tailor.M** of the-**designer.F** always **improvises.F** before the show

Feature: Number

Target match	zaxarti še-ha-xayatim šel ha-me'acev tamid me'alterim likrat ha-tecuga
	I.remember that-the-**tailor.PL** of the-**designer.SG** always **improvise.PL** before the show
No match	* zaxarti še-ha-xayat šel ha-me'acev tamid me'alterim likrat ha-tecuga
	I.remember that-the-**tailor.SG** of the-**designer.SG** always **improvise.PL** before the show
Distractor match	* zaxarti še-ha-xayat šel ha-me'acvim tamid me'alterim likrat ha-tecuga
	I.remember that-the-**tailor.SG** of the-**designer.PL** always **improvise.PL** before the show

Figure 1. Mean acceptability accuracy.
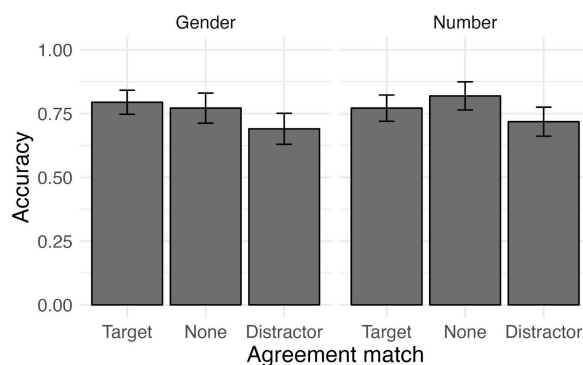
Figure 2. Mean RTs in accurate trials.
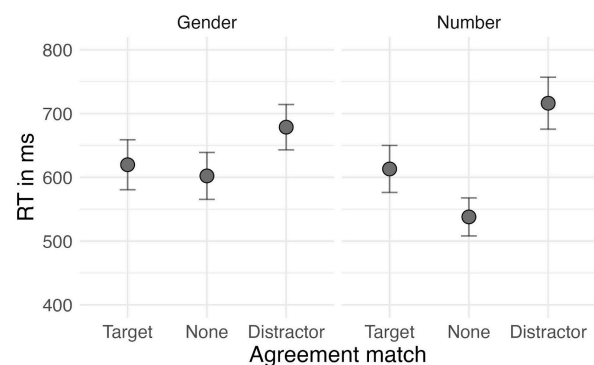
References: **[1]** Bock & Miller (1991) *Cogn. Psychol.* **[2]** Wagers, Lau & Phillips (2009) *JML*; **[3]** Pearlmutter, Garnsey, & Bock (1999) *JML*.; **[4]** Ryskin, Bergen & Gibson (2021), *Preprint*; **[5]** Deutsch & Dank (2009) *JML*; **[6]** Bock, Eberhard, Cutting, Meyer & Schriefers. (2001) *Cogn. Psychol*.; **[7]** Keshev, Cartner, Meltzer-Asscher & Dillon (2024), *CogSci*; **[8]** Tucker, Idrissi & Almeida (2021) *Front. Psychol*.; **[9]** Antón-Méndez, Nicol, & Garrett (2002), *Syntax*; **[10]** Carminati (2005), *Lingua*.