A pragmatic theory of generic language

Michael Henry Tessler and Noah D. Goodman

Department of Psychology, Stanford University

E-mail: mtessler@stanford.edu, ngoodman@stanford.edu

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 formalization. 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 establishing 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.

Keywords: generic language; semantics; pragmatics; categories

Most would agree that *Swans are white*, but certainly not every swan is. This type of utterance conveys a generalization about a category (i.e. SWANS) and is known as a generic utterance (*1, 2*). It is believed that every language can express generic meaning (*3, 4*), and that generics are essential to the growth of conceptual knowledge (*5*) and how kinds are represented in the mind (*2*). Generic language is ubiquitous in everyday conversation as well as in child-directed speech (*6*), and children as young as two or three understand that generics refer to categories and support generalization (*7*). Additionally, generics are the primary way by which speakers discuss social categories, making them key to propagating stereotypes (*8–10*) and impacting motivation (*11*). Despite their cognitive centrality and apparent simplicity, a formal account of generic meaning remains elusive.

One major hurdle to formalizing generic language arises when one tries to describe the conditions by which a generic sentence is true. 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. 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 common 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.

The strong interpretation of novel generics deepens the mystery. While *Mosquitos carry malaria* suggests the generic must in some way be analogous to "some" (i.e. *Some swans are white.*), generics are often interpreted as implying the property is widespread within the kind: Listeners often interpret a novel generic as applying to *nearly all* of a category (*12*) (compare *Some swans have hollow bones* to *Swans have hollow bones*). When asked how common a property would need to be for the generic to apply, both children and adults require lower prevalence than they infer when told a generic sentence and asked how prevalent the property is (*13, 14*), suggesting that communicating with generics can exaggerate evidence.

How can generics have such flexible truth conditions while simultaneously having strong implications? In this paper we offer a mathematical model of generic language in terms of pragmatic inference of the degree of prevalence required to assert a generic. We show that this model predicts the patterns in human endorsement of familiar generic sentences and interpretation of novel generics.

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 dictate that they do) (*15*). This latter perspective emphasizes the structure of generic knowledge (*16*), and views generic utterances as the way of expressing special mental relationships between kinds and properties (*2, 17*). 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, which can result in meanings interpreted with a complex sensitivity to context (*18–20*), recently formalized in probabilisitic models of language understanding—the Rational Speech Acts (RSA) theory (*21, 22*). Perhaps then, the puzzles of generic language can be partly understood as effects of pragmatic reasoning. If this is the case, then a relatively simple semantic theory, phrased in terms of statistical regularities, could be enough to formalize generic language.

For a given kind, K (e.g. ROBINS), and a property, F (e.g. LAYS EGGS), we refer to the probability that an object of kind K has property F, that is $P(F \mid K)$, as the *prevalence* of

F within K.[1] Logical quantifiers can be described as conditions on prevalence (i.e. *some* is $P(F \mid K) > 0$, *all* is $P(F \mid K) = 1$). Extending this, it seems the simplest meaning for generic statements would be a similar threshold on prevalence: $P(F \mid K) > \tau$ (*23*). 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 otherwise a general mechanism of language not specific to interpreting generic statements. We follow the treatment of RSA applied to vague adjectives (e.g. *tall*), using an underspecified threshold criterion (*24, 25*).We imagine a hypothetical, pragmatic listener (L_1) concerned with learning the prevalence of a certain property in a certain category, $x = P(F \mid K)$, who reasons about an informative speaker (S_1), who in turn reasons about a literal listener (L_0):

$$P_{L_1}(x, \tau \mid u) \propto P_{S_1}(u \mid x, \tau) \cdot P(x) \cdot P(\tau) \quad (1)$$

$$P_{S_1}(u \mid x, \tau) \propto P_{L_0}(x \mid u, \tau)^{\lambda} \quad (2)$$

$$P_{L_0}(x \mid u, \tau) \propto \delta_{[\![u]\!](x,\tau)} P(x). \quad (3)$$

The pragmatic listener L_1 (Eq. 1) has heard an utterance u, which is either a generic statement (e.g. *Robins lay eggs.*) or an uninformative *null* utterance. L_1 is trying to resolve how common F is within K (i.e. $x = P(F \mid K) \in [0,1]$); she does so by considering both what she knows about property F in general—the probability distribution $P(x)$ over prevalence of F—and her model of the speaker S_1 as an informed and helpful interlocutor (Eq. 2). L_1 knows that the meaning of the generic utterance, if used, depends on a prevalence threshold, π, but does not know what this threshold is: $\tau \sim$ Uniform([0,1]). Speaker S_1 has a particular π in mind, and is trying to be informative about the prevalence x to the listener in his head—an idealized literal

[1] Because we aim to explain the psycholinguistics of generics, we are generally interested in the subjective probability, not the actual frequency in the world.

listener (L_0, Eq. 3). The literal listener has access to the threshold π, and simply restricts her prior beliefs to situations where the truth-functional denotation of the utterance, $[\![u]\!]$, is true. The generic utterance has a threshold semantics, $[\![\text{K F}]\!](x,\tau) = x > \tau$; while the null utterance is always true, $[\![null]\!](x,\tau) = T$.[2]

Example posterior distributions for $P_{L_1}(x,\tau \mid u)$ 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, what prevalence is a listener likely to infer? We can now imagine a speaker S_2 who reasons about this type of listener:

$$P_{S_2}(u \mid x) \propto \int_{\theta} P_{L_1}(x,\tau \mid u) \tag{4}$$

Speaker S_2 considers the thought-processes of the listener L_1 (Eq. 1) and decides if the

[2] All results reported are similar when the alternative utterance is stipulated to be the negation—*It is not the case that robins lay eggs.*—or the generic of the negation—*Robins do not lay eggs.*—as opposed to the *null* utterance.

generic is a good (even if 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 vague utterance means (i.e. doesn't have access to the threshold τ), but knows that L_1 will have to think about it, and integrates over the likely values she'll consider. We use S_2 (Eq. 4) to model felicity or truth judgments (*26*).

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 predictions, 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) that give rise to interesting generics (described below). Participants ($n = 57$) reported their beliefs about the prevalence of target properties for many different animal kinds, both pre-specified by the experiment (e.g. ROBINS, MOSQUITOS) and self-generated by the participant (see Supplement Section A2 for complete details).[3] Figure 1 (insets) shows five example elicited prior distributions.

While each $P(x)$ is a single distribution on prevalence, it may be highly structured as the result of deeper conceptual knowledge. For instance, if participants believe that some kinds have a causal mechanism 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. (*27*)). Assuming that *when the property is possible*, its prevalence is distributed according to a Beta distribution with mean γ and concentration ξ: $P(x) = \theta \cdot \beta(\gamma, \xi)(x) + (1-\theta) \cdot \delta_{x=0}$.[4] Thus, θ is the potential of a property

[3]Our experiments stay within the animal kingdom because we expect there to be considerably less variability in participants' beliefs about animals than about other types of categories (e.g. social categories).

[4]This is similar in spirit to Hurdle Models of epidemiological data where the observed counts of zeros is substantially greater than one would expect from standard models of count data such as the Poisson model (e.g. adverse events to vaccines) (*28*).

to be present in a kind and γ is the mean prevalence of the property among the kinds with the potential to have it. This statistical model reproduced the prior elicitation data with a low reconstruction error ($r^2 = 0.94$; see Supplement Section A3; Supplement Figure 1), strongly supporting the assumption of a structured prior.

Using this structure, we can explore the prevalence priors for our 21 properties: Figure 1 shows the estimated mixture-parameter θ (the potential of the property to be present in a kind) and the mean prevalence when the property is present, γ. We see significant diversity among these properties in both dimensions. This diversity matters because the model predictions for generic interpretation and felicity depend on the shape of the prior distribution (see insets for example interpretation posteriors). For example, lowering θ increases the relative probability mass at 0% and works to relax the truth conditions by making a lower threshold more informative; increasing γ will tend to make truth conditions stricter, by reducing the range of prevalence values that are higher than the prior expectation.

Generics have flexible truth conditions

We tested the degree to which the S_2 model, Eq. 4, coupled with the empirically-elicited $P(x)$ predicted that a given category–property pair (e.g. ROBINS and LAYS EGGS) would result in a felicitous generic (e.g. *Robins lay eggs.*).[5] We collected human judgments ($n = 100$) about the acceptability of thirty generic sentences. These sentences cover a range of conceptual distinctions (*30*) including 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 (“definitely acceptable”, “definitely unacceptable”, and “uncertain”) and include both acceptable and unacceptable generics with low-prevalence, medium-prevalence, and high-

[5]The model, once $P(x)$ is fixed, has one free parameter: the speaker rationality parameter λ in Eq. 2. Because λ is not of theoretical interest here, it is marginalized in accord with Bayesian data analysis principles (*29*).

prevalence properties.

From the prevalence-prior data 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 Supplement 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; Supplement Figure 2). This is not surprising given the inclusion of high-prevalence true generics (e.g. *Leopards have spots.*) and low-prevalence false generics (e.g. *Leopards have wings.*) in our stimuli. However, large deviations from a purely target-category prevalence account remain: 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 ($r^2_{Q2,3} = 0.029$; MSE = 0.110).

The probabilisitic pragmatics model, using empirically measured priors, does a much better job of explaining the truth judgments ($r^2 = 0.981$; MSE=0.003; Figure 2, Supplement 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, $r^2_{Q2,3} = 0.955$; MSE=0.005; Figure 2, intermediate shades). The probabilistic pragmatics model explains the puzzling flexibility of generic truth-conditions as the result of communicative pressures operating over diverse prior beliefs about the properties.

Interpreting novel generics

Perhaps the most important role for generic language is to provide information about new or poorly understood categories. To study this we must look to *interpretation* of novel generics (e.g. (*12, 13*)). The pragmatic listener model (Eq. 1) describes interpretation of a generic utterance, without previous knowledge of the prevalence within the target-kind. With an uncertain meaning and the pressure to be informative, interpretation is heavily driven by *a priori* beliefs about properties. Classic work in generalization suggests beliefs about the prevalence of properties differ by type of property, including relatively fine distinctions between properties that are all biological in nature (*31*). We used 40 different properties to explore a wide range of *a priori* beliefs about prevalence. These items make up five categories of properties: biological parts (e.g. HAS CLAWS), color adjectives about parts (e.g. HAS GREEN FEATHERS), vague adjectives about parts (e.g. HAS SMALL WINGS), common accidental or disease states (e.g. HAS WET FUR) and rare accidental or disease states (e.g. HAS SWOLLEN EARS)[6]. $P(x)$ was measured empirically ($n = 40$, see Supplement Section C), and the most likely priors were inferred using the same structured, statistical approach used for the familiar generics experiment.

Biological properties are expected to be *a priori* more prevalent (when present) than accidental properties, with fine-grained differences even among types of biological and accidental properties (Figure 3). The shapes of the prevalence distributions for different properties are importantly diverse (Figure 3 insets). Biological properties ("biological parts", "vague parts", "color parts") are bimodal with peaks at 0% and near-100% prevalence. After hearing the generic, the listener (L_1 model, Eq. 1) updates this distribution to a concave posterior peaked at 100% (Figure 3; red, blue and green insets). Accidental properties ("rare" and "common") follow unimodal prior distributions and update to convex posterior distributions, reflecting a higher degree of residual uncertainty after hearing the generic utterance as compared to the biological properties.

We compared the interpretations of the pragmatic listener in Eq. 1 to human judgments ($n = 40$) about the likely prevalence of the property after hearing a generic about an unfamiliar

[6]The distinction between common and rare accidental properties was determined empirically by analyzing the data by item, and performing a median split based on the *a priori* mean prevalence when present, γ, of the property.

kind (e.g. *Lorches have green feathers.*). Human prevalence judgments after reading the generic were affected by the type of property and its corresponding *a priori* mean prevalence when present γ: The more prevalent a property is expected to be *a priori*, the stronger the implications of a generic statement ($\beta = 0.57; SE = 0.08; t(39) = 7.12; p < 0.001$)[7]. The pragmatic listener model closely aligns with human judgments, displaying the same sensitivity of interpretation to the property information ($r^2(40) = 0.89$, MSE=0.006; Figure 4).

There is a surprising décolage between the truth conditions and interpretations of generic language: Novel generics are interpreted as implying a higher prevalence than that required to assent to the exact same generic (*13*). Our model predicts that this effect should hold, but only for properties with most extreme prior beliefs (particularly mean prevalence when present, γ; see Figure 5, right. To explore this prediction, we recruited participants ($n = 40$) to help determine the average prevalence at which a speaker would assent to the generic. We told participants one of five frequencies for each type of property (e.g. *50% of lorches have green feathers.*), and then asked for truth judgments of the corresponding generic sentence (e.g. *Lorches have green feathers.*). For both behavioral data and model predictions (Eq. 4) we computed the average prevalence that led to an assenting judgment, for each property type and participant, following the procedure used by (*13*) (see Supplement Section D for more details). The speaker S_2 model did not make appreciably different predictions as to the average prevalence to accept the generic, consistent with the intuition that generics are acceptable for a wide range of prevalence levels. A similar absence of a gradient was observed in the human data ($\beta = 2.82; SE = 4.02; t(39) = 0.70; p = 0.49$; Figure 5, dotted lines). Interpretations of generic utterances are stronger than their average truth conditions for biological properties but not for accidental properties (Figure 5); the extent of the difference is governed by prior property knowledge (γ, inferred from prior elicitation data). The listener and speaker pair of models predicts human endorsements and

[7]These statistics are the result of a mixed-effects linear regression with a maximal mixed-effect structure: Random by-participant effects of intercept and slope

interpretations of novel generic utterances well ($r^2(10) = 0.921$, MSE = 0.004).

Conclusions

We evaluated a theory of generic language derived from general principles of pragmatic language understanding and 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 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 (*2, 30, 32, 33*). 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. Where in the prevalence-based semantics could such conceptual distinctions come into play? Beliefs about prevalence in our approach are represented as probability distributions; a framework that is useful for representing rich, structured knowledge of the world (*34*). Indeed, we found that empirical prevalence distributions are structured, perhaps reflecting intuitions about causal mechanisms underlying different properties. It is plausible that richer conceptual knowledge also influences these distributions, including higher-order conceptual knowledge about the nature of properties and categories (*32, 35*). However, 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 distributions. This suggests a number of further tests of our model.

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 clearly as possible? To the contrary, such underspecification can be efficient, given that context can be used to resolve the uncertainty (*36*). 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 distribution, but many properties of social categories may as well (*9, 37, 38*).

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 generalizations to each other and down through generations. The model presented here gives one explanation of how we do so, providing a computational perspective on how category generalizations are conveyed and how beliefs play a central role in understanding language.

References and Notes

1. G. N. Carlson, Reference to kinds in english, Ph.D. thesis, University of Massachusetts, Amherst (1977).

2. S.-J. Leslie, *Philosophical Review* **117** (2008).

3. L. Behrens, *Linguistics* **43**, 275 (2005).

4. G. N. Carlson, F. J. Pelletier, *The generic book.* (Chicago University Press, Chicago, IL, 1995).

5. S. A. Gelman, *Learning words for kinds: Generic noun phrases in acquisition* (MIT Press, 2004), pp. 445–484.

6. S. A. Gelman, P. J. Goetz, B. W. Sarnecka, J. Flukes, *Language Learning and Development* **4**, 1 (2008).

7. A. Cimpian, E. M. Markman, *Cognition* **107**, 19 (2008).

8. S. A. Gelman, M. G. Taylor, S. P. Nguyen, C. Leaper, R. S. Bigler, *Monographs of the Society for Research in Child Development* **69**, vii, 116 (2004).

9. M. Rhodes, S.-J. Leslie, C. M. Tworek, *Proceedings of the National Academy of Sciences* **109**, 13526 (2012).

10. S.-J. Leslie, A. Cimpian, M. Meyer, E. Freeland, *Science* **347**, 262 (2015).

11. A. Cimpian, *Developmental psychology* **46**, 1333 (2010).

12. S. A. Gelman, J. R. Star, J. E. Flukes, *Journal of Cognition and Development* **3**, 179 (2002).

13. A. Cimpian, A. C. Brandone, S. A. Gelman, *Cognitive science* **34**, 1452 (2010).

14. A. C. Brandone, S. A. Gelman, J. Hedglen, *Cognitive science* pp. 1–28 (2014).

15. G. N. Carlson, *The Generic Book*, G. N. Carlson, F. J. Pelletier, eds. (University of Chicago Press, Chicago, 1995), chap. 4, pp. 224–38.

16. S. Prasada, *Trends in Cognitive Sciences* **4**, 66 (2000).

17. S. Prasada, L. Hennefield, D. Otap, *Cognitive Science* **36**, 1224 (2012).

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

19. H. P. Grice, *Logic and conversation.* (Blackwell, 1975), chap. 1, pp. 1–10.

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

21. M. C. Frank, N. D. Goodman, *Science* **336** (2012).

22. N. D. Goodman, A. Stuhlmüller, *Topics in Cognitive Science* (2013).

23. A. Cohen, *Linguistics and Philosophy* **22** (1999).

24. D. Lassiter, N. D. Goodman, *Semantics and Linguistic Theory (SALT) 23* (2013).

25. D. Lassiter, N. D. Goodman, *Synthese* (2015).

26. J. Degen, N. D. Goodman, *Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Science Society* (2014).

27. T. L. Griffiths, J. B. Tenenbaum, *Cognitive psychology* **51**, 334 (2005).

28. C. E. Rose, S. W. Martin, K. A. Wannemuehler, B. D. Plikaytis, *Journal of biopharmaceutical statistics* **16**, 463 (2006).

29. M. D. Lee, E. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge, 2014).

30. S. Prasada, S. Khemlani, S.-J. Leslie, S. Glucksberg, *Cognition* **126**, 405 (2013).

31. R. E. Nisbett, D. H. Krantz, C. Jepson, Z. Kunda, *Psychological Review* **90**, 339 (1983).

32. S. A. Gelman, *Essential Child: Origins of Essentialism in Everyday Thought.* (Oxford University Press, 2005).

33. S.-J. Leslie, *Philosophical Perspectives* **21**, 375 (2007).

34. N. D. Goodman, J. B. Tenenbaum, T. Gerstenberg, *Concepts in a probabilistic language of thought* (MIT Press, 2015), chap. 1, pp. 1–10.

35. F. C. Keil, *Concepts, kind, and cognitive development.* (MIT Press, 1992).

36. S. T. Piantadosi, H. Tily, E. Gibson, *Cognition* **122**, 280 (2012).

37. A. Cimpian, E. M. Markman, *Child Development* **82**, 471 (2011).

38. A. Cimpian, Y. Mu, L. C. Erickson, *Psychological Science* **23**, 533 (2012).

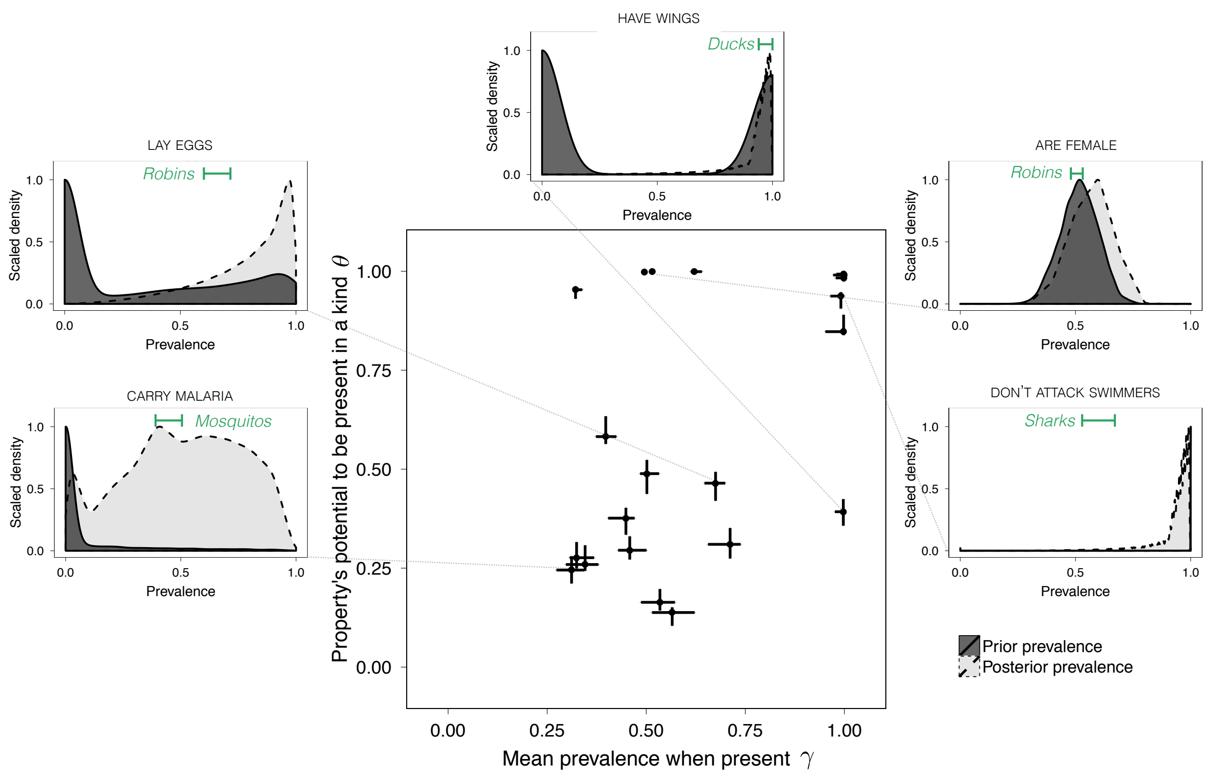

Figure 1: Prevalence prior distributions empirically elicited for twenty-one animal properties. Prior distributions are summarized by θ—-a property's potential to be present in a category—-and γ—-the mean prevalence when it is possible for the property to be present in a category. Inset plots display 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.

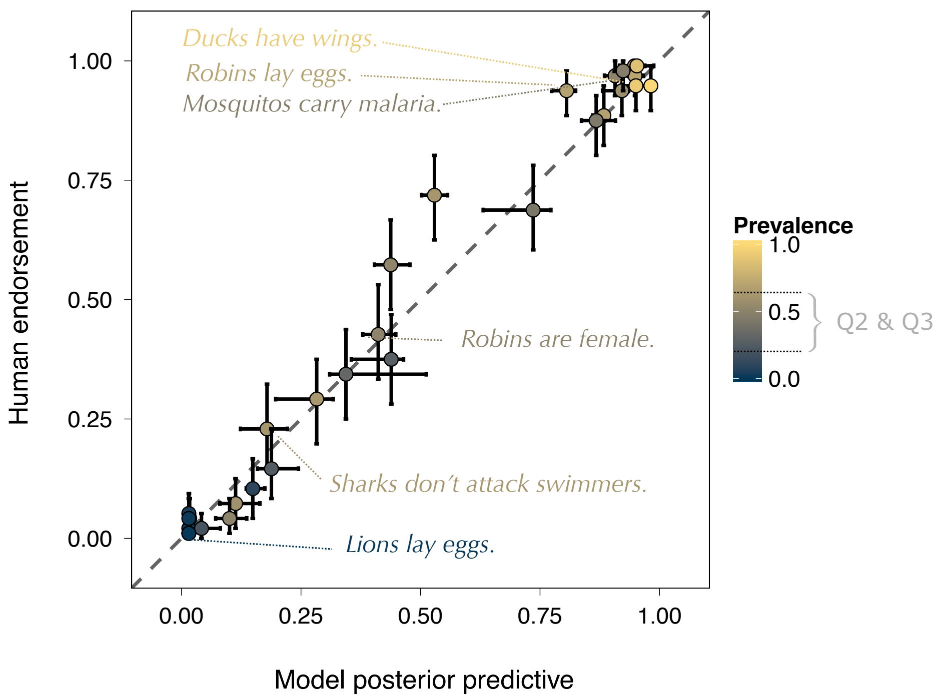

Figure 2: Human acceptability judgments and model predictions for thirty generic utterances about familiar animals and properties. Color denotes target-category prevalence of the property, with darker colors indicating lower prevalence. Intermediate prevalences (quartiles 2 & 3) are in intermediate shades (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.

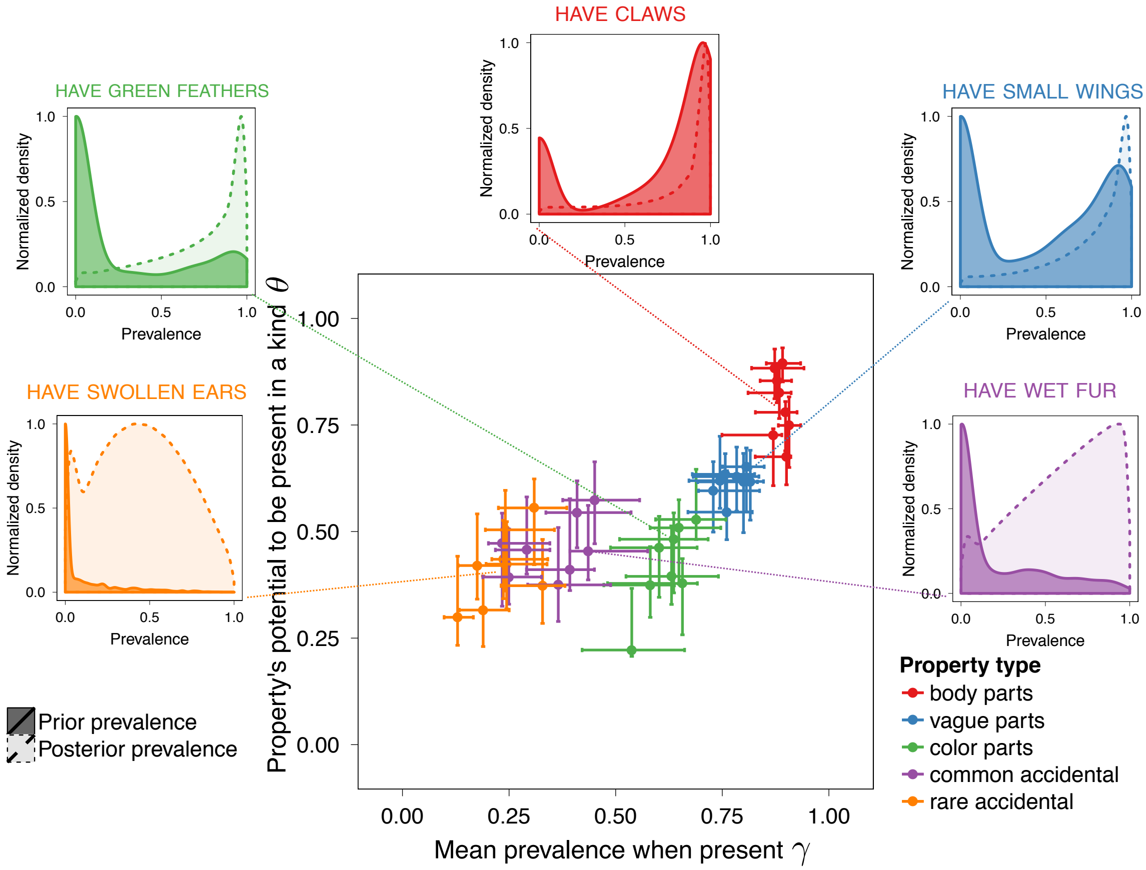

Figure 3: Prevalence prior distributions empirically elicited for 40 animal properties. Parameters of the structured statistical model—θ and γ—reveal quantitative differences in beliefs about the prevalence of conceptually different types of properties (scatterplot). Inset plots show differences in shapes between biological properties (red, green, blue; bimodal) and accidental properties (orange, purple; unimodal). These differences give rise to the variability of interpretations of generic utterances. Error bars denote Bayesian 95% credible intervals.

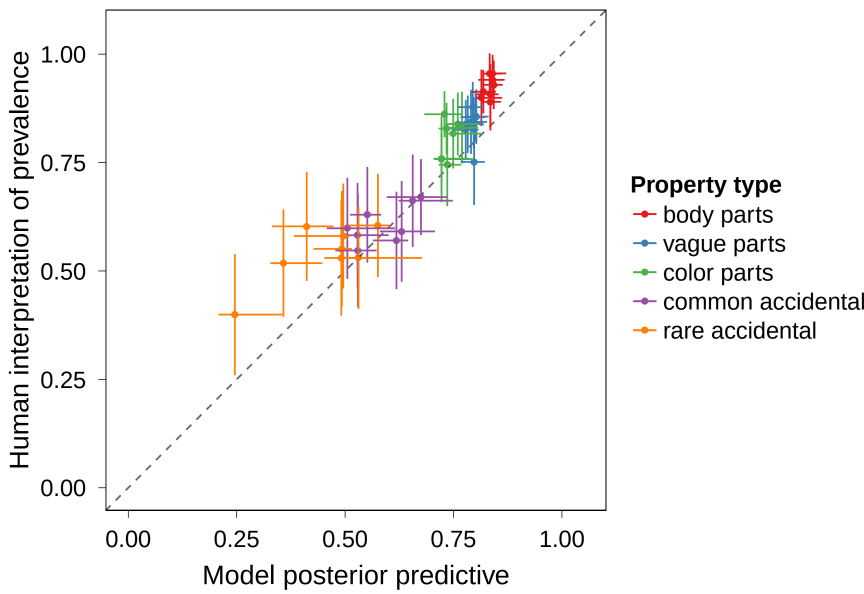

Figure 4: Human interpretation of prevalence upon hearing a generic compared with the L_1 model posterior predictive. Participants display graded endorsements of generics in terms of prevalence based on type of property (which is also associated with *mean prevalence when present* γ, see Figure 3). The model displays the same variability of interpretation, producing strong interpretations for generics of biological properties (red, blue, green) and weaker interpretations of generics of accidental properties (purple, orange). Error bars denote bootstrapped 95% confidence intervals for the data and Bayesian 95% credible intervals for the model.

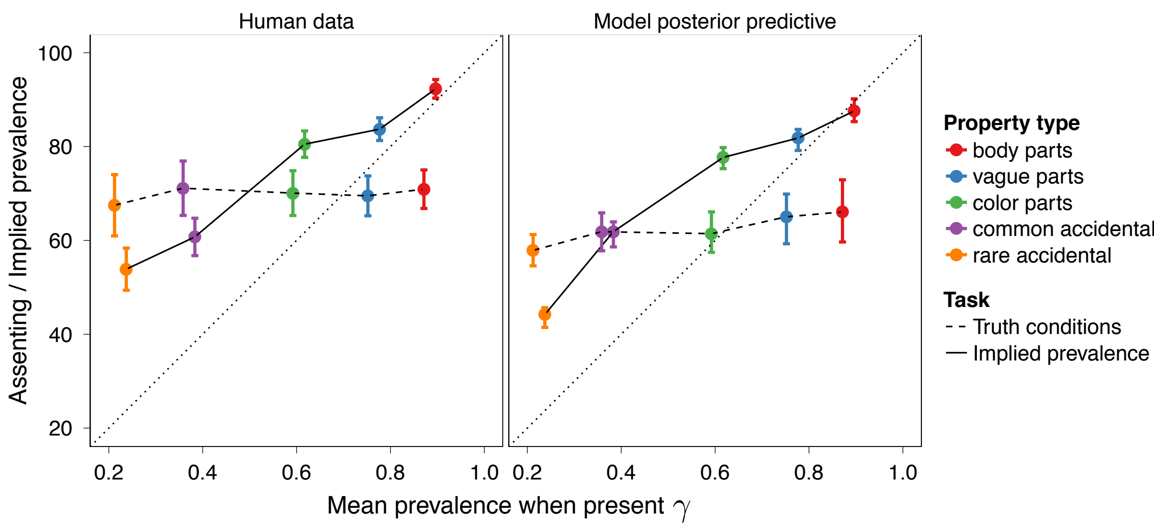

Figure 5: Human judgments and model predictions of prevalence implied by novel generic utterances (implied prevalence task; solid line) and average prevalence that leads to an acceptable generic utterance (truth conditions task; dotted line) as it relates to the *a priori* mean prevalence when present 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 a high proportion of the category has the property, for both human participants 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 implications 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. Error bars denote bootstrapped 95% confidence intervals for the data and Bayesian 95% credible intervals for the model.