

Bayesian models of argument strength (e.g. Hahn and Oaksford, 2007) generally predict that arguments become stronger as more evidence is provided. This study investigates a competing intuition: that arguments can be weakened by excess reasons (“the lady doth protest too much!”). We suggest that this effect is predicted by rational theories of language use.

In order to maximize information exchange under cognitive constraints, pragmatic speakers tend to avoid mentioning predictable information. Listeners expect this production preference and often make inferences to “rationalize” perceived redundancies (Kravtchenko and Demberg, 2022). For example, “Alice went to the grocery store... she paid the cashier!” can generate an inference that Alice doesn’t reliably pay for groceries.

If speakers are similarly pragmatic when constructing arguments, they should provide enough evidence for the conclusion to become acceptable to the listener, and not much more. Elaborate justifications would only be worthwhile when the conclusion is particularly difficult to reconcile with the listener’s other beliefs. Then, if listeners rationalize seemingly superfluous reasons by assuming that the speaker expected a disagreement, reasons that evidentially count in favor of the conclusion might pragmatically count against it.

We develop a recursive pragmatic model of argumentation using the Rational Speech Act framework (Frank and Goodman, 2012). The model is grounded in a literal listener (L0), a simple Bayesian agent who performs no pragmatic reasoning. The pragmatic speaker (S1) optimizes her argument to maximize the simple agent’s credence while minimizing production cost. Intuitively, S1 provides more reasons for an unfavorable prior, and fewer reasons for a favorable prior. By assuming utility-maximizing behavior from S1, the pragmatic listener (L1) can infer which distribution S1 must be treating as L0’s prior. He performs the final Bayesian update from an adjusted prior that rationalizes S1’s utterance.

60 participants were recruited for a behavioral study to test the model’s predictions. Stimuli consisted of text-message dialogues, in which the speaker either justified (“four reasons”) or did not justify (“zero reasons”) a typical or atypical action. Using a 7-point Likert scale, participants rated how favorably the speaker expected the listener to feel about the action (L1’s estimate of the S1/L0 prior) and how favorably the listener would actually feel (L1’s posterior).

Analyses were conducted using Bayesian ordinal models. Participants indeed inferred a lower prior for costly justifications, but only when the justifications were uninformative (i.e. for typical actions). While the posterior was rated similarly in the zero-reasons condition for both typical and atypical actions, only the atypical actions were rated higher in the four-reasons condition. This indicates that the reduction of the prior in response to costly justifications also reduced L1’s posterior in the “typical” condition, in keeping with the predictions of the pragmatic model.

- Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, 336(6084), 998–998.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704–732.
- Kravtchenko, E., & Demberg, V. (2022). Informationally redundant utterances elicit pragmatic inferences. *Cognition*, 225, 105159.