Teleological properties matter more for categorization

David Rose
Stanford University

Siying Zhang
Stanford University

Tobias Gerstenberg
Stanford University

Abstract

What kinds of properties matter most in categorization? Does it matter whether properties are attributed in generic statements (e.g., "Bees are striped") or specific ones (e.g., "This bee is striped")? A standard view about the role of generics in categorization is that we categorize things by considering their essential properties, and generics are the central vehicle by which beliefs about essences are transmitted. Given that generics are supposed to play a role in promoting essentialist thinking, we focus on classic tests of essentialism – transformation, induction and offspring tasks – to address what kinds of properties matter in categorization and whether it matters if they feature in generic or specific statements. In Experiment 1, participants categorized properties as being either behavioral, biological, social, or teleological. In Experiment 2, we used the top four properties from each group to describe a generic noun or a specific individual. Transforming a teleological property had a stronger effect on categorization judgments compared to the other types of properties. Teleological properties also mattered more when categorizing novel creatures (Experiment 3) or offspring (Experiment 4). While for Experiments 2 and 3, it made no difference whether properties were attributed in generic or specific form, Experiment 4 showed stronger effects for generic statements. Overall, our findings suggest that teleological properties play a privileged role in how people categorize and essentialize, and that whether these properties are introduced in generic or specific form might matter less.

Keywords: categorization; teleology; essentialism; generics; transformations.

Introduction

Categories are represented in terms of their properties. For instance, when thinking of bees, something comes to mind that has wings, hair, yellow and black stripes, and six legs. Various behaviors bees engage in might come to mind too, such as pollinating flowers, making honey, and cooperating together in colonies.

Properties not only matter for how we represent categories, they also affect how we talk about them. If asked, for instance, "What is a bee?" a natural response would be one that lists a variety of properties. When GPT-4 is prompted to explain to a toddler what a bee is, it answers: "A bee is a tiny creature with wings that flies from flower to flower. They are very important because they help flowers grow by carrying pollen from one flower to another. You know the sweet, sticky stuff called honey? Bees make that too! But remember, never try to touch a bee because some of them can sting to protect themselves" (OpenAI, 2023, retrieved, September 13, 2023). GPT-4 mentions several properties of bees including their size and wings, that they fly, pollinate flowers and make honey.

Categories have different kinds of properties, such as those describing appearance (e.g., color, size, shape), behavior (e.g., flying, making honey), and so forth. Here, we ask whether some kinds of properties matter more than others for how people categorize things. For example, in order to classify something as a bee, does it matter more whether it has yellow and black stripes or that it makes honey?

Some property types might matter more in categorization. But the way categories and their properties are talked might also affect how we categorize. For instance, one could say something specific, such as "This bee is striped" or something generic, such as "Bees are striped". This subtle difference between specific and generic statements has been shown to influence how people learn categories (Cimpian & Markman, 2009; Gelman, 2009; Gelman & Tardif, 1998; Gelman, Ware, & Kleinberg, 2010; Leslie, 2007; Rhodes, Leslie, & Tworek, 2012): properties attributed in generic statements are more likely to be viewed as causally central (see e.g., Ahn, 1998; Cimpian & Markman, 2009), and are more likely to be generalized to new exemplars (see e.g., Gelman et al., 2010; Rhodes et al., 2012). Experiments that look at how generic statements affect categorization typically include a mixture of property types, spanning biological, behavioral, and social properties (see e.g., Gelman et al., 2010; Rhodes et al., 2012). We explore whether some property types matter more in categorization when being attributed to things in generic and specific statements.

The remainder of the paper is organized as follows. We first discuss prior work on the role of generics in categorization. This work indicates that generics promote generalization, create stronger links between properties and categories, and elevate properties to become causally central to categories. We then review some studies on generics more closely before presenting a series of experiments that investigate how property types (biological, behavioral, social and teleological) affect categorization judgments. We find that teleological properties (properties that express a thing's purpose, or what it is for) matter more for people's categorization judgments than other types of properties. How a thing and its properties are talked about – using generic of specific language – matters in some contexts but not in others. Together, these findings raise questions about the role generic language plays in categorization, and about why some property types matter more than others in categorization.

The role of generic statements in categorization

Properties can be linked to one individual (e.g., “This bee is striped”), several individuals using a quantifier (e.g., “Some bees are striped”), or to a non-quantified number of individuals via a generic statement (e.g., “Bees are striped”). When generic statements are used, they often express category-wide generalizations about a property (e.g., Carlson, 1977; Leslie, 2007). Some work has investigated semantic aspects of generics (e.g., Asher & Pelletier, 2012; Carlson, 1977; Cohen, 1999; Greenberg, 2004) and what makes some generic statements acceptable but not others (e.g., Leslie, 2007; Tessler & Goodman, 2019). Other work has focused on the consequences that generics have for how we think about categories. For instance, in contrast to specific statements, generics help children generalize beyond the present context, create stronger links between categories and properties, and suggest that properties are causally central to a category. We briefly discuss each in turn.

Generics promote generalization. Generics shape category representations and affect generalization (e.g., Brandone & Gelman, 2009; Cimpian & Cadena, 2010; Cimpian & Markman, 2009; Gelman et al., 2010; Haward, Carey, & Prasada, 2021; Haward, Wagner, Carey, & Prasada, 2018; Prasada & Dillingham, 2006). Indeed, some argue that generic statements are central to how children learn about categories: they reveal the structure of abstract categories in a way that goes beyond the present context (e.g., Cimpian & Erickson, 2012; Gelman & Raman, 2003; Gelman et al., 2010; Leslie, 2007, 2008, 2012). For example, if looking at two penguins and asked “Do birds fly?” (generic question) or “Do the birds fly?” (specific question) children are more inclined to say “yes” to the latter question. Even when children see an example of birds that cannot fly, they still say that birds fly, suggesting that children recognize that the generic statement refers to the category of birds more broadly.

Generics create strong links between categories and properties. Properties attributed to categories in generic statements suggests a stronger link between the category and properties than when properties are attributed to categories in specific statements (Gelman, 2009). And the stronger the link between the property and the category, the more diagnostic that property is of category membership. To take just one example, if hearing“Pagons are friendly” (generic statement) or “These Pagons are friendly” (specific statement), regardless of whether hearing these in the presence of two or five exemplars, children only say a novel Pagon is friendly when exposed to the generic statement (Chambers, Graham, & Turner, 2008). This suggests that hearing the property in a generic statement strengthens its connection to the category in a way that using specific statements, even when they involve several individual exemplars, doesn’t.

Generics suggest that properties are causally central to categories. Generic statements are also thought to be important because the properties conveyed in generic statements become central in category representations (see e.g., Cimpian & Markman, 2009; Gelman et al., 2010; Rhodes et al., 2012). For instance, if told “Snakes have holes in their teeth” (generic statement) or “He has holes in his teeth” (specific statement), having holes is explained as a cause of other features when exposed to the generic statement and as an accidental feature when exposed to the specific statement (Cimpian & Markman, 2009). Insofar as properties treated as causes are more central in category representations than those treated as effects (see e.g., Ahn, 1998), this suggests that properties conveyed through generics become more central in category representations.

Properties in generic statements

For any generic or specific statement there is always some property or set of properties that is being attributed to a category. Might some property types end up mattering more in generics?

Some work suggests that dangerous properties make generics more acceptable (Cimpian, Brandone, & Gelman, 2010; Leslie, 2007, 2008). Other work suggests that properties viewed as inborn as opposed to acquired make some generics more acceptable (Gelman & Bloom, 2007). Yet other work suggests that those properties that are viewed as bearing a principled connection to a category (instead of an accidental connection) determine which generics people find acceptable (Haward et al., 2021, 2018; Prasada & Dillingham, 2006). For instance, while participants judge that both of these generics, “Barns have roofs” and “Barns are red” are true, they are more inclined to accept “Barns, in virtue of being barns, have roofs” than “Barns, in virtue of being barns, are red” This “in virtue of” linguistic test (Prasada & Dillingham, 2006) is one indicator that the property bears a principled connection to the kind. And according to this linguistic test, having a roof bears a principled connection to barns while being red doesn’t.

In addition to whether properties are dangerous, inborn or bear a principled connection to a category, some work has investigated generics that involve a mix of different property types, including those that might be thought of as involving appearance-based, biological or behavioral properties. For instance, in Gelman et al. (2010) participants were told that “Zarpies have stripes on the bottom of their feet” and that “Zarpies hop over puddles”. While having striped feet says something about Zarpies’ visual appearance or their biology, hopping over puddles says something about their behavior (see also Rhodes et al., 2012). This work typically includes a mix of properties, but doesn’t attempt to determine whether some of them matter more than others.

One exception is recent work by Noyes and Keil (2019). They investigated generics by classifying properties as biological (e.g., “Vawns can hold their breath really long”), social (e.g., “Vawns value nature”), or neutral (e.g., “Vawns can pick apples quickly”). They gave participants a series of statements about Vawns that attributed one type of these properties to them. Participants were then asked to rate statements that, according to Noyes and Keil (2019), reflect essentialist beliefs – beliefs that category members share some true, underlying nature which makes them what they are and is responsible for similarities among category members (Gelman & Ware, 2012) – about Vawns. These statements included, “Vawns have internal or microscopic properties that cause their characteristic appearance and behavior” and “The boundary between the category Vawn and non-Vawn is something decided by people” among others. Noyes and Keil (2019) found that participants were more inclined to find these essentialist statements acceptable when they had earlier read generic statements about Vawns that attributed putative biological, as opposed to social or neutral, properties to them.

However, Noyes and Keil (2019) assigned the different properties to the three categories themselves and it is possible that people viewed them differently. For example, “Vawns can pick apples quickly” seems similar to what they counted as *biological* but it was classified as a neutral property. And “Vawns work together” is arguably *social* but was classified as *neutral*. Though their work gives some reason for thinking that some property types matter more than others when being used in a generic statement (see also Sutherland, Cimpian, Leslie, & Gelman, 2015), it isn't entirely clear which ones people take to be more relevant. Indeed, instead of stipulating how property types should be categorized, a different approach would be to see how people categorize them.

Our approach

We take a novel approach the question of whether some property types matter more than others for how people categorize things. First, rather than stipulating a mapping of property types, we ask how people classify different kinds of properties, focusing on biological, behavioral, social and teleological property types. Biological and social properties have been investigated in prior work (Cimpian et al., 2010; Gelman & Bloom, 2007; Gelman et al., 2010; Noyes & Keil, 2019, see e.g.,). We include behavioral properties because some of the typical properties used in work on generics, such as "Zarpies hop over puddles" Gelman et al. (2010) and "Vawns can hold their breath really long" (Noyes & Keil, 2019) are arguably behavioral. We include teleological properties because they have independently been shown to play an important role in categorization. For instance, Lombrozo and Rehder (2012) suggest that functional properties matter more for categorization judgments than non-functional properties. Moreover, Korman and Khemlani (2020) show that some functional properties yield more acceptable teleological generics, such as "Cars are for driving", than others even when the property is just as prevalent, as in "Cars are for parking". To determine which property instances people view as falling into these different types, we conduct a norming study where we ask participants to classify different properties.

Second, we use transformation cases as a test for how much the different property types affect categorization judgments. Transformations change the properties of a category member into those from a different category (Keil, 1989). For example, older children and adults judge that a raccoon that has been transformed into an individual with the observable properties of a skunk is still a raccoon (Keil, 1989). But younger children, from kindergartners up through third-graders, judge that a raccoon made to look like a skunk *is* a skunk (Keil, 1989). Younger children categorize largely based on appearance. But once they come to have essentialist beliefs about categories, and believe that essences stay the same despite changes in appearance, children move beyond their reliance on appearance to categorize on the basis of essences.

Generic statements are thought to serve as a vehicle by which essentialist beliefs are transmitted (Gelman et al., 2010; Rhodes et al., 2012, see e.g.,). So far, no work on generics has used transformation cases even though robustness across transformations is one of the best indicators that a kind is essentialized. Here we use transformation cases as a test for how much the different property types affect categorization judgments.

Even though transformation cases provide one of the best tests of essentialist thinking, there are additional indicators that a category is essentialized (Gelman, 2003, see e.g.,). For example, properties that feature in essentialized categories are more likely to be generalized to other exemplars than those that feature in non-essentialized categories (Gelman, 2003). Another indicator is that essential features will be transmitted to offspring (Gelman, 2003; Rose & Nichols, 2019, 2020). Thus, in addition to transformation tasks, we also test which property types might be more important in induction and offspring tasks.

Table 1
Experiment 1. *Percentage of participants who judged that a given property belonged in that category (e.g., 86% of participants judged that "jump" is a behavior). The four bolded properties in each column are the ones for which participants were the most sure that they belonged to the corresponding category. We use the short labels in square brackets in the figures below.*

Behavior	Biology	Purpose	Social
jump (86%)	**warm blooded** [blood] (84%)	pollinate flowers* (62%)	**share food with group members** [share] (82%)
swim (84%)	**pointy ears** [ears] (80%)	**enable decomposition** [decompose] (62%)	**cooperate with group members** [cooperate] (78%)
chew (76%)	**long legs** [legs] (80%)	**purify water** [purify] (58%)	**follow the dominant group member** [follow] (66%)
run (74%)	**hair** (80%)	**make honey** [honey] (54%)	**pair bond** [bond] (66%)
swallow (66%)	sharp teeth (76%)	**aerate soil** [aerate] (52%)	dance before mating (54%)
fly (60%)	tail (74%)	recycle nutrients in soil (48%)	nomadic (54%)
urinate (44%)	spots (72%)	enable nitrogen fixation (48%)	sing (30%)
salivate (44%)	small nostrils (72%)	catch and kill insects (46%)	mark territory (28%)
smell (36%)	large eyes (70%)	produce oxygen (40%)	store resources (16%)
digest slowly (18%)	claws (70%)	eat animal carcasses (22%)	build shelter (12%)

* "Pollinate flowers" was a top-rated item but we used it as an example in the introduction to this experiment and since this may have affected how it was rated we decided not to use this item in subsequent experiments.

In line with work by Lombrozo and Rehder (2012) and Rose and Nichols (2019), we hypothesized that *teleological properties* – those that specify what something is for – carry more weight in categorization. Across three experiments that vary the categorization task, we found support for this. We also predicted, in line with previous research on generics, that there would be stronger effects on categorization when properties were attributed in generic statements. However, we found that this occurred only in some contexts.

All experiments, data, analyses, and links to pre-registrations are available here: `https://github.com/cicl-stanford/teleological_properties`

Experiment 1: Property classification

The goal of this experiment was to identify properties that participants view as biological, behavioral, social, or teleological. Even though these categories aren't mutually exclusive in general (e.g., a property could be both social and behavioral), we treated them this way here because in later experiments, participants are asked to assign properties uniquely to different property types.

Methods

Participants. 50 participants who met our pre-registered inclusion criteria were recruited through Amazon Mechanical Turk (*age*: M = 36, SD = 9; *gender*: 33 male, 15 female, 2 no response/other; *race*: 42 White, 4 no response/other, 2 Black, 1 Asian, 1 Asian Indian, 1 Latino,*ethnicity*: 43 Non-Hispanic, 5 Hispanic, 2 no response/other). Participants received $1 as compensation.

Materials. We created a list of 40 properties that we expected to be viewed as either biological, behavioral, social, or teleological (see Table 1). We selected 10 properties for each property type based on the same kinds of properties included in typical work on generics (see e.g., Gelman et al., 2010; Noyes & Keil, 2019; Rhodes et al., 2012). Teleological properties were based on prior work in Rose and Nichols (2019, 2020).

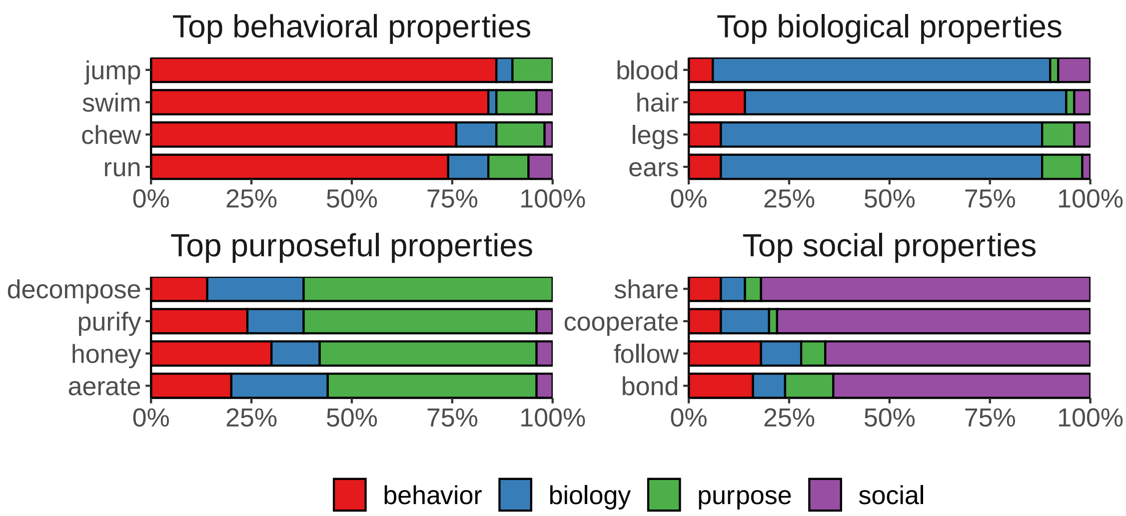


Figure 1. **Experiment 1**: Percentage of participants who assigned a given property (e.g., "jump") to the four different property types.

Procedure. The experiment was programmed using PsiTurk (Gureckis et al., 2016). Participants were first told that they landed on a new island and that they discovered 42 new things. For each thing, they discovered some feature of it and were told that their task was to categorize that feature into one of four types: biology, behavior, purpose, or social. They were then given two examples involving familiar things and how their properties might be categorized (see Figure A1 and Figure A2 in the Appendix). At the end of the instructions, participants were asked two comprehension check questions. If any of the questions was answered incorrectly, participants were directed to read the instructions again. Participants needed to pass both questions in order to proceed.

For the test trials, participants were told that they came across a new thing on the island and decided to give it a name. Each name, such as "Delpons", was a made-up name. Then they were told that after observing it, they noticed that the thing has one of the properties listed in Table 1. They were then asked, "What category does this feature belong to?", with "biology", "behavior", "purpose", or "social" to choose from. Each property type was presented in the center of a square button at the bottom of the screen, running left to right. After clicking on of the buttons, participants proceeded to the next trial.

Design. Participants saw all 40 items from Table 1 and two additional items that served as attention checks. Six participants were removed from the analysis for failing at least one of the attention checks. All items were presented in a random order. For each feature, participants could categorize them as belonging to "biology", "behavior", "purpose", or "social". The order in which the response options were listed was randomized between participants.

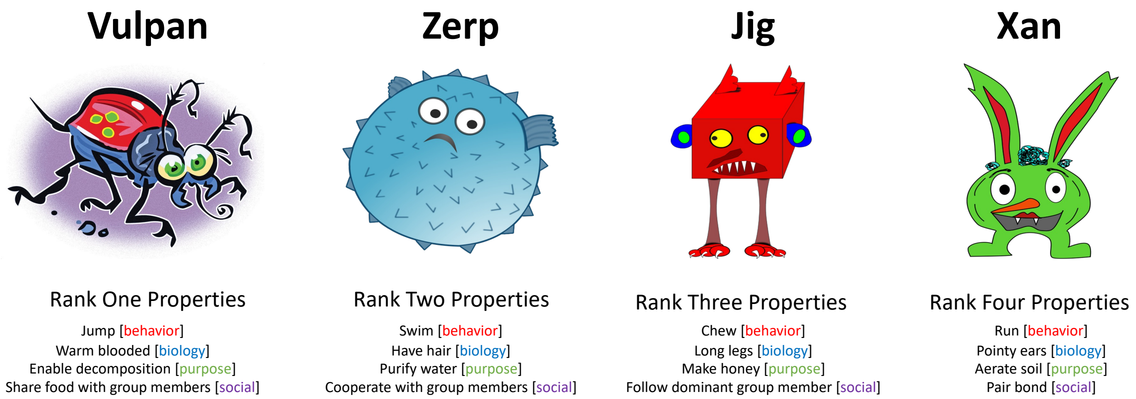


Figure 2. **Experiment 2 items**: Each creature was described as having the same rank properties for each property type. For example, the Vulpan was described as having the top ranked property from Experiment 1. In the specific condition, only one creature was shown when it was introduced and described. In the generic condition, three creatures were shown. These same items were used in Experiments 2–4.

Results and discussion

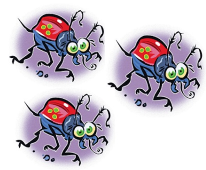
Our main goal was to find properties for which participants agreed that they belonged to a particular property type, so that we could use these as items in subsequent experiments. Figure 1 shows the four responses of each property type for which participants agreed the most that it belonged to that type. For example, most participants judged that “jump” was a behavioral property, and that “warm blooded” was a biological property. Notice that the probability of assigning a property to a given type was somewhat higher for “behavior” and “biology” compared to “purpose” and “social”. To take one example, while most participants judged that “decompose” was a purposeful property, some judged that it better fit “biology” or “behavior”. More generally, 80% of participants categorized the top four behavioral properties into the behavioral category, 81% of participants categorized the top four biological properties into the biological category, 57% of participants categorized the top four teleological properties into the purpose category, and 73% of participants categorized the top four social properties into the social category.[1]

Experiment 2: Categorization after transformation

Experiment 1 identified properties that people systematically categorize into four property types. In Experiment 2, we use a transformation task and ask whether some properties affect categorization judgments more strongly than other properties when predicated in either a generic or specific statement. Our pre-registered hypothesis was that teleological properties would have a bigger impact on categorization when compared individually to

[1] Ratings for teleological properties were lower overall than ratings for other properties. But in subsequent experiments, what matters is not whether they view these particular properties as teleological but that when they view a property as teleological it plays a greater role in their categorization judgments. As it turns out, however, these particular properties were almost always categorized as teleological in our experiments (see Figure B1, Figure C1 and Figure D1 in the Appendix).

Vulpans jump, are warm-blooded, enable decomposition, and share food with group members.



Please categorize the following properties in the table below for **Vulpans.**

For each property, indicate whether you think it is best categorized as:
a biological property, a behavioral property, a social property, or a property that is related to its purpose.

Please select a unique category for each property.

Property	Category			
enable decomposition	○ biological	○ behavioral	○ social	◉ purpose
jump	○ biological	◉ behavioral	○ social	○ purpose
are warm blooded	◉ biological	○ behavioral	○ social	○ purpose
share food with group members	○ biological	○ behavioral	◉ social	○ purpose

Figure 3. **Experiment 2 categorization task**: Participants categorize the properties in the table. Each must be assigned a unique category before participants can proceed. This is then repeated for a second creature.

either biological, social, or behavioral properties.

Methods

Participants. We recruited 100 participants through Prolific who met our pre-registered inclusion criteria (*age*: M = 34, SD = 11; *gender*: 57 female, 38 male, 5 no response/other; *race*: 62 White, 20 Asian, 13 Black, 5 no response/other *ethnicity*: 87 Non-Hispanic, 10 Hispanic, 3 no response/other). Participants were compensated at a rate of $11 per hour.

Materials. We created four creatures – Vulpans, Zerps, Jigs and Xans – based on the properties in Figure 1. Each had the same rank order property from each category: Vulpans had the top ranked property from each category, Zerps had the second ranked property, Jigs had the third, and Xans the fourth. Pictures were included for each creature (see Figure 2).

Procedure. The experiment was programmed using jsPsych (de Leeuw, 2015). An example of the procedure for the generic condition is shown in Figure 3 and Figure 4. Participants in this condition were given a bare plural generic (e.g., Vulpans) followed by four properties (see Figure 3). They were also shown three exemplars of the creature below the generic statement. Participants categorized each property as being either behavioral, biological, social, or related to the thing's purpose. Each property needed to be placed in a unique category. Participants' selected categories in a table at the bottom of the page. This was repeated on a new page for a different creature. They then viewed a table summarizing how they categorized the properties of the two creatures.

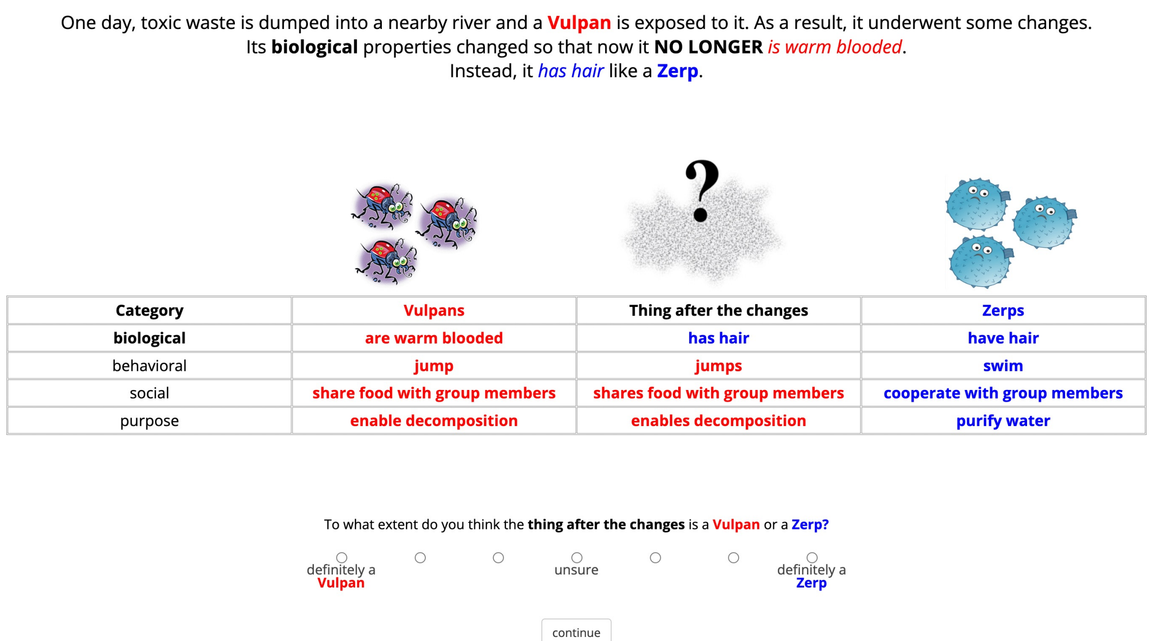


Figure 4. **Experiment 2 judgment task**: Participants judge to what extent the transformed creature is one or the other category. In this example, the property that the participant categorized as being biological in Vulpans was transformed into that of the Zerps.

On the test screen (see Figure 4), participants were told that the first creature they encountered was exposed to toxic waste and underwent some changes. They were told that one of its properties changed and that it had the property of the same type from the second creature of the pair. For example, if one of the properties predicated to the first creature was "warm blooded" and a participant categorized this as biological, and if one of the properties predicated to the second creature was "has hair" and this was categorized as biological, then if the property changed by exposure to toxic waste was biological, participants would be told that the creature wasn't warm blooded like the first creature but instead had hair like the second creature. A table summarizing how each property was categorized for each creature was provided at the bottom of the screen and included a column indicating the properties of the creature after the changes. After learning about the change, participants answered the question "To what extent do you think the thing after the changes is a [name of first creature] or a [name of second creature]?" on a 7pt. The endpoints were labeled with "definitely a [first/second creature]" and the midpoint was labeled "unsure".

Participants were given two pairs of creatures. The pairs and their order were randomized. Creatures were introduced one at a time. When each creature was introduced the order of their properties in the property categorization task was randomized. Participants completed eight trials that involved a property from a creature being transformed into a property of the same type from the other creature in the pair. Property type transformation order was randomized within creature pairs.

Design. In addition to the property type transformation being manipulated within participants, we also manipulated, between participants, whether the properties predicated to the creature were in generic or specific form (we used a definite singular, e.g., "This

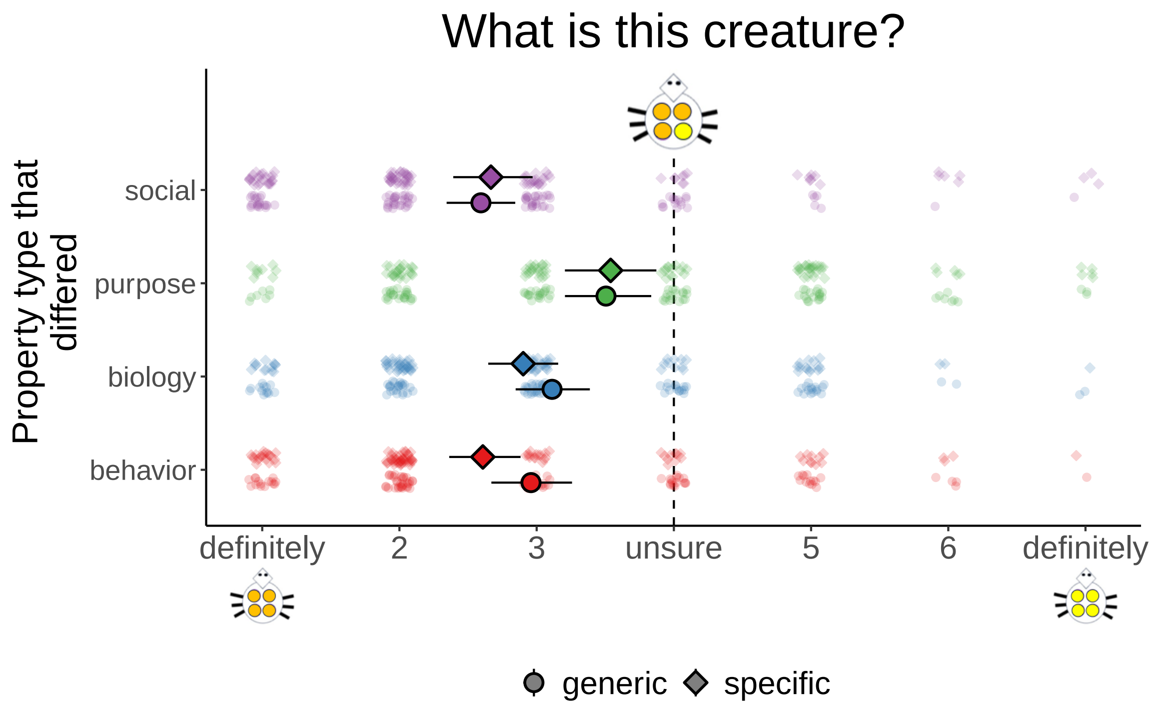


Figure 5. **Experiment 2**: Category judgments based on what property was transformed in the generic (circle) and specific (diamond) conditions. In all figures, higher ratings indicate that the original thing has changed categories after the property change. Large shapes show means with 95% bootstrapped confidence intervals. Small points show individual responses (jittered along both axes for visibility).

Vulpan ...", and depicted a single individual below the statement). In the specific condition, only one creature was displayed on the screen when it was introduced.

Results

Property categorization task. Participants' categorization of the different properties was consistent with what we had found in Experiment 1 (see Figure B1 in the Appendix). Notice that while in Experiment 1, each property was judged independently, this time participants had to assign four properties uniquely to the different categories.

Judgment task. Figure 5 shows participants' category judgments based on the type of property that was transformed and whether a generic or specific statement was used. As we predicted, teleological properties carried more weight in categorization (see Table 2). For each contrast, the median of the posterior distribution was positive and the 95% credible interval excluded 0. We found this to be the case both in the generic condition and the specific condition. Indeed, whether properties were predicated in generic or specific form made no credible difference (Posterior median $= .06$, 95% credible interval $[-.14, .25]$).

Exploratory analysis

Our results indicate that teleological properties carry more weight in categorization. But this raises the question of whether teleological properties only matter because some of the teleological features are more diagnostic of category membership than other types of features. For instance, “makes honey”, probably only brings one thing to mind: a bee. In contrast, “has hair” could bring many things to mind. Perhaps teleological properties are simply more diagnostic of category membership than other properties.

We probed large language models to see what probabilities they assigned to completions. The basic idea is that if a language model is relatively certain how a sentence with a given property continues, then this property is diagnostic of its kind. We probed BERT-base, BERT-large and RoBERTa-large. Each model was given the following prompt: “An animal that [property] is a ...” [property] was replaced with each property used in Experiment 2 (see Figure 1). For each property, we extracted the top five completions and their associated probabilities. Figure 6 shows the results.

Some of the features were strongly diagnostic. As suggested, “making honey” points toward bees. “swimming”, a behavioral property, was strongly diagnostic of fish. So, if mere diagnosticity mattered for categorization, one would expect that “swimming” and “making honey” would have a greater influence on categorization than other properties. But as can be seen in Figure 7, this wasn’t the case. “Making honey”, for example, didn’t have a bigger effect than other teleological properties even though those other properties were less diagnostic. Similarly, even though “swimming” was more diagnostic than other behavioral properties, it didn’t appear to be carrying more weight.

To further explore whether diagnosticity might explain why teleology had the biggest effect on categorization judgements, we treated entropy as a measure of diagnosticity, and computed the Shannon entropy score for the top five completions for each individual property (e.g., “swimming”, “making honey”) for each language model. We then correlated mean participant responses for each individual property with entropy scores computed from each language model, resulting in correlations of $r = -.14$ for RoBERTa-large, $r = -.51$ for BERT-base, and $r = -.69$ for BERT-large.

Since BERT-large showed the strongest correlation between entropy scores and par-

Table 2
Experiment 2*: Posterior distributions of the difference between purpose and other properties for both the generic and specific condition.* Note*: CrI = credible interval.*

contrast	median	lower 95% CrI	upper 95% CrI
generic condition			
purpose - behavior	0.54	0.22	0.90
purpose - biology	0.40	0.07	0.73
purpose - social	0.92	0.57	1.24
specific condition			
purpose - behavior	0.93	0.57	1.25
purpose - biology	0.63	0.32	0.98
purpose - social	0.87	0.54	1.20

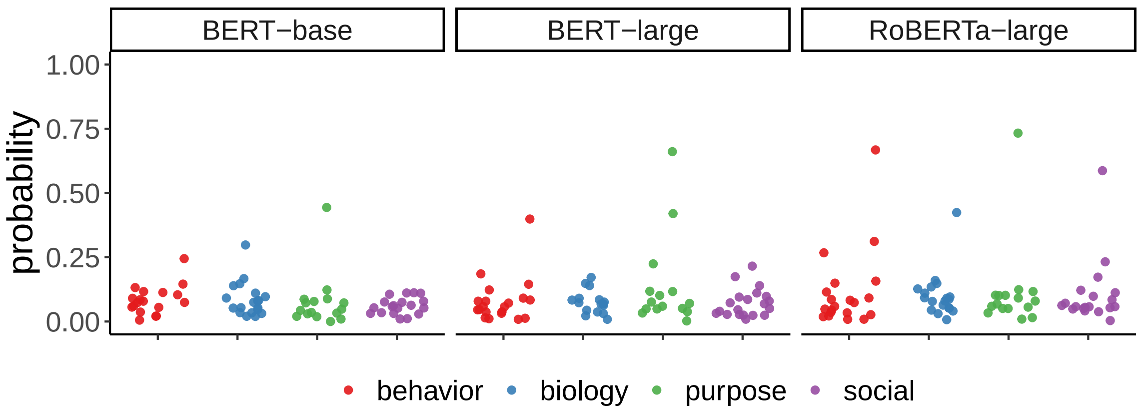


Figure 6. **Experiment 2**: Probabilities of completion assigned by language models separated by property type. Individual points show the probability of each completion (jittered along both axes for visibility).

ticipant responses, we examined whether it still mattered whether a property was teleological or not, once diagnosticity was statistically controlled for. We ran a Bayesian linear mixed effects model with entropy scores from BERT-large as as fixed effect, as well as a dummy-coded variable that encoded whether a property was teleological or not. We also included random intercepts for participants. We found that in that model, entropy scores do not credibly predict participant ratings $-.18$ $[-.53, .15]$, while teleological properties do $.69$ $[.50, .88]$. This suggests that diagnositicity is not driving the effect of teleology and that teleological properties are important in their own right.

Discussion

Generics are supposed to promote essentialist beliefs about categories. But as we suggested, one of the best tests of essentialist thinking involves judgments about things that undergo transformation to their properties. Examining things that undergo property transformation, we find that teleological properties carry more weight when a creature undergoes a transformation, regardless of being predicated in a generic or specific statement. The importance of teleology isn't explained by the individual features included in this property type being more diagnostic (see Figure 6 and Figure 7).

Overall we found that categorization judgments were generally below the midpoint. This isn't surprising since only one feature out of four was changed in the transformation. The important point, though, is that changing teleological properties had a stronger impact on categorization judgments than changing any of the other properties.

Experiment 3: Categorization in induction

Research indicates that generics play a role in induction (e.g., Gelman, 2009; Rhodes et al., 2012). For instance, if children and adults are shown a novel creature with stripes and told that either "Bants have stripes" (generic statement) or "This bant has stripes" (specific statement), they are more likely to say a novel exemplar that possesses the relevant property, e.g., stripes, is a bant when exposed to the generic statement (Hollander, Gelman,

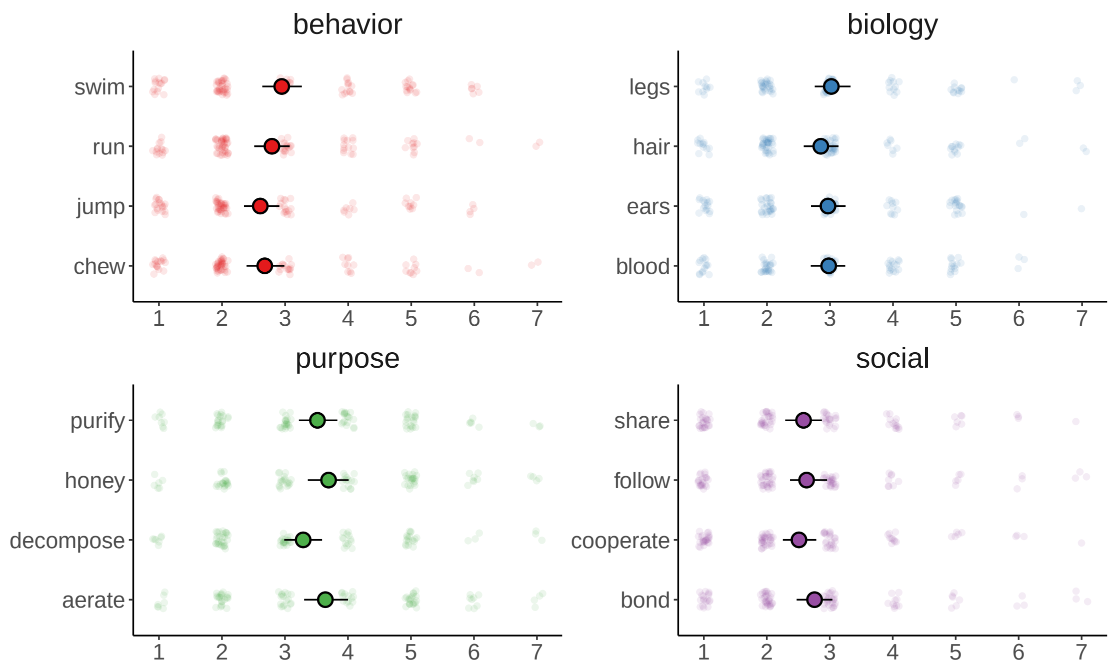


Figure 7. **Experiment 2**: Categorization ratings for each property across all property types. Higher ratings indicate that the original thing has changed categories after the property change. Large points show means with 95% bootstrapped confidence intervals. Small points show individual responses

& Raman, 2009). Generics are thought to strengthen the link between the category and the property, leading to essentialist thinking about the category.

Our main question is whether teleological properties carry more weight in categorization in the context of an induction task. We examine this in the context of generic and specific statements. Our pre-registered hypothesis was that teleological properties will have a bigger impact on categorization than either biological, social, or behavioral properties, and that this will occur for both generic and specific statements.

Methods

Participants. We recruited 100 participants through Prolific who met our pre-registered inclusion criteria (*age*: M = 34, SD = 14; *gender*: 48 female, 48 male, 4 no response/other; *race*: 76 White, 14 Asian, 6 Black, 4 no response/other *ethnicity*: 85 Non-Hispanic, 14 Hispanic, 1 no response/other) through Prolific. Participants received compensation at a rate of $12 an hour.

Materials & Design. The materials and design were largely the same as in Experiment 2.

Procedure. The experiment was programmed using jsPsych (de Leeuw, 2015) and the procedure was the same as in Experiment 2. The only difference was that on the test page, participants didn't read about a creature getting exposed to toxic waste and undergoing transformation. Instead, they were told that one day, they came across a creature

Category	Jigs	The new creature	Vulpans
biological	have long legs	has long legs	are warm blooded
behavioral	chew	chews	jump
social	follow the dominant group member	shares food with group members	share food with group members
purpose	make honey	makes honey	enable decomposition

Figure 8. **Experiment 3 judgment task**: Participants judge to what extent the new creature is a member of one or the other category. In this example, the property that the participant categorized as being social from the first creature is replaced with the property that the participant categorized as social from the second creature.

that had all of the properties of the first creature in the pair except for one. That property was one from the second creature in the pair (see Figure 8). Participants were then asked, “To what extent do you think that this new creature is a [name of first creature] or a [name of second creature]?” on a 7pt. The endpoints were labeled with “definitely a [first/second creature]“ and the midpoint was labeled “unsure”.

Results

Property categorization task. Again, we found that participants’ categorization of the different properties was consistent with what we had found in Experiment 1 (see Figure C1 in the Appendix).

Judgment task. Figure 9 shows participants’ categorization judgments. Teleological properties continued to carry more weight when categorizing in an induction task. They did so whether they were predicated in a generic or specific statement (see Table 3). As with transformations, whether properties were predicated in generic or specific form made no credible difference $-.09$ $[-.29, .10]$.

Discussion

There are a range of indicators that a kind is essentialized. One, which we examined in Experiment 2, concerns robustness across transformations. Another, which we examined here in Experiment 3, concerns generalization. In both cases, we found that teleological properties carry more weight. They carry more weight in situations where a creature undergoes a transformation, and they carry more weight in generalization. Teleological properties

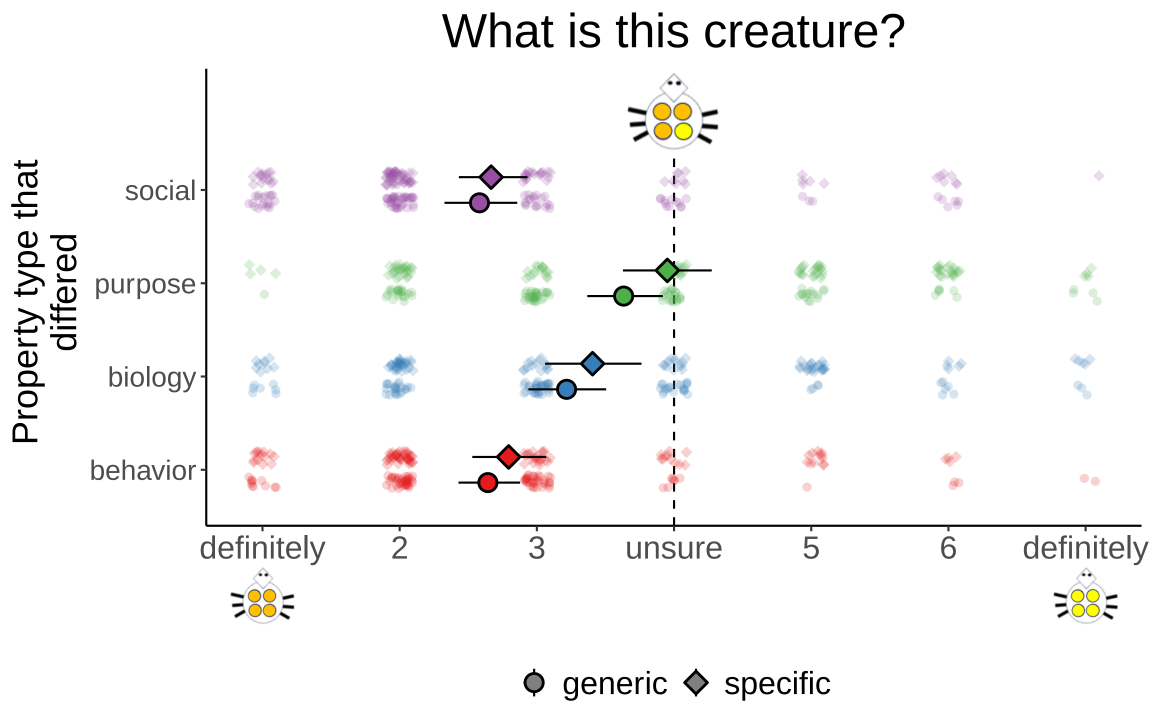


Figure 9. **Experiment 3**: Effect of property type on categorization ratings in the generic (circle) and specific (diamond) conditions. Large shapes show means with 95% bootstrapped confidence intervals. Small points show individual responses (jittered along the x-axis for visibility).

appear to be more projectable and thus play a greater role in determining whether a novel exemplar is a member of a category. And they play this greater role whether they feature in generic or specific statements.

These two indicators – robustness across transformation and generalizability – aren't the only two indicators that a kind is essentialized. Another indicator is that essential

Table 3

Experiment 3*: Posterior distributions of the difference between purpose and other properties for both the generic and specific condition.* Note*: CrI = credible interval.*

contrast	median	lower 95% CrI	upper 95% CrI
generic condition			
purpose - behavior	0.99	0.65	1.30
purpose - biology	0.41	0.07	0.72
purpose - social	1.05	0.75	1.37
specific condition			
purpose - behavior	1.16	0.86	1.49
purpose - biology	0.56	0.26	0.88
purpose - social	1.28	0.96	1.59

features will be transmitted to offspring. We focus on this in our final experiment.

Experiment 4: Categorization of offspring

In our final experiment, we ask whether teleological properties carry more weight when categorizing offspring. Again, we predicted that teleological properties will affect categorization judgments more strongly than biological, social, or behavioral properties, and that they will do so regardless of being predicated in generic or specific statements.

Methods

Participants. We recruited 100 participants through Prolific who met our pre-registered inclusion criteria (*age*: M = 34, SD = 14; *gender*: 57 female, 38 male, 5 no response/other; *race*: 62 White, 20 Asian, 13 Black, 5 no response/other *ethnicity*: 87 Non-Hispanic, 10 Hispanic, 3 no response/other) through Prolific. Participants received compensation at a rate of $12 an hour.

Materials & Design. The materials and design were largely the same as in Experiment 2 and Experiment 3.

Procedure. The experiment was programmed using jsPsych (de Leeuw, 2015) and the procedure was the same as in Experiment 2 and Experiment 3. The only difference was that participants were told that the first creature they were introduced to mated with the second one and they had an offspring. The offspring had all of the properties of the first parent except for one. That property was one from the second parent (see Figure 10. Participants were then asked, “To what extent do you think that the offspring is a [name of first creature] or a [name of second creature]?” on a 7pt. The endpoints were labeled with “definitely a [first/second creature]“ and the midpoint was labeled “unsure”.

Results

Property categorization task. As in our previous two experiments, we found that participants' categorization of the different properties was consistent with what we had found in Experiment 1 (see Figure D1 in the Appendix).

Judgment task. Figure 11 shows participants' categorization judgments. Teleological properties influenced categorization judgments more strongly than other properties. They did so whether they were predicated in a generic or specific statement (see Table 4). But in contrast to transformations (Experiment 2) and induction (Experiment 3), here we found that categorization ratings were higher in the generic compared to specific statement condition .20 [.02, .37].

Discussion

We again find that teleological properties matter more than other properties. And they do so even when considering offspring. Interestingly, we found that there was an effect of whether the statement was generic or specific, with generic statements leading to higher categorization ratings. We discuss a possible explanation of this effect in the General Discussion.

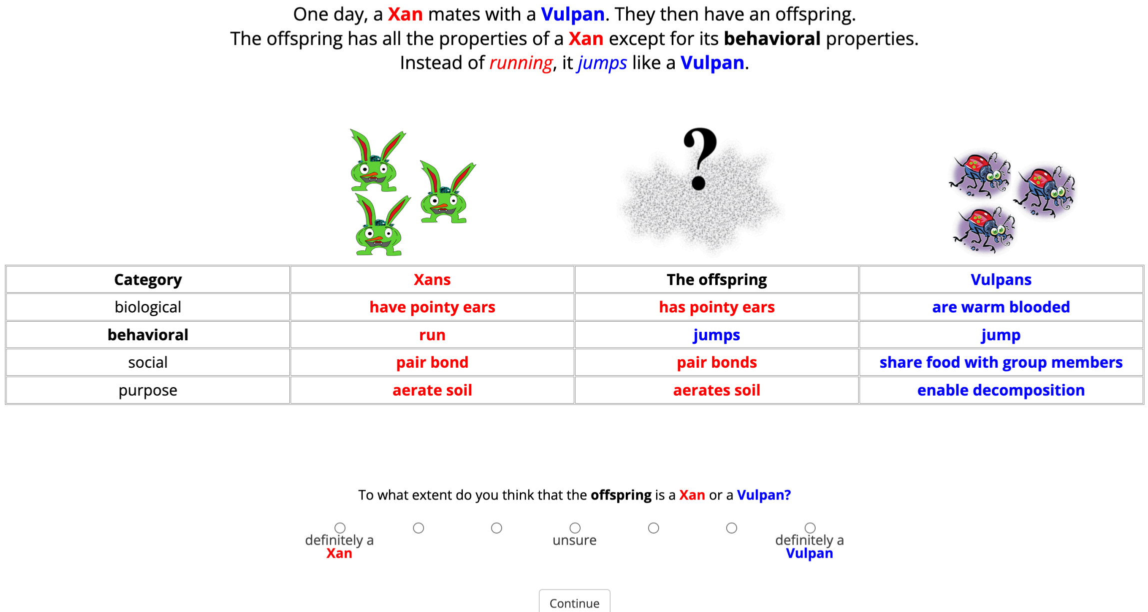


Figure 10. **Experiment 4 judgment task**: Participants judge to what extent the offspring is a member of one or the other category. In this example, the property that the participant categorized as being behavioral from the first creature is replaced with the property that the participant categorized as behavioral from the second creature.

General Discussion

Our main question was whether some types of properties – biological, behavioral, social or teleological – matter more in categorization. We also asked whether the way properties are talked about matters. Specifically, we asked whether using generic or specific statements makes a difference. We found that teleological properties mattered more in categorization than other properties. While this effect was small, it was consistent and

Table 4
Experiment 4: *Posterior distributions of the difference between purpose and other properties for both the generic and specific condition.* Note: *CrI = credible interval.*

contrast	median	lower 95% CrI	upper 95% CrI
generic condition			
purpose - behavior	0.53	0.24	0.82
purpose - biology	0.36	0.09	0.67
purpose - social	0.73	0.44	1.02
specific condition			
purpose - behavior	0.66	0.37	0.96
purpose - biology	0.35	0.07	0.64
purpose - social	0.64	0.36	0.93

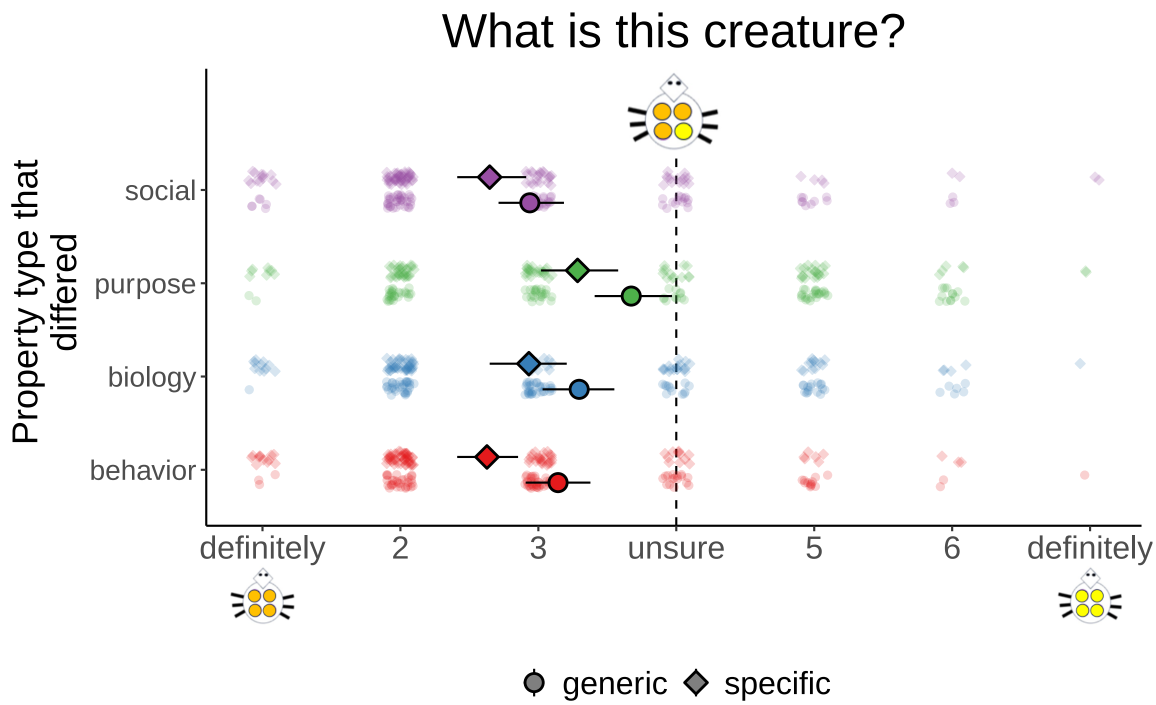


Figure 11. **Experiment 4**: Effect of property type on categorization ratings in the generic (circle) and specific (diamond) conditions. Large shapes show means with 95% bootstrapped confidence intervals. Small points show individual responses (jittered along the x-axis for visibility).

held across a diverse set of tasks. In particular, teleological properties mattered more for categorization when one of a creature's features was transformed, when a new creature was encountered that shared one of the features with one group and the rest with the other, and when two creatures had an offspring.

Whether the statement used to attribute properties was generic or specific didn't matter in transformation and induction tasks. It did, however, matter in tasks involving offspring. There, participants were more inclined to think offspring that had a single property of a category would be a member of that category when generics were used to attribute properties.

Together, our findings raise at least two questions. Why do teleological properties matter more than other property types when categorizing creatures? Why does it only sometimes matter whether the statement is generic or specific?

Why do teleological properties matter more than other properties?

There is a great deal of evidence indicating that teleological considerations play an important role in our thinking about diverse objects, including artifacts and animals. Judgments about the purpose of something play a role in both whether we think there is an object – whether some collection of parts composes an object (Rose, 2017) – and whether we think there continues to be an object (Rose, 2015; Rose, Schaffer, & Tobia, 2020). Rose

(2022) suggests that these effects might be due to people thinking that purposes are essential, that they represent the true nature of things, and make them the kind of thing that they are.

Using classic tests of essentialist thinking, Rose and Nichols (2019) and Rose and Nichols (2020) find that teleological considerations play an important role in people's categorization judgments for artifacts and animals. For instance, if a bee is made to look like a spider but preserves the purpose of bees, making honey and pollinating flowers, people are inclined to think the animal is still a bee. Even if a bee has all of its insides replaced with the insides of a spider, people are inclined to categorize it as a bee if it preserves the purpose of bees. Moreover, if a bee is changed to look like a spider but still has the purpose of bees, people are inclined to think its offspring will be bees.

The fact that teleological considerations play a role in categorization across different kinds, including artifacts and animals, and that they play this role when using classic tests of essentialist thinking, suggests that teleological properties are important because they are associated with category essences. And given that they continue to be more important in the present set of experiments – where these experiments use classic tests of essentialist thinking – than other kinds of properties provides further evidence that teleological properties might be associated with essences. This doesn't suggest that teleology is the only thing that matters when categorizing things (Joo & Yousif, 2022; Neufeld, 2021). But a view that treats teleological properties as essential goes beyond other proposals, such as those that account just for their importance in biological categorization (e.g., Lombrozo & Rehder, 2012), since it provides an account of why they play an important and similar role in categorizing a range of different kinds (Prinzing et al., 2024; Rose, 2017, 2022; Rose, Jamarillo, Nichols, & Horne, 2022; Toorman, 2023; Zhang, She, Gerstenberg, & Rose, 2023).

Why does it only sometimes matter whether the statement is generic or specific?

Recent findings suggest that generics may not play a distinct role in categorization from language signaling high proportions of properties, such as "most" or "many" (Hoicka, Saul, Prouten, Whitehead, & Sterken, 2021; Reuter, Neufeld, & Del Pinal, 2023). For instance, whether children and adults are told either"Zarpies hate ice cream" or "Most Zarpies hate ice cream", makes no difference in the extent to which they think a novel exemplar that is identified as a Zarpie will like ice cream. Indeed this work suggests that both generics and high proportion quantifiers affect generalization – attributing properties to other members of the same group given evidence that some members of the group have the property – but not people's tendency to essentialize a category, where this involves thinking that there is some internal property that is causally responsible for similarities among category members (e.g., Hoicka et al., 2021). That is, generics along with quantifiers like "most", lead people to extend a property to new exemplars. But this doesn't lead them to think that property reflects the true nature or essence of a kind. In fact, generics, or quantified statements using "most" are no more likely to lead to essentializing a kind than hearing specific statements using "this" (e.g., Hoicka et al., 2021).

It may be that generics, at least for adults, promote generalization but not essentializing (see e.g., Tessler & Goodman, 2019). Of course, the situation might be very different for children (though see Noyes & Keil, 2020). That said, our findings cohere with recent work suggesting that, for adults, generics may not be especially important for essentializing categories (see e.g., Hoicka et al., 2021; Noyes & Keil, 2019; Reuter et al., 2023; Tessler & Goodman, 2019). Indeed, our findings build on this work in two key ways.

First, the results from our transformation task indicated that generics were not more important than specific statements. The transformation task is typically taken to be one of the key tests for whether we believe that a category has an essence (e.g., Keil, 1989). Given that we didn't find that generics were more important than specific statements when using what is arguably one of the best tests of essentialist thinking, our results serve as an important addition to work suggesting that generics may not promote essentialist beliefs, at least in adults (for work with children, see e.g., Noyes & Keil, 2020).

Second, generics are supposed to create *stronger* links between a category and properties than specific statements (e.g., Cimpian & Markman, 2009; Gelman, 2009; Gelman et al., 2010; Hollander et al., 2009; Rhodes et al., 2012; Tessler, Bridgers, & Tenenbaum, 2020). And so we might have expected that a creature that has most of the features that were attributed to a category in a generic statement, would be more likely to be judged a member of that category than an exemplar that has those same properties attributed in a specific statement (see e.g., Tessler et al., 2020). But this isn't what we found, either in the transformation or induction tasks. In fact, in the one case where we did find an effect of whether the statement was generic or specific – the offspring task – the effect of generics was exactly opposite of what one would have expected if generics strengthen the connection between categories and properties. Why then might generics make people *less* likely to think a creature is a member of the category that has most of the properties that were attributed to the category, and do so only in tasks involving offspring?

One possibility is that reasoning about offspring involves reasoning about genes in a way that transformation and induction tasks usually doesn't. When reasoning about the traits that animals are likely to have, we might distinguish between those that are likely to be passed on to future generations and maintained in a population and those that likely won't. When properties are mentioned in a generic statement it might imply that they are likely to be passed on to future generations. Moreover, by highlighting a particular property through explicitly mentioning it – as we did in our experiment (see e.g., Figure 10) – people might give it more weight than each of the other features that weren't highlighted. So when reasoning about a creature that has three properties that were attributed to one category, and one property that was attributed to a different category, and also highlighted, this might lead people to think that these properties are likely to recur in future generations. If this collection of features is likely to recur in future generations, then this might lead people to think that speciaton is likely to occur, resulting in a hybrid species that has a mixture of properties. And while most of these features would be shared with one kind of creature, some would be shared with another, resulting in an overall shift in categorization responses away from the category that shares the majority of properties with the exemplar. In contrast, specific statements wouldn't imply anything about what future generations might be like, and so in deciding what kind the offspring might belong to, one might default to the category that shares the majority of features with the exemplar.

Notice that this kind of explanation is one that is consistent with generics playing a special role in promoting essentialist thinking insofar as one views reasoning about genes and whether new kinds might arise due to genetic blending as an essentialist form of reasoning. Even so, generics might play a much more circumscribed role in promoting essentialist thinking than is typically assumed.

Limitations and future directions

Our work builds on research investigating the role of generics in essentialism and moves the discussion forward, in part, by focusing on the importance of different property types that accompany generic and specific statements, and their role in categorization. But our work is also limited in several ways. One limitation of our work is that it uses a only four different properties for each type. Though we used the top four properties that participants categorized into different types (Experiment 1) in our experiments, other work might explore an even broader range of properties.

Another limitation is that our design only allowed people to categorize property instances into one type. We recognize that some properties plausibly fall into multiple types. We restricted participants to selecting only one since our goal was to systematically vary properties that fall under a certain type. But future work might explore the importance of properties that fall under multiple types. We also only used one paradigm. It was developed so that we could manipulate properties that fell into the property types that participants thought were most appropriate. Further work might explore other paradigms that allow this kind of control over how the properties are viewed and how they are changed.

A different limitation is that our work only focused on novel creatures. We limited our attention to creatures since animals are often thought to be the central kinds of things that people essentialize (see e.g., Gelman, 2003; Keil, 1989). Examining other kinds would also be useful. This would, of course, involve changing the property types. Teleological properties can plausibly be extended to biological kinds, artifacts and non-living kinds, like clouds. But the other property types we examined here, such as biological properties, are more restricted. In fact, the scope and flexibility of teleological properties provides an additional reason to think they should be especially important in categorization. But examining their role, alongside other property types, for a broader range of kinds, and in the context of generic and specific statements, would be useful in future work.

Lastly, our current work focused on adults. We think a valuable step forward will be to extend this to children. The main reason is that one leading explanation of how essentialist beliefs are transmitted is that children develop them through exposure to generic statements. But that work also typically uses a mix of property types in generic statements. It may be that some property types are especially important for essentializing, and perhaps even more important than exposure to generic statements.

Conclusion

Categories are represented in terms of their properties. And there are different kinds of properties that we attribute to categories. Some of these might be more important than others in categorization. And the way in which they are attributed, in the form of a generic or specific statement, might also matter for how important they are in categorization. Focusing on various creatures, we explored whether biological, behavioral, social or teleological properties matter more in categorization. Across a range of standard tasks that provide evidence of essentialist thinking, transformation, induction and offspring tasks, we found that teleological properties carried more weight in categorization. Whether they were attributed in generic or specific statements made no difference, with the exception of offspring. The fact that generics didn't tend to be more important in categorization than specific statements adds to a growing body of work suggesting that generics may not specifically promote essentialism. But our finding that teleological properties are privileged, and given that this coheres with a range of findings indicating that teleology plays an important role in categorization, suggest that teleology might play this important role in categorization because purposes are associated with category essences.

References

Ahn, W.-k. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition*, *69*(2), 135–178.

Asher, N., & Pelletier, F. J. (2012). More truths about generic truth. *Genericity*, 312–333.

Brandone, A. C., & Gelman, S. A. (2009). Differences in preschoolers' and adults' use of generics about novel animals and artifacts: A window onto a conceptual divide. *Cognition*, *110*(1), 1–22.

Carlson, G. N. (1977). *Reference to kinds in english: A dissertation* (Unpublished doctoral dissertation). Graduate Linguistic Student Association, Department of Linguistics.

Chambers, C. G., Graham, S. A., & Turner, J. N. (2008). When hearsay trumps evidence: How generic language guides preschoolers' inferences about unfamiliar things. *Language and Cognitive Processes*, *23*(5), 749–766.

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., & Cadena, C. (2010). Why are dunkels sticky? preschoolers infer functionality and intentional creation for artifact properties learned from generic language. *Cognition*, *117*(1), 62–68.

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology*, *64*(3), 161–185.

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*(1), 14–25.

Cohen, A. (1999). Generics, frequency adverbs, and probability. *Linguistics and philosophy*, *22*, 221–253.

de Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Gelman, S. (2009). Generics as a window into children's concepts. In F. J. Pelletier (Ed.), *Kinds, things, and stuff: Mass terms and generics* (pp. 101–121). Oxford Univeristy Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford University Press, USA.

Gelman, S. A., & Bloom, P. (2007). Developmental changes in the understanding of generics. *Cognition*, *105*(1), 166–183.

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child development*, *74*(1), 308–325.

Gelman, S. A., & Tardif, T. Z. (1998). Generic noun phrases in english and mandarin: An examination of child-directed speech. *Cognition*, *66*.

Gelman, S. A., & Ware, E. A. (2012). Conceptual development: The case of essentialism. In *The oxford handbook of philosophy of cognitive science.*

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive psychology*, *61*(3), 273–301.

Greenberg, Y. (2004). *Manifestations of genericity.* Routledge.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., … Chan,

P. (2016). psiturk: An open-source framework for conducting replicable behavioral experiments online. *Behavior research methods*, *48*(3), 829–842.

Haward, P., Carey, S., & Prasada, S. (2021). The formal structure of kind representations. *Cognitive Science*, *45*(10), e13040.

Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition*, *176*, 255–268.

Hoicka, E., Saul, J., Prouten, E., Whitehead, L., & Sterken, R. (2021). Language signaling high proportions and generics lead to generalizing, but not essentializing, for novel social kinds. *Cognitive Science*, *45*(11), e13051.

Hollander, M. A., Gelman, S. A., & Raman, L. (2009). Generic language and judgements about category membership: Can generics highlight properties as central? *Language and Cognitive Processes*, *24*(4), 481–505.

Joo, S., & Yousif, S. R. (2022). Are we teleologically essentialist? *Cognitive Science*, *46*(11), e13202.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* MIT Press.

Korman, J., & Khemlani, S. (2020). Teleological generics. *Cognition*, *200*, 104157.

Leslie, S.-J. (2007). Generics and the structure of the mind. *Philosophical perspectives*, *21*, 375–403.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, *117*(1), 1–47.

Leslie, S.-J. (2012). Generics articulate default generalizations. *Recherches linguistiques de Vincennes*(41), 25–44.

Lombrozo, T., & Rehder, B. (2012). Functions in biological kind classification. *Cognitive psychology*, *65*(4), 457–485.

Neufeld, E. (2021). Against teleological essentialism. *Cognitive Science*, *45*(4), e12961.

Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359.

Noyes, A., & Keil, F. C. (2020). There is no privileged link between kinds and essences early in development. *Proceedings of the National Academy of Sciences*, *117*(20), 10633–10635.

OpenAI. (2023). *Gpt-4 technical report.*

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*(1), 73–112.

Prinzing, M., Rose, D., Zhang, S., Tu, E., Concha, A., Schaffer, J., … Knobe, J. (2024, 1 24). *From artifacts to human lives: Investigating the domain-generality of judgments about purposes.* Retrieved from `https://doi.org/10.31219/osf.io/7enkr`

Reuter, K., Neufeld, E., & Del Pinal, G. (2023). Asymmetry effects in generic and quantified generalizations. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

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.

Rose, D. (2015). Persistence through function preservation. *Synthese*, *192*(1), 97–146.

Rose, D. (2017). Folk intuitions of actual causation: a two-pronged debunking explanation. *Philosophical Studies*, *174*, 1323–1361.

Rose, D. (2022). Mentalizing objects. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy volume 4* (pp. 182–213). Oxford University Press.

Rose, D., Jamarillo, S., Nichols, S., & Horne, Z. (2022). Teleological essentialism across development. *Proceedings of the Annual Meeting of the Cognitive Science Society*(44).

Rose, D., & Nichols, S. (2019). Teleological essentialism. *Cognitive science*, *43*(4), e12725.

Rose, D., & Nichols, S. (2020). Teleological essentialism: Generalized. *Cognitive science*, *44*(3), e12818.

Rose, D., Schaffer, J., & Tobia, K. (2020). Folk teleology drives persistence judgments. *Synthese*, *197*(12), 5491–5509.

Sutherland, S. L., Cimpian, A., Leslie, S.-J., & Gelman, S. A. (2015). Memory errors reveal a bias to spontaneously generalize to categories. *Cognitive science*, *39*(5), 1021–1046.

Tessler, M. H., Bridgers, S., & Tenenbaum, J. B. (2020). How many observations is one generic worth? In *Proceedings of the annual conference of the cognitive science society.*

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological review*, *126*(3), 395.

Toorman, J. (2023). Against arguments from diagnostic reasoning. *Cognitive Science*, *47*(11), e13376.

Zhang, S., She, J. S., Gerstenberg, T., & Rose, D. (2023). You are what you're for: Essentialist categorization in large language models. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Appendix A
Experiment 1: Examples

Instructions

Example 1. Suppose you come across a bee and observe four things about it: it secretes wax, stings, works with other bees, and pollinates flowers. If you had to categorize each of these things you might do the following:

1. "secreting wax" → **biology**: This is like familiar biological processes, such as producing tears or sweating.
2. "stinging" → **behavior**: This is like familiar behaviors such as snapping your fingers.
3. "working with other bees" → **social**: This is characteristically social in the way that, for instance, playing a game with other people is social.
4. "pollinating flowers" → **purpose**: This is the sort of thing that might be one of their purposes in the same way that you might think one purpose of clouds is to rain.

Continue

Figure A1. **Experiment 1**: First example given to participants illustrating how they might view different properties as belonging to different property types.

Instructions

Example 2. Suppose you come across a venus fly trap and observe four things about it: it photosynthesizes, snaps its leaves shut, sends signals to venus fly traps and traps insects.

1. "photosynthesizes" → **biology**: This is like familiar biological processes, such as digesting.
2. "snaps its leaves shut" → **behavior**: This is like familiar behaviors such as chewing.
3. "sending signals to other venus fly traps" → **social**: This is characteristically social in the way that, for instance, communicating with other people is social.
4. "traps insects" → **purpose**: This is the sort of thing that might be one of their purposes in the same way that one of the purposes of vultures is to eat carcasses.

We will now ask you a few comprehension check questions before you can start with the experiment.

Continue

Figure A2. **Experiment 1**: Second example given to participants illustrating how they might view different properties as belonging to different property types.

Appendix B
Experiment 2: Property categorization

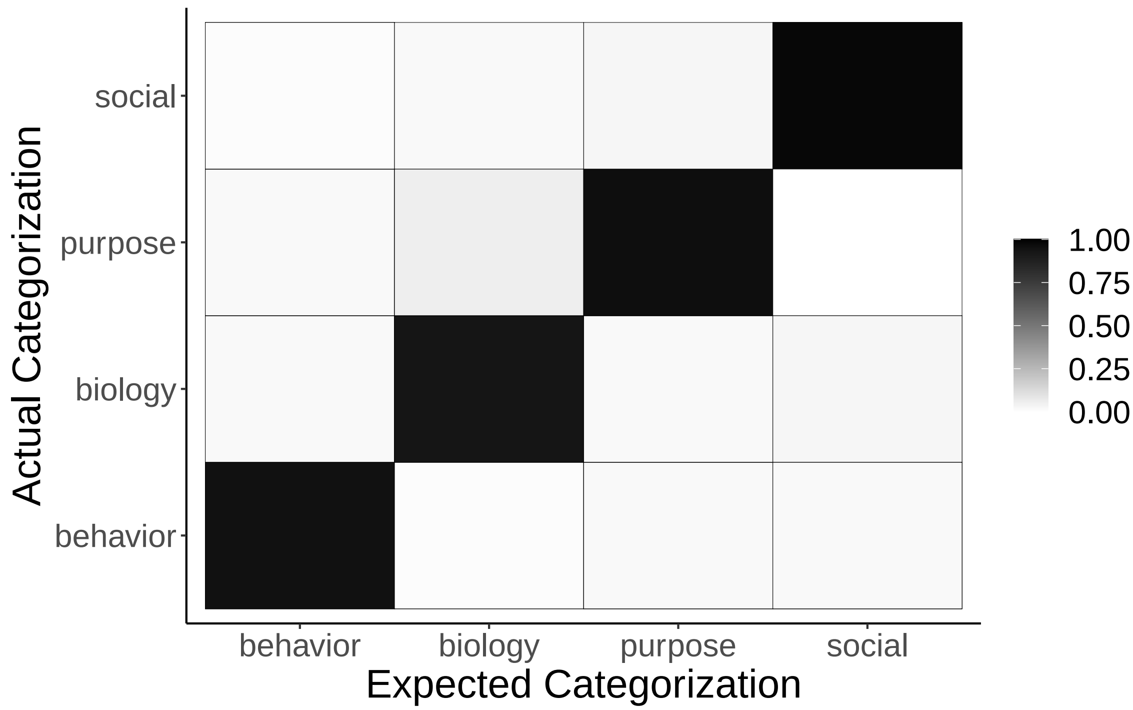


Figure B1. **Experiment 2**: Proportion of expected property categorization (x-axis) versus actual property categorization (y-axis).

Appendix C
Experiment 3: Property categorization

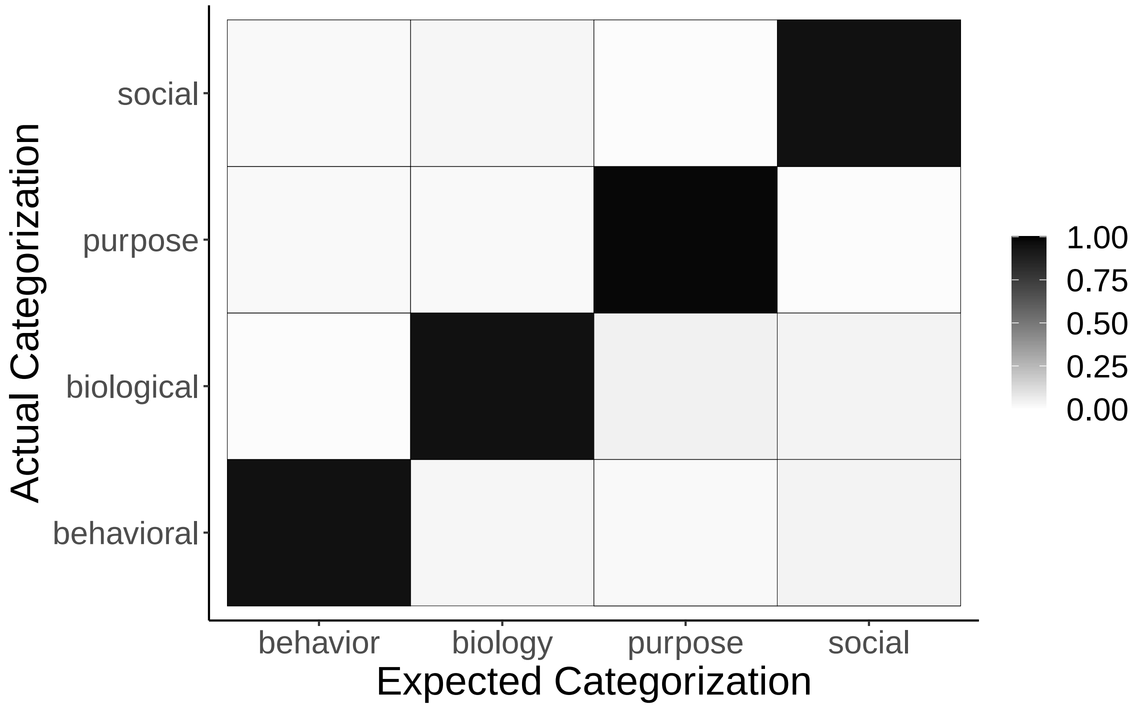


Figure C1. **Experiment 3**: Proportion of expected property categorization (x-axis) versus actual property categorization (y-axis).

Appendix D
Experiment 4: Property categorization

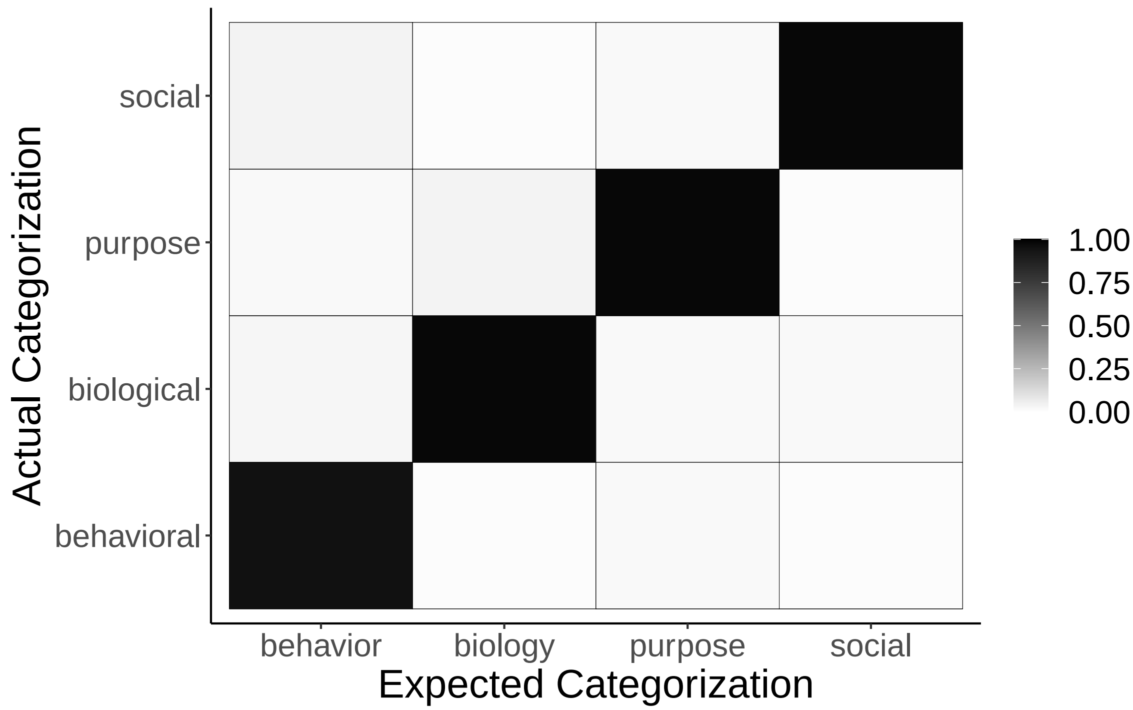


Figure D1. **Experiment 3**: Proportion of expected property categorization (x-axis) versus actual property categorization (y-axis).