COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–24

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12440

Differences in the Evaluation of Generic Statements About Human and Non-Human Categories

Arber Tasimi,[a] Susan A. Gelman,[b] Andrei Cimpian,[c,d] Joshua Knobe[a]

[a]*Department of Psychology, Yale University*
[b]*Department of Psychology, University of Michigan*
[c]*Psychology Department, University of Illinois at Urbana-Champaign*
[d]*Department of Psychology, New York University*

Received 26 August 2015; received in revised form 22 June 2016; accepted 14 July 2016

Abstract

Generic statements (e.g., "Birds lay eggs") express generalizations about categories. Current theories suggest that people should be especially inclined to accept generics that involve threatening information. However, previous tests of this claim have focused on generics about non-human categories, which raises the question of whether this effect applies as readily to human categories. In Experiment 1, adults were more likely to accept generics involving a threatening (vs. non-threatening) property for artifacts, but this negativity bias did not also apply to human categories. Experiment 2 examined an alternative hypothesis for this result, and Experiments 3 and 4 served as conceptual replications of the first experiment. Experiment 5 found that even preschoolers apply generics differently for humans and artifacts. Finally, Experiment 6 showed that these effects reflect differences between human and non-human categories more generally, as adults showed a negativity bias for categories of non-human animals, but not for categories of humans. These findings suggest the presence of important, early-emerging domain differences in people's judgments about generics.

Keywords: Generic language; Concepts; Cognitive development; Psychological essentialism

1. Introduction

Consider the following statement: "Sharks attack people." This is a *generic statement*—that is, a statement that expresses a generalization about an entire category (Carlson, 1977; Carlson & Pelletier, 1995; Gelman, 2003; Leslie, 2008). Many people consider this statement to be true, despite knowing that the vast majority of sharks never attack people.

Correspondence should be sent to Arber Tasimi, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520. E-mail: arber.tasimi@yale.edu

Now, consider the following statement: "Men attack people." In fact, the proportion of men who attack people is greater than the proportion of sharks that do so, yet many people would disagree with this second statement. This intuition illustrates the hypothesis investigated here: namely, that there may be important differences in the acceptability of generic statements that express dangerous, harmful, or threatening information about human versus non-human categories.

Recent theoretical work suggests that because generic sentences serve as a linguistic outlet for our conceptual representations, people should be especially inclined to accept generics that involve dangerous, harmful, or threatening (henceforth, "threatening") information (Leslie, 2008, in press). For example, witnessing a single instance of a shark attacking a person should lead to the conclusion that "Sharks attack people" because undergeneralizing such information could have profound consequences. Initial evidence for this proposal demonstrated that generic statements about non-human categories are indeed sensitive to the content of the properties being generalized (Cimpian, Brandone, & Gelman, 2010). Participants were more likely to accept generics expressing threatening properties of animals (e.g., "Zorbs have venomous purple feathers") than neutral properties (e.g., "Zorbs have purple feathers"), even when the statistical evidence for these statements was perfectly matched (e.g., 30% of zorbs display the relevant property). Thus, threatening information holds a privileged status in how we represent kinds.

The proposal that people have a tendency to rapidly generalize threatening information raises the further question of whether such a tendency also influences how we reason about categories of humans. For example, just as it takes only a few shark attacks for people to endorse the corresponding generic ("Sharks attack people"), does it likewise take the threatening actions of just a few members of a social group (e.g., men attacking individuals) for people to hold a general belief about the entire group in generic form (i.e., "Men attack people")? In other words, is the tendency to readily accept generics about threatening properties a domain-general fact about generic statements or, alternatively, might generics about *human categories* be in some way distinctive?

Consistent with the former possibility, a number of studies have documented that people show a negativity bias in judgments about humans (i.e., bad impressions are quicker to form and are more stable than good ones; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001; Vaish, Grossmann, & Woodward, 2008). Such evidence suggests that, with respect to generics, human categories would be treated like the animal categories investigated in prior work. On the other hand, it might be that people have a distinctive approach to thinking about humans that differs in important respects from the way they think about categories of other types. In particular, people tend to conclude that there is some deeper sense in which humans are fundamentally good (Newman, Bloom, & Knobe, 2014). Even when participants are told explicitly that a particular human being consistently has morally bad desires and performs morally bad actions, they still show a tendency to conclude that, deep down, there is some core essential part of this human being that is good. In combination with the fact that generic statements are typically interpreted as expressing deep, essential properties (e.g., Carlson & Pelletier, 1995; Cimpian & Cadena, 2010; Cimpian & Markman, 2009, 2011; Gelman, 2004;

Lyons, 1977), this may mean that people would not endorse generics that involve threatening properties more than those that involve non-threatening ones for human categories, in contrast with their generic judgments about non-human categories.

In this investigation, we explored the generality of the previously hypothesized tendency to accept generics about threatening properties more easily than other generics. In particular, we asked whether people endorse generic statements about threatening properties more than about non-threatening ones for human categories, in much the same way as they do for non-human categories. Six experiments explored this issue. Experiment 1 tested whether people endorse generics similarly or differently for novel human and non-human (specifically, artifact) categories. This experiment revealed a tendency to accept generics involving threatening information (more than non-threatening information) for novel artifact categories but *not* novel human categories. Experiment 2 examined an alternative hypothesis regarding expectations about base rates in the different domains (i.e., are people assumed to differ from artifacts in how dangerous they are?), and Experiments 3 and 4 served as conceptual replications of the first experiment. Experiment 5 examined preschoolers' endorsement of generic statements and found that children, like adults, show different patterns for human versus artifact categories. Because young children, unlike adults, are generally not concerned with appearing unbiased when explicitly reasoning about social categories (e.g., Abrams, Rutland, Cameron, & Ferrell, 2007; Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008), there is reason to conclude that an absence of a negativity bias for human categories in their responses would not be due to a strategy of avoiding the appearance of prejudice. Finally, Experiment 6 explored whether the effects from the previous experiments are restricted to comparisons between humans and artifacts, or whether they extend to comparisons of humans and non-human categories more generally. This experiment demonstrated that whereas adults once again did not accept generics more for threatening versus non-threatening information for humans, they did do so for categories of non-human animals, thus treating non-human animals in much the same way as artifacts in the previous experiments. Together, these studies suggest important differences in people's evaluation of generics about human and non-human categories.

2. Experiment 1

2.1. Method

2.1.1. Participants

Four-hundred adults (286 men, 114 women; M = 26 years; range = 18–69 years) completed the study online for 10 cents each via Amazon's Mechanical Turk (MTurk).

2.1.2. Procedure

Each participant was assigned to a *valence* (dangerous or wonderful), a *domain* (people or tools), and a *prevalence* (varying from 10% to 100% in increments of 10). We examined opposing valences and chose tools for a non-human category as an extension of previous work that contrasted threatening and neutral information about non-human animal categories (Cimpian et al., 2010). Participants received and evaluated a single statement that embodied a particular combination of the three factors (valence, domain, and prevalence), with reference to a novel category (Krens/krens). For example:

> Imagine that there is a land far away where you can find people (tools) called Krens (krens). Below, you will read some information about Krens (krens).
>
> 30% of Krens (krens) are dangerous (wonderful).
>
> How true is the following sentence about these people (tools)?
>
> Krens are dangerous (wonderful).

After reading the statement, participants evaluated it on a seven-point scale anchored by *not true at all* (1) and *completely true* (7).

2.2. Results and discussion

We conducted a multiple regression with valence, domain, prevalence, and all their two- and three-way interactions as predictors. All predictors were mean centered to facilitate interpretation of the coefficients; we report standardized coefficients. Valence was a significant predictor of participants' truth ratings, $\beta = .16$, $p < .001$, indicating that generic sentences regarding a threatening property ($M = 4.49$) were judged to be true more often than those regarding a non-threatening property ($M = 3.97$). In addition, prevalence significantly predicted truth ratings, $\beta = .63$, $p < .001$, with generics being judged to be true more often as the prevalence level increased. This analysis also yielded a domain $\times$ valence interaction, $\beta = .09$, $p = .018$, which is consistent with the prediction that participants' evaluation of generic statements differed significantly by domain. No other coefficients were significant.

Given the interaction, we conducted a separate regression in each domain. Consistent with prior work, generic statements involving tools were judged to be true more often when they described threatening ($M = 4.69$) than non-threatening ($M = 3.86$) properties, $\beta = .24$, $p < .001$; see Fig. 1A. By contrast, for generics involving people, there was no significant difference between threatening ($M = 4.29$) and non-threatening ($M = 4.07$) properties, $\beta = .07$, $p = .24$; see Fig. 1B.

In total, these findings provide initial support for the idea that people differentiate between human and non-human (tool) categories when evaluating generic sentences involving threatening (dangerous) and non-threatening (wonderful) properties.

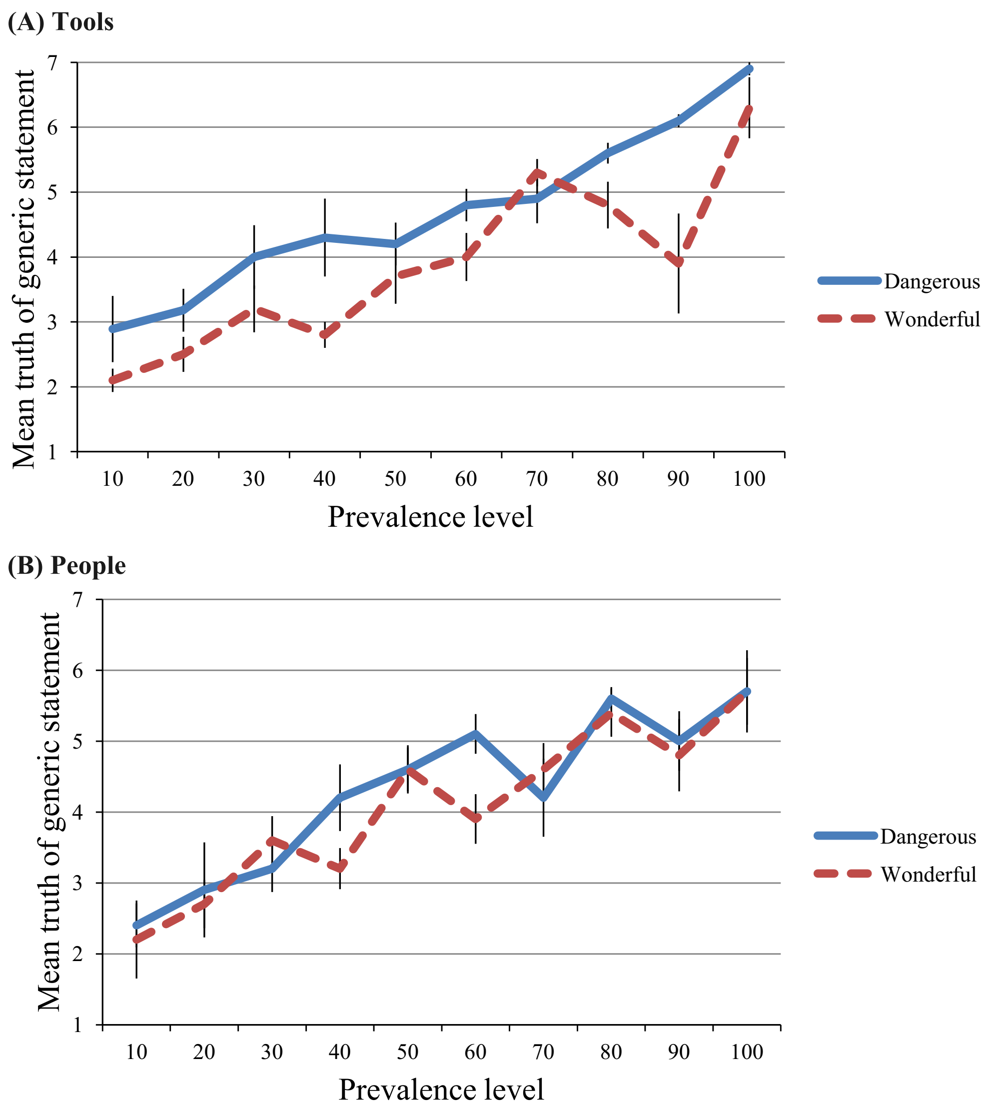

Fig. 1. (A) Participants' mean ratings of the truth of the generic statement, on a scale of 1–7, for the category of "tools" in Experiment 1. Error bars represent standard error. (B) Participants' mean ratings of the truth of the generic statement, on a scale of 1–7, for the category of "people" in Experiment 1. Error bars represent standard error.

3. Experiment 2

Experiment 1 found a difference in how people evaluate generic sentences about human and non-human (tool) categories. It is possible, however, that this finding could simply reflect a difference in base rates of certain properties within human versus non-human categories, rather than fundamental differences in the acceptability of generic statements in these domains. As has been noted previously, people's intuitions about the acceptability of describing a particular category using a generic depend not only on the prevalence of a property within that category but also on its prevalence in other categories (Cohen, 1999). For example, consider the sentence "Bulgarians are good weightlifters." To the extent that people regard this sentence as true, it is not because they think that the absolute percentage of Bulgarians who are good weightlifters is itself high, but rather because they think that the percentage is high *relative* to the percentages found for other nationalities. Thus, if humans are generally assumed to be more dangerous than tools, then the threatening information in Experiment 1 would be relatively more distinctive for the tool categories than for the human categories (relative to their respective baselines), which might, in turn, make the threatening generics about tools (vs. humans) more acceptable (see also Cimpian et al., 2010).

Note, however, that the same difference in base rates could also make the generic *less* acceptable: If humans are generally assumed to be more dangerous than tools, then participants may more readily conclude that a new category of humans is dangerous. Either way, differences in base rates would introduce uncertainty in the interpretation of the results from Experiment 1. To investigate this issue, in Experiment 2, we asked participants to report their baseline expectations about whether tools and people exhibit threatening versus non-threatening properties.

3.1. Method

3.1.1. Participants

Three hundred twenty-three adults (223 men, 100 women; $M = 28$ years; age range = 18–67 years) completed the study online for 10 cents via MTurk.

3.1.2. Procedure

Each participant was assigned to a *valence* (dangerous or helpful) and a *domain* (people, tools, or things). We changed the non-threatening property from "wonderful" to "helpful" because the latter is more closely matched to the threatening property used in our experiments (i.e., both "dangerous" and "helpful" entities have a direct impact on others). In addition, we included *things* as a domain because it is a more superordinate category than *tools*, and it is thus better matched with *people*. This domain could thus be used for a tighter comparison with *people* in subsequent experiments, especially if the base rates are also similar (see Experiments 3–5 below).

Participants responded to a single question asking what percentage of the relevant category's members possesses the relevant property. For example:

> Imagine that there is a land far away where you can find people (things, tools) called Merts (merts). What percentage of Merts (merts) do you think are dangerous (helpful)?

After reading the question, participants were asked to enter a number between 0 and 100.

3.2. Results and discussion

Results are displayed in Table 1. A 3 (domain) × 2 (valence) ANOVA did not yield an interaction between domain and valence, $F(2, 317) = 1.14$, $p = .32$, which argues against domain differences in baseline rates of threatening or non-threatening properties. We nevertheless conducted two follow-up analyses to check for domain differences separately for dangerous (threatening) and helpful (non-threatening) expectations.

When asked to predict what percentage of Merts (merts) are dangerous, there was a significant effect of domain, $F(2, 158) = 3.30$, $p = .039$, $\eta_p^2 = .04$. Participants judged tools (M = 36%) to be more dangerous than people (M = 25%), $t(105) = 2.20$, $p = .03$, and things (M = 25%), $t(105) = 2.11$, $p = .04$. There was no difference between the latter two categories, $t(106) = 0.08$, $p = .94$. In contrast, estimations regarding helpfulness did not differ by domain (people: M = 61%, things: M = 58%, and tools: M = 63%), $F(2, 159) = 0.45$, $p = .64$.

To speculate, the lower base rate of dangerousness for people (vs. tools) may have made it more likely for participants in the previous experiment to agree with generics about human (vs. tool) categories that involve threatening information. For example, learning that 50% of people in a category are dangerous presents a starker contrast to the presumed base rates of dangerousness among humans than learning that 50% of tools in a category are dangerous. This starker contrast could have led participants to readily conclude that this category of people is dangerous, which would have made it easier to find a negativity bias for human categories. In light of these considerations, it may be particularly revealing that we found no negativity bias for these categories. On the other hand, the lower base rate of dangerousness for people (vs. tools) may have made it more likely that participants would judge that a new category of tools is dangerous because tools are generally assumed to be dangerous (at least relative to people).

Regardless, to avoid any interpretive issues due to differences in base rates, in Experiment 3 we provide a more controlled test of the potential differences in participants' evaluation of generics about human versus non-human categories. Specifically, the comparable base rates for the domains of *people* and *things* (see Table 1) permit such a controlled test of people's judgments about generic sentences across domains.

Table 1
Participants' mean estimations, on a scale of 1–100, of the dangerousness and helpfulness of the three domains in Experiment 2

	Dangerous	Helpful
People	25 (22)	61 (18)
Things	25 (27)	58 (27)
Tools	36 (30)	63 (24)

Note. Standard deviations are in parentheses.

4. Experiment 3

Experiment 3 served as a conceptual replication of the first experiment. We contrasted *people* with *things* in this experiment, given their comparable level of generality and their equivalent base rates in Experiment 2. We also contrasted *dangerous* with *helpful*, as these attributes are more closely matched to one another than *dangerous* and *wonderful*.

4.1. Method

4.1.1. Participants

Eight hundred adults (439 men, 361 women; M = 30 years; age range = 18–72 years) completed the study online for 10 cents each on MTurk. The sample size was doubled relative to Experiment 1 to provide a high-powered conceptual replication.

4.1.2. Procedure

The procedure was the same as in Experiment 1, with two exceptions: The non-human category was labeled as *things*, and the non-threatening property was *helpful* instead of *wonderful*.

4.2. Results and discussion

We conducted a multiple regression with valence, domain, prevalence, and all their two- and three-way interactions as predictors. All predictors were mean centered to facilitate interpretation of the coefficients; we report standardized coefficients. Valence was again a significant predictor of participants' truth ratings, $\beta = .08$, $p < .001$, as was prevalence, $\beta = .80$, $p < .001$. Unlike in Experiment 1, this analysis did not yield a significant domain × valence interaction, $\beta = .03$, $p = .15$.[1] No other coefficients were significant.

Despite the non-significant domain × valence interaction, we looked separately at the results for each domain. As in Experiment 1, generic statements involving non-human entities (*things*) were judged to be true more often when they described threatening (M = 4.71) than non-threatening (M = 4.36) properties, $\beta = .11$, $p < .001$; see Fig. 2A. For generics involving people, there was no significant difference between threatening (M = 4.50) and non-threatening (M = 4.34) properties, $\beta = .05$, $p = .09$; see Fig. 2B.

Taken together, these findings provide additional support for the idea that people show a negativity bias in judgments about categories of artifacts, but not categories of humans.

5. Experiment 4

Experiment 4 investigated adults' judgments about generics for human and non-human categories using a visual task that could be employed with children (see Experiment 5).

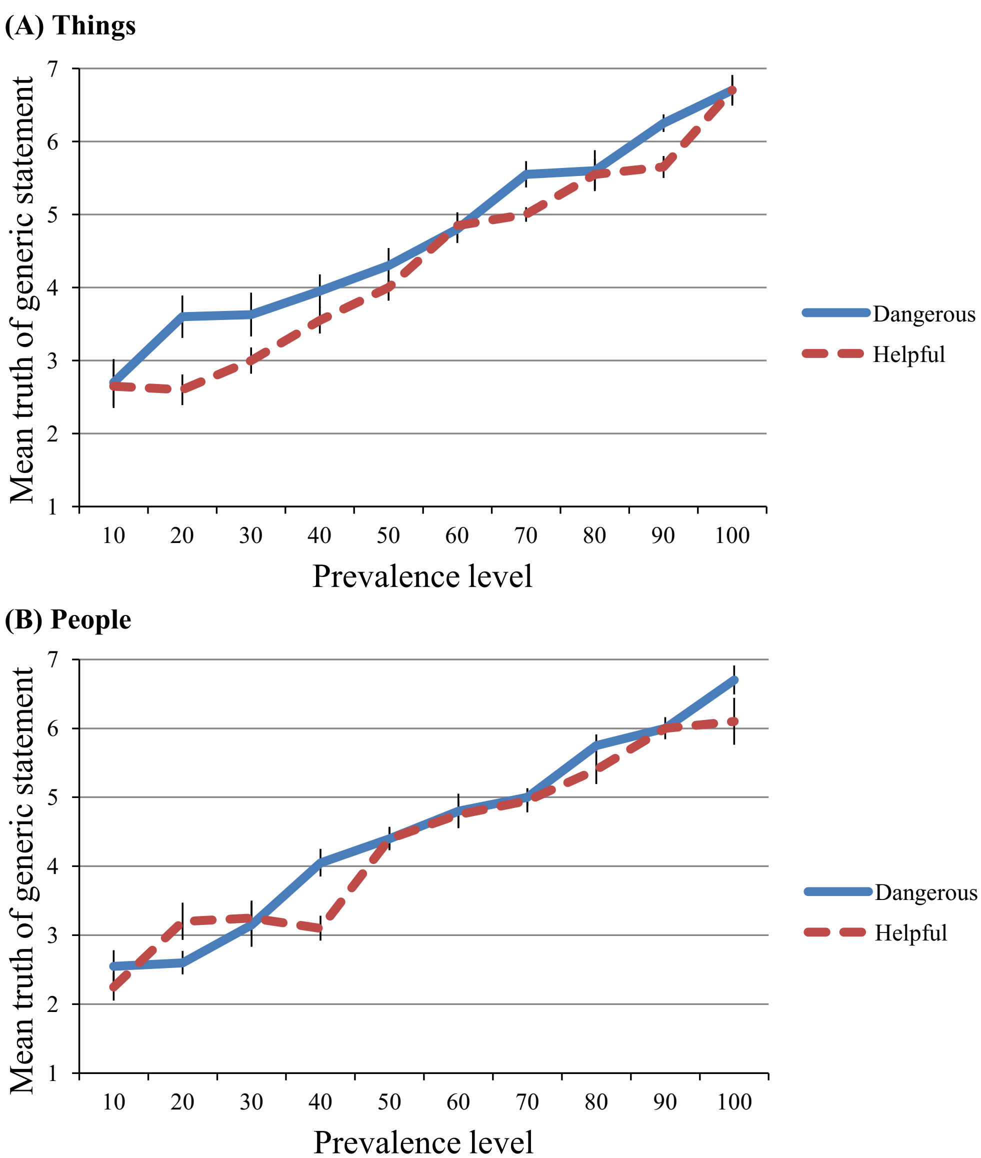

Fig. 2. (A) Participants' mean ratings of the truth of the generic statement, on a scale of 1–7, for the category of "things" in Experiment 3. Error bars represent standard error. (B) Participants' mean ratings of the truth of the generic statement, on a scale of 1–7, for the category of "people" in Experiment 3. Error bars represent standard error.

5.1. Method

5.1.1. Participants

Sixty-four adults (28 men, 36 women; M = 23 years; range = 18–52 years) from the New Haven community participated for $2 each.

5.1.2. Procedure

Participants were tested in person and individually on the campus of Yale University. We adapted a method from Brandone, Gelman, and Hedglen (2015) that was used to examine preschoolers' and adults' intuitions regarding the semantics of generic statements. Each participant was assigned to a *domain* (people or things). The study consisted of two blocks differing in *valence* (dangerous vs. helpful). These blocks were separated with a distractor task (the memory game Simon), which participants played for 2 min. Within each block, there were four different, novel kinds. For each kind, six exemplars were depicted (see Figs. 3 and 4). The number of exemplars within each sample exhibiting the property involved in the generic (dangerous or helpful) varied, with four prevalence levels: 0 of 6 (0%), 2 of 6 (33%), 4 of 6 (67%), and 6 of 6 (100%). Although our main focus was on the intermediate prevalence levels (33% and 67%), we included the 0% and 100% prevalence levels as a way of ascertaining that participants properly understood the task. In other words, we expected participants to largely disagree with the generic at the 0% prevalence level and largely agree with the generic at the 100% prevalence level. The novel kinds were rotated throughout the blocks, across participants (e.g., "krens" were presented at each prevalence level equally often, across participants). Participants were asked to circle whether a corresponding statement (e.g., "Krens are dangerous") was "right" or "wrong" about each kind. Block order was counterbalanced using a Latin Square design.

At the beginning of each block, participants were provided with a sheet of instructions explaining which exemplars corresponded to which attributes (e.g., "A person that looks like this is dangerous; he has a dangerous face"; "A person that looks like this is helpful; he has a helpful face"; "A thing that looks like this is dangerous; it has sharp spikes"; "A thing that looks like this is helpful; it has a soft brush"). Exemplars lacking the relevant properties were described as not being dangerous (e.g., "A person that looks like this is not dangerous; he does not have a dangerous face") or helpful (e.g., "A person that looks like this is not helpful; he does not have a helpful face").

5.2. Results and discussion

As expected, participants largely disagreed with the generic at the 0% prevalence level (M = 100% "wrong" responses) and largely agreed with the generic at the 100% prevalence level (M = 97% "right" responses).

Because the design involved a dichotomous dependent measure, a repeated-measures binary logistic regression (RM-BLR) was conducted, with domain (people vs. things; between subjects), valence (dangerous vs. helpful; within subject), prevalence (33% and

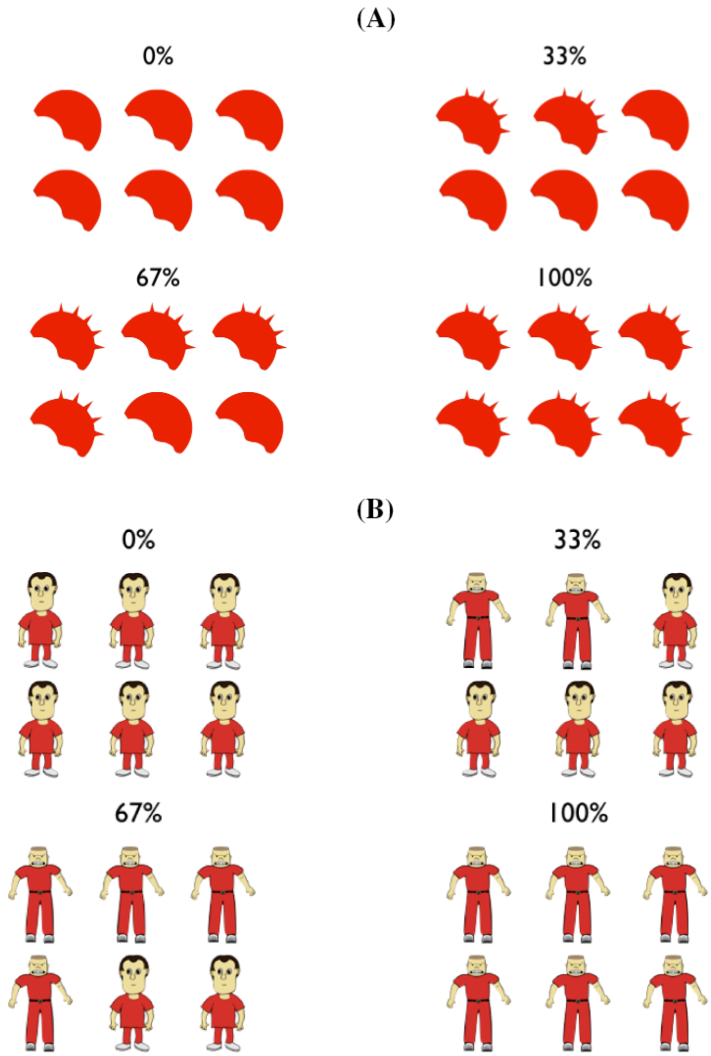

Fig. 3. (A) Sample category of things ("krens") showing target feature ("dangerous") at each of the four prevalence levels (0%, 33%, 67%, and 100%). (B) Sample category of people ("Krens") showing target feature ("dangerous") at each of the four prevalence levels (0%, 33%, 67%, and 100%).

67%; within subject), and their two- and three-way interactions as predictors. The RM-BLR revealed a main effect of domain, Wald $\chi^2 = 11.16$, $df = 1$, $p = .001$, indicating that participants were more willing to endorse generics about things ($M = 65\%$) than people ($M = 39\%$), as well as a significant effect of prevalence, Wald $\chi^2 = 60.79$, $df = 1$, $p < .001$, indicating that generic sentences were more acceptable for higher than lower

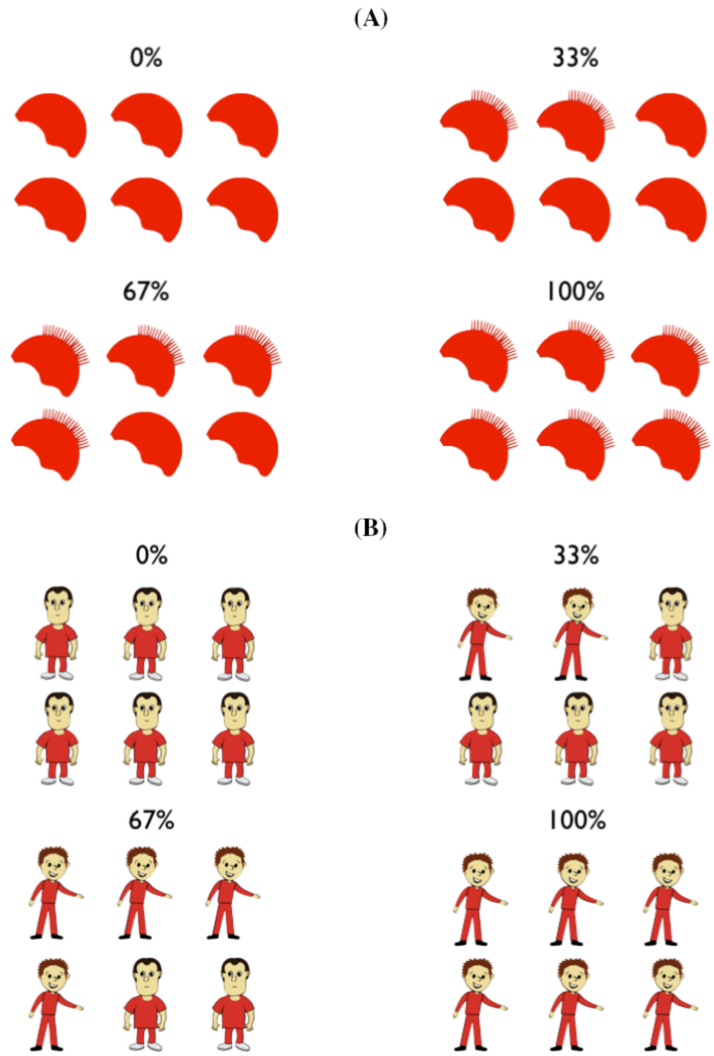

Fig. 4. (A) Sample category of things ("krens") showing target feature ("helpful") at each of the four prevalence levels (0%, 33%, 67%, and 100%). (B) Sample category of people ("Krens") showing target feature ("helpful") at each of the four prevalence levels (0%, 33%, 67%, and 100%).

prevalence levels. There was no significant effect of valence ($M_{\text{dangerous}} = 57\%$; $M_{\text{helpful}} = 47\%$), Wald $\chi^2 = 3.21$, $df = 1$, $p = .073$. Importantly, this analysis also yielded the predicted interaction between domain and valence, Wald $\chi^2 = 7.58$, $df = 1$, $p = .006$; see Fig. 5. No other effects were significant.

Given the domain × valence interaction, we looked separately at the results for each domain. For generic sentences about things, statements involving a threatening property

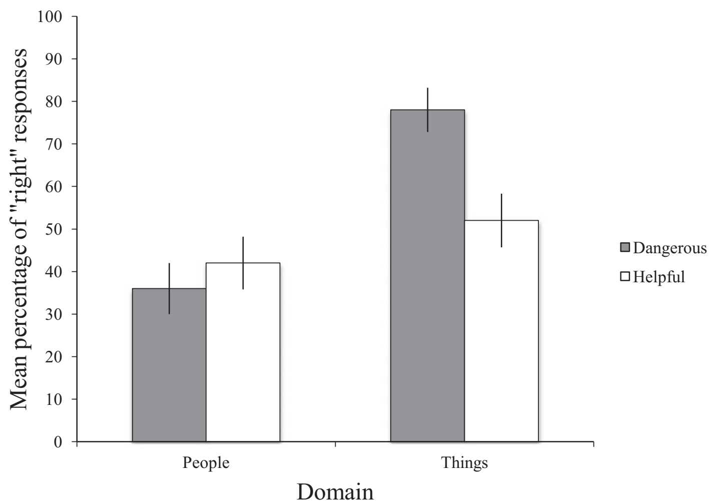

Fig. 5. Mean percentage of "right" responses in Experiment 4, by domain and valence. Error bars represent standard error.

(M = 78%) were endorsed more than statements involving a non-threatening property (M = 52%), Wald $\chi^2 = 8.87$, $df = 1$, $p = .003$. By contrast, for generic sentences about people, there was no difference between threatening (M = 36%) and non-threatening (M = 42%) properties, Wald $\chi^2 = 0.55$, $df = 1$, $p = .46$. This asymmetry between the acceptability of threatening (vs. non-threatening) generics about human and non-human categories replicates the findings reported in Experiments 1 and 3.

In sum, these findings provide further evidence that adults treat generic sentences differently for categories of humans and non-humans, as in Experiments 1 and 3. Next, we investigate whether young children also show differences in their evaluations of generics for human and non-human categories.

6. Experiment 5

Experiments 1, 3, and 4 find that adults' judgments concerning generic statements differ between human and non-human categories. We have suggested that this result reflects conceptual differences in the kinds of generalizations that people make across domains. An alternative interpretation, however, is that participants in the previous experiments were simply concerned about appearing biased and were thus unwilling to (openly) endorse generics involving threatening information about categories of people. To explore this possibility, we tested young children in Experiment 5 because they are generally far less concerned than adults with appearing unbiased when explicitly reasoning about social categories (e.g., Abrams et al., 2007; Apfelbaum et al., 2008). Thus, if children show the same domain difference in their judgments about generics as adults did, it seems less likely that such an asymmetry could be attributed to concerns about appearing unbiased.

6.1. Method

6.1.1. Participants

Sixty-four preschoolers (31 boys, 33 girls; $M = 4.81$ years; age range = 4.18–5.99 years) participated in the study. Participants were recruited from the greater New Haven, Connecticut, area and tested individually in a quiet room at their preschool. Two additional children were tested but excluded because they provided the same response across all eight trials.

6.1.2. Procedure

The same procedure and materials from Experiment 4 were used, with several modifications to make the task more appropriate for young children. First, we framed the study as a game. We introduced Newton, a puppet from outer space who gets confused, so sometimes he says things that are right and sometimes he says things that are wrong. Children were told that their job in the game was to decide if what Newton says is right or wrong. Second, the task began with four practice trials used to convey the options of "right" and "wrong" in the context of the task (e.g., the experimenter showed a picture of a banana, which Newton said was an apple, and children were asked if Newton was right or wrong). Third, we included a training phase at the beginning of each block in which children were told which items depicted dangerous (or helpful) items. For children assigned to the domain of things, dangerous things were described as having sharp spikes and non-dangerous things as not having sharp spikes (see Fig. 3A); helpful things were described as having a soft brush and non-helpful things as not having a soft brush (see Fig. 4A). For children assigned to the domain of people, dangerous people were described as having a dangerous face and non-dangerous people as not having a dangerous face (see Fig. 3B); helpful people were described as having a helpful face and non-helpful people as not having a helpful face (see Fig. 4B). The experimenter then showed children four new types of things (or people) and asked children to identify whether each item was dangerous or helpful. Training ended only after the child responded to each item correctly. Fourth, we read the generic statements to the children (e.g., "Krens are dangerous"), rather than having children read them (as adults did in the previous experiment); children were then asked to identify each statement as "right" or "wrong." Finally, we introduced a child-friendly distractor game, which participants played on an iPad for 2 min in between the two blocks.

6.2. Results and discussion

As expected, participants largely disagreed with the generic at the 0% prevalence level (M = 87% "wrong" responses) and largely agreed with the generic at the 100% prevalence level (M = 92% "right" responses).

As in Experiment 4, a RM-BLR with domain (people vs. things; between subjects), valence (dangerous vs. helpful; within subject), prevalence (33% and 67%; within subject), and their two- and three-way interactions as predictors was conducted. The RM-BLR did not reveal a significant effect of domain (M_{things} = 66%; M_{people} = 59%), Wald $\chi^2 = 1.41$, $df = 1$, $p = .23$, suggesting that children did not accept generic statements more in one domain than another. In addition, there was a marginal effect of prevalence, Wald $\chi^2 = 3.37$, $df = 1$, $p = .066$, and no significant effect of valence ($M_{\text{dangerous}}$ = 59%; M_{helpful} = 66%), Wald $\chi^2 = 1.02$, $df = 1$, $p = .31$. This analysis also revealed an interaction between valence and prevalence, Wald $\chi^2 = 3.97$, $df = 1$, $p = .046$, and, importantly, the predicted interaction between domain and valence, Wald $\chi^2 = 5.59$, $df = 1$, $p = .018$; see Fig. 6. No other effects were significant.

Given the domain × valence interaction, we looked separately at the results for each domain. Children did not differentiate between threatening (M = 70%) and non-threatening statements (M = 63%) when judging generics about things, Wald $\chi^2 = 0.92$, $df = 1$, $p = .34$. However, when judging generics about people, children accepted non-threatening statements (M = 70%) more than threatening statements (M = 47%), Wald $\chi^2 = 5.70$, $df = 1$, $p = .017$.

Taken together, these findings suggest that children, like adults, show an asymmetry in how they think about categories of humans and non-humans. However, the pattern of children's responses in this experiment differed from that displayed by adults in the previous experiments. For adults, the valence effect was within the domain of artifacts, whereby generics involving threatening information were endorsed more than those involving non-threatening information. By contrast, for children, the valence effect was within the domain of humans, whereby generics involving non-threatening information

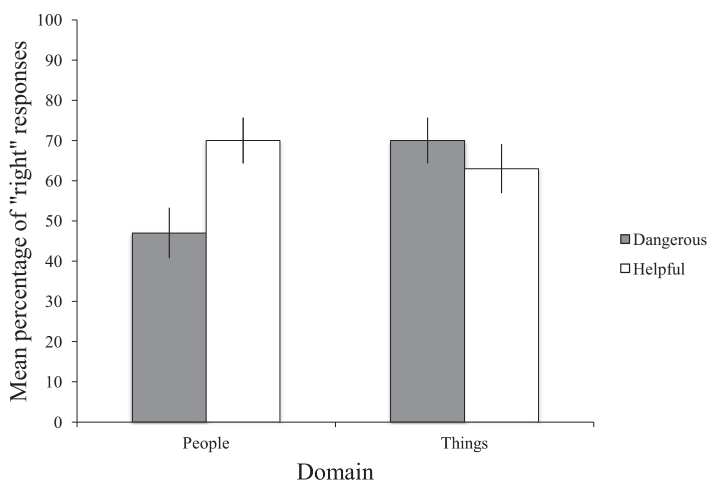

Fig. 6. Mean percentage of "right" responses in Experiment 5, by domain and valence. Error bars represent standard error.

were endorsed more than those involving threatening information. This positivity advantage among children is consistent with previous work showing a positivity bias in their reasoning about personality traits, whereby children generalize positive information more readily than negative information about other people (Boseovski, 2010).

A potential alternative explanation for these findings is that perhaps children thought that the neutral human characters looked more likely to be capable of being helpful than dangerous, which could explain why children were more likely to endorse generics involving non-threatening information for human categories. However, this account would predict that at the 0% prevalence level, children should also be more likely to endorse the non-threatening generic than the threatening generic for humans. In fact, however, there was no difference at the 0% prevalence level between the threatening generic (1 of 32 children said "right") and the non-threatening one (2 of 32 children said "right").

Moreover, it is notable that children did not show a negativity bias in their generic judgments about artifacts; indeed, children accepted generic statements involving threatening and non-threatening properties at comparable rates. One explanation of this null difference is that the artifacts used in this study were unfamiliar to children, who may not have known what to think of them. Moreover, the use of the label "things" might have increased the novelty of the artifacts and, as a result, children may not have been able to effectively reason about them, unlike human categories that are familiar to children. Of course, it may also be that the absence of a negativity bias speaks to an absence of a negativity bias in children's generic judgments more generally. Although additional research is needed to address this issue, these findings suggest the presence of early-emerging domain differences in people's judgments about generic statements.

7. Experiment 6

The experiments reported thus far demonstrate consistent domain differences in the evaluation of generic statements, but the precise nature of this domain difference remains unclear. Experiments 1–5 presented a rather stark contrast between humans on the one hand and artifacts on the other, a distinction that is consistent with a variety of conceptual distinctions (e.g., living vs. non-living, animate vs. inanimate, human vs. non-human), all of which are available to both adults and young children (e.g., Hirschfeld & Gelman, 1994). An important next step is to clarify the basis of the demonstrated effects. In this context, animals provide a critical contrast because they are distinct from humans but, like humans, are both living and animate. Contrasting humans with non-human animals provides a minimal pair that will shed light on the conceptual basis of the phenomenon established in the prior studies. Thus, in Experiment 6, we assess adults' generic interpretations concerning novel categories of humans versus non-human animals. In addition, we included a broader range of threatening and non-threatening properties, to assess the generality of the effects.

7.1. Method

7.1.1. Participants

Two hundred adults (121 men, 79 women; $M = 35$ years; age range = 18–72 years) completed the study online for 60 cents each on MTurk.

7.1.2. Procedure

Each participant was assigned to a *domain* (people or animals). The study consisted of two blocks differing in *valence* (threatening vs. non-threatening). These blocks were separated with an anagram task, which participants played for 2 min. At the beginning of each block, participants were asked to imagine faraway lands where they could find people or animals. Within each block, there were five different, novel kinds. Five different properties were used in the threatening block (dangerous, harmful, hostile, mean, and threatening), and five different properties were used in the non-threatening block (comforting, friendly, gentle, helpful, and nice). The percentage of the kind exhibiting the property involved in the generic (e.g., hostile) varied, with five prevalence levels: 10%, 30%, 50%, 70%, and 90%. The novel kinds were rotated throughout the blocks, across participants (i.e., each property was presented at each prevalence level equally often, across participants). Participants were asked to indicate whether a corresponding statement (e.g., "Krens are gentle") was "true" or "false" about each kind. Block order was counterbalanced using a Latin Square design.

7.2. Results and discussion

Participants' true/false judgments were analyzed with a multilevel logistic regression model that allowed each subject's intercept to vary randomly. Domain (dichotomous), valence (dichotomous), and prevalence (continuous), as well as all their two- and three-way interactions, were included as independent variables. This analysis revealed a main effect of valence, $b = .34$, $SE = .14$, $z = 2.39$, $p = .017$, indicating that participants were more willing to endorse generics about threatening ($M = 59\%$) than non-threatening properties ($M = 54\%$), as well as a significant effect of prevalence, $b = .09$, $SE = .004$, $z = 20.91$, $p < .001$, indicating that generic sentences were more acceptable for higher than lower prevalence levels. There was no significant effect of domain (*M*s = 56% and 57% for humans and animals, respectively), $b = .09$, $SE = .27$, $z = 0.31$, $p = .75$. Critically, this analysis also revealed the predicted interaction between domain and valence, $b = .64$, $SE = .29$, $z = 2.23$, $p = .026$. No other effects were significant.

Given the domain × valence interaction, we looked separately at the results for each domain. Consistent with prior work (Cimpian et al., 2010), generic statements about non-human animals were judged true more often when the properties were threatening ($M = 60\%$) than when they were non-threatening ($M = 53\%$), $b = .08$, $SE = .02$, $z = 3.76$, $p < .001$; see Fig. 7A. In contrast, and as predicted by our hypothesis, the bias for threatening information did not hold when participants evaluated generic statements

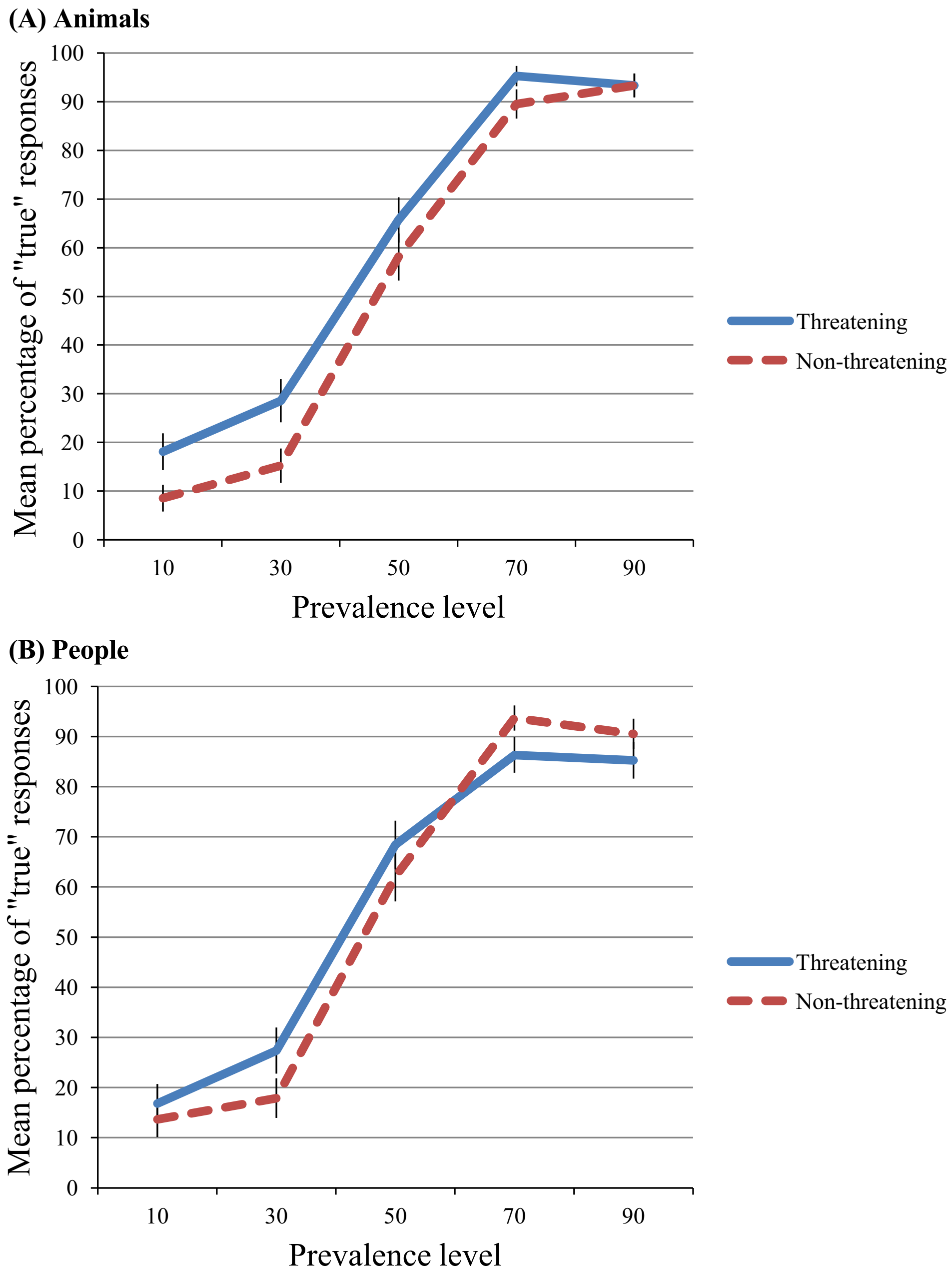

Fig. 7. (A) Mean percentage of "true" responses, by prevalence and valence, for the category of "animals" in Experiment 6. Error bars represent standard error. (B) Mean percentage of "true" responses, by prevalence and valence, for the category of "people" in Experiment 6. Error bars represent standard error.

about people (*M*s = 57% and 56% for threatening and non-threatening properties, respectively), $b = .02$, $SE = .02$, $z = 0.69$, $p = .49$; see Fig. 7B.

Taken together, these findings support the interpretation that domain differences in people's evaluation of generic statements reflect a difference between human and non-human categories, and not either an animate/inanimate or living/non-living distinction. Moreover, given the range of properties tested, it seems that the current findings hold across the sets of threatening and non-threatening properties as a whole.

8. General discussion

The current experiments suggest that people's judgments about generic statements for human categories are systematically different from their judgments about generic statements for non-human categories. For non-human categories, people are more inclined to accept generics involving threatening properties than non-threatening properties even when those properties have precisely the same prevalence levels. However, this difference does not arise for human categories. Instead, for human categories, adults accepted generic statements involving threatening and non-threatening information at comparable rates (Experiments 1, 3, 4, and 6). Domain differences in people's evaluation of generics were not merely due to differences in assumed base rates for threatening versus non-threatening properties across human and non-human categories (Experiments 3 and 4), or were they likely due to social desirability: Even 4-year-olds' endorsement of generic statements showed domain differences; in fact, children were *more* willing to accept non-threatening than threatening information in generic form about human categories (Experiment 5).

Although the current findings consistently show that people evaluate generic statements differently for human versus non-human categories, it is notable that the size of the effect varied across our experiments. The domain × valence interaction was small (Experiments 1 and 6) and non-significant (Experiment 3) for the studies conducted on MTurk, but larger and quite robust for the studies conducted in person (Experiments 4 and 5). One potential explanation for this difference is that Experiments 1, 3, and 6 were conducted online and, as a result, may have reduced concerns about appearing biased. However, this explanation is inconsistent with the finding that even preschoolers show the effect, as they are unlikely to be concerned about appearing biased. Another potential explanation for this difference is that Experiments 1, 3, and 6 provided neither pictures nor descriptions of the novel entities in question (as in Experiments 4 and 5), so all that was known was their membership in a superordinate category (animals, people, things, or tools). Without further information, participants may have felt hard pressed to make firm judgments of the novel categories. (This is in contrast to previous work, which provided participants with descriptions of the novel category members; Cimpian et al., 2010.) In contrast, participants in Experiments 4 and 5 were provided with pictures, which may have facilitated more stable category representations.

8.1. Explaining the effect

We turn next to possible explanations for the differences observed between human and non-human categories. One possibility stems from a dual-process framework suggesting that intuition and reflection interact to produce decisions (Frederick, 2005; Kahneman, 2011; Sloman, 1996). Stereotypes are automatically activated but can be overridden with sufficient motivation (Devine, 1989). Perhaps, in the context of our task, participants' immediate intuitions about human categories showed the same negativity bias found for non-human categories, but they were then overridden using a more controlled, analytic form of cognition. On this account, participants truly disagreed with generics involving threatening information about human categories (rather than just pretending that they disagreed to appear unbiased), but they may have only reached this conclusion after overriding their initial impulse to regard those generics as correct.

However, the current results provide at least some evidence against this hypothesis. Across a variety of phenomena, researchers have found that when adults are drawn toward one response by intuition and to another response by careful reasoning, children tend to be drawn more toward the response that is characteristic of intuition in adults (e.g., Cimpian & Steinberg, 2014; Eidson & Coley, 2014; Epley, Morewedge, & Keysar, 2004; Kelemen & Rosset, 2009). Strikingly, these experiments do not find that children differ from adults by being more inclined to endorse generic statements involving threatening properties about human categories. This developmental result provides at least some evidence against the hypothesis that the effect observed in adults arises from a process whereby participants used controlled reasoning to overcome initial intuitions. Still, it would be fruitful for future research to further investigate this dual-process explanation (e.g., by looking at responses under cognitive load or at speeded reactions).

Another possibility is that, even at the level of immediate intuition, people do not endorse generic statements in the same way for human and non-human categories. In other words, it might be that people's intuitive way of making sense of human categories is different in some important respect from their way of making sense of non-human categories. Then, as a result, it might be that people's intuitions truly do not show the same negativity bias for human categories as they show for non-human categories. For example, existing research indicates that people show a tendency to think that, deep down, human beings are drawn to behave in ways that are morally good (Newman et al., 2014). Of course, people recognize that human beings often behave in ways that are morally bad, but even in such cases, they show a tendency to posit a deeper "true self" that is morally good (Newman, De Freitas, & Knobe, 2015). Perhaps it is this belief about humans' fundamental goodness that explains the difference we observe between human and non-human categories, especially given that generic statements are assumed to convey deep, essential properties (Carlson & Pelletier, 1995; Cimpian & Cadena, 2010; Cimpian & Markman, 2009, 2011; Gelman, 2004; Lyons, 1977). Importantly, it seems that children may show this belief to an even greater extent than adults do. For example, children say that another's goodness is more stable than their badness (Heyman & Dweck, 1998) and that a person is good, despite all evidence suggesting otherwise (Rholes &

Ruble, 1986). If this belief is indeed more robust in childhood than adulthood, that might explain the findings in Experiment 5, where children were more likely to accept generics involving non-threatening rather than threatening properties about human categories.

8.2. Generics and stereotyping

Finally, an important question to consider is how to reconcile the current results with the pervasiveness of prejudice and negative stereotyping in everyday life. Stereotypes can be thought of as generic judgments about human categories (Gelman, Taylor, & Nguyen, 2004), so the current findings may seem at odds with this negative aspect of social cognition.

To begin with, it is important to emphasize that the present results do not in any way call into question existing findings about prejudice and negative stereotypes. Rather, what these results suggest is that there is something about the cognitive processes underlying generic generalizations in particular such that negative stereotypes do not affect these processes in the same way they affect other aspects of cognition. For example, it seems plausible that many people hold a negative stereotype of Italians as mobsters, and that they would show many of the effects that social psychologists have identified as indexing stereotyping and prejudice. However, we suspect that few people would endorse the generic statement, "Italians are mobsters." If this gap between stereotypes and generic endorsement does turn out to be the case, it would not give us reason to reject the hypothesis that people have negative stereotypes about Italians, but rather would provide evidence that these negative stereotypes do not affect generic generalizations in the same way they affect other aspects of cognition.

Why should generics differ from other aspects of cognition? One possibility may follow from the observation that generics are specifically understood to express deep, essential properties (Carlson & Pelletier, 1995; Cimpian & Cadena, 2010; Cimpian & Markman, 2009, 2011; Gelman, 2004; Lyons, 1977). Recent research has found that people have a tendency to think that humans are essentially good (i.e., that there is some deeper essence within humans drawing them to do the right thing; Newman et al., 2014, 2015). Strikingly, this tendency arises even when reasoning about members of outgroups who are negatively stereotyped. Even when people hold clearly negative views about members of such outgroups, they still show a tendency to think that, deep down, there is something more essential in these outgroup members that is calling them toward the good (De Freitas & Cikara, 2016). If this idea of a "good essence" is an aspect of how people think about outgroups, and if generic generalizations have a privileged connection with this essentialist idea, then perhaps it is not surprising that generics about social groups are less negative than other types of generic judgments.

Further research could ask whether there are any conditions under which this effect does not arise. Perhaps the typical negativity of social judgments might emerge even in the context of the current task if participants received additional information about the novel social categories in question. For example, providing explicit information about the outgroup status of these categories or the possibility that they would compete for resources or status with participants' ingroup (e.g., Rhodes & Brickman, 2011) might be sufficient to elicit the same level of prejudice seen in many social psychological studies, as well as everyday contexts.

9. Conclusion

Further research will be necessary to explore the cognitive processes underlying these effects, but regardless of the outcome, the present experiments indicate that people's judgments about generic statements differ depending on whether the target category is human or non-human. Generic judgments about human categories do not exhibit the same negativity bias that generic judgments about non-human categories do.

Acknowledgments

We thank the children, families, and staffs of the following schools: Carrot Patch Early Learning Center, Chase Collegiate School, Cheshire Nursery School, The Children's Center of New Milford, Children's Village of Wolcott, KIDCO Child Care Center of Newington, Kiddie Korner Nursery School, and Wallingford Community Day Care Center. The first author would also like to thank Michael Mezzo for correcting his tendency to split infinitives.

Note

1. At the 100% prevalence level, participants (unsurprisingly) showed near-universal endorsement of the corresponding generic statements, thereby potentially masking differences by domain and valence. A multiple regression on participants' truth ratings that excluded the 100% prevalence level yielded the predicted domain $\times$ valence interaction, $\beta = .05$, $p = .046$. Again, generics involving things were judged to be true more often when they described threatening ($M = 4.49$) than non-threatening ($M = 4.09$) properties, $\beta = .13$, $p < .001$. By contrast, for generics involving people, there was no significant difference between threatening ($M = 4.26$) and non-threatening ($M = 4.14$) properties, $\beta = .04$, $p = .25$.

References

Abrams, D., Rutland, A., Cameron, L., & Ferrell, J. (2007). Older but wilier: In-group accountability and the development of subjective group dynamics. *Developmental Psychology*, *43*, 134–148. doi:10.1037/0012-1649.43.1.134

Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: When older children underperform in social categorization. *Developmental Psychology*, *44*, 1513–1518. doi:10.1037/a0012835

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370. doi:10.1037/1089-2680.5.4.323

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives*, *4*, 212–218. doi:10.1111/j.1750-8606.2010.00149.x

Brandone, A. C., Gelman, S. A., & Hedglen, J. (2015). Children's developing intuitions about the truth conditions and implications of novel generics versus quantified statements. *Cognitive Science*, *39*, 711–738. doi:10.1111/cogs.12176

Carlson, G. N. (1977). *Reference to kinds in English*. Doctoral dissertation, University of Massachusetts, Amherst, MA

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

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, *34*, 1452–1482. doi:10.1111/j.1551-6709.2010.01126.x

Cimpian, A., & Cadena, C. (2010). Why are *dunkels* sticky? Preschoolers infer functionality and intentional creation for artifact properties learned from generic language. *Cognition*, *117*, 62–68. doi:10.1016/j.cognition.2010.06.011

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*, 14–25. doi:10.1016/j.cognition.2009.07.004

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development*, *82*, 471–492. doi:10.1111/j.1467-8624.2010.01525.x

Cimpian, A., & Steinberg, O. D. (2014). The inherence heuristic across development: Systematic differences between children's and adults' explanations for everyday facts. *Cognitive Psychology*, *75*, 130–154. doi:10.1016/j.cogpsych.2014.09.001

Cohen, A. (1999). *Think generic!: The meaning and use of generic sentences*. Stanford, CA: Center for the Study of Language and Information

De Freitas, J., & Cikara, M. (2016). *Deep down my enemy is good: Thinking about the true self reduces intergroup bias*. Unpublished manuscript, Harvard University

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. doi:10.1037/0022-3514.56.1.5

Eidson, R. C., & Coley, J. D. (2014). Not so fast: Reassessing gender essentialism in young adults. *Journal of Cognition and Development*, *15*(2), 382–392. doi:10.1080/15248372.2013.763810

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentricism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760–768. doi:10.1016/j.jesp.2004.02.002

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42. doi:10.1257/089533005775196732

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

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*, 404–409. doi:10.1016/j.tics.2004.07.001

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother–child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*, 1–127

Heyman, G. D., & Dweck, C. S. (1998). Children's thinking about traits: Implications for judgments of the self and others. *Child Development*, *69*, 391–403. doi:10.1111/j.1467- 8624.1998.tb06197.x

Hirschfeld, L. A., & Gelman, S. A. (Eds.) (1994). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge, UK: Cambridge University Press

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*, 138–143. doi:10.1016.j.cognition.2009.01.001

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

Leslie, S. J. (in press). The original sin of cognition: Fear, prejudice, and generalization. *Journal of Philosophy*.

Lyons, J. (1977). *Semantics: I*. Cambridge, UK: Cambridge University Press

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, *40*, 203–216. doi:10.1177/0146167213508791

Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, *39*, 96–125. doi:10.1111/cogs.12134

Rhodes, M., & Brickman, D. (2011). The influence of competition on children's social categories. *Journal of Cognition and Development*, *12*, 194–221. doi:10.1080/15248372.2010.535230

Rholes, W. S., & Ruble, D. N. (1986). Children's impressions of other persons: The effect of temporal separation of behavioral information. *Child Development*, *57*, 872–878. doi:10.2307/1130364

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320. doi:10.1207/S15327957PSPR0504_2

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. doi:10.1037/0033-2909.119.1.3

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403. doi:10.1037/0033-2909.134.3.383