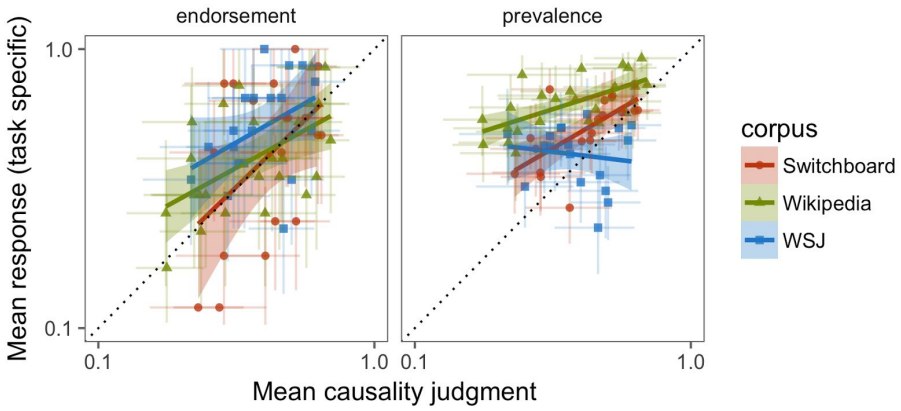
People are crazy: Investigating naturally occurring generics

Generalizations about categories (*generics;* e.g., “People are crazy”) are ubiquitous in everyday conversation [1-3]. Despite their ubiquity and relative morpho-syntactic simplicity, generics’ truth conditions have proven notoriously difficult to characterize. The data to be explained are complex but generally agreed-upon patterns of truth value or pragmatic felicity judgments of examples, typically made up by researchers. However, recent developments in experimental pragmatics highlight the importance of understanding the natural input that listeners receive as a source of constraints on linguistic theorizing [4]. Thus inspired, we ask: **(1) Is there variability in listeners’ interpretation of generics?** If so, this would suggest that the data situation is even more complex than hitherto assumed but would also provide a clear constraint on theories of generics: they must be relativized to *individuals’* idiosyncracies like their prior beliefs about the world. Given that world knowledge is likely to vary across genres, we thus also ask: **(2) Is there variability in inferred property prevalence and generic endorsement across genres (corpora)?** Finally, many theories of generics include mechanisms by which prevalence could deviate from endorsement (e.g., “Mosquitos carry malaria” is true, despite the low prevalence of malaria among mosquitos). One generally established mediator of this variability is perceived causality: there is something about mosquitos that causes them to carry malaria [5-6]. We thus also ask: **(3) Does subjective causality predict prevalence inferences and endorsement rates?**

Methods. We compared generics from the Switchboard (telephone conversations between strangers), Wikipedia (encyclopaedic knowledge), and Wall Street Journal corpus (news texts on mostly financial topics). From each corpus, we extracted all disfluency-free occurrences of bare plurals followed by a present tense (non-progressive, non-deontic) verb phrase. We randomly sampled 20 cases from each corpus. An example from each is shown in (6) - (8).
(6) People are crazy. (7) Art indexes track winners, not losers. (8) Affected persons show loosening of associations.
In subsequent experiments on Amazon’s Mechanical Turk, each item, presented with six lines of context, received 20 ratings each on three different tasks: **implied prevalence** (“How many people are crazy?”, continuous), **endorsement** (“Do you agree with the statement?”, binary), and **perceived causality** (“Is there something about people that causes them to be crazy?”, continuous).

Results. To evaluate the effect of causality and genre on judgments (see Figure) we conducted two exploratory mixed-effects regression analyses (endorsement--logistic; prevalence--linear) with fixed effects of causality judgment, corpus, and their interaction, and random by-participant and by-item intercepts. The only significant predictor of endorsement was the main effect of causality ($\beta = 5.2$, SE = 1.9, $p < 0.01$). In contrast, implied prevalence was greater in the Wikipedia corpus than in the WSJ ($\beta = 0.2$, SE = 0.05, $t = 3.4$), and there was an interaction of corpus and causality such that causality was a better predictor of prevalence in Switchboard ($\beta = 0.4$, SE = 0.2, $t = 2.2$) and Wikipedia ($\beta = 0.7$, SE = 0.2; $t = 3.0$) than in the WSJ.

Conclusion. The observed variability in inferred prevalence and endorsement both within items and across corpora supports theories that treat production and interpretation of generics as influenced by multiple features (e.g., statistical and causal features, or by an integration of listeners’ subjective beliefs about the world with assumptions of speaker cooperativity; [6-7]). The methods presented here showcase an exciting avenue for collecting novel, natural, data that constrain theory-building.



References.
[1] Carlson, G. N. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, *1*, 413–457. [2] Carlson, G. N., & Pelletier, F. J. (Eds.). (1995). *The generic book*. University of Chicago Press. [3] Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A., & Keil, F. C. (1998). Beyond Labeling: The Role of Maternal Input in the Acquisition of Richly Structured Categories. *Monographs of the Society for Research in Child Development*, *63*(1), 1–159. [4] Degen, J. (2015). Investigating the distribution of “some” (but not “all”) implicatures using corpora and web-based methods. Semantics and Pragmatics, 8(11), 1–55. [5] Prasada, S., Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition*, *126*(3), 405–22. [6] Tessler, M. H., & Goodman, N. D. (2016). A pragmatic theory of generic language. *arXiv*, *abs/1608.02926* [7] Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press, USA.