Running head: PRAGMATIC THEORY OF GENERIC LANGUAGE

A pragmatic theory of generic language

**Michael Henry Tessler** (mtessler@stanford.edu) **Noah D. Goodman** (ngoodman@stanford.edu) Department of Psychology, Stanford University

CORRESPONDING AUTHOR

MICHAEL HENRY TESSLER

STANFORD UNIVERSITY, DEPARTMENT OF PSYCHOLOGY

450 SERRA MALL, BLDG. 420, RM. 316 STANFORD, CA 94305

[MTESSLER@STANFORD.EDU](mailto:MTESSLER@STANFORD.EDU)

757 - 561- 7071

**Abstract**

Generalizations about categories are central to human understanding, and generic language (e.g. *Dogs bark.*) provides a simple and ubiquitous way to communicate these generalizations. Yet the meaning of generic language is philosophically puzzling and has resisted precise for- malization. We explore the idea that the core meaning of a generic sentence is simple but underspecified, and that general principles of pragmatic reasoning are responsible for estab- lishing the precise meaning in context. Building on recent probabilistic models of language understanding, we provide a formal model for the evaluation and comprehension of generic sentences. This model explains the puzzling flexibility in usage of generics in terms of diverse prior beliefs about properties. We elicit these priors experimentally and show that the resulting model predictions explain almost all of the variance in human judgments for both common and novel generics. This theory provides the mathematical bridge between the words we use and the concepts they describe.

Most would agree that *Swans are white*, but certainly not every swan is. This type of utter- ance conveys a generalization about a category (i.e. SWANS) and is known as a generic utterance (Carlson, 1977; Leslie, 2008). It is believed that every language can express generic meaning (Behrens, 2005; Carlson & Pelletier, 1995), and that generics are essential to the growth of concep- tual knowledge (Gelman, 2004) and how kinds are represented in the mind (Leslie, 2008). Generic language is ubiquitous in everyday conversation as well as in child-directed speech (Gelman, Goetz, Sarnecka, & Flukes, 2008), and children as young as two or three understand that generics refer to categories and support generalization (Cimpian & Markman, 2008). Additionally, generics are the primary way by which speakers discuss social categories, making them key to propagat- ing stereotypes (Gelman, Taylor, Nguyen, Leaper, & Bigler, 2004; Rhodes, Leslie, & Tworek, 2012; Leslie, Cimpian, Meyer, & Freeland, 2015) and impacting motivation (Cimpian, 2010). De- spite their psychological centrality and apparent simplicity, a formal account of generic meaning remains elusive.

The major issue in formalizing generic language is determining what makes a generic sentence true or false. At first glance, generics seem like universally-quantified statements as in *All swans are white*, but unlike universals, generics are resilient to counter-examples (e.g. *Swans are white* even though there are black swans). Interpreting the generic as meaning “most” (i.e. *Most swans are white*) captures many cases, but cannot explain why *Robins lay eggs* and *Mosquitos carry malaria* are so intuitively compelling: Only adult female robins lay eggs and a very tiny fraction of mosquitos actually carry malaria. Indeed, it appears that any explanation in terms of how com- mon the property is within the kind violates intuitions — for the robins, laying eggs is practically synonymous with being female (i.e., the properties are present in the same proportion), yet *Robins are female* is not a reasonable utterance while *Robins lay eggs* is fine.

How generic language is interpreted is also a mystery. *Mosquitos carry malaria* suggests the generic must in some way be analogous to “some” (i.e. *Some swans are white.*). Yet generics are often interpreted as implying the property is widespread within the kind: Consider the difference between *Some swans have hollow bones* and *Swans have hollow bones* (Gelman, Star, & Flukes, 2002). When asked how common a property would need to be for the generic utterance to be true, both children and adults require the property to be less widespread than they infer when told the generic (Cimpian, Brandone, & Gelman, 2010; Brandone, Gelman, & Hedglen, 2014), suggesting that communicating with generics can exaggerate evidence.

How can generics have such flexible truth conditions while simultaneously carrying strong im- plications? In this paper we resolve these philosophical and empirical puzzles using a mathematical model that understands generic language by pragmatic inference about the degree of prevalence required to assert the generic. We show that this model predicts the patterns in both human en- dorsement of familiar generic sentences and interpretation of novel generic sentences.

## Computational method: The semantics and pragmatics of generic language

Generics express a relation between a kind (e.g. ROBINS) and a property (e.g. LAYS EGGS). Semantic accounts of generics are usually given by appealing to either the statistics of the world (e.g. *Barns are red* because most barns are) or to structured, conceptual representations (e.g. *Bishops move diagonally* not because most bishops do but because the laws of chess mandate that they do) (Carlson, 1995). This latter perspective emphasizes the structure of generic knowledge (Prasada, 2000), and views generic utterances as the way of expressing special mental relationships between kinds and properties (Leslie, 2008; Prasada, Hennefield, & Otap, 2012). The puzzles of

generic language then reduce to puzzles about mental representation of kind-property relations.

However, generic language is not unique in its flexibility. Language understanding in general depends on assumptions interlocutors make about each other and can result in manifold interpre- tations with a complex sensitivity to context (Clark, 1996; Grice, 1975; Levinson, 2000). Can the puzzles of generic language be understood as effects of pragmatic reasoning? If so, it may be un- necessary to encode abstract mental representations into the semantics of generics. We thus begin with a relatively simple semantic theory, phrased in terms of statistical regularities, measured with probability.

For a given kind *K* (e.g. ROBINS) and property *F* (e.g. LAYS EGGS), we refer to the probability

that an object of kind *K* has property *F*, that is *P*(*F | K*), as the *prevalence* of *F* within *K*.1 Logical quantifiers can be described as conditions on prevalence (i.e. *some* is *P*(*F | K*) *>* 0, *all* is *P*(*F |*

*K*) = 1). Extending this, it seems the simplest meaning for generic statements would be a similar threshold on prevalence: *P*(*F | K*) *>* τ (Cohen, 1999). However, no fixed value of the threshold, τ, would allow for the extreme flexibility generics exhibit (e.g. *Robins lay eggs* vs. *Robins are*

*female*; *Mosquitos carry malaria*).

We suggest that this threshold is not a fixed property of the language, but is established by pragmatic inference. This inference depends on property and category knowledge, but is other- wise a general mechanism of language not specific to interpreting generic statements. Pragmatic reasoning can be formalized in the Rational Speech Acts (RSA) theory—a probabilisitic model of language understanding—as recursive Bayesian inference between speaker and listener (Frank & Goodman, 2012; Goodman & Stuhlmu¨ller, 2013). We follow the treatment of RSA applied to

1Because we aim to explain the psycholinguistics of generics, we are generally interested in the subjective proba- bility, not the actual frequency in the world.

vague adjectives (e.g. *tall*), using an underspecified threshold criterion (Lassiter & Goodman, 2013, 2015). We imagine a hypothetical, pragmatic listener (*L*1) concerned with learning the prevalence

of a certain property in a certain category, *x* = *P*(*F | K*), who reasons about an informative speaker

(*S*1), who in turn reasons about a literal listener (*L*0):

*PL*1 (*x,* τ *| u*) ∝ *PS*1 (*u | x,* τ) *· P*(*x*) *· P*(τ) (1)

1 (*u | x,* τ) ∝ *P*  0 (*x | u,* τ)

*PS L* λ

(2)

*PL*0 (*x | u,* τ) ∝ δ [[u]] (*x,*τ)*P*(*x*)(3)

The pragmatic listener *L*1 (Eq. 1) is trying to resolve how widespread F is within K (i.e. *x* = *P*(*F | K*) ~[0*,* 1]); she does so by considering both what she knows about property F in general— her prior distribution *P*(*x*) over prevalence of F—and her intuitive theory of the speaker *S*1 as an informed and helpful interlocutor (Eq. 2). *L*1 assumes the meaning of the generic utter- ance expresses something about how common the property is (she knows it is a threshold on prevalence), but doesn’t know exactly what it conveys (she has uncertainty about the threshold: τ *~* Uniform([0,1])). She believes the speaker *S*1 knows what τ is, and is trying to be informative about the prevalence *x* to the listener he is thinking about—an idealized literal listener (*L*0, Eq. 3). The literal listener has access to the threshold τ, and simply restricts her prior beliefs to situations where the truth-functional denotation of the utterance, [[ *u*]] , is true.

Pragmatic interpretation of language assumes the production of an utterance was a deliberate decision on behalf of the speaker who could have said other things but didn’t. In our case, we assume the speaker had the options of uttering the generic sentence or staying quiet—the *null utterance* alternative carries no information (i.e.[[ *null*]] (*x,* τ) = *T* ) and produces a posterior distri-

bution in the listener identical with the prior.2 The contrast with the *null* utterance alternative gives the generic its communicative force, making it a *speech-act*. Throughout this paper we will treat

the bare plural construction as generic utterances with a threshold semantics:[[ K F]] (*x,* τ) = *x >* τ.

The bare plural construction may additionally be ambiguous between a generic interpretation and a specific plural predication (consider *dogs are on my lawn* compared to *dogs have fur*), and dis- ambiguation itself may be a pragmatic process (Cimpian & Markman, 2008). In this paper we are concerned with the interpretation of a truly generic utterance, once disambiguated.

Example interpretation distributions for *PL*1 (*x,* τ *| u*) upon hearing a generic utterance can be

seen in Figure 1. Also shown are the corresponding prior beliefs, *P*(*x*), about the prevalence of the property (which are also the posteriors upon hearing the null utterance). We see that the interpretation of the generic depends a great deal on the shape of the prior. When the prior is very left-skewed as in the case of CARRIES MALARIA, then τ can plausibly be quite low while still being informative, since a low threshold still rules out many possible alternative kinds (and their corresponding degree of prevalence). If the prior is right-skewed (e.g. DOESN’T ATTACK SWIMMERS), even an intermediate threshold would not result in an informative utterance (as not many kinds would be ruled out), and so the generic is unlikely to be used by speaker *S*1 unless the property is practically-universal within the target category. Priors for properties that are unimodal with low variance (e.g. IS FEMALE) are present in every kind in almost exactly the same proportion and thus are too obvious and certain to allow for a realistic, informative utterance: The posterior is not very different from the prior.

The pragmatic listener *L*1 (Eq. 1) is a model of generic interpretation: Upon hearing a generic,

2This alternative can be realized in at least two other ways: the speaker could have said the negation of the utterance (i.e. *It is not the case that robins lay eggs.*) or the negative generic (i.e. *Robins do not lay eggs.*). All results reported are similar for these two alternatives, and we use the alternative of the *null* utterance for simplicity of demonstration.

what prevalence is a listener likely to infer? We can now imagine a speaker *S*2 who reasons about this type of listener:

*PS*2 (*u | x*) ∝

*∫ PL*1 (*x,* τ *| u*) (4)

τ

Speaker *S*2 is a model of felicity or truth judgments (Degen & Goodman, 2014). The speaker considers the thought-processes of listener *L*1 (Eq. 1) and decides if the generic is a good (albeit, vague) way to describe the prevalence *x*. *S*2’s decision is with respect to the alternative of saying nothing: He will choose to produce the generic when the true prevalence *x* is more likely under

*L*1’s posterior than under her prior. Critically, speaker *S*2 doesn’t actually know what the generic means (i.e. doesn’t have access to the threshold τ), but knows that *L*1 will have to consider it, and integrates over the likely values she’ll consider3.

# Flexible truth conditions

Any theory of generic language must explain their puzzling flexibility of usage with respect to prevalence. For instance, *Mosquitos carry malaria* and *Birds lay eggs*, but *Birds are female* is not a good thing to say. To test the ability of the pragmatic *S*2 model, Eq. 4, we constructed thirty generic sentences that cover a range of conceptual distinctions (Prasada, Khemlani, Leslie, & Glucksberg, 2013): characteristic (e.g. *Ducks have wings.*), minority (e.g. *Robins lay eggs.*), striking (e.g. *Mosquitos carry malaria.*), false generalization (e.g. *Robins are female.*), and false (e.g. *Lions lay eggs.*). The sentences were also chosen to elicit a range of acceptability judgments (“acceptable”,

3A fully specified version of the generics model, as well as the structured prior model (described below), can be found at [http://forestdb.org/models/generics.html.](http://forestdb.org/models/generics.html)

“unacceptable”, and “uncertain”) with low-, medium-, and high-prevalence properties. In Expt. 1b we collect human judgements about the acceptability of these sentences. In order to compare the results to the model we will need appropriate prior beliefs about the relevant properties; in Expt. 1a we empirically measure the prior distribution over prevalence of these properties.

# Experiment 1a: The structure of prevalence priors

The prior *P*(*x*) (in Eqs. 1, 3) describes the belief distribution on the prevalence of a given property (e.g. LAYS EGGS) across relevant categories. The shape of this distribution affects model predic- tions, but may vary significantly among different properties. We measured it empirically for a set of properties (e.g. LAYS EGGS, CARRIES MALARIA; 21 in total) used in our target sentences.

## Method

### Participants

We recruited 60 participants over Amazon’s crowd-sourcing platform Mechanical Turk (MTurk). We chose this number of participants based on intuition with similar experiments and model com- parison; since this is a quantitative experiment with no planned comparisons, power analysis in not appropriate. Participants were restricted to those with US IP addresses and with at least a 95% MTurk work approval rating (the same criteria apply to all experiments reported).

### Procedure and materials

On each trial of the experiment, participants filled out a table where each row was an animal cat- egory and each column was a property. In order to alleviate the dependence of the distribution

on our animal categories of interest, half of the animal categories were self-generated by the par- ticipant; the other half were randomly sampled from a set corresponding to the generic sentences used in Expt. 1b (e.g. ROBINS, MOSQUITOS). Participants were asked to fill in each row with the percentage of members of each of the species that had the property (e.g. “50%”). Each participant reported on sixteen properties (see *SI Section A* for more details). We used a set of properties as- sociated with generics of theoretical interest (twenty-one properties in total), as described above. For a full list of the properties, and generic sentences used in Expt. 1b, see Table S1.

## Data analysis and results

While each *P*(*x*) is a single distribution on prevalence, it may be structured as the result of deeper conceptual knowledge. For instance, if participants believe that some kinds have a causal mech- anism that *could* give rise to the property, while others do not, then we would expect *P*(*x*) to be structured as a mixture distribution (cf. Griffiths and Tenenbaum (2005)). If a kind can have the property, we assume the prevalence follows a Beta distribution with mean γ and concentration ξ. If a kind cannot, the prevalence can only be 0%. The relative contribution of these two compo- nents is governed by mixture parameter θ, inferred from the data.4 Thus, θ is the potential of a property to be present in a kind and γ is the mean prevalence of the property among the kinds with the potential to have it. The prevalences given by participants would then be distributed as: *P*(*d*) = θ *·* Beta(*d |* γ*,* ξ) + (1 *−* θ) *·* δ*d*=0. We put uninformative priors over all the parameters, θ *~* Uniform(0*,* 1), γ *~* Uniform(0*,* 1), ξ *~* Uniform(0*,* 50), and implemented this Bayesian sta- tistical model using the probabilisitic programming language WebPPL (Goodman & Stuhlmüller,

4This is similar in spirit to Hurdle Models of epidemiological data, where the observed count of zeros is of- ten substantially greater than one would expect from standard models, such as the Poisson (e.g. adverse events to vaccines)(Rose, Martin, Wannemuehler, & Plikaytis, 2006)

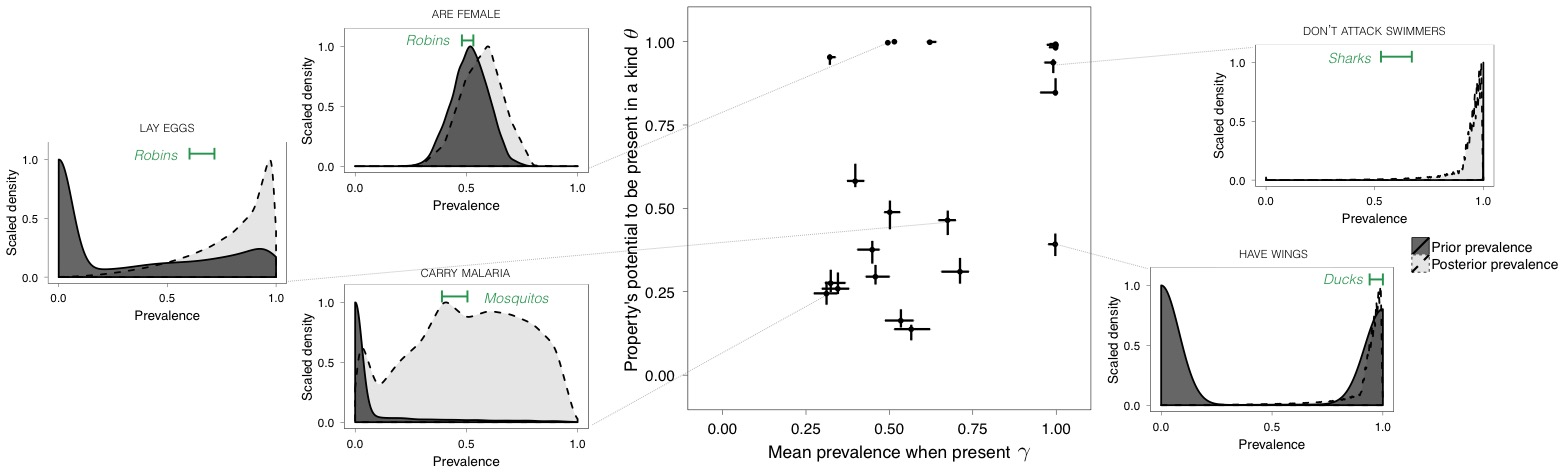


Figure 1: Prevalence prior distributions empirically elicited for twenty-one animal properties. Prior distributions are summarized by θ—-a property’s potential to be present in a category—-and γ—- the mean prevalence when it is possible for the property to be present in a category. Inset plots dis- play example empirical prior distributions over prevalence together with corresponding *L*1 model predictions: the posterior after hearing a generic utterance. Intervals on the top of plots show human beliefs about the prevalence of the property within a target category. Felicitous generic utterances result when the target prevalence is more likely under the posterior than under the prior.

Error bars denote 95% Bayesian credible intervals (same for Figure 3).

2014). This statistical model reproduces the prior elicitation data with a low reconstruction error (*r*2 = 0*.*94; see *SI Fig. 1*), strongly supporting the assumption of a structured prior.

Using this structure, we can explore our twenty-one properties: Figure 1 shows the estimated mixture-parameter θ (the potential of the property to be present in a kind) and the mean preva- lence when the property is present, γ. We see significant diversity among these properties in both dimensions, resulting in priors over prevalence with dramatically different shapes (insets). This diversity matters because the model predictions for generic interpretation and felicity depend on the shape of the prior distribution (insets show example *L*1 interpretation posteriors). Lowering θ effectively makes the property more distinctive by increasing the relative probability mass at 0%; this relaxes the truth conditions by making a lower threshold more informative. Increasing γ means the property is expected to be present in more members of the category; this tends to make truth conditions stricter, by reducing the range of prevalence values higher than the prior expectation.

# Experiment 1b: Truth judgments of familiar generics

We tested the degree to which the *S*2 model, Eq. 4, coupled with the empirically-elicited prevalence priors from Expt. 1a predicted that a given category–property pair (e.g. ROBINS and LAYS EGGS) would result in a felicitous generic (e.g. *Robins lay eggs.*).

## Method

### Participants

We recruited 100 participants over MTurk. We chose this number of participants based on intuition from similar experiments which were designed primarily to test a quantitative model.

### Procedure and materials

Participants were shown thirty generic sentences in bare plural form one after another. They were asked to press one of two buttons to signify whether they agreed or disagreed with the sentence (see *SI Table 2* for complete list and *SI Section B* for more details). The thirty sentences were presented in a random order between participants and cover a range of conceptual distinctions and acceptability levels, as described above.

### Analysis and results

From the prevalence-prior data (Expt. 1a) we can estimate participants’ beliefs about the prevalence of a property *for a given kind* (e.g. the percentage of ROBINS that LAY EGGS; see green intervals on Figure 1 and *SI Table 2*). As a simple baseline hypothesis, we first explore whether these prevalence

values themselves predict generic endorsement (e.g. does the fraction of ROBINS that LAY EGGS

predict the felicity of *Robins lay eggs*?). A little over half of the variance in truth judgments data is explained by the prevalence of the property within the kind alone (*r*2 = 0*.*599; MSE=0.065; *SI Figure 2*). This is not surprising given our inclusion of high-prevalence true generics (e.g. *Leopards have spots.*) and low-prevalence false generics (e.g. *Leopards have wings.*). However, large deviations remain from an account based purely on target-category prevalence: Generics in which the target category has intermediate prevalence (prevalence quartiles 2 and 3: 20% *<*

*prevalence <* 64%), are not at all explained by prevalence within those categories (*r*2

*Q*2*,*3

= 0*.*029;

MSE = 0.110).

The speaker model, *S*2 in Eq. 4, predicts an endorsement probability for a generic sentence, given prior beliefs for the property and a target prevalence. We use the measured within–kind prevalence as the target prevalence that speaker *S*2 is trying to communicate, and use the empiri- cally inferred priors from Expt. 1a. The model has one remaining parameter: the speaker optimality parameter λ (in Eq. 2). We integrate out the likely values of this parameter using Bayesian data analytic techniques (Lee & Wagenmakers, 2014) (see SI Section B for full details). As we see in Figure 2, the pragmatic speaker model *S*2, using empirically measured priors, does a very good job of explaining human truth judgments (*r*2 = 0*.*981; MSE=0.003; Figure 2, *SI Figure 4*). Generics that received definitive agreement or disagreement are predicted to be judged as such by the model (corners of Figure 2), including items for which target-category prevalence is not a good indicator

of the acceptability (for prevalence quartiles 2 and 3, *r*2

*Q*2*,*3

= 0*.*955; MSE=0.005; Figure 2, inter-

mediate shades). The probabilistic pragmatics model explains the puzzling flexibility of generic truth-conditions as the result of communicative pressures (*be truthful*, *be informative*) operating over diverse prior beliefs about the properties.

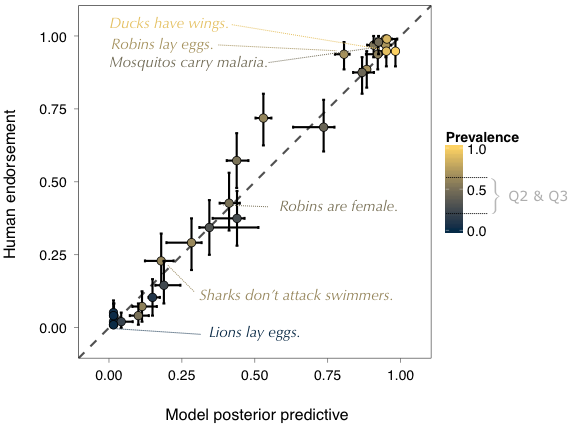


Figure 2: Human acceptability judgments and model predictions for thirty generic utterances about familiar animals and properties. Color denotes target-category prevalence of the property, with darker colors indicating lower prevalence. Intermediate prevalences (quartiles 2 & 3) are in inter- mediate shades (marked on color bar). Error bars correspond with 95% bootstrapped confidence intervals for the participant data and 95% highest probability intervals for the model predictions (same for Figures 4 and 5).

# Strong interpretations for novel categories

One of the most important roles for generic language is to provide learners information about new or poorly understood categories. This role depends on how unfamiliar generic sentences are in- terpreted (e.g. Gelman et al., 2002; Cimpian, Brandone, & Gelman, 2010). The pragmatic theory we present includes such a theory of generic comprehension: the listener model (Eq. 1) describes interpretation of a generic utterance—*Feps* HAVE PROPERTY—without previously knowing the prevalence of the property within this kind. With an uncertain meaning and the pressure to be in- formative, interpretation in the theory is heavily driven by *a priori* beliefs about properties. Classic work in generalization suggests beliefs about the prevalence of properties differ by type of prop- erty, including relatively fine distinctions among properties that are all biological in nature (Nisbett, Krantz, Jepson, & Kunda, 1983). We leverage these strong but diverse expectations, using forty different properties to explore a wide range of *a priori* beliefs about prevalence. These items make up four categories of properties: body parts (e.g. HAS CLAWS), body parts of a particular color (e.g. HAS GREEN FEATHERS), body parts described vaguely (e.g. HAS SMALL WINGS), body parts in accidental or disease states (e.g. HAS WET FUR, HAS SWOLLEN EARS). We first use a novel method for measuring *a priori* beliefs about the prevalence of these properties (Expt. 2a). We then test the predictions of the pragmatic listener model *L*1 using these empirically derived priors against human interpretations of novel generic sentences (Expt. 2b). Finally, we explain a pre- viously reported empirical asymmetry between truth conditions and interpretations by comparing the speaker *S*2 and listener *L*1 models in the same experimental context (Expt. 2c).

# Experiment 2a: Prevalence priors for unfamiliar categories

To measure the prior over prevalence using *familiar* categories (Expt. 1a), participants filled out a table with rows corresponding to different animal kinds and columns corresponding to different properties. This method is not applicable for unfamiliar kinds (where nothing would distinguish the rows). We instead leverage the latent structure uncovered in Expt. 1a, where we decomposed prevalence priors into the property’s potential to be present in a kind and the expected prevalence when present. This suggests a two-stage elicitation procedure.

## Method

### Participants

We recruited 40 participants over MTurk. We again chose this number of participants based on intuition from similar experiments which were designed primarily to test a quantitative model.

### Procedure and materials

We constructed forty different properties to explore a wide range of *a priori* beliefs about preva- lence. These items make up four categories of properties: body parts (e.g. HAS CLAWS), body parts of a particular color (e.g. HAS GREEN FEATHERS), body parts described vaguely (e.g. HAS SMALL WINGS), body parts in accidental or disease states (e.g. HAS WET FUR, HAS SWOLLEN EARS). Because pilot testing revealed more variability for items in the accidental category rela- tive to the other types of properties, we used twice as many exemplars of accidental properties, yielding a better test of the quantitative predictive power of the *L*1 interpretation model. We used 8

exemplars of each of the three non-accidental properties (“parts”, “colored parts”, “vague parts”), and 16 exemplars of accidental properties, building on a stimulus set from Cimpian, Brandone, and Gelman (2010). All materials are shown in Table S3.

Participants were introduced to a data-collection robot that was tasked with learning about properties of animals. Participants were told the robot randomly sampled an animal to ask the participant about (e.g. The robots says: “We recently discovered animals called feps.”). The robot then asks the participant two questions, aimed to measure the two components of the structured prior model: the potential of the property to be present in a kind and the expected prevalence when present. The robot asked how likely it was that “there was *a* fep with PROPERTY” (potential to be present), to which participants reported on a scale from “unlikely” to “likely”. For example, it is very likely that there is a fep that is female, less likely that there is a fep that has wings, and even less likely that there is a fep that has purple wings. The robot then asked, “Suppose there is a fep that has wings. What percentage of feps do you think have wings?” (expected prevalence when present). Participants completed a practice trial to make sure they understood the meaning of these two questions.

## Data analysis and results

We used the same structured, statistical model from Expt. 1a. The only difference from Expt. 1a. is that our experimental data comes from inquiring about the parameters of the priors directly, as opposed to asking about particular samples (i.e. particular kinds) as was done in Expt. 1a. We assume these two measurements follow Beta distributions (*dpotential ~* Beta(γ1*,* ξ1); *dexpected ~* Beta(γ2*,* ξ2)), and construct single prevalence distributions, *P*(*x*), by sampling from the posterior

predictive distribution of prevalence as we did before: *P*(*x*) =∫[θ *·* Beta(*x |* γ2*,* ξ2)+(1 *−* θ) *·* δ*x*=0] *·*

Beta(θ *|* γ1*,* ξ1)*d*θ. We used the same parameter priors as in Expt. 1a.

Figure 3 shows a summary of the elicited priors, in terms of the diversity of *dpotential* and *dexpected* . Biological properties are expected to be *a priori* more prevalent within a kind when present than accidental properties, with additional fine-grained differences within biological and accidental properties. Like the priors elicited using familiar categories, these priors elicited us- ing unfamiliar categories have diverse shapes (see insets). Biological properties (“biological”, “vague”, and “color” body parts) have prevalence distributions that are bimodal with peaks at 0% and near-100% prevalence. Interpretations of generics about these properties (*L*1 model, Eq. 1) up- date these distribution to concave posteriors peaked at 100% (Figure 3; red, blue and green insets); this predicts strong interpretation of novel generics for these properties (Gelman et al., 2002). Ac- cidental properties (both “rare” and “common”) follow unimodal prior distributions and update to convex posterior distributions, predicting weaker and less certain interpretations of novel generics for these properties (Cimpian, Gelman, & Brandone, 2010).

# Experiment 2b: Interpretations of novel generics

We tested the degree to which the *L*1 listener model, Eq. 1, coupled with the empirically-elicited priors, *P*(*x*), from Expt. 2a predicted the interpretation of a generic sentence consisting of a novel category with one of the forty properties described above.

Macintosh HD:Users:mht:Documents:research:generics:manuscript:figures:prevalence-asymmetry-scatterwDists-byItem3.pdf

Figure 3: Prevalence prior distributions empirically elicited for 40 animal properties. Parameters of the structured statistical model—θ and γ—reveal quantitative differences in beliefs about the prevalence of conceptually different types of properties (scatterplot). Inset plots show differences in shapes between biological properties (red, green, blue; bimodal) and accidental properties (or- ange, purple; unimodal). These differences give rise to the variability of interpretations of generic utterances.

## Method

### Participants

We recruited 40 participants over MTurk to determine how widespread different properties are believed to be upon hearing a novel generic. The experimental design is very similar to Cimpian, Brandone, and Gelman (2010), and we chose to have a sample size at least twice as large as the original study (original n=15). This is a quantitative experiment with only quantitative comparisons planned.

### Procedure and materials

Participants were told they were the resident zoologist of a team of scientists on a recently discov- ered island with many unknown animals; their task was to provide their expert opinion on questions

about these animals5. Participants were supplied with the generic (e.g. “Feps have yellow fur.”) and asked to judge prevalence: “What percentage of feps do you think have yellow fur?”. Partic- ipants completed in randomized order 25 trials: 5 for each of the biological properties and 10 for the accidental (described in Expt. 2a).

## Analysis and results

The pragmatic listener *L*1 model predicts the prevalence implied by a generic utterance based on the prior distribution of prevalence (measured in Expt. 2a) and the fact a speaker chose to say a generic sentence. We integrate out the single free parameter the using the same Bayesian model analysis as before (details in SI Section D). We first explore two important trends predicted by the pragmatic listener model. In Figure 5, solid lines, we see the implied prevalence judgments are predicted (at the property class level) to vary linearly with the *a proiri* expected prevalence. A mixed-effects linear model with random by-participant effects of intercept and slope indeed reveals the more widespread a property is expected to be *a priori*, the stronger the implications of a generic statement (β = 0*.*57; *SE* = 0*.*08; *t*(39) = 7*.*12; *p <* 0*.*001). The prevalence implied by a generic is also predicted to be greater than the *a proiri* expected prevalence. A mixed-effects linear model with random by-participant effects of intercept and random by-item effects of intercept and condition reveals implied prevalence after hearing a generic is significantly greater than the *a priori* prevalence (β = 0*.*17; *SE* = 0*.*018; *t*(39) = 9*.*7; *d* = 0*.*64; *p <* 0*.*001). Over all, the pragmatic listener model predictions closely align with the human judgments of prevalence for novel generics (*r*2(40) = 0*.*89, MSE=0.006), displaying sensitivity of interpretation to details of the property

5The experiment in full can be viewed at <http://stanford.edu/>˜mtessler/generics/experiments/ asymmetry/asymmetry-2.html

(Figure 4).

# Experiment 2c: The asymmetry between truth conditions and interpretations

There is a surprising décolage between the truth conditions and interpretations of generic language: Interpretations are characteristically strong while truth conditions are flexible. Experimentally, Cimpian, Brandone, and Gelman (2010) have shown that novel generics are interpreted as implying a higher prevalence than that required to assert the exact same generic. We use a procedure adapted from these authors to test both speaker *S*2 and listener *L*1 models in the same experiment.

## Method

### Participants

We recruited 40 participants over MTurk. We chose a sample size at least twice as large as the original study by Cimpian, Brandone, and Gelman (2010) (original n=20), and to match Expt. 2b.

### Procedure and materials

The cover story and materials were the same as in Expt. 2b. We told participants one of five frequencies for each property type (e.g. *50% of feps have yellow fur.*), and then asked for truth judgments of the corresponding generic sentence (e.g. *Feps have yellow fur.*). Prevalence varied between 10, 30, 50, 70, and 90%. The experiment consisted of twenty-five trials: 5 trials for each of 5 types of properties measured in Expt. 2a.

## Analysis and results

For both behavioral data and model predictions (Eq. 4) we computed the average prevalence that led to an assenting judgment (the *average prevalence score*), for each property type and partici- pant, following the procedure used by Cimpian, Brandone, and Gelman (2010). For example, if a participant agreed with the generic whenever the prevalence was 70% or 90% and disagreed at the other prevalence levels, that participant received an *average prevalence score* of 80%.

After integrating out the one parameter of the speaker *S*2 model, we subjected our model to the same procedure. The speaker model *S*2 returns a posterior probability of producing the generic, for each level of prevalence6. We sample a response from this posterior distribution for each prevalence level, simulating a single subject’s data. Just as with the human data, we took the trials

where the model agreed with the generic, and took the mean of the prevalence levels corresponding to those trials, giving us the average prevalence at which the model assented to the generic. We repeated this for each type of property 40 times to simulate a sample of 40 participants. We repeated this procedure 1000 times to bootstrap 95% confidence intervals.

The speaker *S*2 model predicted *average truth conditions* that did not vary appreciably across the different types of properties: Generics are acceptable for a wide range of prevalence lev- els for all property types. A similar absence of a gradient was observed in the human data

(β = 2*.*82; *SE* = 4*.*02; *t*(39) = 0*.*70; *p* = 0*.*49; Figure 5, dotted lines). Interpretations of generic utterances are stronger than their average truth conditions for the biological properties but not for

6We assume here that the prevalence told to the participant is the subjective probability that that speaker *S*2 is trying to communicate. This assumption is mostly likely incorrect due to theory-driven considerations on behalf of the speaker. For example, if the participant believes the prevalence being reported in the experiment is describing a temporary or accidental state and not one that is likely to be predictive of the future prevalence, the speaker may derive a subjective probability substantially less than that stated verbally in the experiment. Addressing this issue is beyond the scope of this article, and is not necessary for the relatively coarse-grained comparison of *average truth conditions* to *implied prevalence*

1.00

●

●

●

●

●●●

●

●

●●

●

●

●

●

● ●

●●

● ●●

●

●●

●

●

●

●

● ●

●

● ●

● ●

●

●

**Property type**

● body parts

● vague parts

● color parts

● common accidental

●

rare accidental

Human interpretation of prevalence

0.75

0.50

0.25

0.00

0.00 0.25 0.50 0.75 1.00

Model posterior predictive

Figure 4: Human interpretation of prevalence upon hearing a generic compared with the *L*1 model posterior predictive. Participants and the model interpret generics differently for different prop- erty types: Generics of biological properties (red, blue, green) have strong interpretations while generics of accidental properties (purple, orange) are weak.

the accidental properties (Figure 5), replicating Cimpian, Brandone, and Gelman (2010) with both human data and the model; the extent of the difference is governed by prior property knowledge (mean prevalence when present γ, from Expt. 2a). The listener and speaker pair of models pre- dicts human endorsements and interpretations of novel generic utterances well (*r*2(10) = 0*.*921, MSE = 0.004). Thus, our model predicts that the asymmetry between truth conditions and implied prevalence should hold, but only for properties with the most extreme prior beliefs.

# Discussion

We evaluated a theory of generic language derived from general principles of pragmatic language understanding using a simple but uncertain basic meaning—a threshold on property prevalence. Our formal model is a minimal extension of the RSA theory of language understanding, together

Macintosh HD:Users:mht:Documents:research:generics:manuscript:figures:asym-lines-data-model.pdf

Figure 5: Human judgments and model predictions of prevalence implied by novel generic ut- terances (implied prevalence task; solid line) and average prevalence that leads to an acceptable generic utterance (truth conditions task; dotted line) as it relates to the *a priori* mean prevalence when present γ. Expectations of prevalence are higher after hearing a generic than before hearing

it (solid line compared to *y* = *x* line; both for human data and model). Generic statements about

biological properties, imply that the property is widespread in the category, for both human par-

ticipants and the model (solid line: red, blue and green). Generics about accidental properties do not result in such a high implied prevalence (solid line: purple and orange). While the implica- tions of generic utterances are highly variable across the different types of properties, the average prevalence that leads to an acceptable generic does not vary, for participants or the model.

with an underspecified threshold semantics. The model was able to explain two qualitative puzzles of generics: their extremely flexible truth conditions and the contrastingly strong interpretation of novel generics, both of which were revealed to depend in systematic ways on prior knowledge about properties. The model predicted the quantitative details of participants’ judgments with high accuracy.

Previous psychological and philosophical work on generics has looked beyond prevalence and focused on conceptual distinctions and relations (Gelman, 2005; Prasada et al., 2013; Leslie, 2007, 2008). Prasada et al. has argued for a distinction between *characteristic* properties (e.g. *Diapers are absorbent.*) and *statistical* properties (e.g. *Diapers are white.*). Leslie suggests information that is striking (e.g. *Tigers eat people.*) is useful and thus permitted to be a generic. Gelman outlines how generics express *essential* qualities that are relatively timeless and enduring. Our approach makes the strong claim that beliefs about prevalence are the connective tissue between conceptual knowledge and generic language. That is, the effect of conceptually meaningful differences on generic language is predicted to be mediated by differences in corresponding prevalence distribu- tions. Indeed, we found that empirical prevalence distributions are structured in a way that reflects intuitions about causal mechanisms underlying different properties. It is plausible that richer con- ceptual knowledge also influences these distributions, such as higher-order conceptual knowledge about the nature of properties and categories (Gelman, 2005; Keil, 1992).

It is also important to note that our approach is based on *subjective* probability, and not mere frequency. This difference would be most apparent when abstract intuitive theories lead us to reject observed frequencies in forming our subjective probabilities. For instance, because we believe that social security numbers have no influence on selection for the Supreme Court, even if we find out that all Supreme Court justices have even social security numbers, we will assign about 50%

subjective probability to the next justice having an even number; we will thus reject the generic *Supreme Court Justices have even social security numbers*. (If we learned much more surprising information, such as all justices having prime numbers, we might revise our theory, for instance appealing to a conspiracy, and then accept the generic.) Further research will be needed to explore the predicted relationship between conceptual distinctions, subjective probabilities for prevalence, and generic language.

It might seem paradoxical that a part of language that is so common in communication and central to learning should be vague. Shouldn’t speakers and teachers want to express their ideas as crisply as possible? To the contrary, underspecification can be efficient, given that context can be used to resolve the uncertainty (Piantadosi, Tily, & Gibson, 2012). In our work, context takes the form of a listener and speaker’s shared beliefs about the property in question. By leveraging this common ground, generics provide a powerful way to communicate and learn generalizations about categories, which would be difficult or costly to learn through direct experience. The dark side of this flexibility is the potential for miscommunication or deceit: A speaker might assert a generic utterance that he himself would not accept, conveying a too-strong generalization to a naïve listener. Our model predicts this potential particularly for properties which, when present, are widespread in a category—we showed that biological properties are believed to have this distribu- tion, but many properties of social categories may as well (Cimpian & Markman, 2011; Cimpian, Mu, & Erickson, 2012; Rhodes et al., 2012).

Categories are inherently unobservable. You cannot see the category DOG, only some number of instances of it. Yet we easily talk about these unobservables, conveying hard-won generaliza- tions to each other and down through generations. The theory presented here gives one explanation of how we do so, providing a computational perspective on how category generalizations are con-

veyed and how beliefs play a central role in understanding language.

### Author contributions

M. H. Tessler and N. D. Goodman developed theory and the study concept and design. M. H. Tessler performed research and analyzed data. M. H. Tessler and N. D. Goodman wrote the paper.

### Acknowledgements

This work was supported in part by National Science Foundation Graduate Research Fellow- ship DGE-114747 (to M.H.T.), by John S. McDonnell Foundation Scholar Award 220020252 (to N.D.G.) and Office of Naval Research Grant N00014-13-1-0788 (to N.D.G.)

# References

Behrens, L. (2005). Genericity from a cross-linguistic perspective. *Linguistics*, *43*(2), 275–344. Brandone, A. C., Gelman, S. A., & Hedglen, J. (2014). Children’s Developing Intuitions About the

Truth Conditions and Implications of Novel Generics Versus Quantified Statements. *Cognitive science*, 1–28.

Carlson, G. N. (1977). *Reference to kinds in english* (Unpublished doctoral dissertation). Univer- sity of Massachusetts, Amherst.

Carlson, G. N. (1995). Truth conditions of generic sentences: Two contrasting views. In G. N. Carl- son & F. J. Pelletier (Eds.), *The generic book* (p. 224-38). Chicago: University of Chicago Press. Carlson, G. N., & Pelletier, F. J. (1995). *The generic book.* Chicago, IL: Chicago University

Press.

Cimpian, A. (2010). The impact of generic language about ability on children’s achievement motivation. *Developmental psychology*, *46*(5), 1333–1340. doi: 10.1037/a0019665

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive science*, *34*(8), 1452–1482.

Cimpian, A., Gelman, S. A., & Brandone, A. C. (2010). Theory-based considerations influence the interpretation of generic sentences. *Language and Cognitive Processes*, *25*(2), 261–276. doi: 10.1080/01690960903025227.Theory-based

Cimpian, A., & Markman, E. M. (2008). Preschool children’s use of cues to generic meaning.

*Cognition*, *107*, 19–53. doi: 10.1016/j.cognition.2007.07.008

Cimpian, A., & Markman, E. M. (2011). The Generic/Nongeneric Distinction Influences How

Children Interpret New Information About Social Others. *Child Development*, *82*(2), 471–492. doi: 10.1111/j.1467-8624.2010.01525.x

Cimpian, A., Mu, Y., & Erickson, L. C. (2012). Who Is Good at This Game? Linking an Activity to a Social Category Undermines Children’s Achievement. *Psychological Science*, *23*(5), 533–541. doi: 10.1177/0956797611429803

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Cohen, A. (1999). Generics, Frequency Adverbs, and Probability. *Linguistics and Philosophy*, *22*.

Degen, J., & Goodman, N. D. (2014). Lost your marbles? the puzzle of dependent measures in experimental pragmatics. In *Proceedings of the thirty-sixth annual conference of the Cognitive Science Society.*

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games.

*Science*, *336*(6084).

Gelman, S. A. (2004). Learning words for kinds: Generic noun phrases in acquisition. In *Weaving a lexicon* (p. 445-484). MIT Press.

Gelman, S. A. (2005). *Essential child: Origins of essentialism in everyday thought.* Oxford University Press.

Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic Language in Parent-Child Conversations. *Language Learning and Development*, *4*(1), 1–31. doi: 10.1080/ 15475440701542625.Generic

Gelman, S. A., Star, J. R., & Flukes, J. E. (2002). Children’s Use of Generics in Inductive Inferences. *Journal of Cognition and Development*, *3*(2), 179–199.

Gelman, S. A., Taylor, M. G., Nguyen, S. P., Leaper, C., & Bigler, R. S. (2004). Mother-child

conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*(1), vii, 116–127. doi: 10.1111/j.1540-5834

.2004.06901001.x

Goodman, N. D., & Stuhlmu¨ller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*.

Goodman, N. D., & Stuhlmu¨ller, A. (2014). *The Design and Implementation of Probabilistic Programming Languages.* [http://dippl.org.](http://dippl.org/) (Accessed: 2015-7-17)

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language

of thought. In *The conceptual mind: New directions in the study of concepts.* MIT Press. Grice, H. P. (1975). Logic and conversation. In *Readings in language and mind.* Blackwell.

Griffiths, T. L., & Tenenbaum, J. B. (2005, December). Structure and strength in causal induction.

*Cognitive psychology*, *51*(4), 334–84. doi: 10.1016/j.cogpsych.2005.05.004 Keil, F. C. (1992). *Concepts, kind, and cognitive development.* MIT Press.

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and Linguistic Theory (SALT) 23.*

Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation.

*Synthese*.

Lee, M. D., & Wagenmakers, E. (2014). *Bayesian cognitive modeling: A practical course*. Cam- bridge: Cambridge University Press.

Leslie, S.-J. (2007). Generics and the Structure of the Mind. *Philosophical Perspectives*, *21*(1), 375–403.

Leslie, S.-J. (2008, July). Generics: Cognition and acquisition. *Philosophical Review*, *117*(1). Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie

gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. doi: 10.1081/ E-EWS

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. The MIT Press.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*(4), 339–363. doi: 10.1037/0033-295X

.90.4.339

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. doi: 10.1016/j.cognition.2011.10.004

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, *4*(2), 66–72. doi: 10.1016/S1364-6613(99)01429-1

Prasada, S., Hennefield, L., & Otap, D. (2012). Conceptual and Linguistic Representations of Kinds and Classes. *Cognitive Science*, *36*(7), 1224–1250. doi: 10.1111/j.1551-6709.2012.01254

.x

Prasada, S., Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2013, March). Conceptual distinctions amongst generics. *Cognition*, *126*(3), 405–22. doi: 10.1016/j.cognition.2012.11.010

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism.

*Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531. doi: 10.1073/pnas

.1208951109

Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of bio- pharmaceutical statistics*, *16*(4), 463–481.