Human/animal contrast in patient independently motivates passive production across Japanese dialects

A sentence production study with a picture description task

Masataka Ogawa

The University of Tokyo
Japan Society for the Promotion of Science

2021/11/04

Introduction

Target languages

- Japanese dialects
  - Tohoku (Northeast Japan)
  - Tokyo (The Capital)
  - Kansai (Southern-central Japan)

Target phenomenon

Passive production

- Animacy contrast (Human/animal)
- Difference between Japanese dialects reported by previous research

Main claim

Human/animal contrast in patient independently and uniformly motivates passive production across Japanese dialects

The condition of passive use in Japanese

Passives describe an event where a lower entity in the hierarchy acts on a higer one

(Kuno 1979; Shibatani 2006)

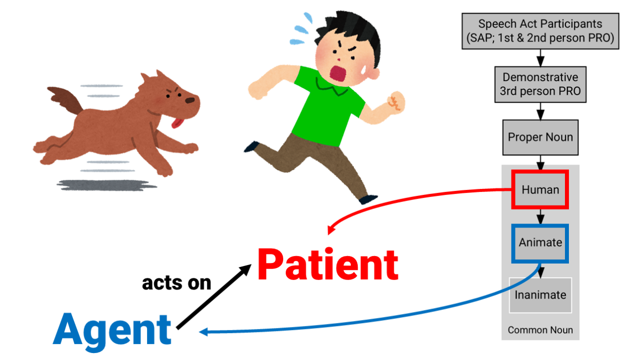


Figure 1: Schematic diagram of Relevance Hierarchy [Mainly from Dixon (1994), pp.84–85; cf. DeLancey (1981) for the treatment of SAP; TOP: Typically Agentive Entities / BOTTOM: Typically Patientive Entities] and a *chasing* event where a lower entity in the hierarchy acts on a higer one

Q: Can the animacy of either agent or patient individually contribute to passive production?

Animacy manipulation in Japanese sentence production studies

The following studies manipulated the animacy of agent and patient to elicit active or passive sentences:

Animate versus Inanimate

Montag et al. (2017), Tanaka et al. (2011)

Human beings versus Animals (both *animate*)

Hidaka (2002, 2016a,b)

- not well evaluated since only one pair of items was used
- **Q: Does human versus animal contrast have a reliable effect on passive production?**

Regional difference in Japanese passive production

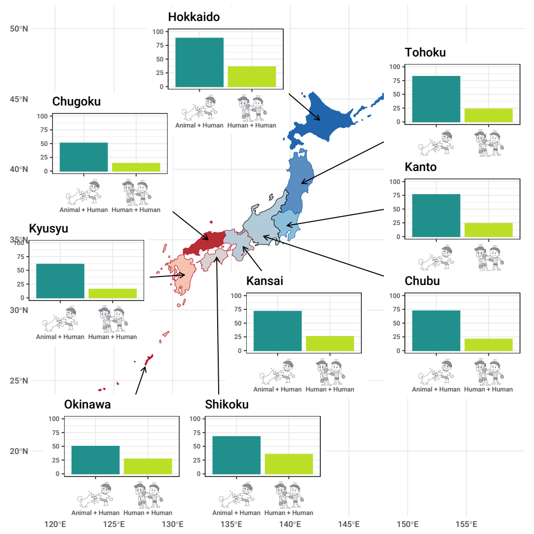


Figure 2: Proportion of passives in each region reported by Hidaka (2016,a,b)

Eastern dialects speakers produced more passives than Western dialects speakers (Fig. 2, Hidaka, 2002, 2016a,b)

- Animacy and structural preference may be unevenly linked across Japanese dialects
- However, only one item pair is used
- **Q: Is such cross-dialectal difference replicable when more items are used in an experiment?**

Methods

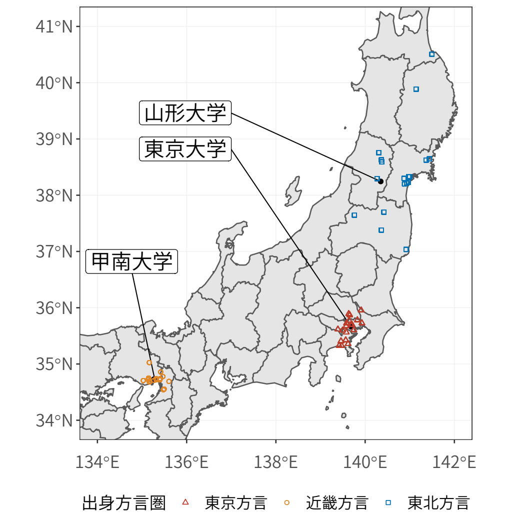
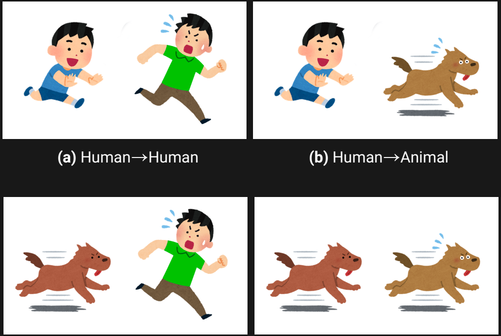


Figure 3: Dialect/birth region where the participants have lived for the longest time up to 18 y.o. and the experimental hosts

Picture description task in the three regions

1. Tohoku
   - one of Eastern dialects
   - the same dialect as Hidaka (2002)
2. Tokyo (≈ Standard Japanese)
3. Kansai
   - one of Western dialects
   - the same dialect as Hidaka (2002)

Settings of the Experiment



(a) Human→Human (b) Human→Animal

(c) Animai→Human (d) Animal→Animal

Figure 4: A sample set of the experimantal items ("chase")

Animacy Manipulation

Human v. Animal (Fig. 4)

(a) **Human→Human**
(b) **Human→Animal**
(c) **Animal→Human**
(d) **Animal→Animal**

Items (Verbs; transitive events)

- *chase, hit, kick, pat, pull, push, rescue, scratch, tickle, wake*

Overall results

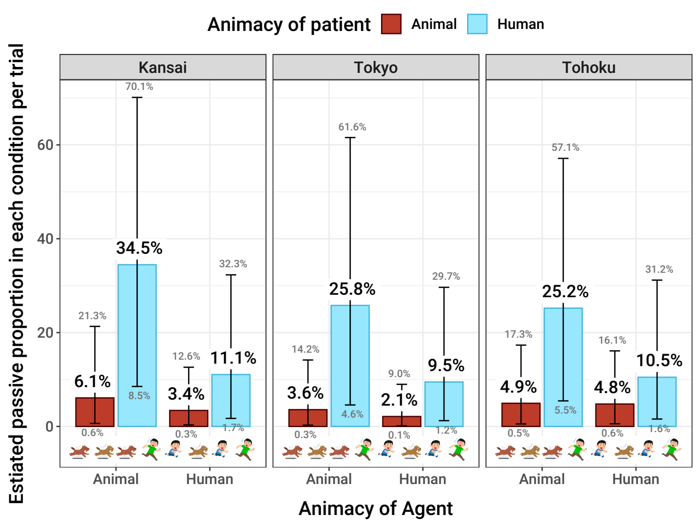


Figure 5: Passive proportion in each condition per trial of the picture description task estimated by the Bayesian mixed effects logistic regression model with a maximally specified random structure (Error bar: 95% ***Credible*** Interval)

- When the patient was a human being, more passive were produced
- Human → human condition elicited more passives than animal → animal condition.
- ∴ The patient's animacy alone can motivate passive production
- No difference between dialects

Conclusion

- Human/animal distinction affects the voice choice
- The patient's animacy independently influences voice selection
- The current results were consistent across Japanese dialects we examined

- DeLancey, Scott (1981). “An Interpretation of Split Ergativity and Related Patterns”. In: *Language* 57.3, pp. 626–657.
- Dixon, Robert M. W. (1994). *Ergativity*. Cambridge studies in linguistics 69. Cambridge University Press.
- Gabry, Jonah and Rok Češnovar (2020). *cmdstanr: R Interface to 'CmdStan'*. https://mc-stan.org/cmdstanr, https://discourse.mc-stan.org.
- Gronau, Quentin F., Henrik Singmann, and Eric-Jan Wagenmakers (2020). “bridgesampling: An R Package for Estimating Normalizing Constants”. In: *Journal of Statistical Software* 92.10, pp. 1–29.
- Hidaka, Mizuho (2002). “Voice (Judou bun wo chusin ni) [Voice: especially on passive]”. In: *Hogen Bumpou Chousa Guidebook [Guidebook of Investigation of Dialectal Grammar: Research Result of Grants-in-Aid for Scientific Research (KAKENHI)]*. Ed. by Takuichiro Onishi. Japanese. NINJAL, pp. 37–63.
- — (2016a). “136 Naguru / Nagurareru [136 Beat / Beaten]”. In: *Shin Nihon Hogen Chizu [New Linguistic Atlas of Japan]*. Ed. by Takuichiro Onishi. Japanese. Asakura Publishing Co., Ltd., pp. 272–273.

Hidaka, Mizuho (2016b). “137 Kamu / Kamareru [137 Bite / Bitten]”. In: *Shin Nihon Hogen Chizu [New Linguistic Atlas of Japan]*. Ed. by Takuichiro Onishi. Japanese. Asakura Publishing Co., Ltd., pp. 274–275.

Jeffreys, Harold (1939/1998). *The Theory of Probability*. Oxford University Press.

Kuno, Susumu (1979). “On the interaction between syntactic rules and discourse principles”. In: *Explorations in Linguistics: Papers in honor of Kazuko Inoue*. Ed. by George Bedell, Eichi Kobayashi, and Masatake Muraki. Kenkyusha, pp. 279–304.

Lee, Michael David and Eric-Jan Wagenmakers (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Montag, Jessica L. et al. (2017). “Language specific and language general motivations of production choices: A multi-clause and multi-language investigation”. In: *Collabra: Psychology* 3(1).20.

Nicenboim, Bruno, Shravan Vasishth, and Frank Rösler (2020). “Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data”. In: *Neuropsychologia* 142, p. 107427.

- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Schad, Daniel J., Michael Betancourt, and Shravan Vasishth (2020). "Toward a principled Bayesian workflow in cognitive science". In: *Psychological Methods*.
- Shibatani, Masayoshi (2006). "On the conceptual framework for voice phenomena". In: *Linguistics* 44.2, pp. 217–269.
- Stan Development Team (2020). *RStan: the R interface to Stan*. R package version 2.21.2.
- Tanaka, Mikihiro N et al. (2011). "Conceptual influences on word order and voice in sentence production: Evidence from Japanese". In: *Journal of Memory and Language* 65.3, pp. 318–330.

Appendix

Steps of the current Bayesian analyses

1. Create the basic generalised linear mixed effect model
2. Set priors (math expression of what has been revealed before/a hypothesis)
   - Prior Means
   - Prior Standard Deviations (SDs)
3. Calculate Bayes Factors
   - quantitative test which of the two models is better
     - usually a model with a certain factor (alternative model) and a model without the factor in question (null model)

$$\overbrace{\pi(\theta|y)}^{\text{Posterior of the parameter given the data}} \approx \overbrace{\pi(\underbrace{y}_{\text{data}} \mid \underbrace{\theta}_{\text{parameter}})}^{\text{Model/Likelihood}} \overbrace{\pi(\theta)}^{\text{Prior of the parameter}}$$

Mixed effects logistic regression

GLMM variant for a response variable with two levels

- Active v. Passive was produced in one trial
- p: Probability of the passive production over active production in one trial

$$
\begin{aligned}
\log\left(\frac{p}{1-p}\right) &= z \\
&= \beta_{\text{ic.gm}} + \beta_{\text{agt}} x_{\text{agt}} + \beta_{\text{pat}} x_{\text{pat}} + (\beta_{\text{R1}} x_{\text{R1}} + \beta_{\text{R2}} x_{\text{R2}}) \\
&\qquad + \beta_{\text{ia1}} x_{\text{agt}} x_{\text{pat}} \\
&\qquad + (\beta_{\text{ia21}} x_{\text{agt}} x_{\text{R1}} + \beta_{\text{ia22}} x_{\text{agt}} x_{\text{R2}}) \\
&\qquad + (\beta_{\text{ia31}} x_{\text{pat}} x_{\text{R1}} + \beta_{\text{ia32}} x_{\text{pat}} x_{\text{R2}}) \\
&\qquad + (\beta_{\text{ia41}} x_{\text{agt}} x_{\text{pat}} x_{\text{R1}} + \beta_{\text{ia42}} x_{\text{agt}} x_{\text{pat}} x_{\text{R2}})
\end{aligned}
$$

where

$$
x_{\text{agt}} = \begin{cases} -0.5 & \text{if an agent is an animal} \\ 0.5 & \text{if an agent is a human being} \end{cases}
$$

$$
x_{\text{pat}} = \begin{cases} -0.5 & \text{if a patient is an animal} \\ 0.5 & \text{if a patient is a human being} \end{cases}
$$

Table 1: Sum contrast coding for the three regions

	x_{R1} (REGION1)	x_{R2} (REGION2)
Tokyo	-1	-1
Tohoku	1	0
Kansai	0	1

Set priors (Choice of the Means)

prior means of βs: $-4 \leq \beta \leq 4$

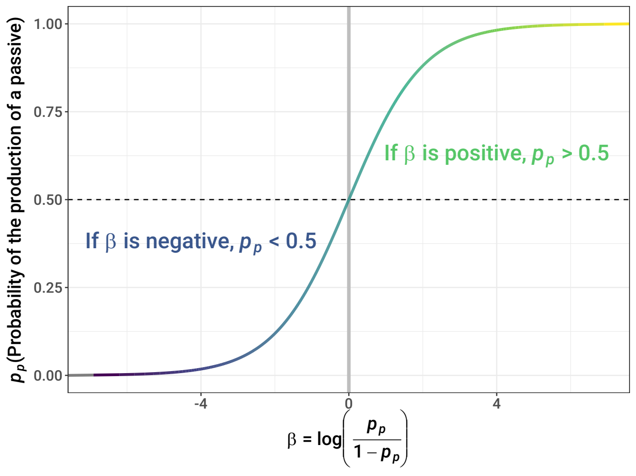


Figure 6: Relation between $\beta = \log\left(\frac{p}{1-p}\right)$ and the probability p

Set priors (Choice of the Means)

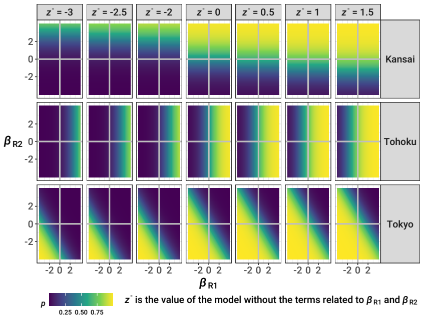


Figure 7: Simulated estimates of the dialectal difference and probability of the passive production

Results of the previous research predict

1. Larger β_{R1} = higher probability of the passive production in Tohoku
2. Smaller β_{R2} = lower probability in Kansai

Current project;

1. β_{R1} is not large = probability of the passive production in Tohoku is not high
2. β_{R2} is not small = probability in Kansai is not low

Set priors (Choice of the SDs)

Deviation of the priors indicates the certainty of the priors

- 'Defining increasingly informative priors amounts to defining smaller standard deviations for the relevant prior' (Nicenboim, Vasishth, and Rösler 2020, p.7L)

$$\beta \sim N(\overbrace{0}^{\text{Ave. }\beta\text{ is 0}}, \overbrace{0.01}^{\text{Definitely }\beta=0}), \quad \beta \sim N(\overbrace{0}^{\text{Ave. }\beta\text{ is 0}}, \overbrace{0.1}^{\beta=0\text{ is certain}}),$$

$$\beta \sim N(\overbrace{0}^{\text{Ave. }\beta\text{ is 0}}, \overbrace{1}^{\beta=0\text{ is probable}}), \quad \beta \sim N(\overbrace{0}^{\text{Ave. }\beta\text{ is 0}}, \overbrace{2}^{\beta=0\text{ is not so likely}})$$

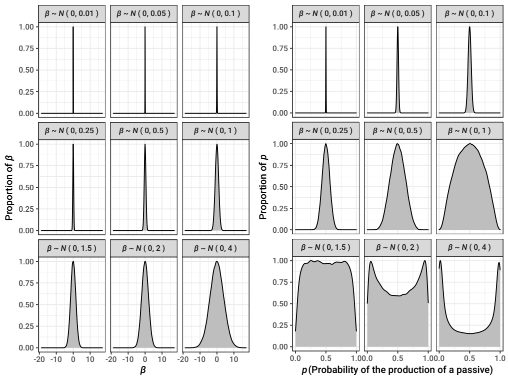


Figure 8: Simulated proportion of the probability of the passive production conditioned by the priors

Calculate Bayes Factors

Bayes Factors

‘Bayes factors provide a way to quantify the evidence that some data provide in favor of one model over another model. It evaluates the model and the prior by evaluating its prior predictive accuracy’ (Schad, Betancourt, and Vasishth 2020, p.20L)

Calculate Bayes Factors: BFs and predictions

Comparing two models $\mathscr{M}_1$ and $\mathscr{M}_0$

1. $\mathscr{M}_1$: Model with the factors in question
2. $\mathscr{M}_0$: Model missing one of the factors; all other components are identical to $\mathscr{M}_1$

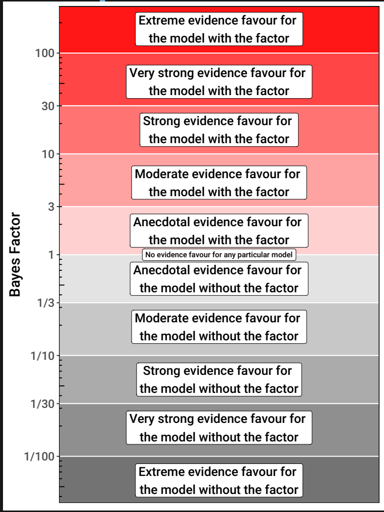


Figure 9: Bayes Factors comparing the presence/absence of a explanatory factor; The interpretation follows Jeffreys (1939/1998) à la labelling of Lee and Wagenmakers (2013, p.105)

Calculate Bayes Factors: Comparison in the current project

1. $\mathscr{M}_1$ v. $\mathscr{M}_0$ without β_{R1}
   - Do Tohoku people prefer passives?
2. $\mathscr{M}_1$ v. $\mathscr{M}_0$ without β_{R2}
   - Do Kansai people avoid passives?
3. $\mathscr{M}_1$ v. $\mathscr{M}_0$ without β_{agt}
   - Is the Human v. Animal contrast in agent effective?
4. $\mathscr{M}_1$ v. $\mathscr{M}_0$ without β_{pat}
   - Is the Human v. Animal contrast in patient effective?
5. $\mathscr{M}_1$ v. $\mathscr{M}_0$ without β_{ia1}
   - Is the animacy interaction between the agent and patient effective?

Calculate Bayes Factors: Prediction on the dialectological difference

Prediction from the previous research

If the previous results are replicable;

1. The more $\mathscr{M}_1$ is favoured over $\mathscr{M}_0$ without β_{R1} for Tohoku, the larger β_{R1} is, even with low certainty
2. The more $\mathscr{M}_1$ is favoured over $\mathscr{M}_0$ without β_{R2} for Kansai, the smaller β_{R2} is, even with low certainty

Prediction from the current project

If the **current project**'s view is accurate, $\mathscr{M}_0$ is almost always favoured

Computational conditions for the parameters of the regions (β_{R1}, β_{R2})

Both β_{R1} and β_{R2} were estimated with various normally distributed priors

Prior *mean* of the parameters

Mean $= \{0, \pm 0.5, \pm 1, \pm 2, \pm 4\}$

If the observation of the previous research holds;

1. The more $\mathscr{M}_1$ is favoured over $\mathscr{M}_0$ without β_{R1}, the larger β_{R1} is, even with low certainty
2. The more $\mathscr{M}_1$ is favoured over $\mathscr{M}_0$ without β_{R2}, the smaller β_{R2} is, even with low certainty

If the **current project**'s view is accurate, $\mathscr{M}_0$ is almost always favoured

Prior *SD* of the parameters

SD $= \{0.01, 0.05, 0.1, 0.25, 0.5, 1, 1.5, 2, 4\}$

- 0.01: **Extremely high certainty** (searching for the posterior in a 'pinpoint' range)
- 1.5: **moderate certainty** (searching for the posterior in a relatively wide range)
- 4: **Extremely low certainty** (searching for the posterior in a vast range)

Computational conditions for the animacy parameters ($\beta_{\text{agt}}, \beta_{\text{pat}}, \beta_{\text{ia1}}$)

β_{agt}, β_{pat} and β_{ia1} were estimated with various normally distributed priors

Prior *mean* of the parameters

Mean $= \{0\}$
If the animacy effect on the agent (β_{agt}) and patient (β_{pat}), and the effect of the interaction of the two β_{ia1} exist;

1. non-zero effect can be found
2. $\mathscr{M}_1$ for each parameter is favoured over its corresponding $\mathscr{M}_0$

Prior *SD* of the parameters

SD $= \{0.01, 0.05, 0.1, 0.25, 0.5, 1, 1.5, 2, 4\}$

- 0.01: **Extremely high certainty** (searching for the posterior in a 'pinpoint' range)
- 1.5: **moderate certainty** (searching for the posterior in a relatively wide range)
- 4: **Extremely low certainty** (searching for the posterior in a vast range)

Computational Methods

executed with 4.0.2 (2020-06-22) (R Core Team 2020)

1. Estimate the coefficients (βs), compiling the models and sampling by `brms::brm()` (**R-brms**) with `cmdstanr` (Gabry and Češnovar 2020, Ver.0.4.0) backend
   - 2000 iterations for warming-up, 50000 for sampling
2. Compile the model again by `brms::brm()` with `rstan` (Stan Development Team 2020, Ver.2.21.2) backend
3. Bridge-sampled by `bridgesampling::bridge_sampler()` (Gronau, Singmann, and Wagenmakers 2020, Ver.1.1.2)

- See also the discussion on GitHub and The Stan Forum for further details of these calculation steps
  - https://github.com/quentingronau/bridgesampling/issues/27
  - https://discourse.mc-stan.org/t/how-to-do-bridge-sampling-and-calculate-bayes-factor-with-brms-with-cmdstanr-backend/18873

Results of the posterior estimation: β_{R1} and β_{R2}

Posterior coefficient of β_{R1} and β_{R2}

Contrary to the prediction from the previous research,

1. β_{R1} is ***not consistently large*** and may take either positive or even ***negative*** values.
2. β_{R2} is ***not consistently small*** and may take either negative or even ***positive*** values.

Bayes Factor comparing the models with/without β_{R1} and β_{R2}

Contrary to the prediction from the previous research, BF shows;

1. $\mathscr{M}_0$s without β_{R1} or β_{R2} are favoured
   - especially compared to $\mathscr{M}_1$s with the large absolute effect size
   - i.e. $\beta_{\text{R1}}, \beta_{\text{R2}} = |\pm 1, \pm 2, \pm 4|$
   - no matter how much certainty (from extremely certain [SD $= 0.01$] to extremely uncertain [SD $= 4$]) the priors represent
2. $\mathscr{M}_0$s and $\mathscr{M}_1$s do not differ considerably
   - especially $\mathscr{M}_0$s v. $\mathscr{M}_1$s with small effect size and extreme certainty
   - i.e. $\mathscr{M}_1$s with $\beta_{\text{R1}}, \beta_{\text{R2}} = 0, |\pm 0.5|$

Estimation of β_{R1}

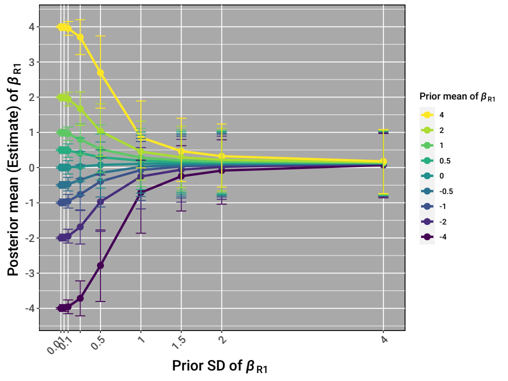


Figure 10: Prior and posterior estimates of β_{R1} (Error bar: 95% ***Credible*** Interval)

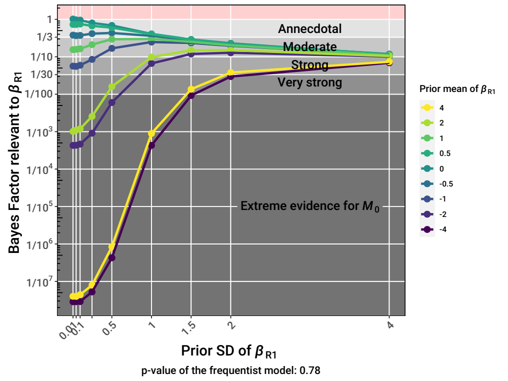


Figure 11: Bayes Factor relevant to β_{R1}

With a moderate certainty (SD $= 1.5$), the posteriors of β_{R1} vary mainly between -0.246 and 0.466

- β_{R1} is **not consistently large** and can be a negative value

Bayes Factor indicated that $\mathscr{M}_0$ without β_{R1} was favoured over $\mathscr{M}_1$

- Not merely 'anecdotally', but also with '**extreme evidence**'

Estimation of β_{R2}

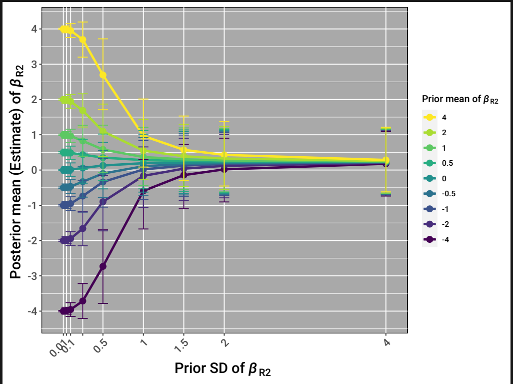


Figure 12: Prior and posterior estimates of β_{R2} (Error bar: 95% ***Credible*** Interval)

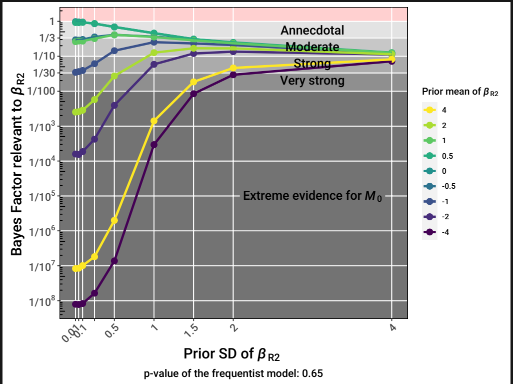


Figure 13: Bayes Factor relevant to β_{R2}

With a moderate certainty (SD $= 1.5$), the posteriors of β_{R2} vary mainly between -0.144 and 0.572

- **Not stably small β_{R2}**

Bayes Factor indicated that $\mathscr{M}_0$ without β_{R2} was favoured over $\mathscr{M}_1$

- Again, not merely 'anecdotally', but also with '**extreme evidence**'

Results of the posterior estimation: β_{agt}, β_{pat} and β_{ia1}

Posterior coefficient of β_{agt}, β_{pat} and β_{ia1}

1. β_{agt} and β_{ia1} are ***relatively small***
   - i.e. The effect of agent animacy and the interaction of the animacy of event participants are ***relatively small***
2. β_{pat} is ***reliably large***
   - i.e. The effect of patient animacyis ***relatively large***

Bayes Factor comparing the models with/without β_{agt}, β_{pat} and β_{ia1}

BF shows;

1. In regard to the **patient** animacy, $\mathscr{M}_1$s considerably differ from $\mathscr{M}_0$s and are favoured
   - especially in moderate (i.e. reasonable) SD condition
2. As for other parameters, $\mathscr{M}_1$s are anecdotally favoured over $\mathscr{M}_0$s without the parameters

Estimation of β_{agt}

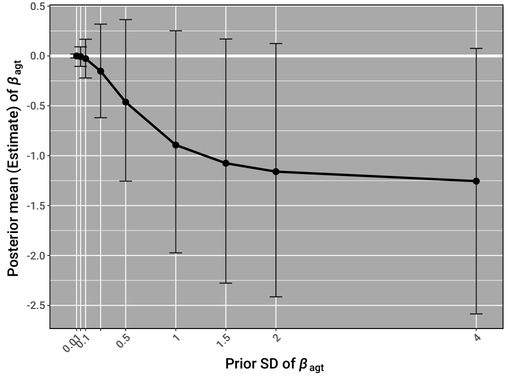


Figure 14: Prior and posterior estimates of β_{agt} (Error bar: 95% ***Credible*** Interval)

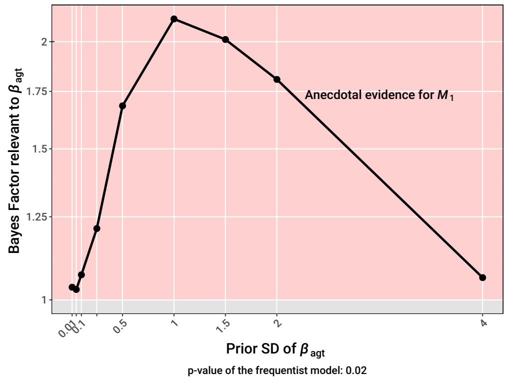


Figure 15: Bayes Factor relevant to β_{agt}

With a moderate certainty (SD $= 1.5$), the posterior of β_{agt}'s 95% CrI lies between -2.277 and 0.17

- Passive is preferred more with an animal agent than with a human agent

Bayes Factor indicated that $\mathscr{M}_1$ was favoured over $\mathscr{M}_0$ without β_{agt}

- But merely 'anecdotally'

Estimation of β_{pat}

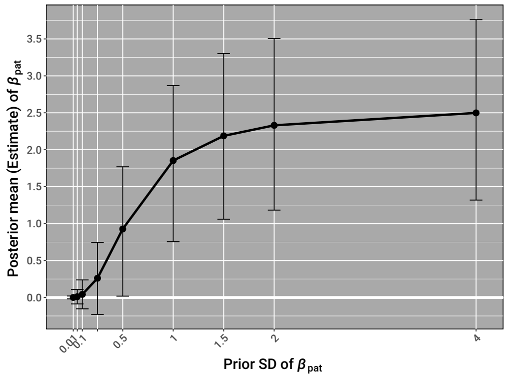


Figure 16: Prior and posterior estimates of β_{pat} (Error bar: 95% ***Credible*** Interval)

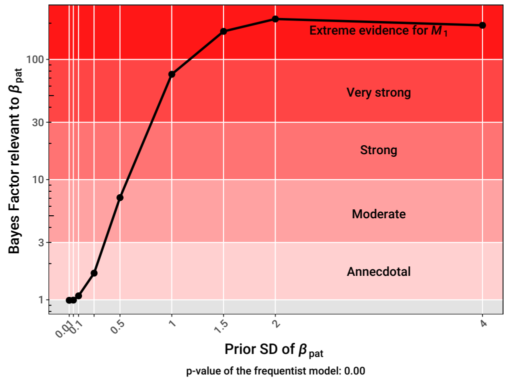


Figure 17: Bayes Factor relevant to β_{pat}

With a moderate certainty (SD $= 1.5$), the posterior of β_{agt}'s 95% CrI lies between 1.06 and 3.301 (95% CrI does not cross 0)

- Passive is considerably preferred more with a human patient than with an animal patient

Bayes Factor indicated that $\mathscr{M}_1$ was favoured over $\mathscr{M}_0$ without β_{pat}

- Not merely 'anecdotally', but also with '**extreme evidence**'

Estimation of β_{ia1}

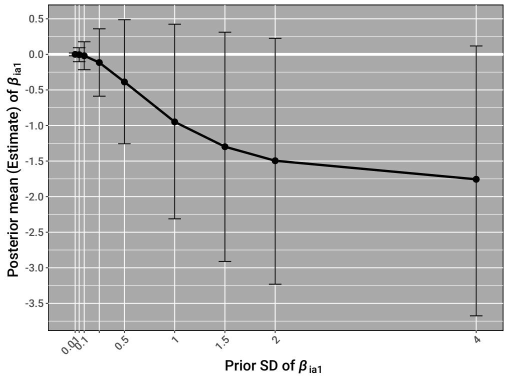


Figure 18: Prior and posterior estimates of β_{ia1} (Error bar: 95% ***Credible*** Interval)

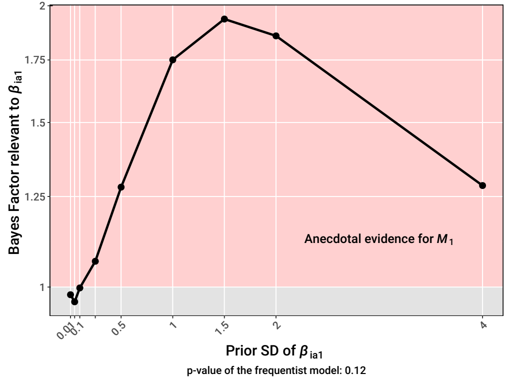


Figure 19: Bayes Factor relevant to β_{ia1}

With a moderate certainty (SD $= 1.5$), the posterior of β_{ia1}'s 95% CrI lies between -2.911 and 0.311

- Passive preference is not so much regulated by the interaction between the animacy of an agent and patient

Bayes Factor indicated that $\mathscr{M}_1$ was favoured over $\mathscr{M}_0$ without β_{ia1}

- But merely 'anecdotally'

(*) Comparison to the frequentist approach

Table 2: Estimates of the Bayesian and Frequentist mixed effects logistic regression model

	Bayesian				Frequentist			
	Coefficients	Estimated Error	2.5% CrI	97.5% CrI	Coefficients	Standard Error	z-value	p-value
Intercept (Grand Mean; $\beta_{\text{ic.gm}}$)	-2.67	0.50	-3.68	-1.71	-2.47	0.46	-5.36	0.00
Animacy of Agent (β_{agt})	-0.89	0.56	-1.97	0.25	-0.99	0.41	-2.40	0.02
Animacy of Patient (β_{pat})	1.85	0.53	0.76	2.87	2.14	0.43	4.97	0.00
Interaction between Animacy of A and P (β_{ia1})	-0.95	0.70	-2.31	0.42	-1.26	0.81	-1.57	0.12
Tohoku or Other Regions (β_{R1})	0.11	0.42	-0.72	0.92	0.10	0.37	0.28	0.78
Kansai or Other Regions (β_{R2})	0.19	0.42	-0.63	1.02	0.17	0.38	0.46	0.65
Interaction between Animacy of A and R1 (β_{ia21})	0.30	0.54	-0.76	1.35	0.39	0.54	0.73	0.46
Interaction between Animacy of A and R2 (β_{ia22})	-0.23	0.57	-1.37	0.89	-0.18	0.58	-0.31	0.76
Interaction between Animacy of P and R1 (β_{ia31})	-0.35	0.53	-1.39	0.69	-0.50	0.54	-0.94	0.35
Interaction between Animacy of P and R2 (β_{ia32})	0.03	0.56	-1.07	1.14	0.16	0.57	0.28	0.78
Interaction among A, P and R1 (β_{ia41})	-0.22	0.75	-1.69	1.24	-0.66	1.07	-0.62	0.54
Interaction among A, P and R2 (β_{ia42})	-0.01	0.77	-1.52	1.50	0.16	1.14	0.14	0.89

Note:
The numbers in red indicate the smaller of the coefficients of the Bayesian model or the frequentist model. The numbers in orange indicate the smaller of the errors of the Bayesian model or the frequentist model.

[1] The Bayesian model was fit with varying both intercepts and slopes with correlation (the maximal model). The prior for the intercept was $N(0, 1.5)$. The priors for the $betas$ (group-level effects) were $N(0, 1)$. The priors for the random intercepts and slopes were $N(0, 1)$. The priors for the correlations between the random intercepts and slopes were $lkj(2)$.

[2] The frequentist model was fit with varying intercepts only. The other frequentist models, including the maximal model, failed to converge.