Modeling the context dependence of artifact nouns

Brandon Waldon, Cleo Condoravdi, Beth Levin, and Judith Degen (Stanford)

Whereas the context dependence of gradable adjectives (*tall*, *open*) has received extensive treatment in the computational pragmatics literature [7, 5, 6], modeling the context dependence of vague artifact nouns (*vehicle*, *electronic device*) is an under-developed enterprise. We show that when interpreting rules that express requirements/prohibitions on behavior, contextual information as to the rule's purpose modulates interpreter beliefs about artifact noun category boundaries. We support this claim with a quantitative comparison of Bayesian probabilistic pragmatic models of linguistic interpretation couched within the Rational Speech Act (RSA) framework [1].

Experiment: Participants (n = 188) completed 12 trials, each of which contained a rule featuring an artifact noun. Participants were assigned to 1 of 4 possible 'goal' conditions, which differed as to the relevant authority's motivation for issuing the rule. In the 'None' goal condition, participants read the rule with no preceding context; the other 3 conditions identified the authority's goal in passing the rule. For example, for the *No electronic devices...* rule, one of the 'goal' conditions featured the following text above the rule: "The managers of a theater are concerned that certain objects, when brought into the theater, emit light that could distract audience members and performers." Each rule was associated with a set of 12 images. For prohibition-expressing rules, participants were instructed to select each item that would violate the rule; for rules expressing requirements, the instruction was to select each item that would satisfy the rule.

Norming studies: N1: Objects were normed for category membership (n = 40). For example, participants reported whether they believed an object was an *electronic device*, see Fig. 2. **N2:** Beliefs about policy goal-relevant features of experimental items (e.g. whether or not an object emits light or could be used to record live performances) were elicited in a feature attribution norming study (n = 120, see Fig. 3). **N3, N4:** Rules (n = 40) and goals (n = 120) were normed for plausibility. **N5:** Images (12 per rule; 144 total) were normed for nameability (n = 40).

Results: Object selection rates across goal conditions for 1 of the 12 tested rules are shown in Fig. 1. Mean selection rates correlated with category membership norms (Fig. 4), but a priori category membership cannot explain any observed variance between the goal conditions (Fig. 5).

Computational model: An RSA model that incorporates contextual information as to the signaler's policy goal outperforms a baseline model that does not (see also, e.g., [3] for a similar model applied to metaphor). At the base of the RSA recursion, the 'literal' L_0 interpreter observes a rule and a policy goal g and infers whether an object o is in the scope of prohibition/requirement. f^g returns normed feature attribution values from **N2**, while $P_{NOM}(o)$ is defined via **N1** (Fig. 2).

Model details and example:

- $L_0(o \text{ prohibited}|\text{"No elec. devices..."}, g) \propto f^g(o) \cdot P_{\text{elec.-device}}(o)$
- $L_0(o\ \underline{\text{not}} \text{ prohibited}|\text{"No elec. devices..."}, g) \propto \big(1 - f^g(o)\big) \cdot \big(1 - P_{\text{elec.-device}}(o)\big)$
- $S_1(u|g,s) \propto exp\big((\alpha \cdot log(L_0(s|g,u)) - C(u))\big)$
  (where $u \in \{\text{"No elec. devices..."}, silence\}$; $s \in \{o \text{ prohibited}, o\ \underline{\text{not}} \text{ prohibited}\}$)
- $L_1(o \text{ prohibited}|\text{"No elec. devices..."}, g) \propto S_1(\text{"No elec. devices..."}|g, o \text{ prohibited}) \cdot P_{\text{elec.-device}}(o)$

This model has high overall predictive accuracy (Fig. 6) and outperforms a baseline $L_{1-\text{no-goal}}$ model (Bayes Factor > 18), which is identical save for the fact that its literal interpreter $L_{0\text{-no-goal}}$ encodes only prior beliefs from $P_{NOM}(o)$ and no information about goal-relevant object features. This result suggests that contextual information as to the rules' goals modulates interpretation of the rules themselves. This work thus a) informs the empirical landscape regarding the context dependence of artifact noun interpretation (see also [4, 2, 8]); and b) offers a first computational pragmatic foray into a complex problem of linguistic vagueness.

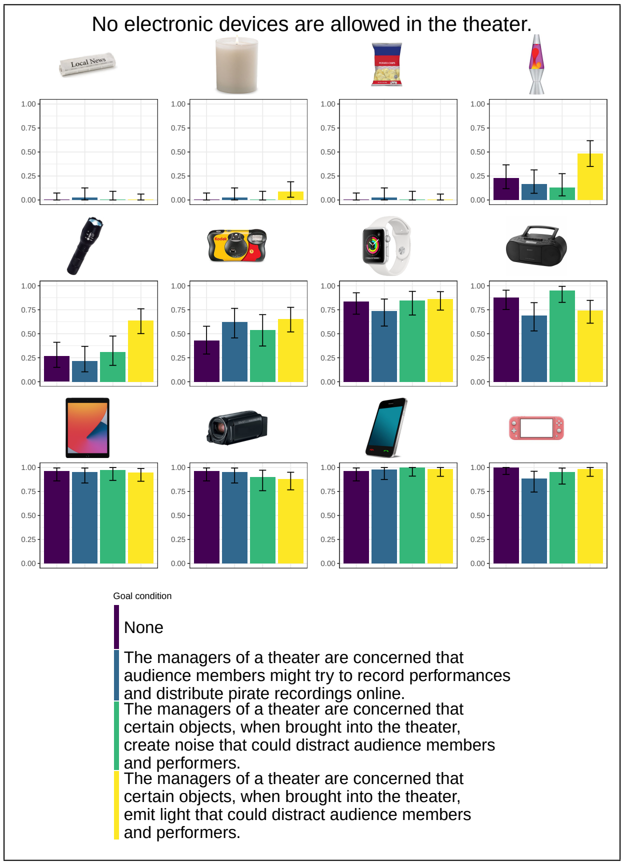
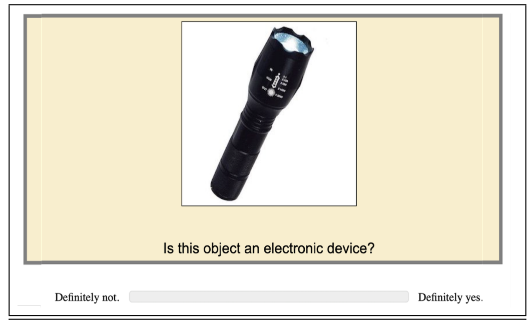


Figure 1: Results from 1 of 12 scenes of the study. Y-axis: proportion of participants that selected each object in each of the 4 goal conditions associated with the scene; error bars are 95% binomial confidence intervals.



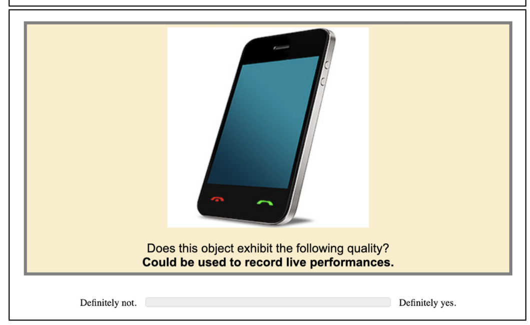


Figure 2 (top): sample screen from the category membership norming study N1; Figure 3 (bottom): sample screen from the feature attribution norming study N2.

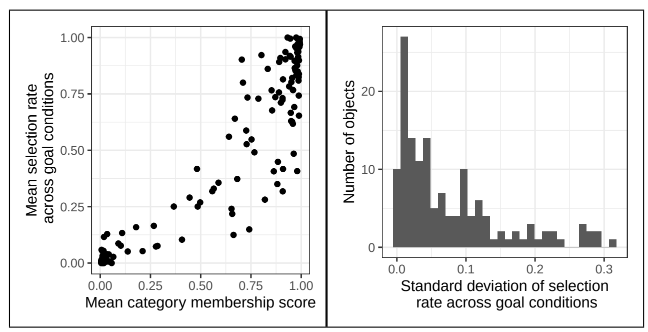


Figure 4 (left): Mean responses from the category membership norming study N1, plotted against mean selection rates from the main experiment (each point represents a single object seen in the study); Figure 5 (right): standard deviation of object selection rates across goal conditions.

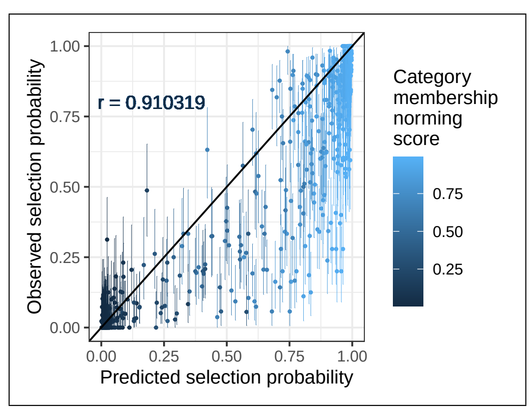


Figure 6: prediction accuracy of L_1. Every point represents a single object shown in a single goal condition.

[1] M. C. Frank and N. D. Goodman. "Predicting pragmatic reasoning in language games". In: *Science* 336.6084 (2012). [2] S. Grimm and B. Levin. "Who Has More Furniture? An Exploration of the Bases for Comparison". Paper presented at Mass/Count in Linguistics, Philosophy and Cognitive Science Conference. 2012. [3] J. Kao, L. Bergen, and N. Goodman. "Formalizing the pragmatics of metaphor understanding". In: *Proceedings of Cog Sci*. Vol. 36. 36. 2014. [4] W. Labov. "The Boundaries of Words and their Meanings". In: *Fuzzy Grammar: A Reader*. Ed. by B. Aarts. Oxford University Press, 2004. [5] D. Lassiter and N. D. Goodman. "Adjectival vagueness in a Bayesian model of interpretation". In: *Synthese* 194.10 (2017). [6] C. Qing. *Semantic Underspecification and Its Contextual Resolution in the Domain of Degrees*. Stanford University, 2020. [7] C. Qing and M. Franke. "Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model". In: *Semantics and linguistic theory*. Vol. 24. 2014. [8] G. Scontras et al. "Who has more? The influence of linguistic form on quantity judgments". In: *Proceedings of the Linguistic Society of America* (2017).