German pronoun interpretation follows Bayesian principles

Clare Patterson (University of Cologne), Petra B. Schumacher (University of Cologne), Bruno Nicenboim (Tilburg University), Johannes Hagen (University of Cologne), Andrew Kehler (University of California, San Diego)

The Bayesian Model for pronouns (Kehler et al 2008 et seq.) predicts that pronoun production and comprehension are related by Bayes' rule: P(*referent* | *pronoun*) ∝ P(*pronoun* | *referent*)P(*referent*). P(*referent* | *pronoun)* represents the comprehension bias: the probability that a particular referent is being referred to by a pronoun. The likelihood term *P(pronoun* | *referent)* represents the production bias: the hearer's estimate of the probability that speaker will use a pronoun to refer to a particular referent. The prior term, *P(referent)*, represents the next-mention bias: the probability that a particular referent will get mentioned next, regardless of the referring expression used. Values for the prior and likelihood terms are estimated from passage completion experiments with free prompt conditions, yielding a predicted comprehension bias that can be compared to the actual comprehension bias measured using pronoun-prompt conditions with the same contexts. The Bayesian Model has been quantitatively examined in English (Rohde & Kehler 2014) and Mandarin Chinese (Zhan et al 2020) by comparing its predictions against the predictions from two competing models: a Mirror Model, a normalized P(*pronoun* | *referent*), and the Expectancy Model, a normalized P(*referent*). In this study, we further the cross-linguistic support for the Bayesian Model by applying it to German personal and demonstrative pronouns, and provide novel quantitative support for the model by assessing model performance in a Bayesian statistical framework that allows implementation of a fully hierarchical structure, providing the most conservative estimates of uncertainty. Applying the Bayesian model to German provides new cross-linguistic evidence because both personal and demonstrative pronouns can refer to human entities. Additionally, the referential biases for the demonstrative *dieser* are not well understood, but demonstratives are thought to be more rigid in their interpretation than the personal pronoun (Kaiser 2011, inter alia), making them a good test for the Bayesian Model.

Two passage completion studies were conducted with items consisting of a context sentence followed by one of three prompt types: personal pronoun (*er*), demonstrative pronoun (*dieser*), and free prompt (a blank line). To explore the effects of syntactic and semantic context factors, Experiment 1 (N=48) compared contexts with active-accusative verbs (1) and dative-experiencer verbs (2) and Experiment 2 (N=40) compared contexts with experiencer–stimulus verbs (3) and stimulus–experiencer verbs (4). Each model (Expectancy, Mirror, and Bayes) was fit with Bernoulli likelihoods for the referent and categorical likelihoods for the expression type, with weakly regularizing priors. Observation-level predictions for each model were made based on the free-prompt data and fitted against the held out observations from the pronoun-prompt data. Model fit was evaluated graphically with holdout predictive check, and numerically using holdout validation (Vehtari & Ojanen 2012).

Overall, the Bayesian Model makes more accurate predictions than both the Expectancy and Mirror Models in both experiments (see table and figures, which compare the predictive accuracy of the models with respect to pronoun interpretation). Furthermore, the model accounts for the demonstrative pronoun *dieser* as well as the personal pronoun, despite its more rigid resolution preferences. We further confirmed that semantic factors (implemented as a verb-type contrast) affect the prior term P(*referent*) to a much greater extent than the likelihood term P(*pronoun*|*referent*), underlining the separation of pronoun-related biases from form-independent expectations about the upcoming referent (Kehler & Rohde 2013).

As an ensemble, the results for German pronouns strongly support the predictions of the Bayesian Model, according to which comprehenders reverse engineer the speaker's referential intentions using Bayesian principles.

(1) Vorletzte Nacht hat der Hund den Papagei geärgert. Er/Dieser/__
The night before last the dog (nom.masc.) annoyed the parrot (acc.masc.). He/DEM/__
(2) Gestern ist dem Feuerwehrmann der Polizist aufgefallen. Er/Dieser/__
Yesterday the firefighter (dat.masc.) noticed the police officer (nom.masc.). He/DEM/__
(3) Der Dieb fürchtete den Polizisten. Er/Dieser/__
The thief (nom.masc.) feared the police officer (acc.masc.). He/DEM/__
(4) Der Fußballer erstaunte den Manager. Er/Dieser/__
The footballer (nom.masc.) astonished the manager (acc.masc.). He/DEM/__

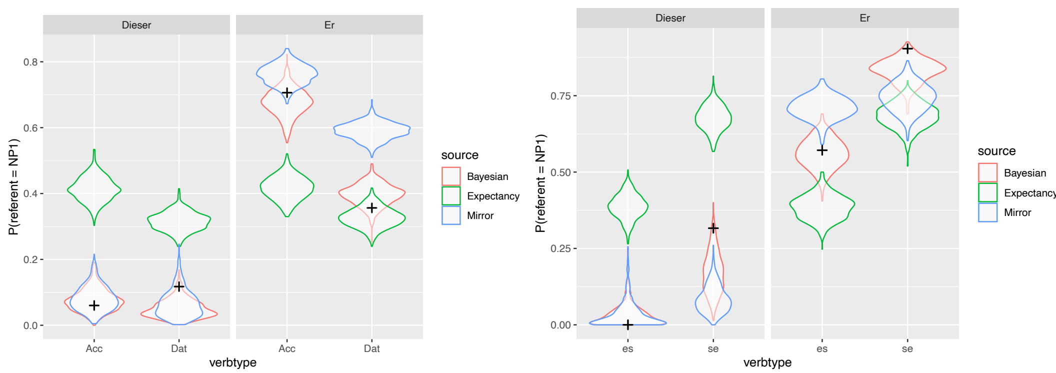

Figure 1. Crosses show observed proportion of NP1 interpretations for Experiment 1 (left plot) and Experiment 2 (right plot) (from held out data); violin plots depict distribution of simulated proportions based on model predictions.

	Experiment 1					**Experiment 2**				
	elpd	SE elpd	elpd_diff	SE diff	weight	elpd	SE elpd	elpd_diff	SE diff	weight
B	-728	27	0	0	0.89	-368	19	0	0	0.9
M	-860	27	-132	14	0.00	-467	24	-98	13	0.0
E	-966	16	-238	24	0.11	-578	16	-209	23	0.1

Table 1. B = Bayesian Model, M = Mirror Model, E = Expectancy Model. A higher expected log-predictive density (elpd) indicates better predictive accuracy. The highest scoring model is the baseline for elpd difference (elpd_diff) and difference Standard Error (SE). Weight columns represent weights of the individual models that maximize the total elpd score of all the models.

Kaiser E (2011). On the Relation between Coherence Relations and Anaphoric Demonstratives in German. In Reich I et al (Eds.), *Proceedings of Sinn & Bedeutung 15* (pp. 337-351). Saarland University Press.
Kehler A, Kertz L, Rohde H, & Elman J (2008). Coherence and Coreference Revisited, *Journal of Semantics* 25:1, 1-44.
Kehler A & Rohde H (2013). A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation, *Theoretical Linguistics* 39, 1-37.
Rohde H & Kehler A (2014). Grammatical and Information-structural Influences on Pronoun Production. *Language, Cognition and Neuroscience* 29, 912-927.
Vehtari A, & Ojanen J (2012). A Survey of Bayesian Predictive Methods for Model Assessment, Selection and Comparison. *Statistics Surveys* 6, 142–228.
Zhan M, Levy R & Kehler A (2020). Pronoun Interpretation in Mandarin Chinese Follows Principles of Bayesian Inference. *PLoS ONE* 15:8, e0237012.