On the context dependence of artifact noun interpretation[1]

Brandon WALDON — *Stanford University*
Cleo CONDORAVDI — *Stanford University*
Beth LEVIN — *Stanford University*
Judith DEGEN — *Stanford University*

Abstract. The context dependence of artifact noun category boundaries is underexplored relative to interpretational context dependence in other linguistic domains (e.g., gradable adjectives). Taking inspiration from a normative debate over the role of context in legal interpretation (in particular, the role of legislators' policy goals), we show experimentally that contextual information as to a rule's purpose systematically modulates interpreter beliefs about the category boundaries of artifact nouns contained within the rule. We propose a Bayesian pragmatic model of the context-dependent resolution of artifact noun extensions, which we compare against context-independent baselines. Our experimental and modeling results suggest the need for an explicitly context-sensitive, multi-dimensional degree semantics for artifact nouns.

Keywords: computational pragmatics, vagueness, degree semantics, nominal semantics

1. Introduction

(1) "A legal rule forbids you to take a vehicle into the public park. Plainly this forbids an automobile, but what about bicycles, roller skates, toy automobiles? ... Are these, as we say, to be called 'vehicles' for the purpose of the rule or not?" (Hart 1958: 607).

As (1) shows, *vehicle* is vague in that it admits of borderline cases (e.g., bicycles and roller skates). By virtue of this vagueness, we expect to find "contextual variability in truth conditions" of sentences featuring the expression (Kennedy 2007: 2). This paper will offer experimental evidence – and a formal linguistic analysis – of this vagueness and context dependence.

Theorists have already extensively explored the ways in which context influences the interpretation of artifact nouns (such as *vehicle*): phenomena such as copredication (e.g., *My car is no longer manufactured but still runs great*) have motivated lexical semantic analyses that account for the range of senses that a single artifact noun may invoke in context (Pustejovsky, 1995; Asher, 2011). However, when we restrict our focus to the 'physical object' sense of these artifact nouns – the sense invoked by Hart in (1) – we find that there is limited work that formally characterizes the context dependence of their category boundaries.

Conversely, some other linguistic domains are relatively well-explored in this regard, especially single-dimensional gradable adjectives such as *tall*, *open*, and *dry*.[2] For our purposes, artifact nouns are notable in the following respect: although *tall* clearly makes reference to a single dimension (height), there is no single dimension to which *vehicle* intuitively refers. On the whole, multi-dimensional predicates have been less studied than their single-dimensional

---

[1]We would like to thank audiences at SuB 27, XPrag 2022, CUSP 13, Stanford, and Georgia Tech for valuable feedback and discussion. We are grateful as well to our anonymous SuB reviewers for their insightful commentary. All remaining errors are our own. Corresponding author: Brandon Waldon `bwaldon@stanford.edu`. Data and code associated with this project can be found at `https://github.com/bwaldon/artifactnouns`.

[2]With regards to single-dimensional gradable adjectives, there exists a rich literature in formal semantics (Kamp, 1975; Klein, 1980; Kennedy, 2007); experimental semantics/pragmatics (Aparicio et al., 2016; Qing et al., 2018); and computational pragmatics (Qing and Franke, 2014; Lassiter and Goodman, 2017).

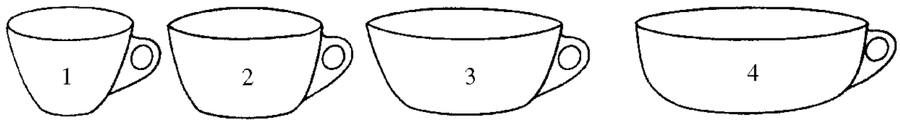
counterparts when it comes to vagueness and context dependence (though see e.g., Klein 1980 and Sassoon 2013 for important exceptions pertaining mostly to multi-dimensional adjectives).

This paper seeks to redress this disparity. We first present an experiment which shows that highlighting the function(s) of artifact noun objects can serve as a cue for how to resolve the category boundaries of artifact nouns in context, in line with previous findings reviewed below. We then present a context-sensitive, multi-dimensional semantics for artifact nouns, which we embed in a probabilistic model of artifact noun interpretation. The model predicts our experimental data, and does so more accurately than a context-insensitive baseline model.

2. Producing and understanding artifact nouns in context

We will use the term 'artifact nouns' in contrast to natural kind-denoting nouns, where the former denotes entities of human invention and/or entities that, through some assimilative procedure, come to serve some human-intended function.[3] Many artifact nouns are also identified via diagnostics that distinguish them from natural kinds. Artifact nouns, for example, exhibit systematically different compounding strategies from natural kinds (Downing, 1977; Wisniewski and Love, 1998; Levin et al., 2019). While modifiers in natural kind compounds (e.g., <u>*three-toed*</u> *sloth*) tend to denote essential characteristics which specify a subordinate kind within the kind denoted by the bare head (such as "perceptual features or typical habitat," Levin et al. 2019: 464), modifiers in artifact compounds (e.g., <u>*coffee*</u> *mug*) tend to specify "the event in which the artifact is created or the event in which it is intended to be used" (ibid: 463).

Artifact category judgments have been extensively studied by cognitive psychologists (see below discussion as well as Murphy 2002 and references therein). However, relatively little work to our knowledge has investigated effects of local interpretive context on category boundary judgments for artifact nouns. To the extent this question has been asked, prior work suggests that judgments can shift when a 'purpose' or 'function' that entities are expected to achieve is highlighted in context. Labov (1973) had participants name objects with handles that were of various widths and fixed heights: (Labels provided by participants included *cup*, *bowl*, etc.)



In what was called the 'neutral' context of presentation, Labov found that participants provided the label *bowl* only for fairly wide objects. However, in a second context (the 'food' context), people were first asked to envision that the above objects contained mashed potatoes. Objects of relatively narrow widths received the *bowl* label more frequently in this second context.[4]

In more recent work, Grimm and Levin (2012) asked participants to judge comparatives featuring artifactual aggregate nouns such as *furniture* or *jewelry*. Participants compared two aggregations, a high-cardinality but homogenous aggregation (e.g., 5 chairs) vs. a low-cardinality but

[3]The second disjunct captures the observation that, despite not having been created by humans, entities clearly belonging in the extension of *gourd* (a natural kind) may also belong in the extension of *decoration* (an artifact). However, see Sperber (2007) for discussion of cases that would problematize even this very broad sense of *artifact noun*, and perhaps any attempt to invoke 'artifact' as a theoretically useful concept.

[4]See also Feist and Gentner (2003) and Coventry et al. (1994) for studies demonstrating that preposition choice between *in* and *on* in *The X is __ the Y* depends not only on the spatial relationship between *X* and *Y* but also on the canonical function of *Y* when *Y* is an artifact. Feist and Gentner (2003), for example, find that with spatial relations held constant, *in* is more frequent in the environment *The X is __ the bowl* than in *The X is __ the plate*.

heterogeneous aggregation (e.g., a sofa, a chair, a coffee table, and a bookcase – 4 items total). In what they called a 'function-neutral' context, the authors found that the higher-cardinality aggregation tended to be judged as *more furniture* than the lower-cardinality aggregation. However, in a context that highlighted a purpose associated with the artifact noun (e.g., completely furnishing a room), lower-cardinality, heterogenous aggregations could count as *more furniture*.

The current investigation, inspired by contexts of legal decision-making, builds on the above findings. Artifact noun categories – in particular, superordinate categories – group together objects that share in abstract characteristics often described as 'functional' properties (Rosch et al., 1976; Tversky and Hemenway, 1984). For example, objects nameable as 'vehicles' intuitively share in common the ability to transport people and/or things (though as Malt and Johnson 1992 note and as will be relevant to our experiment and analysis, additional characteristics may have to be selectively invoked to further distinguish superordinate category members from non-members in context, e.g., the property of having wheels or of relying on fuel sources that pollute the environment).[5] The generality of superordinate categories (Wisniewski and Murphy, 1989; Murphy, 2002), coupled with the observation that their unifying features may be relevant to the goals of policy-making entities (e.g., the goal to limit pollution), often make such categories natural targets of rules that express prohibitions or requirements, as in (1).

We exploit this property of contexts in which rules targeting (especially superordinate) artifact noun categories are produced and consider whether policy goals, in turn, influence interpreter beliefs regarding the domain of applicability of such rules. That is, we consider whether evidence of the goals motivating the formulation of a legal rule can have a similar modulating effect on interpretation as did the artifact noun functions invoked by previous studies.

In this sense, our study is also inspired by a normative debate in American jurisprudence. On one side, purposivists consider the policy goals of legislators to be relevant to legal interpretation and readily consult evidence of such goals beyond the text of the law itself. By contrast, textualists tend to minimize the relevance of such goals (Brannon, 2018). Following recent work in *experimental jurisprudence* (Struchiner et al. 2020, Hannikainen et al. 2022, inter alia), we consider which of these normative frameworks more closely aligns with how laypeople actually interpret natural language in settings that simulate legal decision-making. In the final section of the paper, we consider the implications of our results for this normative debate.

3. Experiment: resolution of artifact noun extensions

The experiment investigates two hypotheses regarding the role of a particular contextual cue – the policy objective behind a legal rule – in how interpreters resolve the extensions of artifact nouns. The **Goal Insensitive** hypothesis is a null hypothesis which states that policy goals are orthogonal to interpretation. On the **Goal Sensitive** hypothesis, interpreters systematically integrate beliefs about policy goals when interpreting rules that target artifact noun categories (e.g., *No vehicles...*). This hypothesis predicts that when a particular goal is made salient in a context, we should see a corresponding reflex in how the artifact noun category boundary is resolved, such that the rule is construed in a manner that advances the policy goal.

[5] Invoking such properties might serve to exclude non-members (e.g., escalators) but may also exclude objects that one might want to treat as members of the category in some contexts (e.g., sailboats). Malt and Johnson take this tension as evidence that artifact noun categories cannot be characterized by a single functional 'core' – we take it as evidence that the properties that are relevant for determining category boundaries may vary across contexts.

3.1. Methods

Participants. We recruited 200 particpants via Prolific (all self-reported native English speakers; US-based IP addresses; 100 minimum submissions on Prolific; minimum 95% prior approval rating). Participants were paid $3, and median completion time was approximately 13 minutes, for a median compensation rate of approximately $14 per hour.

Procedure and materials. Participants completed twelve trials, where each trial featured a rule randomly selected without replacement from a list of twelve researcher-designed rules.[6] These rules were separately normed for a priori plausibility.[7] Of the twelve rules, which each targeted an artifact noun category, seven expressed prohibitions (e.g., *No electronic devices are allowed in the theater*) and five expressed requirements (e.g., *Shoes must be worn in the courtyard).*

For each rule, participants were randomly assigned to one of four possible 'goal' conditions, which varied with respect to the availability and nature of contextually-available evidence as to the relevant authority's motivation for issuing the rule. In the 'goal-neutral' condition, participants read the rule with no preceding context; the remaining three 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.* These goals were normed for plausibility in a separate norming study.[8] We ensured that each participant saw exactly 1/3 of trials in the 'goal-neutral' condition; assignment of trials to goal condition was otherwise random.[9]

Each rule was associated with a unique set of twelve images. Images were normed for nameability.[10] On prohibition-expressing rule trials, participants were instructed to "Select each item that, CONDITION, would violate the rule"; on requirement-expressing rule trials, the instruction was to "Select each item that, CONDITION, would satisfy the rule", where CONDITION was tailored to the particular trial. For example, the instruction for the *No electronic devices...* rule was *Select each item that, if brought into the the theater, would violate the rule.*

Participants were required to select at least one object in order to complete the trial. After every trial, participants completed two types of attention checks. On 'image' attention checks, participants were shown one of the twelve images from the preceding scene and three distractor images that never featured on any of the main trials. Participants were asked to identify the image that appeared in the preceding trial. On 'goal' attention checks, participants completed a forced choice task recalling the goal that was featured in the preceding scene. (The competitors were the other two goals that could have been combined with the rule in that scene). Participants who failed more than two image attention checks and/or more than two goal attention checks, as well as those who prematurely exited the study, were excluded from the analysis.

We ran two additional norming studies on the images from the main experiment, and we use the results from these studies in the model comparison reported in Section 4.2. In the **feature**

[6]Preregistered at `https://osf.io/5whfy`. We first planned for $n = 100$. Subsequently, we preregistered a plan to recruit 100 more participants to estimate response patterns with greater confidence; see `https://osf.io/ad32c`.
[7]See `https://osf.io/9njdu` for more details.
[8]See `https://osf.io/k74h3` for more details.
[9]The rules and their associated goals are presented in Appendix A.
[10]See `https://osf.io/682q7` for more details.

attribution norming study, participants were asked to use a sliding scale to rate the extent to which the objects used in the main study exhibited characteristics that corresponded to the policy goals tested in the study. For example, participants were asked to report the extent to which they believed the objects shown in Fig. 1 were capable of emitting light.

In the **category membership** norming study,[11] participants used a sliding scale to rate the extent to which the objects featured in the images were members of the relevant artifact noun category. For example, participants were asked to report the extent to which they believed the objects shown in Fig. 1 were in the category *electronic device*.

For each rule, we selected a diverse set of objects according to the following qualitative typology. For each subclass, we also report the proportion of objects that exhibit a corresponding mean category membership score from the category membership norming study.[12]

Clear members (51%) were objects that were relatively uncontroversial members of the artifact noun category targeted by the rule. For example, the *No electronic devices...* item featured an image of a tablet computer. (Mean category membership scores between 0.75 and 1).

Edge cases (17%) were more controversial members of the category. For example, the *No electronic devices* item featured a flashlight, whose status as an 'electronic device' is was reported to be less certain than that of a tablet. (Scores between 0.25 and 0.75).

Clear non-members (32%) were outside the category of interest. (Scores less than 0.25).

3.2. Results

Data from 12 participants were excluded based on the exclusion criteria described in Section 3.1. To illustrate the results we obtained in our study, rates of object selection across goal conditions for 4 of 12 tested objects in the *No electronic devices...* rule scenario are displayed in Fig. 1. These include two clear members, one edge case, and one clear non-member. A visual inspection of the graphs suggests that not every object exhibited 'goal sensitivity' to the same degree. With the *No electronic devices...* rule, for example, rate of selection of the flashlight – an edge case 'electronic device' – was highest when presented alongside the goal of limiting distracting light from the theater. However, no such modulation was observed for the candle, another light-emitting object that, unlike the flashlight, clearly is not in the extension of *electronic device*. The candle was virtually never selected, irrespective of goal condition.

Some canonical electronic devices exhibted substantial modulation across goal conditions. For example, the boombox was selected most frequently in the goal-neutral condition and in the condition where participants were told that "[t]he managers of a theater are concerned that certain objects, when brought into the theater, **create noise** that could distract audience members and performers" (emph. added). However, other clear category members (e.g., the tablet) exhibited very little modulation across goal conditions.

Fig. 2 shows by-object mean responses from the 'category-membership' norming study against by-object mean (left panel) and standard deviation (SD, center panel) of selection rates from the main experiment (across the four goal conditions in which each object appeared). As the

[11] For details of the category membership and feature attribution norming studies, see `https://osf.io/hu8v7`.
[12] To calculate these mean scores, we transformed slider scale responses to values on the interval [0,1].

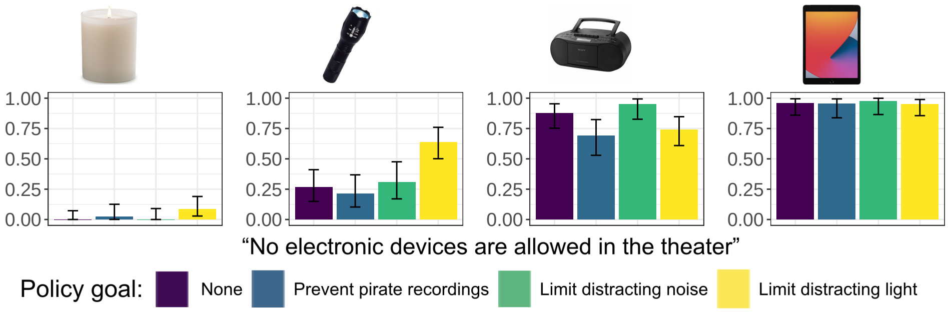


Figure 1: Mean selection rates for 4 of the 12 objects seen in the *No electronic devices...* scenario. From left to right: 1 clear non-member (candle), 1 edge case (flashlight), 2 clear members (boombox, tablet). Error bars denote 95% binomial confidence intervals.

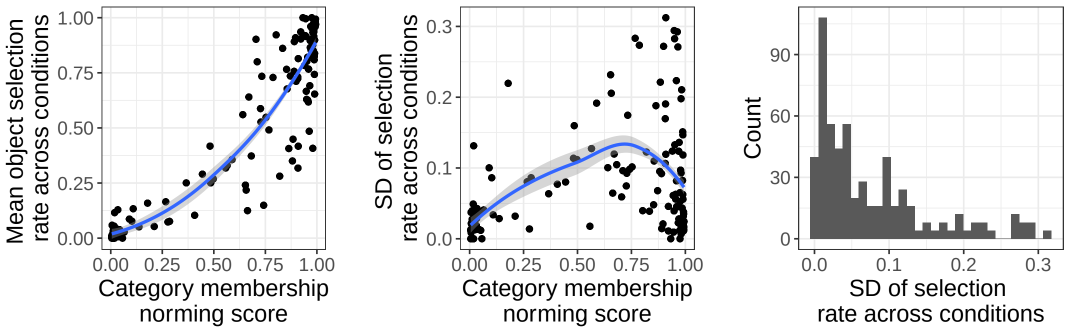


Figure 2: Left: mean object selection rate plotted against mean object category membership score (as measured in the category membership norming study). Center: between-condition standard deviation (SD) of object selection rates plotted against mean object category membership score. Right: distribution of by-object selection rate variability (as measured by SD).

left panel indicates, category membership scores and selection rates are clearly positively correlated. We conducted a Bayesian mixed effects logistic regression analysis using the `brms` package in R, predicting log odds of object selection from a fixed effect of category membership score, with by-participant and by-item (i.e., by-rule) random intercepts as well as by-item random slopes for category membership score.[13] We found evidence of a positive relationship between category membership score and object selection ($\beta = 6.65$, 95% CrI = [5.82, 7.51]).

The smoothing line fit to the center-panel plot has a weakly inverse-parabolic shape, suggesting that edge-case objects exhibited greater variability in selection rate across the four goal conditions relative to the clear category members and non-members. However, there are exceptions: some of the objects that received high scores on the category-membership norming study also exhibited substantial selection variability across conditions. We conducted a Bayesian mixed effects polynomial regression analysis predicting object selection variability (as measured by SD) from fixed effects of category membership score and category membership score squared, with by-item random intercepts. We found strong evidence of an inverse quadratic relationship between category membership and selection variability ($\beta = -0.38$, 95% CrI = [-0.43, -0.32]) as well as strong evidence of a positive linear relationship ($\beta = 0.44$, 95% CrI = [0.38, 0.49]).

Finally, note that the distribution of variabilities in object selection rate is heavily left-skewed (right panel of Fig. 2), with most objects ($\sim$72%) exhibiting a between-condition selection rate

[13] The maximal random effects structure justified by the experiment design that allowed the model to converge.

SD of <0.1. We return to this point in Section 4.2 when modeling our experimental data.

For now, we claim that the results of the experiment suggest the need for an analysis that accounts for the variability we observed among conditions that highlighted different relevant dimensions of interpretation. We will pursue an account that explains these patterns as variability in how category boundaries of artifact nouns are resolved in context. Accordingly, this account will rely on a semantics for artifact nouns whereby the extensions of such nouns are a function of multiple interpretive dimensions which may vary in importance across contexts. We will then demonstrate that, when coupled with a Bayesian probabilistic model of pragmatic interpretation, this semantics can predict the interpretive variability observed in the experiment.

4. Analysis

4.1. Semantic proposal

Our analysis extends a proposal from Sassoon and Fadlon (2017), who identify artifact nouns as a subset of a broader category of nominal *social predicates* whose denotations are degree functions that are multi-dimensional, weighted, and additive. A denotation for *vehicle* in the spirit of their analysis is provided in (2):

(2) $[\![\text{vehicle}]\!]_{<d,<e,t>>} = \lambda d \lambda x. \big(w_{F_1} f_{F_1}(x) + \cdots + w_{F_N} f_{F_N}(x)\big) > d$

On Sassoon and Fadlon's analysis, artifact nouns are multi-dimensional in the sense that their denotations reference a set of measure functions $f_{F_1} + \cdots + f_{F_N}$, where each function measures entities along a distinct dimension. These measurements are weighted by scalars $w_{F_1}, \ldots, w_{F_N}$, and the truth conditions of the predicate depend upon a weighted composite of the entity's measurements along the relevant dimensions. Following prior cognitive psychology research on concepts and categorization (Rosch 1973; Hampton 1998; Murphy 2002; Hampton et al. 2009), this composite is taken to be additive in nature; natural kinds, by comparison, are analyzed as multiplicative concepts as in (3), whereby failure to exhibit any single relevant dimension significantly penalizes an entity vis à vis category membership.

(3) $[\![\text{duck}]\!]_{<d,<e,t>>} = \lambda d \lambda x. \big(w_{F_J} f_{F_J}(x) * \cdots * w_{F_K} f_{F_K}(x)\big) > d$

[adapted from Sassoon and Fadlon 2017:29]

Sassoon and Fadlon's proposal is intended to account for their finding that artifact nouns are of intermediate acceptability in dimensional constructions (4) and degree constructions (5), in contrast to multi-dimensional adjectives (acceptable) and natural kind nouns (degraded):

(4) **Dimensional constructions (quantifying over dimensions):**

a. # This tree is a pine in {some, most, every} respect(s). *natural kind*
b. ? This place is a church in {some, most, every} respect(s). *artifact*
c. This tree/place is safe in {some, most, every} respect(s). *multidim. adjective*

[adapted from Sassoon and Fadlon 2017:18]

(5) **Degree constructions (comparatives):**

a. # This tree is more a pine than {that one, an oak}. *natural kind*
b. ? This place is more a church than {that one, an art gallery}. *artifact*
c. This tree/place is more safe than {that one, dangerous}. *multidim. adjective*

[adapted from Sassoon and Fadlon 2017:21]

Sassoon and Fadlon's theory links the acceptability of a predicate in dimensional and degree environments to that predicate's 'dimension accessibility,' i.e. the "tendency to base exemplariness judgments and classification…on dimension-counting" (2017: 2). multi-dimensional gradable adjectives such as *healthy* strongly exhibit this property because their scales correspond to "the number of dimensions in which [the entity] is within the norm" (2017:28). For example, *Mary is healthier than Bill* is true iff Mary meets or exceeds the standards for a greater number of 'health'-relevant dimensions than does Bill.[14]

Additive multi-dimensional nouns (such as artifacts) exhibit dimension accessibility, though more weakly than do multi-dimensional adjectives: "upon a shift…to a flattened representation of dimensions (with binary scales consisting of 0 and 1 and equal weights), [additive nouns] denote dimension-counting relations" (2017: 29). (Such a shift is less available for multiplicative nouns such as natural kinds; thus, natural kinds exhibit less dimension accessibility).

The analysis in (2) – which leaves open whether features and weights are lexicalized or supplied by context – is consistent with a theory whereby the interpretation of artifact nouns is context sensitive in simple predicative sentences (*X is a vehicle*) as well as in the sorts of quantificational environments featured in the rules in the experiment reported above (e.g., *No vehicles are allowed…*). For purposes of exposition, we decompose (2) into a measure function (following Kennedy 2007) that combines with a silent *pos* morpheme to yield a context-sensitive $< e,t >$ function whose interpretation depends on a contextual standard of comparison:[15]

(6) $$[\lambda x.[w_{F_1} f_{F_1}(x) + \cdots + w_{F_N} f_{F_N}(x) \succeq \mathbf{s}(\text{vehicle})]]_{<e,t>}$$

$${}_{deg}pos_{<<e,d>,<e,t>>} \quad \lambda g \lambda x.g(x) \succeq \mathbf{s}(g)$$

$$\text{vehicle}_{<e,d>} \quad \lambda x.w_{F_1} f_{F_1}(x) + \cdots + w_{F_N} f_{F_N}(x)$$

However, there are at least two reasons to believe that the nature of *vehicle*'s context dependence cannot be adequately accounted for using a single context-sensitive threshold. First, we observe that a single policy goal could lead to increased selection rates for some objects relative to the goal-neutral baseline condition but decreased rates for others. To illustrate the problem this poses, call the threshold of inclusion for the category of *electronic device* in the goal-neutral baseline condition θ_B. When alerted to the goal of limiting distracting light, participants in our experiment were more inclined to select the flashlight relative to baseline (suggesting a threshold lower than θ_B in that condition); however, participants were also less inclined to select objects such as the boombox in that condition (suggesting a threshold higher than θ_B). This discrepancy suggests that the value of some additional parameter shifts between contexts.

A second reason emerges from a closer examination of the comparative constructions in (5b), cf. the single-dimensional comparatives in (7):

(7) This tree is taller (more tall) than {that one, it is wide}.

The degree analysis of Kennedy (2007), inter alia, correctly predicts that the comparatives in (7) are not context-sensitive, in the sense that the sentences are readily evaluated for truth or falsity. The lack of context dependence is a consequence of the semantics of the comparative, which

[14]This analysis finds independent empirical support in the work of Sassoon (2013). However, we acknowledge that comparatives involving multi-dimensional degree constructions constitute a complicated domain. In the future, we plan to experimentally investigate the role of context in the interpretation of multi-dimensional comparatives.
[15]Following Kennedy (2007), **s** is context sensitive and maps measure functions to comparison standards.

orders the two degrees returned by applying the sentence's constituent measure function(s) (e.g., *tall*, *wide*) to its constituent entities (e.g., *this tree*, *that one*).

However, an analogous analysis yields the wrong results for (5b) (as well as for comparisons involving multi-dimensional adjectives such as *healthy* / *optimistic*), whose interpretations are context dependent despite the presence of the comparative. Intuitively, before evaluating (5b), the interpreter must determine which dimensions of *church* / *art gallery* are relevant for interpretation, and/or the relative importance of those dimensions for the purposes of the context. That is, context sensitivity persists in artifact noun comparatives, in a manner unobserved in the case of single-dimensional gradable adjectives.

An analysis that accounts for this property of artifact nouns is offered in (8):

(8) $[\![\text{vehicle}]\!]_{<e,d>} = \lambda x. \sum_{f \in \mathbf{F}(\text{vehicle})} f(x) * \mathbf{W}(\text{vehicle}, f)$

…where **F** is a context-sensitive function that maps $< e,d >$-type measure functions to sets of $< e,d >$-type measure functions. For example, **F**(*vehicle*) might return a set of contextual measure functions including a function f_P which measures entities' propensity to pollute the air.[16] **W** is a context-sensitive function that maps ordered pairs of $< e,d >$-type functions to scalars, which weight the importance of contextual measure functions in interpretation (e.g., the importance of pollution for interpreting *vehicle*). Thus, when applying (8) to an entity `car`, one summation term will be $f_P(\texttt{car}) * W(\textit{vehicle}, f_P)$ if f_P is a relevant measure function.[17]

As (9) demonstrates, we predict that in environments where *vehicle* composes with *pos*, the interpretation depends on valuing three parameters: **F**, **W**, and **s**. For example, in quantificational environments of the sort observed by participants in the experiment, we posit that the measure function composes with *pos* to form the restrictor argument of the quantifier; thus, interpretation hinges on the valuation of all three contextual parameters.[18] Finally, we posit that the **F** and **W** parameters project in the case of the comparative construction as in (11).

(9) $_{deg}POS(\text{vehicle}) = \left[\lambda x.[\sum_{f \in \mathbf{F}(\text{vehicle})} f(x) * \mathbf{W}(\text{vehicle}, f)] \succeq \mathbf{s}(\text{vehicle})\right]_{<e,t>}$

(10) a. Every vehicle is prohibited from the park.

b. $\forall x \left[[\sum_{f \in \mathbf{F}(\text{vehicle})} f(x) * \mathbf{W}(\text{vehicle}, f)] \succeq \mathbf{s}(\text{vehicle}) \Rightarrow \text{prohibited-from-the-park}(x)\right]$

(11) a. ? [This object is] more (of) a vehicle [than that one].[19]

b. $\lambda y \lambda x. \left[\sum_{f \in \mathbf{F}(\text{vehicle})} f(x) * \mathbf{W}(\text{vehicle}, f)\right] \succ \left[\sum_{f \in \mathbf{F}(\text{vehicle})} f(y) * \mathbf{W}(\text{vehicle}, f)\right]$

$more_{<<e,d>,<e,<e,t>>>}$ $\lambda g \lambda y \lambda x. g(x) \succ g(y)$ $\text{vehicle}_{<e,d>}$ (8)

[16]We remain as noncommittal as possible regarding the precise sub-lexical dimensions that factor into multi-dimensional interpretation. However, for the purposes of fitting an interpretive model to our data, we operationalize these dimensions using data from the feature attribution norming study, as detailed in Section 4.2.

[17]Unlike (6), which posits that the function **s** that provides the comparison standard is the only context-sensitive parameter relevant to artifact noun interpretation, (8) straightforwardly accounts for the observation that a single policy goal cue can lead to increased selection rates for some objects relative to the goal-neutral baseline condition but decreased rates for others. (Not all objects will exhibit contextually-relevant properties to the same degree).

[18]We revisit the consequences of this analytic move in Section 6.

[19]We implicitly assume here that *of* and *a* both denote $<< e,d >,< e,d >>$-type identity functions.

4.2. Interpretive model

Following Lassiter and Goodman (2017), Qing and Franke (2014), and Tessler and Goodman (2019), inter alia, we posit that listeners who observe semantically underspecified expressions jointly infer speaker-intended meanings along with the values of the contextual parameters upon which interpretation depends (**F**, which returns the set of contextually-relevant measure functions; **W**, which returns the relative weight assigned to each such function; **s**, which returns the inclusion threshold for the artifact noun category). In line with this work, we take probability theory to be a natural formalism for linguistic inference under persistent uncertainty, which we model as an instance of Bayesian belief update (Frank and Goodman 2012, et seq):

(12)
$$\underbrace{L(o \text{ is prohibited}, \mathbf{F}, \mathbf{W}, \mathbf{s} | \text{“No electronic devices…”})}_{\substack{\text{Joint posterior probability that an object } o \text{ is prohibited and} \\ \text{that context parameter values = } \mathbf{s, F, W} \text{ (given observation of rule)}}} \propto$$
$$\underbrace{{}_{deg}pos^{\mathbf{s}}([\![\text{elec. device}]\!]^{\mathbf{F},\mathbf{W}})(o)}_{\substack{= 1 \text{ if } o \text{ meets standards of artifact noun} \\ \text{category targeted by the rule, given} \\ \text{parameter valuation } \mathbf{s, F, W} \text{ (= 0 otherwise)}}} * \underbrace{P(o \text{ is prohibited})}_{\substack{\text{Prior belief that} \\ o \text{ prohibited}}} * \underbrace{P(\mathbf{s}) * P(\mathbf{F}) * P(\mathbf{W})}_{\substack{\text{Prior probability of} \\ \text{parameter values}}}$$

To evaluate the Goal Sensitive hypothesis against its null competitor, we will consider whether there is evidence that participants in our experiment assigned additional importance to interpretational dimensions that the policy goal manipulation made salient in the experiment. To answer this question, we fit the model in (12) to our experimental data, with a number of simplifying assumptions. First, we assume that the prior over the prohibition status of any object o is uniform such that $P(o$ is prohibited$)$ = $P(o$ is not prohibited$)$ = 0.5. Next, we assume a prior over values of **s** such that for any artifact noun predicate p, the probability that $\mathbf{s}(p) = i$ is equal to the probability that $\mathbf{s}(p) = j$ for all i, j on the interval [0,1].[20]

Finally, we assume that for each condition of the experiment, there is no a priori interpreter uncertainty regarding how to parameterize the functions **F** and **W**, as (13) exemplifies:[21]

(13) **Parameter values (*No electronic devices* scene, goal = “Limit distracting light”):**
$$\mathbf{F}(\text{elec. device}) = \left[\texttt{cat}^{\textit{elec.device}},\ \texttt{emit-light}\right]$$
$$\mathbf{W}(\text{elec. device},\ \texttt{cat}^{\textit{elec.device}}) = \gamma \qquad \mathbf{W}(\text{elec. device},\ \texttt{emit-light}) = 1 - \gamma$$

In conditions where participants are exposed to an explicit policy goal (e.g., the goal of limiting distracting light from the theater in the *No electronic devices…* scene), **F** maps the measure function denoted by the artifact noun predicate to two component measure functions: one that measures the extent to which entities in the domain possess the feature relevant to the mentioned goal (e.g., `emit-light`, which maps entities to a continuous scale that tracks the extent to which entities emit light), parameterized by the **feature attribution** norming study described in Section 4.1.1; and another measure function $\texttt{cat}^{\textit{artifact-noun}}$ that measures the extent to which entities are believed to possess the quality of being a member of the relevant artifact noun category independent of local interpretive context. (This second measure function is parameterized

[20]This amounts to a uniform prior over standards of comparison.

[21]A consequence of these simplifying assumptions is that our implementation of (12) is not, in fact, a joint inference model insofar as no update occurs for the **F** and **W** variables and, moreover, **s** value probabilities are uniform on both the prior and posterior distribution. In our implementation, only beliefs regarding the meaning space (the prohibition/requirement status of objects) are updated; however, (12) demonstrates how the model could be implemented under more realistic assumptions about interpreter uncertainty regarding context parameter values.

by the **category membership** norming study, also described in Section 3.1).

The status of the `cat` measure function warrants additional discussion. Though it is a primitive of our implementation of the interpretive model in (12), it is best understood in our framework as a proxy for a weighted composite of several atomic dimensions relevant to the interpretation of the artifact nouns under investigation. In particular, we posit that `cat` approximates those dimensions (and their relative importance) in a context where there is no rule-issuing authority and hence no policy objectives that might influence the evaluation of artifact noun categories.

With these preliminaries in mind, we can employ the modeling framework in (12) to compare the two hypotheses which motivated the experiment. On the Goal Insensitive hypothesis, we expect that observation of an explicit policy goal should not systematically modulate interpreter beliefs regarding the dimensions that are relevant to the evaluation of nouns such as *vehicle* or *electronic device*. In other words, the dimensions that interpreters believe to be relevant – and the relative importance of those dimensions – should be faithfully approximated by `cat`. On the other hand, on the Goal Sensitive hypothesis, we expect that the dimensions that reflect policy objectives (e.g., `emit-light`) are considered additionally relevant in contexts where overt policy objectives are observed.

In our framework, these hypotheses reflect possible valuations of **W**, the function that returns the weight assigned to dimensions of an artifact noun in interpretation. For contexts in which a policy objective is directly observed, as in (13), a free parameter $\gamma \in [0,1]$ weights the relative importance of the composite `cat` dimension. A γ value of 1 reflects a parameterization of **W** that exclusively identifies `cat` as relevant, consistent with the Goal Insensitive hypothesis; γ values less than 1 reflect parameterizations consistent with the Goal Sensitive hypothesis.

γ plays a similar role when modeling interpretation in the 'goal-neutral' condition, as in (14):

(14) **Parameter values (*No electronic devices* scene, goal-neutral condition):**

$$\mathbf{F}(\text{elec. device}) = \left[\texttt{cat}^{elec.device},\ \texttt{emit-light},\ \texttt{emit-noise},\ \texttt{can-record}\right]$$
$$\mathbf{W}(\text{elec. device},\ \texttt{cat}^{elec.device}) = \gamma$$
$$\mathbf{W}(\text{elec. device},\ \texttt{emit-light}) = p(\text{“Limit distracting light”}) * (1-\gamma)$$
$$\mathbf{W}(\text{elec. device},\ \texttt{emit-noise}) = p(\text{“Limit distracting noise”}) * (1-\gamma)$$
$$\mathbf{W}(\text{elec. device},\ \texttt{can-record}) = p(\text{“Prevent pirate recordings”}) * (1-\gamma)$$

In the 'goal-neutral' condition, we posit that interpreters consider `cat` alongside each goal-relevant dimension according to the a priori plausibility of the dimensions' corresponding policy objectives. The plausibility function p maps goals to probabilities such that, e.g., p(“Limit distracting light”)+p(“Limit distracting noise”)+p(“Prevent pirate recordings”) = 1.[22]

4.3. Parameter estimation and model assessment

With the parameter assumptions outlined above, there remains one free parameter: γ. We estimate the value of γ using Bayesian Data Analysis (BDA). In BDA, a posterior probability distribution $P(M|D)$ over models M, given observed experimental data D, is inferred using Bayes' Rule, as shown in part (a) of (15). For our purposes, the space of possible models includes the different possible parameterizations of γ for the model described in (12), plus fixed parameter value assumptions for the other parameters. This is shown in part (b) of (15).

[22] These values were estimated from the norming study in which policy goals were normed for plausibility.

(15) a. $P(M|D) = P(D|M) * P(M)$ b. $P(\gamma = i|D) = P(D|\gamma = i) * P(\gamma = i)$

c. $$P(\gamma = i|D) \propto \forall o \big[P(\underbrace{\text{selection freq. of } o}_{\substack{\text{\# of times } o \text{ selected in} \\ \text{experimental condition } C.}} | \underbrace{B(n, \underbrace{L^{\gamma=i}(o \text{ prohibited}|\text{rule})}_{\substack{\text{Marginal } L \text{ posterior probability of} \\ o\text{'s prohibition in } C, \text{ with } \gamma \text{ set to } i.}})}_{\text{Binomial distribution, where } n = \text{ the number of participants assigned to } C.} \big] * P(\gamma = i)$$

Our procedure for computing a posterior distribution over γ values is more concretely formalized in part (c) of (15). A value for γ is sampled from a prior distribution, which we take to be uniform on the interval [0,1]. This allows us to compute, for each object in each associated condition, the *L* posterior probability that the object is prohibited (required) in that condition. We assume that the probability of any participant actually selecting an object in the experiment is equal to this probability. Since we can assume that the behavior of any individual participant is independent of the behavior of the others, we link *L*'s output to a binomial distribution, where the number of observations is equal to the number of participants who viewed an object in a particular condition and the distribution over 'successes' can be thought of as a distribution over predicted selection frequencies for that object in that condition. Some γ values lead to binomial distributions on which the observed selection frequency is improbable; thus, these values are considered unlikely a posteriori. Other values yield distributions on which the observed selection frequency is not surprising; therefore, these values are relatively likely a posteriori.

To efficiently compute a posterior distribution over γ values, we sample from the prior using an MCMC sampling procedure implemented in WebPPL (Goodman and Stuhlmüller, 2014). As the left panel of Fig. 3 shows, the γ value of 1 is outside the 95% credible interval of the posterior distribution (maximum likelihood estimate = 0.758, 95% CrI = [0.756, 0.758])[23], indicating strong evidence for $\gamma < 1$ and hence strong evidence for the Goal Sensitive hypothesis.

The Goal Sensitive hypothesis is further supported by comparing the predictions of the model given the maximum-likelihood γ value against those of the model when γ is set to 1. The former γ value yields a model with slightly higher predictive accuracy of individual object selection rates across experimental conditions than does the latter (correlation of predictions and observations r = 0.919 vs. r = 0.896, respectively). However, as mentioned in Section 3.2, objects differed in the extent to which they exhibited between-condition variation in their selection rates, with many objects exhibiting little variation. Therefore, in a post-hoc analysis, we investigated whether higher degrees of variation were explained by higher degrees of goal sensitivity as measured by γ. To explore this, we used the procedure in part (c) of (15) to separately fit the model to each object seen in the experiment. The individual γ estimates for each object are plotted against objects' between-condition selection variability in the center cell of Fig. 3. As evident from the figure, γ estimates and selection variability are inversely correlated, suggesting that the selection behavior for high-variability objects is best modeled by models that posit high degrees of goal sensitivity (as reflected by relatively low γ values).[24]

Unlike the Goal Insensitive model, the Goal Sensitive model makes different predictions across

[23]The point estimate was obtained by computing a Gaussian kernel density estimate using the `density` command in R, and the credible intervals were estimated via bootstrap sampling from the computed density function.

[24]We further investigated this relationship with a Bayesian mixed effects regression model predicting object-level γ point estimates from a fixed effect of selection variability (SD), with by-item random intercepts. We find strong evidence of a negative relationship between variability and γ estimates ($\beta = -2.41$, 95% CrI = [-2.82,-2.00]).

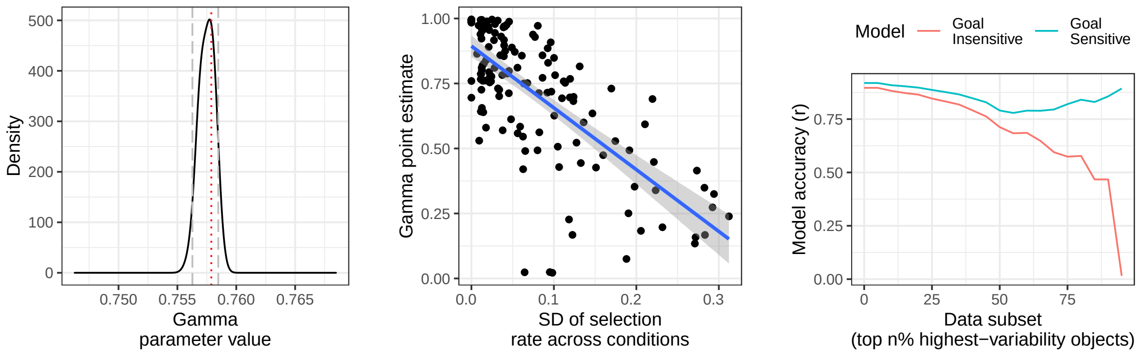


Figure 3: Left: distribution of posterior samples computed via the procedure in part (c) of (15). Dotted line indicates the maximum likelihood estimate; dashed lines indicate the bootstrapped 95% credible interval. Center: object-level γ estimates plotted against object-level variability (SD) in selection rates across conditions. Right: accuracy of the Goal Sensitive ($\gamma < 1$) and Goal Insensitive ($\gamma = 1$) models on progressively higher-variability subsets of the data.

goal conditions; however, we sought to determine whether these predictions bore any systematic relationship to the variation we actually observed in the objects that were sensitive to the goal condition manipulation. Therefore, in a second post-hoc analysis, we investigated whether the predictive advantage of the Goal Sensitive model over the Goal Insensitive model varies as a function of the between-condition variability exhibited by the experimental data against which the models are evaluated. To test this, we used the procedure in part (c) of (15) to fit the model to the top $n\%$ highest-variability objects in the overall dataset, where $n = [95, \ldots, 5, 0]$.

For each data subset, applying the procedure in part (c) of (15) resulted in a maximum likelihood estimate of the γ parameter, which was then used to parameterize the model and predict response patterns for that subset. As the righthand cell of Fig. 3 shows, the accuracy of the Goal Insensitive model is degraded when restricting to data subsets with progressively higher between-condition selection variation. Compare this result to that of the Goal Sensitive model, which does not exhibit the same degradation even when focusing on very high-variance data.

5. Contra alternative accounts

We demonstrated that a model of pragmatic interpretation grounded in a context-sensitive, multi-dimensional semantics for artifact nouns can accurately predict the data observed in the experiment. However, one might still wonder how we can be sure that goal information actually modulated beliefs about the category boundaries of artifact nouns. On an 'imprecision'-based account, policy goals are orthogonal to how the extension of an artifact noun category is determined; rather, beliefs about such goals signal how 'loosely' to interpret artifact nouns. That is, policy goals serve to modulate the 'pragmatic slack' associated with such nouns, where 'slack' measures distance from truth tolerated in context (following Lasersohn, 1999).

On this account, we expect that for each artifact noun category examined in the experiment, a context-invariant 'core' set of objects should always be treated as within the category for the purposes of interpretation; policy goals, however, may selectively allow for objects strictly outside the category to be treated as 'close enough' to the category for the purposes of interpretation. A conceptual challenge for this account is to articulate the conditions of category membership such that (for example) cell phones are *electronic devices* but flashlights are not.

This account also faces an empirical challenge: many objects that were judged to be category members in norming exhibited relatively low selection rates in certain goal conditions.[25] (The boombox, for example, received an *electronic device* category membership score close to ceiling; comparable to the scores of a camcorder and smart watch). To explain this pattern, the imprecision account is committed to the claim that such objects are, strictly speaking, outside the category but merely treated as such under certain conditions.

On a domain restriction (DR) analysis, policy goal information merely suggests that the rule is limited to a certain restricted domain of entities (but the extension of, e.g., *electronic device* is invariant across contexts). DR almost certainly plays a part in rule interpretation. (For example, had a pacemaker been shown to participants, it is plausible that many participants would have considered it an *electronic device* but an exception to the *No electronic devices ...* rule).

However, an analysis that focuses entirely on DR is in tension with the observation that the decision-making contexts tested in the experiment are susceptible to meta-linguistic commentary as to what 'counts as' a member of the category denoted by the noun as in (16a), similar to contexts where thresholds of comparison for gradable adjectives may be explicitly negotiated as in (16b). This suggests that in (16a) and (16b), resolving uncertainty as to what 'counts as' an *N* is relevant for satisfying the speaker's wishes. The first sentence in (16c), however, intuitively involves DR to a discourse-salient set of bottles (ones bought for the party), but what 'counts as' a bottle of Heineken is not up for negotiation (nor is it negotiated via the DR):

(16) a. No electronic devices are allowed in the theater. (By the way: for our purposes, a flashlight counts as an electronic device.)
b. Get me a long ladder. (By the way: for our purposes, 20 feet counts as long for a ladder.)
c. (Planning a party): Put all the bottles of Heineken in the fridge. (# By the way: for our purposes, nothing that my neighbors bought for their party counts as a bottle of Heineken.)

We conclude that when it comes to the interpretation of artifact nouns in rule contexts, goal sensitivity reflects (at least in part) context-sensitive resolution of category boundaries.

6. Discussion and conclusion

The results of our experiment support the claim that highlighting the functions of artifact noun objects can serve as a cue for how interpreters resolve artifact noun category boundaries in context. This observed context sensitivity motivated a semantic proposal whereby the dimensions invoked by an artifact noun – as well as the relative importance of those dimensions – can vary across contexts. This proposal was, in turn, supported by a computational pragmatic model of interpretation that accurately predicts the experimental data by virtue of being grounded in an explicitly multi-dimensional and context-sensitive truth-conditional semantics.

One might object that the semantics we have offered is too context sensitive, in that it suggests that artifact nouns lexicalize no minimal necessary conditions for category membership. Compare our proposal in (8) to that of Grimm and Levin (2017), whose analysis integrates an abilitative modal semantics for artifact nouns. On their analysis, a nominal predicate is true of

[25] This is reflected in the upper-right quadrant of the center panel of Figure 2 and is also supported by the results of the polynomial regression analysis, which suggests that selection variance increases linearly with category membership score in addition to exhibiting an inverse parabolic relationship with that variable.

an entity iff in a maximally-close accessible world, the entity participates in a 'minimal' event e_{min} that exhibits lexically-specified properties (e.g., being used for transport):

(17) a. $[\![\text{ABLE}(P(x))]\!]^{w,g,h_x,j} = 1$ iff $\exists w' \in W$ s.t.: (i) w' is accessible from w given h_x; (ii) w' is maximally close to the ideal established by $j(w)$, and (iii) $< w', d > \in [\![P]\!]$
b. $[\![\text{vehicle}]\!] := \lambda y[\text{ABLE}[x \text{ uses } y \text{ for transport in } e_{min}]]^{w,g,h_y,j}$
c. $[\![\text{car}]\!] := \lambda y[\text{ABLE}[x \text{ uses } y \text{ for transport in } e_{min} \wedge \text{has-four-wheels}(y)]]^{w,g,h_y,j}$
[Grimm and Levin (2017): pp. 4 & 7]

Grimm and Levin's (2017) analysis can straightforwardly capture entailment relations between artifact noun superordinates (e.g., *vehicle*) and basic-level expressions (e.g., *car*). On the proposal in (8), capturing these patterns requires stipulating constraints on how free parameters are valued for superordinate artifact nouns and their basic-level counterparts.

However, the assumption of total context sensitivity can be revised by imposing lexical restrictions on multi-dimensional interpretation. An option we will briefly consider privileges a primary class of essential, lexicalized dimensional properties (E_1 through E_n) over a context-dependent class of properties (in **F**) whose relevance is free to vary across contexts.

(18) $[\![\text{vehicle}]\!]^{\mathbf{F,W}}_{<e,d>} = \lambda x. \sum_{f \in \mathbf{F}(\text{vehicle})} E_1(x) * \cdots * E_N(x) * f(x) * \mathbf{W}(\text{vehicle}, f)$
... where E_1 through E_N are essential properties of vehicle-hood.

Because the denotation in (18) includes an additive measure function – similar to (2) and (8) – this analysis predicts that artifact nouns will still exhibit dimension accessibility of the sort that Sassoon and Fadlon (2017) argue explains the intermediate acceptability of such nouns in dimensional and degree constructions. However, 'essential' properties compose in a multiplicative manner with context-sensitive properties, such that failure to exhibit any single essential property significantly penalizes the composite measure value.

What are an artifact noun's 'essential' properties? Following Grimm and Levin (2017), such properties might in some cases include function-based potentialities (e.g., 'able to be used for transport'), which those authors represent with a semantics that invokes an 'associated event.' Such an event may, for some nouns, reflect a lexically-specified purpose (the 'telic' role of Pustejovsky 1995, see also Nichols 2008); in other cases, such as *cake* and *statue*,"the event of creation or manufacture... may be as important in human interaction with a particular artifact as the event of use" (Levin et al. 2019: 434, see also the 'agentive' role of Pustejovsky 1995).

A second question for a (18)-type analysis regards entailment relations between superordinate and basic-level categories. One option would be to stipulate that superordinate category membership is an essential property of corresponding basic-level category membership (e.g., to be a *car*, something must also be a *vehicle*). Another option is to say that essential properties of a basic-level category are a superset of the essential properties of the superordinate category:

(19) $[\![\text{car}]\!]^{\mathbf{F,W}}_{<e,d>} = \lambda x. \sum_{f \in \mathbf{F}(\text{car})} E_1(x) * \cdots * E_N(x) * \cdots * E_M(x) * f(x) * \mathbf{W}(\text{car}, f)$
... where E_1 through E_N are essential properties of vehicle-hood.
... E_{N+1} through E_M are essential properties of car-hood.

However, unlike Grimm and Levin's (2017) analysis, this analysis falls short in that if X is a car, it is merely likely true that X is also a vehicle by virtue of exhibiting the properties that are essential to the latter category. To rectify this, further constraints on the valuation of the basic-level term's free parameters could be imposed in order to guarantee that the basic-level term asymmetrically entails the superordinate term. Such constraints could be thought of as meaning postulates or merely as descriptive generalizations of how such terms are interpreted in context. (For example, contexts where 'pollution' is relevant for interpretation of *car* may tend to also be contexts where it is relevant for interpretation of *vehicle*).

This project also explored an underappreciated dialectical relationship between formal semantics/pragmatics, computational linguistics, and the law. In search of an empirical domain in which to systematically investigate the context dependence of artifact noun interpretation, we identified a systematic relationship between context – in particular, the policy objectives of lawmakers – and the production of context-sensitive natural language. Inspired by a normative debate in jurisprudence regarding the role of such context in legal interpretation, we conducted an experiment to test how laypeople decide the meaning of legal rules. Our findings appear to align more closely with the prescriptions of purposivists (for whom policy goals are often germane to legal interpretation) than with those of textualists, who tend to de-emphasize cues to legislative purpose that lie beyond the letter of the law.[26]

Should the way(s) in which laypeople read the law inform how lawyers and judges read the law? As linguists, we will remain noncommittal. However, our results and analysis speak to the purposivism/textualism debate in a more nuanced manner. To see how, consider how U.S. Supreme Court Justice Antonin Scalia (a textualist) derides the legacy of *Church of the Holy Trinity vs. US*, a case which Scalia sarcastically laments "is cited to us whenever counsel wants us to ignore the narrow, deadening text of the statute, and pay attention to the life-giving legislative intent. It is nothing but an invitation to judicial lawmaking" (Scalia 1997: 49).

Scalia's quote presupposes a dichotomy between textual meaning and the policy objectives of the text's drafters: his concern as a textualist is that judges 'override' textual meaning by giving consideration to the aims that legislators sought to achieve. However, if our analysis is on the right track (namely, that artifact nouns are semantically underspecified and that policy objectives serve as a cue for how to resolve semantic underspecification in rule interpretation contexts), then artifact nouns may be a case in which a dichotomy between 'text' (along with its conventional meaning) and 'purpose' lacks coherence.[27] If so, then the jurisprudential debate will at a minimum have to be reconceived to accommodate these linguistic facts.

A. Rules and goals tested in the experiment

No electronic devices are allowed in the theater. (The managers of a theater are concerned that . . .)
1: certain objects, when brought into the theater, create noise that could distract audience members and performers.
2: certain objects, when brought into the theater, emit light that could distract audience members and performers.
3: audience members might try to record performances and distribute pirate recordings online.
Festival attendees must wear a face covering. (The organizers of an outdoor street festival. . .)
1: desire that attendees appear in costume.
2: are concerned about the possibility that airborne illnesses might spread among attendees.
3: are concerned that attendees may develop frostbite due to cold weather.
No outside food allowed in the garden. (The managers of a museum are concerned. . .)

[26]See also Grinsell (2017) for a linguistic argument in support of considering policy context in the interpretation of multi-dimensional expressions (especially adjectives such as *reasonable*) as they appear in the law.

[27]This conclusion is reminiscent of Carston (2013), who argues that legislative purpose may be part of "the relevant contextual assumptions [that] are available to all ordinary mentally able citizens and [that] would be immediately activated by the text itself" (2013: 30-31). Even if interpreters lack access to explicit evidence of policy goals, legal text may implicitly articulate such goals, which in turn serve as cues for how to resolve linguistic uncertainty.

1: that certain objects, when brought into the museum's garden area, will attract nuisance insects. *2:* about visitors littering in the museum's garden area. *3:* that the restaurant located in the museum's garden area is losing business to outside competition.
Furniture brought to the waste management center can be disposed of for free. (The operators of a waste management center typically charge a fee for processing drop-offs. However, . . .) *1:* they recently received a request from a local homeless shelter for objects that can be used as storage, that can be comfortably sat/slept upon, or otherwise could function as furniture for those in need. *2:* they recently devised a plan to resell viable secondhand objects at auction, rather than disposing of the objects in a landfill. *3:* they recently devised a plan to break down objects into valuable raw materials such as metal and lumber, rather than disposing of the objects in a landfill.
Students are prohibited from wearing jewelry at school. (The administrators of a public high school . . .) *1:* are concerned that certain objects, when worn, pose a safety risk to students. Loose, dangling objects in particular could get caught on exercise equipment in gym class, on machinery in the school's wood shop class, or on certain appliances in the school's home economics class. *2:* are concerned about theft from lockers and common areas. In response to a recent uptick in such thefts, they seek to discourage students from bringing objects of value to school. *3:* are concerned that certain students tend to dress ostentatiously, which undermines a sense of equality among students and distracts from learning in the classroom.
No outside bags are allowed in the convenience store. (The managers of a convenience store are concerned. . .) *1:* about patrons contaminating food items sold at the store. It is especially important to the managers to keep food free of germs and potential allergens. *2:* about shoplifting. *3:* that certain objects, when brought into the store, are too large for patrons to safely carry without potentially knocking into shelves and creating messes in the aisles.
No pets are allowed in the apartment building. (The managers of an apartment building. . .) *1:* are concerned that airborne allergens in the apartments will make the units less appealing to prospective tenants in the future. *2:* have received numerous noise complaints from residents. *3:* want to protect furniture, walls, and carpeting in units and common areas from potential damage.
The restaurant is prohibited from selling sandwiches. (The owners of a strip mall have recently leased one of their storefronts to a new restaurant. However, they are concerned that the restaurant's business may directly compete with other businesses in the strip mall, including. . .) *1:* a Mexican restaurant. *2:* a bagel shop. *3:* a hot dog stand.
Shoes must be worn in the courtyard. (The managers of a hotel are . . .) *1:* concerned that guests may accidentally step on rocks or other sharp objects when walking around in the hotel's courtyard area. *2:* concerned that when guests walk around in the hotel's courtyard area, their feet may be bitten by garden snakes. *3:* attempting to impose a dress code for the hotel's courtyard area.
Participants in the archeological dig must bring a tent. (The organizers of a planned multi-day archeological dig want to ensure. . .) *1:* that participants are protected against the intense sun, to minimize the risk of developing severe sunburns. *2:* that participants are protected from insects, scorpions, and other potentially dangerous pests in the area. *3:* that participants and their belongings are protected from the rain.
Customers wearing a uniform receive a 20% discount on all orders. (The managers of a restaurant seek to bolster the restaurant's image. . .) *1:* as a pro-military establishment. *2:* as an establishment that supports local first responders. *3:* as an establishment that caters to local working folk in the community.
No vehicles are allowed in the town square. (A small town is concerned . . .) *1:* that air pollution is becoming a problem in their town square. *2:* that there is too much ambient noise in their town square. *3:* for the safety of pedestrians in their town square.

References

Aparicio, H., M. Xiang, and C. Kennedy (2016). Processing gradable adjectives in context: A visual world study. *SALT 25*, 413–432.

Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge University Press.

Brannon, V. C. (2018). Statutory interpretation: Theories, tools, and trends. *Congressional Res. Service Reports*.

Carston, R. (2013). Legal texts and canons of construction: A view from current pragmatic theory. In M. Freeman and F. Smith (Eds.), *Law and Language: Current Legal Issues*, Volume 15, pp. 8–33. Oxford University Press.

Coventry, K. R., R. Carmichael, and S. C. Garrod (1994). Spatial prepositions, object-specific function, and task requirements. *J. Semant. 11*(4), 289–309.

Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 810–842.

Feist, M. I. and D. Gentner (2003). Factors involved in the use of in and on. *CogSci 25*, 390–395.

Frank, M. C. and N. D. Goodman (2012). Predicting pragmatic reasoning in language games. *Science 336*(6084), 998–998.

Goodman, N. D. and A. Stuhlmüller (2014). The design and implementation of probabilistic programming languages. `http://dippl.org`.

Grimm, S. and B. Levin (2012). Who has more furniture? Paper presented at Mass/Count in Linguistics, Philosophy and Cognitive Science Conference.

Grimm, S. and B. Levin (2017). Artifact nouns: Reference and countability. *NELS 47*, 55–64.

Grinsell, T. W. (2017). *Semantic indecision*. PhD. Diss, The University of Chicago.

Hannikainen, I. R., K. P. Tobia, G. d. F. de Almeida, N. Struchiner, M. Kneer, P. Bystranowski, V. Dranseika, N. Strohmaier, S. Bensinger, K. Dolinina, et al. (2022). Coordination and expertise foster legal textualism. *Proc. Natl. Acad. Sci. U.S.A. 119*(44).

Hart, H. L. A. (1958). Positivism and the separation of law and morals. *Harv. L. Rev. 71*, 593.

Kamp, J. A. W. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language*, pp. 123–155. Cambridge University Press.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguist. Philos. 30*(1), 1–45.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguist. Philos. 4*(1), 1–45.

Labov, W. (1973). The boundaries of words and their meanings. In *New ways of analyzing variation in English*.

Lasersohn, P. (1999). Pragmatic halos. *Language 75*(3), 522–551.

Lassiter, D. and N. D. Goodman (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese 194*(10), 3801–3836.

Levin, B., L. Glass, and D. Jurafsky (2019). Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds. *Linguistics 57*(3), 429–471.

Malt, B. C. and E. C. Johnson (1992). Do artifact concepts have cores? *J. Mem. Lang. 31*(2), 195–217.

Murphy, G. (2002). *The big book of concepts*. MIT Press.

Nichols, L. (2008). Lexical semantic constraints on noun roots and noun borrowability. *Studies in Language 32*(3), 683–700.

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.

Qing, C. and M. Franke (2014). Gradable adjectives, vagueness, and optimal language use. *SALT 24*, 23–41.

Qing, C., D. Lassiter, and J. Degen (2018). What do eye movements in the visual world reflect? A case study from adjectives. *CogSci 40*, 2297–2302.

Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem (1976). Basic objects in natural categories. *Cogn. Psychol. 8*(3), 382–439.

Sassoon, G. W. (2013). A typology of multidimensional adjectives. *J. Semant. 30*(3), 335–380.

Sassoon, G. W. and J. Fadlon (2017). The role of dimensions in classification under predicates predicts their status in degree constructions. *Glossa 2*(1).

Scalia, A. (1997). *A matter of interpretation*. Princeton University Press.

Sperber, D. (2007). Seedless grapes: nature and culture. In E. Margolis and S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation*, pp. 124–137. Oxford University Press.

Struchiner, N., I. Hannikainen, and G. Almeida (2020). An experimental guide to vehicles in the park. *Judgm. Decis. Mak. 15*(3).

Tessler, M. H. and N. D. Goodman (2019). The language of generalization. *Psychol. Rev. 126*(3), 395.

Tversky, B. and K. Hemenway (1984). Objects, parts, and categories. *J. Exp. Psychol. 113*(2), 169–193.

Wisniewski, E. J. and B. C. Love (1998). Relations versus properties in conceptual combination. *J. Mem. Lang. 38*(2), 177–202.

Wisniewski, E. J. and G. L. Murphy (1989). Superordinate and basic category names in discourse: A textual analysis. *Discourse Process. 12*(2), 245–261.