Learning about Inductive Potential from Generic Statements

Marianna Y. Zhang
Department of Psychology
New York University
marianna.zhang@nyu.edu

Sarah-Jane Leslie
Department of Philosophy
Center for Statistics and Machine Learning
Princeton University
sjleslie@princeton.edu

Marjorie Rhodes*
Department of Psychology
New York University
marjorie.rhodes@nyu.edu

Mark K. Ho*
Department of Psychology
New York University
mark.ho@nyu.edu

Abstract

Generic statements (e.g., "Climbers drive Subarus") shape what categories people take as meaningful bases for generalization. After hearing a generic, people not only learn about the *prevalence* of a feature in a category (e.g., how many climbers drive Subarus), but also about the *inductive potential* of the category (e.g., that climbers share many features). Here, we propose a Bayesian model of how people infer inductive potential from generics. To test our model, we introduced adults ($n = 284$) to nothing (baseline), or to members of a novel social category accompanied by *generic* statements (e.g., "Zarpies sleep in trees") or *specific* statements (e.g., "This Zarpie sleeps in trees"). We then measured inferred inductive potential by eliciting the prevalence of *novel* features. As predicted, generics increased while specific statements decreased the category's inductive potential, relative to baseline. Our account explains how generics facilitate the cultural transmission of social categories believed to be bases for generalization.

Keywords: social cognition; categories; generics; inductive potential; rational speech act theory; hierarchical Bayesian modeling

Introduction

People rely on categories to make sense of the social world (Murphy, 2004; Smith & Medin, 1981). Consider the following categories of people: New Yorkers, librarians, Geminis, climbers, pedestrians, and bystanders. These categories all vary in their *inductive potential* – how well each category supports generalization across members of the group.[1] For example, in the US context, the category "climbers" is often thought of as a relatively coherent and inductively potent category whose members share many similarities (e.g., in personality, lifestyle). As a result, if you meet a climber and learn some fact about them, you might readily generalize that fact to climbers in general, because the category "climber" is seen as providing a solid basis for generalization. In contrast, if you see someone on the sidewalk (i.e., a pedestrian) and learn some fact about them, you are unlikely to generalize that fact to other pedestrians, since the category of "pedestrian" is a relatively minimal category with low inductive potential.

How do people learn the inductive potential of categories? Beliefs about social categories, including about their inductive potential, are acquired and refined not only through direct observations—e.g., observing a climber driving a Subaru—but also via mechanisms of cultural transmission—e.g., being told "climbers drive Subarus." Such statements, called *generic language*, serve as a powerful source of beliefs about categories. Generic statements are frequent in everyday speech across a wide variety of languages (Carlson & Pelletier, 1995; Gelman et al., 2008; Leslie, 2008), and are especially frequent when speakers talk about categories they *essentialize*, that is, categories considered to be natural and high in inductive potential. For example, adults told that a novel social category is a distinct kind of people with distinct biological and cultural traits subsequently produced more generic statements about the social category (Rhodes et al., 2012). Generic statements are thus readily available as sources of information about categories.

When listeners hear generic language, they not only learn about the prevalence of the mentioned feature, but also about the category as a whole: that the category has high inductive potential. For example, when hearing a generic statement about a novel category (e.g., "Zarpies eat flowers"), versus a specific statement about an individual (e.g., "This Zarpie eats flowers"), children and adults not only expect a new category member to exhibit the mentioned feature (*eating flowers*) (Pronovost & Scott, 2022), but also expect any *novel* features of the new category member to be broadly held by category members (e.g., if the new Zarpie is scared of shadows, then relatively many Zarpies must be scared of shadows) (Benitez et al., 2022). In other words, the category is taken as having high *inductive potential*, the ability to support wide-ranging generalizations. People therefore take generics as evidence that the category itself is highly coherent—i.e., that it provides a strong basis for generalization.

Why does generic language suggest something about the category itself, beyond the particular feature mentioned? While generic language (e.g., "Climbers drive Subarus.") clearly indicates a link between a feature and a category (e.g., *driving Subarus* and climbers), it is not obvious how generic language might suggest something about the nature of the category itself (e.g., that the category "climbers" is high

*Joint senior authorship.

[1] Inductive potential can feed into other ideas, such as how similar category members are seen to each other (e.g., climbers are thought of as sharing many similarities), how informative category membership is thought to be (e.g., knowing that someone is a climber is thought to be revealing of what they are likely like), and how essentialized a category is (e.g., whether climbers are characterized by some innate essence). As a result, inductive potential is a component of other concepts in the literature, such as how "kind-like" (Noyes & Keil, 2019) or how "entatitive" (Yzerbyt et al., 2001) a category is perceived to be.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

in inductive potential). Although this phenomenon is well-established in the literature (Rhodes et al., 2024), it has to date eluded formal explanation. Understanding precisely how generic statements lead to inferences about category structure requires considering how language interpretation interacts with concept learning and leads to inferences about novel features. Our goal is to develop a computational model that allows us to precisely specify the contribution of each of these components.

Here, we present a computational model that successfully accounts for how language like generic statements affects inferences about the nature of categories. While existing computational models capture related phenomena, such as how generic language leads to context-sensitive inferences about the prevalence of the feature mentioned in the generic (Tessler & Goodman, 2019), they do not capture how generic language shapes inferences about the category as a whole, e.g., about inductive potential, which could support inferences about features *beyond those mentioned in the generic*. To capture such inferences, we develop a model that combines hierarchical Bayesian inference with a model of the semantics and pragmatics of generics (Kemp et al., 2007). We then empirically validate the predictions in an experiment with adult participants.

Computational Model

Our account combines hierarchical Bayesian inference with the Rational Speech Act (RSA) framework (N. D. Goodman & Frank, 2016; Kemp et al., 2007). We first introduce a basic formalization of the meaning of generic and specific statements in the context of hierarchical inference, before turning to modeling the effects of pragmatic inference. Then, we present model predictions about how people learn about the inductive potential of categories from generic versus specific statements as hypotheses for our experiment.

Setting up the basic meaning of statements

Imagine learning about a novel social category, *Zarpies*. Over time, you may observe individual category members (e.g., various Zarpies) and features of each individual Zarpie (e.g., one Zarpie *eats flowers*, another Zarpie *lives in caves*, a third Zarpie *is afraid of bugs*, ...). An observation of an individual Zarpie may be accompanied by someone else observing the same scene and uttering a generic statement (e.g., "Zarpies eat flowers") or specific statement (e.g., "This Zarpie eats flowers"). What might you learn about Zarpies from such observations and utterances? In particular, what is the inductive potential of Zarpies as a kind: are they relatively high (similar to e.g., "climbers") or relatively low (similar to e.g., "pedestrian") in inductive potential?

To begin our formalization of this task, we define a set of features under consideration, $\mathbb{F}$ (e.g., *eats flowers*, *likes to sing*, ...). $\mathbb{F}$ is built up as the set of all features observed so far among individuals x_i, all of whom are members of the category k (e.g., Zarpies). We task the model with inferring which features among the observed features $\mathbb{F}$ are *kind-linked*; this subset we denote as $\mathcal{F}_k$. The inductive potential of a category, θ, is formalized as *coherence*, an overhypothesis that designates the prior probability that any single feature is kind-linked, i.e., the probability that any feature f in $\mathbb{F}$ is a member of $\mathcal{F}_k$: $\theta = P(f \in \mathcal{F}_k)$.[2]

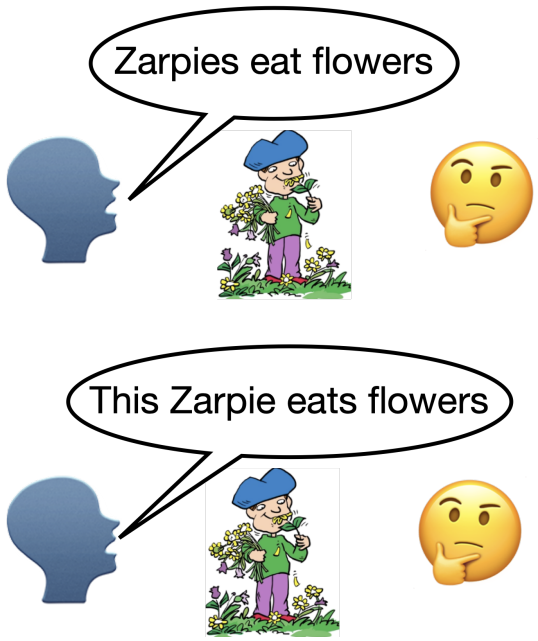

Figure 1: A speaker (left) talking to a listener (right) says either a generic (top) or specific (bottom) statement after jointly observing an individual Zarpie with some feature.

Next, we can use our concept of kind-linked features to introduce a high-level formalization for the meaning of generic statements. Specifically, we treat generic statements (e.g., "*k*s are *f*") as true if and only if the mentioned feature f is in $\mathcal{F}_k$, the set of features kind-linked to the mentioned category k. (Note that our model is agnostic to what exactly it means for a feature to be kind-linked. Thus, different analyses of the semantics of generics (e.g., Leslie, 2008; Tessler & Goodman, 2019) are all compatible with this model.)

In contrast, specific statements (e.g., "This k is f") are true if the mentioned feature f is in the set of observed features of x_i, the individual being observed. Formally, the meaning of an utterance u_i in the context of an individual x_i, a member of category k with kind-linked features $\mathcal{F}_k$, is:

$$[\![u_i]\!](\mathcal{F}_k, x_i) = \begin{cases} \mathbb{1}[f \in \mathcal{F}_k] & \text{if } u_i = \text{"}k\text{s are } f\text{"} \\ \mathbb{1}[f \in x_i] & \text{if } u_i = \text{"This } k \text{ is } f\text{"} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $\mathbb{1}[X]$ is an indicator function that evaluates to 1 when X is true and 0 otherwise.

[2]We use stochastic memoization (N. Goodman et al., 2014) to ensure the same feature is treated with the same kind-linked status across observations, allowing the feature set to grow dynamically as new features are encountered.

Learning from generic statements

To model people's representations of categories, including their inductive potential, we turn to hierarchical Bayesian inference. Hierarchical Bayesian inference (Kemp et al., 2007) provides a natural framework for expressing the relationships between individuals and their features (e.g., whether an individual Zarpie eats flowers), hypotheses (e.g., whether Zarpies in general eat flowers), and overhypotheses (e.g., how likely a feature is to be a kind-linked feature of Zarpies).

Here, inductive potential is formalized as *coherence*, the prior probability that a feature is kind-linked. We posit a single coherence value for a kind, which acts as an *overhypothesis*, i.e., a default hypothesis, about whether a new feature will be kind-linked, a hypothesis which may be further modified by feature-specific knowledge. Being kind-linked is binary. Although there are many accounts of what it means for a feature to be kind-linked, here we simply assume that a feature that is kind-linked has greater prevalence among members of the kind than if it were not kind-linked (though crucially, the change in prevalence can be context-dependent; see (Novoa et al., 2023; Tessler & Goodman, 2019)).

Formally, inferring the coherence of a category is analogous to guessing the bias of a coin. If a coin lands heads 16 out of 16 times, the coin is likely to be heads-biased. Analogously, if you hear 16 different generics about the same category (e.g., "Climbers drive Subarus," "Climbers have short nails", etc.), you can infer the category is likely high in coherence.

Taken together, high coherence allows one to infer that a feature of an individual kind member is likely also a feature of other kind members. In this hierarchical structure, beliefs about the general inductive potential of the category (e.g., how similar do Zarpies tend to be to one another?) govern and interact with beliefs about the prevalence of features among category members (individual Zarpies and their features). As a result, hierarchical Bayesian inference provides a way to model a kind's inductive potential.

Next, we consider how a speaker and listener communicate about categories. When a speaker and a listener jointly observe an individual category member (e.g., a Zarpie), the speaker may choose a generic utterance (e.g., "Zarpies eat flowers") to convey to the listener that that feature of the individual category member is kind-linked. Alternatively, the speaker may choose a specific statement ("This Zarpie eats flowers") to simply note that the individual Zarpie has the feature. A listener then reasons about the speaker's choice of utterance to infer a representation of the category, specifically jointly inferring the category's inductive potential (e.g., how coherent Zarpies are as a kind) and what observed features are kind-linked.[3].

Formally, we define a *literal listener* who reasons jointly about a category's inductive potential $\theta = P(f \in \mathcal{F}_k)$ and features linked to the category $\mathcal{F}_k$ from observed category members $\mathbf{x} = \langle x_1, x_2, ..., x_n \rangle$ and accompanying utterances $\mathbf{u} = \langle u_1, u_2, ..., u_n \rangle$:

$$\text{Lit}(\mathcal{F}_k, \theta \mid \mathbf{x}, \mathbf{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i) \quad (2)$$

Here, the prior $P(\theta)$ represents the listener's prior beliefs about the coherence of the category, θ; the likelihood $P(\mathcal{F}_k \mid \theta)$ represents the likelihood of a given set of kind-linked features $(\mathcal{F}_k)$ given the category's inductive potential (θ); and the product term represents the semantic likelihood of producing the utterances heard $\mathbf{u}$ about individuals seen $\mathbf{x}$ (as defined in Equation 1).

Learning from specific statements

Having formally specified a literal interpretation for generic and specific statements, we can now model the *pragmatic* interpretation of speaker utterances. To do so, we use the Rational Speech Act framework (RSA), which formalizes pragmatic inferences by a listener who performs Bayesian inference, assuming that a speaker is intentionally choosing utterances to align the listener's beliefs with the speaker's (N. D. Goodman & Frank, 2016).

We consider a *speaker* whose goal is to communicate which, among the features of the mutually observed individual x_i, are features of the kind $(\mathcal{F}_k \cap x_i)$. To communicate the kind-linked observed features $\mathcal{F}_k$ to the literal listener, the speaker selects an utterance u_i, either a generic statement or a specific statement, and considers what kind-linked features the literal listener would infer from that $(\mathcal{F}_k^*)$.[4]. Such a speaker has the following utility function:

$$\text{Utility}(u_i, x_i, \mathcal{F}_k^*) = \sum_{\mathcal{F}_k} \text{Lit}(\mathcal{F}_k \mid x_i, u_i) \cdot \text{Similarity}(\mathcal{F}_k^* \cap x_i, \mathcal{F}_k \cap x_i) \quad (3)$$

where $\text{Lit}(\mathcal{F}_k \mid x_i, u_i) = \int \text{Lit}(\mathcal{F}_k, \theta \mid x_i, u_i) d\theta$ and the similarity between the intended and inferred set of kind-linked features is calculated via Jaccard similarity[5]. The speaker can then be defined as trying to maximize such a utility function:

$$\text{Sp}(u_i \mid \mathcal{F}_k^*, x_i) \propto \exp\left\{\beta \cdot \text{Utility}(u_i, x_i, \mathcal{F}_k^*)\right\} \quad (4)$$

where β is a rationality or inverse temperature parameter that controls the degree to which the speaker tries to maximize utility.

Finally, to capture pragmatic inferences, we can define a pragmatic listener that reasons about category inductive potential and kind-linked features by reasoning about what a

[3] See Degen et al., 2015; Kravtchenko and Demberg, 2022 for an analogous approach where the listener revises their prior beliefs about the state of the world to accommodate something unexpected in what the speaker said. Here, the listener revises their beliefs about the kind, including its coherence and linked features, from what the speaker said.

[4] The predictions are relatively unchanged if silence, which is always true, is added as a third alternative option.

[5] The Jaccard similarity ranges between 0 and 1 for two non-empty sets A and B: $\text{Jaccard}(A,B) = \frac{|A \cap B|}{|A \cup B|}$. If both sets are empty, the similarity is 1.

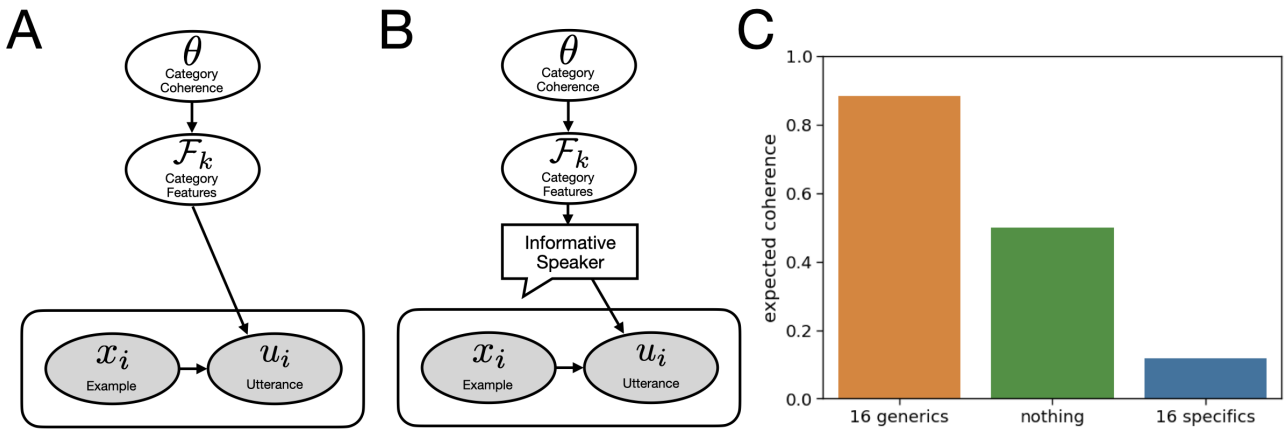

Figure 2: Model of learning from generics. (A) Graphical model of literal listener. Greyed out nodes indicate observations. (B) Graphical model of pragmatic listener. (C) Expected value for coherence from pragmatic listener after hearing 16 generics, nothing, or 16 specifics.

speaker is likely to say, rather than what is semantically true. Formally, we simply replace the literal semantic likelihood in the literal listener (Equation 2) with the speaker likelihood (Equation 4) to define the pragmatic listener:

$$\text{Prag}(\mathcal{F}_k, \theta \mid \mathbf{x}, \mathbf{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i) \quad (5)$$

As a result, hearing generic versus specific statements about a category respectively *increase* and *decrease* the pragmatic listener's inferred coherence of the category. To demonstrate this consequence, we modeled $P(\theta)$ with a uniform distribution, set the rationality parameter β to 20, and calculated $\text{Prag}(\mathcal{F}_k, \theta \mid \mathbf{x}, \mathbf{u})$ for hearing 16 generic statements, 16 specific statements, or nothing (which effectively returns the prior). As shown in Figure 2C, hearing only generics causes the pragmatic listener to infer that the category has relatively high coherence compared to the prior, while hearing only specifics depresses the inferred coherence from the prior. In short, the pragmatic listener learns from language how meaningful and coherent Zarpies are as a category.

The latter prediction about the interpretation of specific statements is a pragmatic effect that only emerges in the pragmatic listener model, and not the literal listener model. To a literal listener, specific statements are completely uninformative, since the semantics of specific statements have nothing to do with the category. However, since the pragmatic listener assumes that the speaker is choosing utterances to be informative about kind-linked features, the pragmatic listener interprets the specific utterance to mean that the generic utterance was *not* chosen, and consequently infers that the mentioned feature is *not* kind-linked and that the category has *low* inductive potential.

To compare model predictions of coherence to human judgments, we asked participants in the following experiment to imagine a new category member with a novel feature and elicited judgments of the prevalence of the novel feature among the category.

To link the model's predictions of coherence to estimates of a particular feature's prevalence, we used a linking function that estimates prevalence as a weighted combination of the feature's prevalence if it were kind-linked, versus the feature's prevalence if it were not kind-linked, with the weight being the kind's coherence (similar to the linking function used by Tessler and Goodman, 2019). We fit two separate beta distributions for each feature in the following experiment, one for its kind-linked prevalence and another for its non-kind-linked prevalence. In sum, features are presumed to have higher prevalence if they are kind-linked than if they were not kind-linked.

As a result, based on the pragmatic listener model, we expect that (1) hearing generic statements about a social category will lead people to infer that the category has *high* inductive potential, and therefore, novel features will be thought to be higher in prevalence among category members than if one had not heard such statements. Our model also predicts that (2) hearing an alternative statement, such as a specific statement (e.g., "This Zarpie eats flowers"), will lead listeners to infer that the category has *low* inductive potential, and therefore, novel features will be lower in prevalence among category members, compared to baseline.

The model implementation and a demo of the model predictions are available on this project's Github repository.

Experiment

We aim to test two predictions made by the model in an experiment with adults. The first is a replication of an effect previously reported in the literature, in which hearing generic statements leads listeners to infer that a novel category has high inductive potential, which results in high generalization of novel features (Rhodes et al., 2012). The second is the novel prediction that hearing specific statements will lead listeners to infer that the category has low inductive potential, which leads to low generalization of novel features.

Design, materials, and procedure

The sample included 284 adult participants based in the United States (48.9% female, 49.6% male, 1.4% prefer not to respond; 71.4% White, 8.1% Black, 8.1% multiracial, 7.7% Asian, 3.2% Other, 1.4% prefer not to respond; age $M = 41.1$ years, $SD = 14.4$ years). An additional 16 participants were excluded due to failing an attention check.

Participants were recruited via Prolific, an online research platform, and were over the age of 18, located in the United States, and reported English as a first language. The experiment was approved by the New York University IRB (IRB-FY2023-6812), and was not preregistered.

Participants were introduced to a novel group of people called Zarpies, shown as cartoon people varying in appearance with similar clothing to each other (see middle image in Fig. 1). Participants were randomly assigned to either the *generic* ($n = 95$), *specific* ($n = 90$), or *baseline* ($n = 99$) condition. Participants in the *generic* and *specific* conditions watched a narrated video about a book titled "All About Zarpies" or "Look at this Zarpie!", respectively (see Rhodes et al. (2012) for the books used). The books contained 16 *training trials*, each a page illustrating a Zarpie with a certain feature (e.g., a Zarpie eating flowers). Each training trial featured a different image and physical or behavioral feature; training trials were presented in fixed order. On each trial, the narrator said, "Look at this Zarpie!" and uttered either a generic statement (e.g., "Zarpies love to eat flowers.") in the *generic* condition or a specific statement (e.g., "This Zarpie loves to eat flowers.") in the *specific* condition (see Leshin et al. (2021) for the videos used). Participants in the *baseline* condition did not receive any training trials and proceeded directly to the test phase.

To assess participants' inferences about the inductive potential of Zarpies, all participants then experienced 16 *test trials* where they were asked to imagine seeing a Zarpie with a *novel* feature *not mentioned* in the training trials (e.g., "Imagine you see a Zarpie painting their hands yellow."), and to estimate the prevalence of Zarpies with that novel feature (e.g., "What percentage of Zarpies do you think paint their hands yellow?"). Each test trial used a different behavioral feature; test trials were presented in randomized order (see Figure 3B) for the features used). Participants responded using a slider ranging from 0% to 100%, with labels every 25%, and initialized at 0%.

Responses were rescaled and truncated between .01 and .99 to directly compare with model predictions, which range from .01 to .99, due to the use of beta distributions.

Results

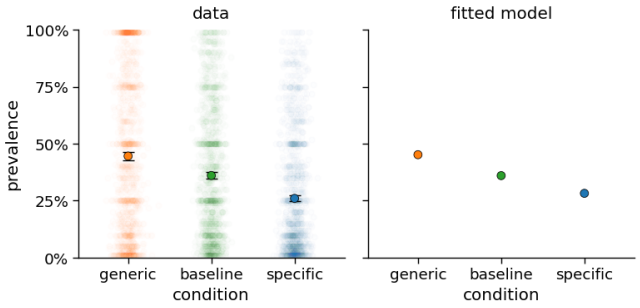

Figure 3: Experiment results: inferred prevalence in each condition, collapsing across test trials. Left facet depicts empirical data: small transparent dots indicate individual responses, and dots with error bars indicate means with 95% confidence intervals. Right facet depicts mean model predictions after fitting each test trial using an inverse temperature of $\beta = 20$.

We tested the main predictions of the model against participants' responses in the test trials. If generics lead to the inference that the category is high in inductive potential, then the generic condition should show *higher* generalization of novel features compared to the baseline condition. If specifics lead to the inference that the category is low in inductive potential, then the specific condition should show *lower* generalization of novel features compared to the baseline condition.

To test both hypotheses, we ran a mixed effects beta regression, predicting prevalence ratings, as a function of condition, with random intercepts per test trial and per participant. Participants generalized differently across conditions, as indicated by their prevalence ratings differing across conditions ($\chi^2(2) = 41.73, p < .001$) (Figure 3).

To break down the main effect of condition, we conducted pairwise comparisons between conditions (with Bonferroni correction). Participants generalized more, i.e., gave higher prevalence estimates, in the generic condition compared to the baseline condition ($\beta = 0.37, SE = .12, p = .01$). Participants also generalized significantly less, i.e., gave lower prevalence estimates, in the specific condition compared to the baseline condition ($\beta = -0.46, SE = .13, p < .001$), confirming each hypothesis, respectively.

Discussion

We proposed a hierarchical Bayesian model of how people learn about categories from hearing generic statements, and confirmed several key predictions. In an experiment with adults, hearing generic statements caused people to generalize novel features of a category member to many category members, which the model explains as an inference that the

category is relatively coherent. In contrast, people were less likely to do so after hearing specific statements, which the model explains as a pragmatic effect of the speaker selecting a specific statement rather than a generic statement.

Learning about categories from generic and specific statements

Generic statements are known to not only shape our beliefs about the prevalence of specific features of a category, but also about the whole category itself, with the category coming to be seen as a meaningful and cohesive kind that supports induction. Our model provides a rational account of this process: people reason not only about the prevalence of various features, but also about categories themselves, including about what features are kind-linked, and about the category's overall inductive potential. Category representations have a hierarchical structure where beliefs about the inductive potential of the category govern and interact with beliefs about the prevalence of specific features among category members. Note that speakers need not intentionally seek to communicate the inductive potential of a category; the (inferred) intent to communicate kind-linked features is sufficient to support listeners' inferences about category-wide inductive potential. As a result, generic language may be one mechanism by which people learn which social categories are considered meaningful and coherent in their society.

This work also contributes to our understanding of *specific statements*, which to date have mostly been used as a contrast for generic statements, rather than studied for their effects themselves. It is not obvious that specific statements would suggest that the category is low in inductive potential. In fact, one might think that specific statements (e.g., "This Zarpie eats flowers.") would *heighten* inductive potential by invoking the category label ("Zarpie"). However, the data support a more complex interpretation in which the specific statement is interpreted as a choice *against* using a more informative generic statement, which leads to inferences that the category is *low* in inductive potential. By integrating semantics and pragmatics, our model presents a sophisticated view of how language shapes our beliefs about categories.

Implications and future directions

If generics induce a sense that a category is highly coherent, what is it that coheres the category together? Our study does not directly address this question, but a possibility is that people posit some placeholder to be filled in that binds the category together. For example, particularly in the absence of an alternative causal theory about how features of category members relate to each other, people may posit that Zarpies share some unknown essence that makes Zarpies similar to one another. Indeed, inductive potential is commonly included as a component of *essentialism*, a wider-ranging belief that categories denote fundamental kinds defined by underlying essences (Medin & Ortony, 1989). While essentialism also includes other components, such as discrete boundaries and a commitment to innateness, generics may pave the path for essentialism by inducing a sense of inductive potential.

The computational model presented also provides additional predictions beyond those tested here. In ongoing work, we confirm the fine-grained predictions of the model with human judgments after hearing varying numbers of generic versus specific statements, where the particular proportion of generics and specifics has a graded effect on the inferred inductive potential. Such a result also rules out alternative explanations for the results presented here, such as different speaker goals inferred from the different book covers; in ongoing work, no book covers were presented.

The current model could also be extended to examine learning about categories in richer, more complex settings. In ongoing work, we consider how coherence may vary based on the feature under consideration, modeling coherence as a function over feature space rather than as a single fixed value. As a result, hearing a generic about Zarpies having feature A, may increase the inferred coherence of Zarpies for features semantically close to A, but have less of an effect on features semantically distant from A. In addition, the model could be extended to account for multigroup or multi-individual environments, individuals with multiple features or group memberships, and multiple utterances about an observation, each of which effectively expand the alternative utterance space. Accounting for these aspects could provide a more naturalistic account of how we learn from generics.

Finally, future research could examine how prior knowledge about domains and features shape inferences about categories from language. Prior knowledge might include domain knowledge (e.g., if Zarpies are a kind of people, bringing to bear knowledge of people in general), knowledge of other categories in the domain (e.g., other kinds of people tend to be highly coherent), and knowledge about the feature itself (e.g., knowledge about *eating flowers* as a behavior: its possible functions, its possible causes, etc.; knowledge that *being 20 years old* is less likely to be linked to any kind). These beliefs or *over*-overhypotheses could inform priors about the coherence of a category, priors about the likelihood of a particular feature being kind-linked, and the process by which the feature set under consideration. Modeling such interactions could contribute not just to our understanding of how language shapes our beliefs about basic-level categories, but also of how our highest-level beliefs about superordinate categories and domains are formed.

Conclusion

The everyday categories we group people by, such as New Yorkers, climbers, and pedestrians, vary in their perceived inductive potential, i.e., how strongly they license generalization across individuals in that category. Using a pragmatic hierarchical Bayesian model, we provide an account of how language shapes our beliefs about the inductive potential of a category, which suggests that language may provide a route for acquiring these fundamental beliefs about categories in the first place.

Acknowledgments

We thank Jess Stephenson for her help with conducting the study, members of the Conceptual Development and Social Cognition Lab for feedback on this paper, and those on Prolific who participated in the study for their time. MYZ was supported by a National Science Foundation SBE Postdoctoral Research Fellowship (NSF SPRF).

References

Benitez, J., Leshin, R. A., & Rhodes, M. (2022). The influence of linguistic form and causal explanations on the development of social essentialism. *Cognition*, *229*, 105246. https://doi.org/10.1016/j.cognition.2022.105246

Carlson, G. N., & Pelletier, F. J. (Eds.). (1995). *The generic book*. University of Chicago Press.

Degen, J., Tessler, M. H., & Goodman, N. D. (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd. *Proceedings of the Cognitive Science Society*, 6.

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. https://doi.org/10.1080/15475440701542625

Goodman, N., Mansinghka, V., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2014, July). Church: A language for generative models. https://doi.org/10.48550/arXiv.1206.3255

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Kravtchenko, E., & Demberg, V. (2022). Informationally redundant utterances elicit pragmatic inferences. *Cognition*, *225*, 105159. https://doi.org/10.1016/j.cognition.2022.105159

Leshin, R. A., Leslie, S.-J., & Rhodes, M. (2021). Does It Matter How We Speak About Social Kinds? A Large, Preregistered, Online Experimental Study of How Language Shapes the Development of Essentialist Beliefs. *Child Development*, *0*(0), 1–17. https://doi.org/10.1111/cdev.13527

Leslie, S.-J. (2008). Generics: Cognition and Acquisition. *The Philosophical Review*, *117*(1), 1–47. https://doi.org/10.1215/00318108-2007-023

Medin, D., & Ortony, A. (1989). Psychological essentialism. In A. Ortony & S. Vosniadou (Eds.), *Similarity and Analogical Reasoning* (pp. 179–196). Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.009

Murphy, G. (2004, January). *The Big Book of Concepts*. MIT Press.

Novoa, G., Echelbarger, M., Gelman, A., & Gelman, S. (2023). Generically partisan: Polarization in political communication. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(47). https://doi.org/10.1073/pnas.2309361120

Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359. https://doi.org/10.1073/pnas.1900105116

Pronovost, M. A., & Scott, R. M. (2022). The influence of language input on 3-year-olds' learning about novel social categories. *Acta Psychologica*, *230*, 103729. https://doi.org/10.1016/j.actpsy.2022.103729

Rhodes, M., Gelman, S. A., & Leslie, S.-J. (2024). How generic language shapes the development of social thought. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2024.09.012

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *PNAS*, (27), 2–7. https://doi.org/10.1073/pnas.1208951109

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Harvard University Press.

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review*, *126*(3), 395–436. https://doi.org/10.1037/rev0000142

Yzerbyt, V., Corneille, O., & Estrada, C. (2001). The Interplay of Subjective Essentialism and Entitativity in the Formation of Stereotypes. *Personality and Social Psychology Review*, *5*(2), 141–155. https://doi.org/10.1207/S15327957PSPR0502_5