Consolidating linguistic tendencies at the level of individuals and populations for Zipf's Law of Abbreviation

Thomas Brochhagen

LINGUAE, 29/04/2021

Background

- Zipf's Law of Abbreviation: frequent forms are shorter
- Inverse relationship between *use* and *linguistic reduction*
- Kanwal et al. (2017): study what drives this relationship by controlling for communicative pressures

Background

Zipf's Law of Abbreviation: frequent forms are shorter
at the level of languages

Zipf (1935): The Psycho-Biology of Language

Sigurd et al. (2004): Word length, sentence length and frequency – Zipf revisited

Piantadosi et al. (2011): Word lengths are optimized for efficient communication

Ferrer-i-Cancho et al. (2013): Compression as a Universal Principle of Animal Behavior

…

Background

Inverse relationship between *use* and *linguistic reduction*
across strata

Clark & Wilkes-Gibbs (1986): Referring as a collaborative process

Kim et al. (2011): Phonetic convergence in spontaneous conversations as a function of interlocutor language distance

Pickering & Ferreira (2008): Structural priming: A critical review

Hawkins et al (2017): Convention-formation in iterated reference games

...

Background

Kanwal et al. (2017) study what drives the relationship between frequency and form by controlling for communicative pressures **at the level of dyads**

Kanwal et al. (2017): Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*

The study I'll build on today

Modest re-analysis of Kanwal et al. (2017) to

- explicitly model subjects' behavior
- gain insight into how population-level, dyadic-level and individual-level relate to each other; and how to possibly begin to link them

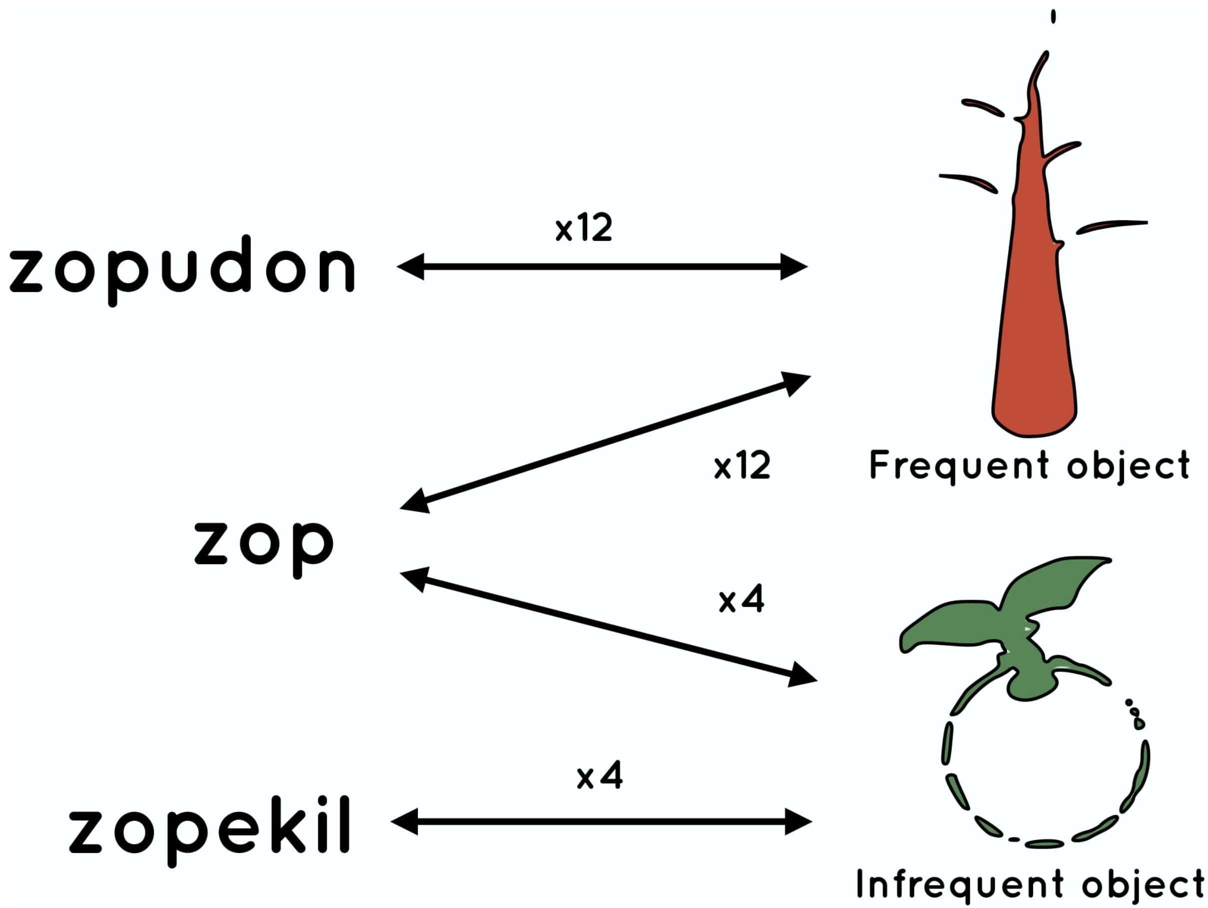
---

Brochhagen (2021): Brief at the risk of being misunderstood: Consolidating population- and individual-level tendencies. *Computational Brain & Behavior*

Experimental setup

Kanwal et al. (2017) condition 4

Training (miniature lexicon acquisition)



Testing (condition 4)

- Pair subjects and let them interact for 62 trials

  32 sender trials, 32 receiver trials (alternating)

- Same object frequencies as in training

- Message transmission time is proportional to message length
  $\rightarrow$ Sending ambiguous *zop* takes $\frac{3}{7}$ of the time of alternatives

- The fastest & most accurate pairs win a prize
  $\rightarrow$ Pressure to use short form if you believe your interlocutor will get it

Kanwal et al's analysis

- Logistic regression with short name as binary response
- Object frequency (frequent/infrequent), trial number, and their interaction as fixed effects
- By-participant intercepts and slopes as random effects for object frequency and trial number

Finding: Positive effect of trial number on short form use only for frequent object

Some open questions

1. How did individual speakers behave?
    - Change over time
    - On what strategies did they converge
2. How did their behavior compare to that of their partners?
3. How do population-level predictions compare to (1) and (2)?
4. How can we relate trends from individuals and dyads to a community of speakers?

Some model desiderata

- Accommodate for multiple association patterns (cf. Parikh 2000)
- Accommodate for signaling behavior that can change as a function of interlocutor's behavior

Parikh (2000): Communication, Meaning, and Interpretation

Brochhagen (2017): Signalling under Uncertainty: Interpretative Alignment without a Common Prior

Hawkins et al (2017): Convention-formation in iterated reference games

Speaker model

We return to model identifiability between Brochhagen (2017) and Hawkins et al. (2017) at the end

Intuitions

- RSA-style choice functions (*soft-max*; *Luce's rule*; *rationality parameter*)

  "If I believe you expect one object over the other (by a large enough margin) I will use the ambiguous but preferred form to convey that object. Otherwise, I'll play it safe."

- Players can change beliefs over time, based on past success/failure using an ambiguous form

  "You interpreted my use of an ambiguous form in a certain way before. Consequently, I believe you expected this object (and therefore I will be more likely to signal this object with this expression later)"

Parameters to estimate from the data

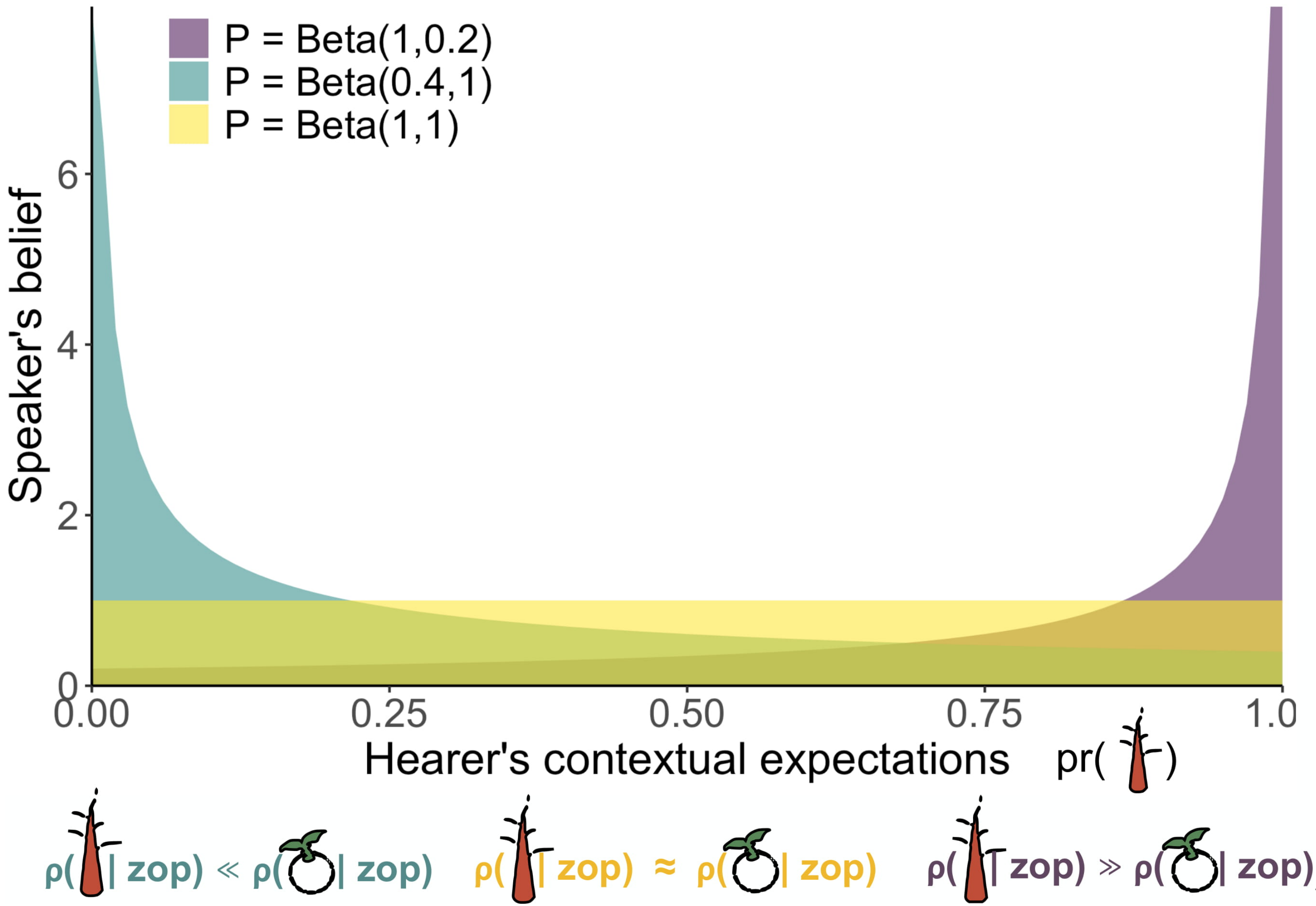
λ usual soft-maximization parameter, $\lambda \geq 0$

$pr \sim P$

less usual beliefs over (non-common) prior

$P = \text{Beta}(\alpha, \beta)$

Latent beliefs over expectations



The full speaker model

$$\rho(r \mid m; pr) \propto L(r, m)\ pr(r),$$
$$\sigma(m \mid r; P) \propto \exp(\lambda((\int P(\theta)\rho(r \mid m; \theta) d\theta) - c(m)))$$

and

$$P_{t+1}(pr \mid w(r); m) \propto (\sum_{r' \in w(r)} \rho(r' \mid m; pr) P_t(pr)),$$

with $w(r) = \{r\}$ if the interaction was successful and $R \setminus \{r\}$ otherwise.

Parameters to estimate from the data

λ	soft-maximization parameter, $\lambda \geq 0$
$P = \mathrm{Beta}(\alpha, \beta)$	beliefs over (non-common) prior

Models

Overview

Model	Population parameters	Individual parameters
NoPool	--	$\lambda_i, \alpha_i, \beta_i$
FullPool λ, α, β	α, λ, β	--
FullPool λ	λ	α_i, β_i
HM λ	μ_λ	$\lambda_i \sim N(\mu_\lambda, 2)$ α_i, β_i
HM λ, α, β	$\mu_\lambda, \mu_\alpha, \mu_\beta$	$\lambda_i \sim N(\mu_\lambda, 2)$, $\alpha_i \sim N(\mu_\alpha, 2)$ $\beta_i \sim N(\mu_\beta, 2)$

Results

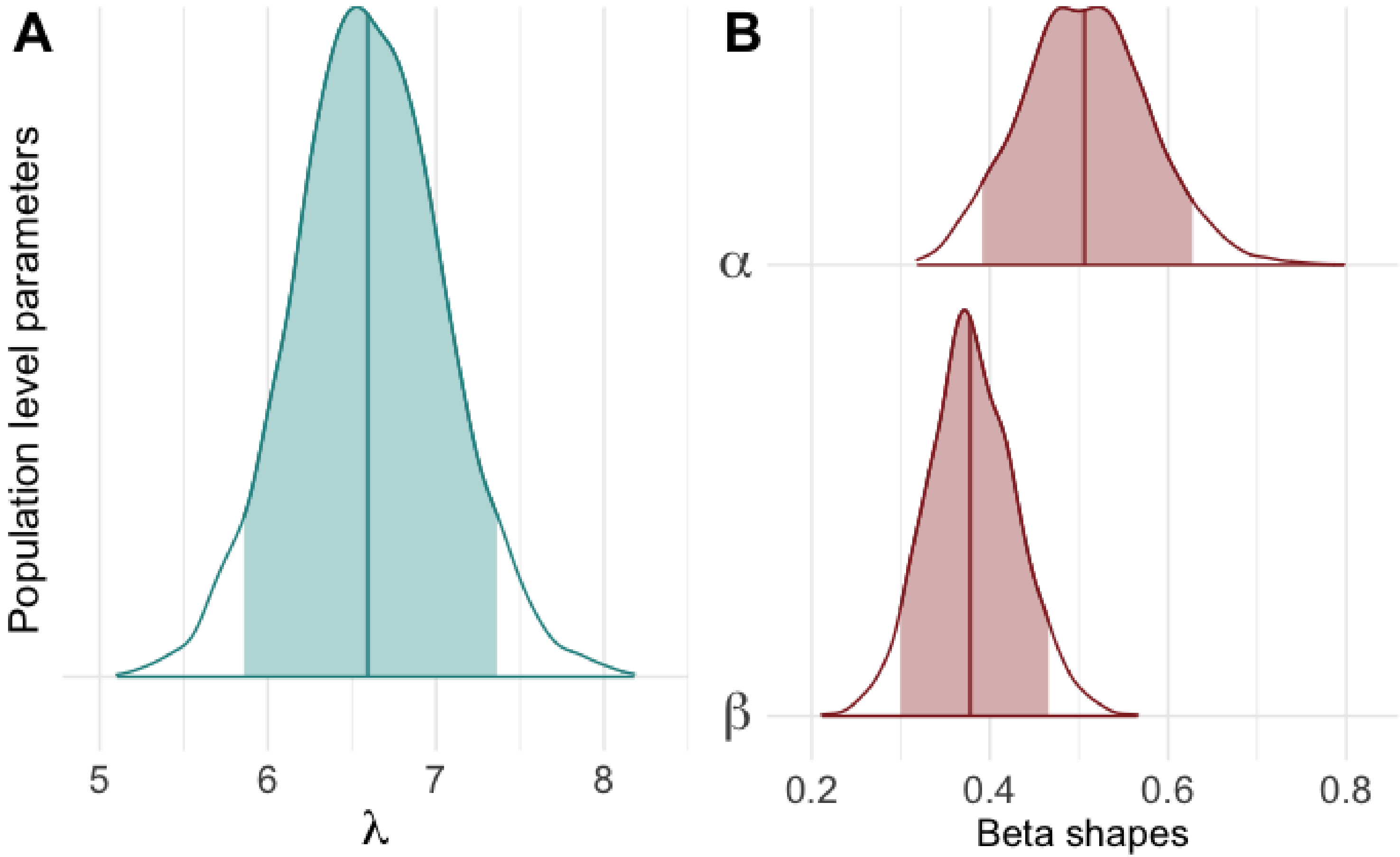
All models were diagnosed to rule out pathologies, and cross-validated.

Model comparison (full dataset)

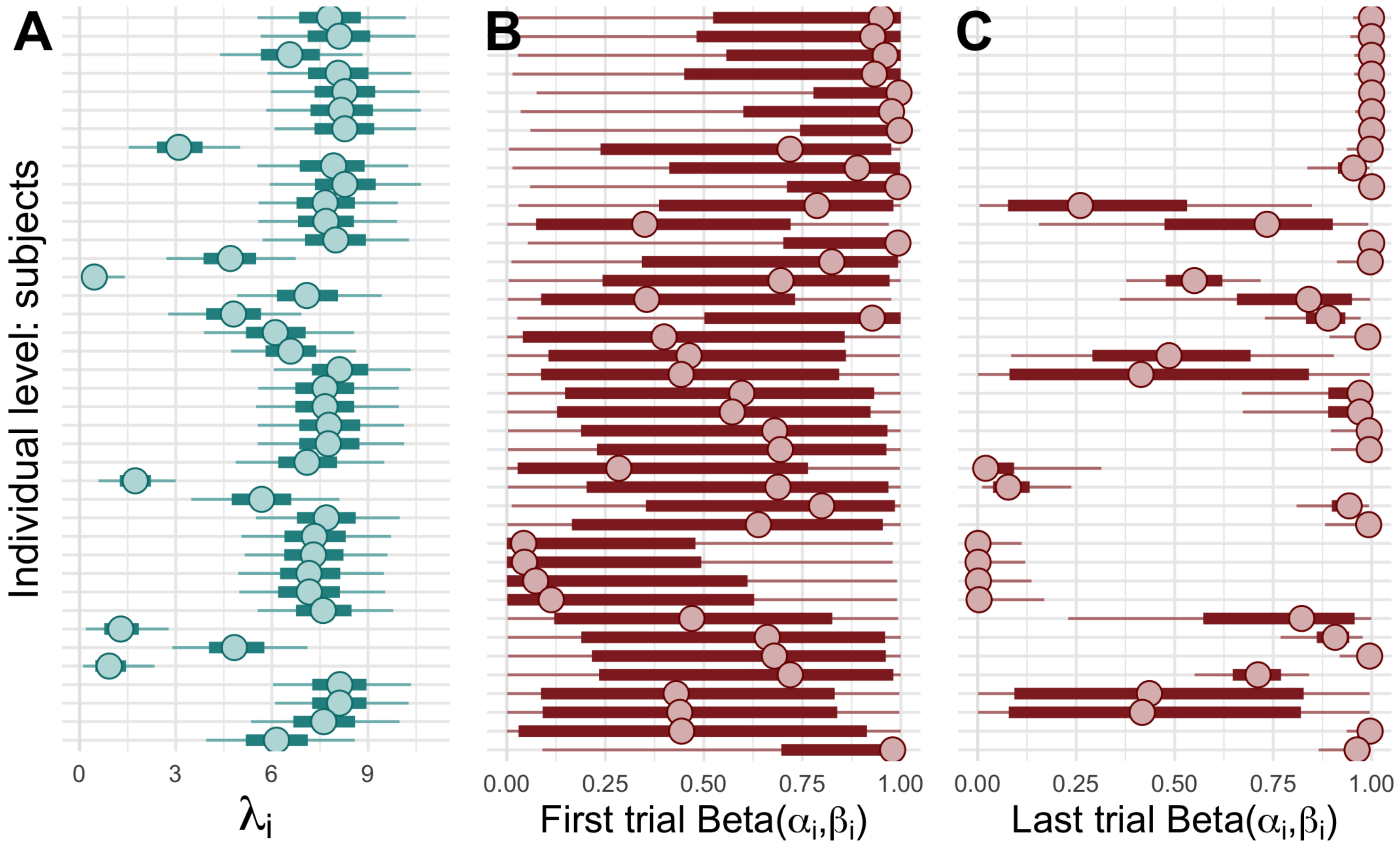
	ELPD diff (SE)	ELPD (SE)
HM λ, α, β	0.00 (0.00)	-471.63 (18.74)
HM λ	-38.16 (5.74)	-509.79 (18.24)
FullPool λ, α, β	-148.24 (12.17)	-619.88 (22.23)
FullPool λ	-154.21 (11.61)	-625.84 (21.24)
NoPool	-201.02 (34.55)	-672.65 (51.02)

Finding The best model, $HM_{\lambda,\alpha,\beta}$, is ranked first, also across other data subsets and when splitting dyads up

Population-level estimates



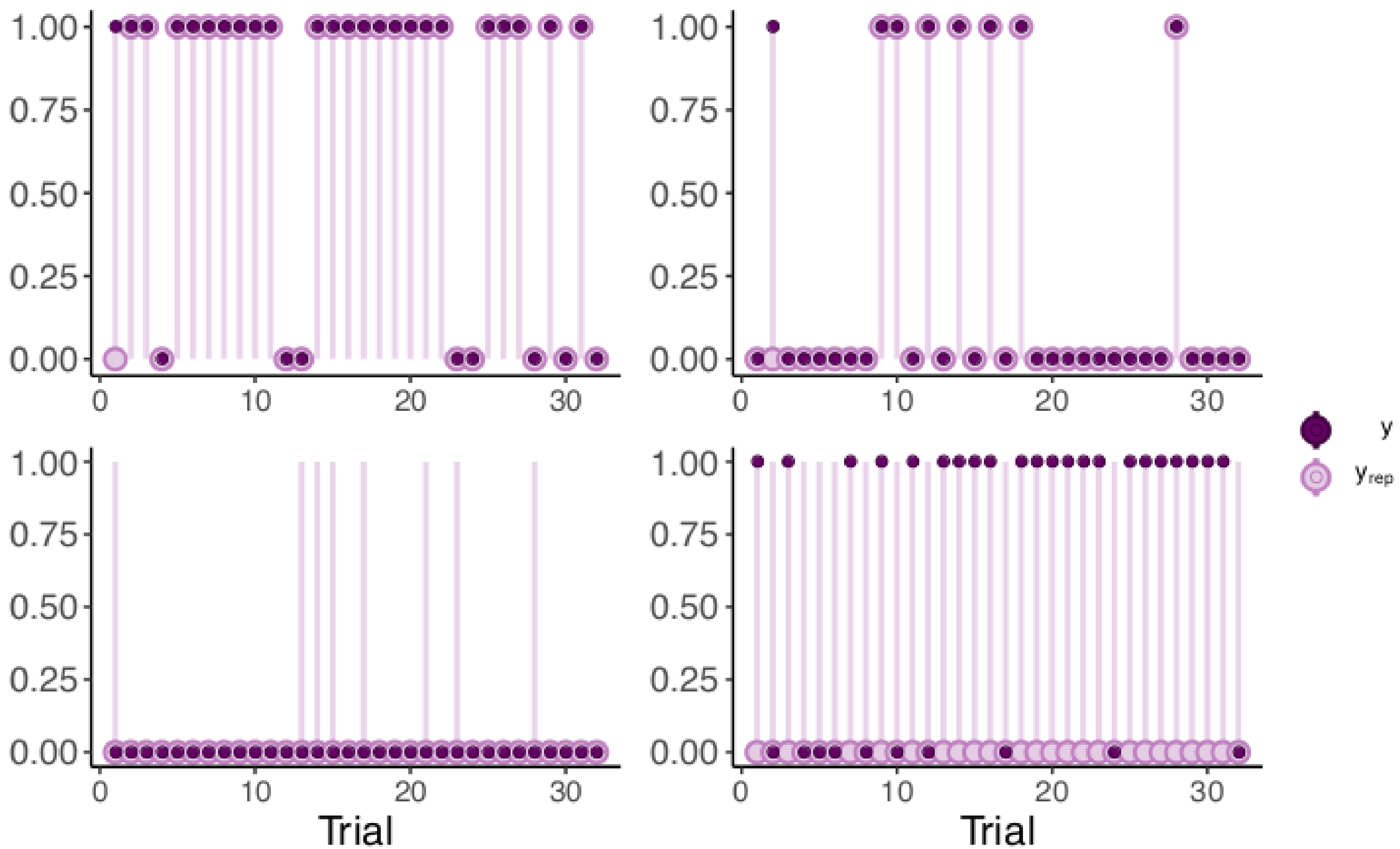
Individual-level estimates



Other relationships

- Individuals' expected rationality, $E[\lambda_i]$, correlates with number of successes ($r \approx 0.84$)
- Uncertainty about partner's expectations decrease over time ($r \approx -0.55$ between trial number and width of $P_i' s\ \ 0.89\%$ HPDI)
- Individual's beliefs diverge from population but grow closer to their partner's over time
($r \approx 0.56$ and $r \approx -0.23$, measured as Kullback-Leibler divergence)
- Subject's own beliefs stabilize over time
($r \approx -0.22$ between trial number and KL divergence)
- Neither divergence from population belief nor width of the P_i's HPDI are related to individuals' rates of success ($r \approx 0.004$ and $r \approx 0.02$)

Predictions



Accuracy

- $HM_{\lambda,\alpha,\beta}$ has RMSE of 0.32
- Always predicting short for frequent has RMSE of 0.63 (cf. Parikh 2000)
- Always predicting short for infrequent has RMSE of 0.78
- Always avoiding it has RMSE of 0.67

RMSE of best model reduces to 0.28 or 0.24 when excluding 4 or 8 worst-faring subjects

Taking stock

Narrower remarks

- Population-level trend is held up: short(er) patterns with more frequent
  - Rooted in expectations carried over from training
- Over time, individuals' established conventions of all flavors
  - frequent $\rightarrow$ shorter
  - infrequent $\rightarrow$ shorter
  - (in)frequent $\rightarrow$ longer
- This is one possible explanation of Zipf's Law of Abbreviation, but
  - Link between dyads and population missing
  - In particular: neutral alternatives

Broader remarks

1. Where possible and called for
    - model variation at multiple levels
    - make relationship between levels explicit

Broader remarks

1. Where possible and called for
    - model variation at multiple levels
    - make relationship between levels explicit
2. Population-level trends can often hide individual-level variation, but the former ultimately draw from the latter.

Broader remarks

1. Where possible and called for
    - model variation at multiple levels
    - make relationship between levels explicit
2. Population-level trends can often hide individual-level variation, but the former ultimately draw from the latter.
3. Many individual-level patterns go against theoretical predictions (Zipf's Law of Abbreviation, Horn's division of pragmatic labor, Parikh's disambiguation model, …). The reasons for this are likely manifold.

Broader remarks

1. Where possible and called for
   - model variation at multiple levels
   - make relationship between levels explicit
2. Population-level trends can often hide individual-level variation, but the former ultimately draw from the latter.
3. Many individual-level patterns go against theoretical predictions (Zipf's Law of Abbreviation, Horn's division of pragmatic labor, Parikh's disambiguation model, …). The reasons for this are likely manifold.
4. How dyadic-conventions find their way into the population is still a major and fascinating open question (cf. Hawkins et al. *forthcoming*)

Broader remarks

1. Where possible and called for
    - model variation at multiple levels
    - make relationship between levels explicit
2. Population-level trends can often hide individual-level variation, but the former ultimately draw from the latter.
3. Many individual-level patterns go against theoretical predictions (Zipf's Law of Abbreviation, Horn's division of pragmatic labor, Parikh's disambiguation model, ...). The reasons for this are likely manifold.
4. How dyadic-conventions find their way into the population is still a major and fascinating open question (cf. Hawkins et al. *forthcoming*)
5. Whether the association of the ambiguous form with the (in)frequent meaning is semantic or pragmatic in nature is another open question

Thank you

Data: http://datashare.is.ed.ac.uk/handle/10283/2702

Code: https://osf.io/7m9np/

thomasbrochhagen@gmail.com
https://brochhagen.github.io