Memory Errors Reveal a Bias to Spontaneously Generalize to Categories

Shelbie L. Sutherland

Andrei Cimpian

University of Illinois at Urbana–Champaign

Sarah-Jane Leslie

Princeton University

Susan A. Gelman

University of Michigan, Ann Arbor

Author Note

This research was supported in part by an NSERC predoctoral fellowship to Sutherland, Spencer Foundation grant 201100111 to Cimpian, NSF grant BCS-1226942 to Leslie, and NICHD grant HD-36043 to Gelman. We thank Larisa Hussak, Sneha Gaikwad, Alana Glickman, and JoAnn Park for their help in coding the data and the members of the University of Illinois Cognitive Development Lab for helpful discussion.

Correspondence concerning this article should be addressed to Shelbie L. Sutherland, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: ssuther2@illinois.edu

Abstract

Much evidence suggests that, from a young age, humans are able to generalize information learned about a subset of a category to the category itself. Here, we propose that—beyond simply being able to perform such generalizations—people are *biased* to generalize to categories, such that they routinely make spontaneous, implicit category generalizations from information that licenses such generalizations. To demonstrate the existence of this bias, we asked participants to perform a task in which category generalizations would distract from the main goal of the task, leading to a characteristic pattern of errors. Specifically, participants were asked to memorize two types of novel facts: quantified facts about sets of kind members (e.g., facts about *all* or *many stups*) and generic facts about entire kinds (e.g., facts about *zorbs* as a kind). Moreover, half of the facts concerned properties that are typically generalizable to an animal kind (e.g., eating fruits and vegetables), and half concerned properties that are typically more idiosyncratic (e.g., getting mud in their hair). We predicted that—because of the hypothesized bias—participants would spontaneously generalize the quantified facts to the corresponding kinds, and would do so more frequently for the facts about generalizable (rather than idiosyncratic) properties. In turn, these generalizations would lead to a higher rate of quantified-to-generic *memory errors* for the generalizable properties. The results of four experiments ($N = 449$) supported this prediction. Moreover, the same generalizable-versus-idiosyncratic difference in memory errors occurred even under cognitive load, which suggests that the hypothesized bias operates unnoticed in the background, requiring few cognitive resources. In sum, this evidence suggests the presence of a powerful bias to draw generalizations about kinds.

Keywords: concepts; memory; generic knowledge; quantifiers

Memory Errors Reveal a Bias to Spontaneously Generalize to Categories

Humans conceive of the world as being populated not just by unique individuals (e.g., the tall leafy thing in the front yard) but also by *kinds of things* (e.g., trees). What's more, we routinely acquire and store knowledge at the level of these abstract kinds, and we use this knowledge with amazing flexibility to communicate with one another, explain the world around us, and predict future outcomes (e.g., Gelman, 2003; Markman, 1989; Murphy, 2002; Smith & Medin, 1981). These achievements are all the more remarkable considering that we do not in fact have perceptual access to kinds per se—only to particular samples. To some researchers, the accumulated evidence in the psychology of concepts has suggested that, beyond being merely *capable* of reasoning about kinds, human cognition may actually be structured so as to *privilege* the processing of information at this general level (e.g., Cimpian & Erickson, 2012; Gelman, 2010; Hampton, 2012; Hollander, Gelman, & Star, 2002; Leslie, 2008, 2012).

According to these arguments, reasoning about kinds requires few cognitive resources compared to reasoning about sets of comparable scope. Consider some of the developmental evidence on this point. Children's ability to evaluate claims about entire kinds (e.g., "Do *girls* have curly hair?") is adult-like starting at around the age of 3, whereas their ability to evaluate claims about similarly broad quantified sets (e.g., "Do *all girls* have curly hair?") has a much more protracted developmental course (Hollander et al., 2002; Mannheim, Gelman, Escalante, Huayhua, & Puma, 2011; Tardif, Gelman, Fu, & Zhu, 2012). This developmental pattern—which has been found in children learning languages from three different language families (English, Mandarin, and Quechua)—is particularly striking when taking into account the fact that, from the perspective of formal semantics, statements about kinds are more complex than quantified statements. To illustrate the formal complexity of statements about kinds (or generic statements), consider that one can truthfully say that *mosquitoes carry malaria* but not that *books are paperbacks*, even though the majority of books are paperbacks, and only a tiny percentage of mosquitoes carry malaria (e.g., Prasada, Khemlani, Leslie, & Glucksberg, 2013). Because of puzzling examples such as these, a formal account of the truth conditions of generic statements has eluded semanticists for over 40 years (e.g., Carlson & Pelletier, 1995; Lawler, 1973; but see Leslie, 2008). In contrast, specifying the truth conditions for quantified statements is often formally simple (e.g., the truth of a universally quantified statement is determined by a clear rule: the statement is true if and only if every single category member has the described property). Thus, the cognitive ease with which children understand generics (which are formally complex), coupled with the cognitive difficulties children encounter with quantified claims (which are formally simple), is suggestive of a bias in the architecture of our cognitive systems—a bias that enables reasoning about kinds to be so effortless that even young children can perform such formally complex reasoning competently.

These ease-of-processing claims (i.e., that reasoning about kinds requires few cognitive resources) are not restricted to children. For instance, when adults have to evaluate or remember quantified facts, they often respond as if these facts were about kinds (Leslie & Gelman, 2012; Leslie, Khemlani, & Glucksberg, 2011; Meyer, Gelman, & Stilwell, 2011; see also Hampton, 2012; Jönsson & Hampton, 2006). To illustrate, Leslie and Gelman (2012) asked children and adults to remember both generic facts (e.g., “Bees have five eyes”) and quantified facts (e.g., “All bees have five eyes”) for a later memory test. At both ages, participants were more likely to mistakenly recall quantified statements as generics than to mistakenly recall generics as quantified statements. This result, which was bolstered by follow-up studies ruling out alternative explanations, seems consistent with the ease-of-processing argument above: If quantified information is more cognitively challenging to process and store than the corresponding generic information, then participants may inadvertently default to the latter and thus recall quantified information as generic.

In the present paper, we investigate another potential cognitive bias that may privilege reasoning and learning about kinds: namely, a bias to make generalizations about kinds. That is, we propose that, whenever people encounter evidence that could reasonably be extended to a kind, they will routinely formulate implicit generalizations that take this evidence and apply it to the kind as a whole. Moreover, these generalizations are hypothesized to be spontaneous, occurring without any sort of external encouragement or prompt. Similar to the ease-of-processing bias described above, this *generalization bias* gives rise to many kind representations that we would not have formed otherwise. However, the process by which it does so is quite different: The kind representations created through the hypothesized generalization bias are not the byproducts of an inability to process quantified information—they are not the side effects of our cognitive limitations. Rather, they are the outcome of inferences (inductive generalizations, to be more precise) that our cognitive system spontaneously performs "behind the scenes" when encountering information about sets of objects in the world.

Our proposal of a generalization bias builds on the extensive research suggesting that people are able to draw kind-wide conclusions from relatively sparse evidence. For example, adults often judge that a property that is present in a minority of the members of a kind (e.g., 30% of morseths have silver fur) is likely to be true of the kind as a whole (e.g., morseths, as a kind, have silver fur; Cimpian, Brandone, & Gelman, 2010). Similarly, the developmental literature on inductive inferences has suggested that even very young children can generalize information from one member of a kind to another arbitrary member—and thus, arguably, to the entire kind (e.g., Keates & Graham, 2008; Gelman & Markman, 1986, 1987; Graham, Kilbreath, & Welder, 2004; Sutherland & Friedman, 2012, 2013; see also Cimpian & Park, 2014; Csibra & Gergely, 2009).

These findings, however, do not provide evidence for the stronger claim of a *bias* to generalize to kinds. These previous results suggest that people will often draw conclusions about kinds when they are provided with explicit opportunities to do so. For instance, Cimpian and colleagues' (2010) data show that people generalize certain quantified facts to the level of kinds when they are asked whether these generalizations are warranted. It is unclear, however, if people would have drawn the generic conclusions they did *without* the experimenter's prompt. The same point applies to the developmental evidence: The experiments that explored children's generalizations typically provided clear, explicit opportunities for children to make such generalizations. For example, children might be given some information about one member of a kind (e.g., this bird feeds its babies mashed-up food) and then asked if another member of that kind also possesses that feature (e.g., does this other bird feed its babies mashed-up food?; e.g., Gelman & Markman, 1986, 1987). Stronger evidence is needed to demonstrate the existence of a *bias* to generalize to categories. That is, we would need evidence that people make category generalizations spontaneously, in the absence of any external prompts or incentives—or, perhaps, even in the presence of *dis*incentives.

In the studies reported here, we tested whether people spontaneously generalize quantified evidence to the level of a kind in a context where such generalizations actually distract participants from the goal of the task at hand (because generalizing leads to incorrect answers). The task, modeled after Leslie and Gelman (2012), is ostensibly about people's memory for generic and quantified facts about novel animals. In reality, however, our reason for adopting it was that it can reveal whether people use the evidence provided by the quantified facts (e.g., *all zorbs* eat fruits and vegetables) to draw spontaneous generalizations about the relevant kinds (e.g., *zorbs* eat fruits and vegetables).[1] If participants did so, their gist memory for the quantified facts would arguably be influenced by these generalizations; as a result, they may misremember some of the facts that were originally quantified as being generic on a memory test. Such generalization-induced memory errors, if they occurred, would be both spontaneous (rather than externally prompted) and counter to the incentive structure of the task (where accurate memory was the only criterion for success). Thus, from the novel perspective outlined here, the evidence from this task could speak to the presence of a bias to generalize to categories.

The use of this memory paradigm, however, gives rise to a problem: The prediction of our generalization bias is, at this point, indistinguishable from that of the ease-of-processing bias. That is, both accounts predict frequent conversions from quantified to generic form—albeit for different reasons. While our account suggests that these conversions are a result of spontaneous generalizations, the ease-of-processing account suggests that the conversions would be a result of the resource-intensive nature of processing quantified information, which should lead participants to inadvertently fall back on the easier-to-process kind representations. To circumvent this ambiguity, we manipulated the content of the facts participants were asked to remember. That is, half of the facts described properties that are typically *generalizable* to an entire animal kind (e.g., diet, habitat), whereas the other half described properties that are typically *idiosyncratic* to a particular individual (e.g., temporary states, accidents). There is

[1] We use the term *generalization to a category* as roughly synonymous with *inference about a category*. Thus, the hypothesized bias to spontaneously generalize to categories may also be understood as a bias to draw spontaneous inferences about the features of categories. While there may be differences of opinion as to whether using a universally quantified statement to derive a generic conclusion counts as a generalization per se, this is certainly a non-trivial inductive inference, as quantified statements do not express facts about kinds (e.g., Carlson & Pelletier, 1995; Cimpian et al., 2010; Leslie, 2008).

extensive evidence that kind generalizations are sensitive to the content of the property being generalized (e.g., Cimpian et al., 2010; Cimpian & Markman, 2008; Gelman, 1988; Gelman & Markman, 1986). This sensitivity to content is, in part, rooted in what are known as *overhypotheses* (e.g., Dewar & Xu, 2010; Goodman, 1955; Shipley, 1993). Overhypotheses are abstract beliefs that specify the types of properties that are likely to be uniform across members of a certain type of category (e.g., categories of animals have uniform diets: horses eat hay, birds eat seeds, etc.). We chose our *generalizable* properties so as to fit under common overhypotheses people might have about animal kinds (e.g., diet: "All zorbs eat fruits and vegetables"), which might thus facilitate kind generalizations from quantified evidence concerning these properties. In turn, because of these generalizations, people should be significantly more likely to misremember quantified facts as generic when the facts are about generalizable properties than when they are about idiosyncratic properties.

Importantly, this prediction is distinct from the prediction of the ease-of-processing claim, which provides no reason to expect an asymmetry in memory errors for generic and quantified facts based on the content of the properties they describe. The ease-of-processing account would instead predict that a quantified fact, regardless of what type of property it is about, is still quantified and thus would be more cognitively taxing to remember, causing people to fall back on an easier-to-process kind representation. (Therefore, in the context of the present task, the ease-of-processing account predicts that we should find more conversions from quantified facts to generic facts for *both* property types. If this prediction is supported, the present studies would also serve as a replication of Leslie and Gelman [2012] and thus provide additional support for the ease-of-processing account.)

In addition to providing evidence for a generalization bias, the present research also sought to investigate some of its characteristics. In particular, we investigated the extent to which this bias can operate in the background, without diverting cognitive resources away from the focus of one's attention. We explored this issue empirically by placing half of the participants under a cognitive load while they were encoding the quantified and generic statements. If the bias to generalize to kinds operates without requiring much cognitive effort, then the participants who are under cognitive load should also misremember the quantified facts as generic more often when these facts are about generalizable (vs. idiosyncratic) properties. Such a result would speak to the low-demand nature of this bias, as well as to the power it has to shape our conceptual knowledge without interfering with ongoing cognitive activities.

To summarize, we proposed that people have a bias to make generalizations about categories. If such a bias were in place, then one symptom of it should be a tendency to make spontaneous kind generalizations even in contexts where such generalizations are unwelcome. In the current memory paradigm, such spontaneous generalizations should lead people to mistakenly recall quantified facts as generic, and these mistakes should be more frequent when the facts are about generalizable properties (which facilitate the unwanted kind generalizations) than when the facts are about idiosyncratic properties. By manipulating whether participants had to perform a concurrent task while encoding the generic and quantified facts, we were also able to test whether this generalization bias requires only minimal cognitive resources to operate. Experiments 1 to 4 provided consistent support for our proposal of a generalization bias. In addition, Experiments 2 to 4 addressed two alternative explanations for the findings.

Experiment 1

Method

Participants. The participants were 187 undergraduate students from a large public university in the Midwestern US. All were native English speakers. The reward for participation was course credit or $5. Participants were randomly assigned to either a No Load (n = 93) or a Cognitive Load (n = 94) condition.

Items. We used 16 facts about novel animals (see Table 1), each of which could be presented either as universally quantified (e.g., “All zorbs eat fruits and vegetables”) or as generic (e.g., “Zorbs eat fruits and vegetables”). However, the same fact was never presented in both forms to the same participant. In addition, half of the facts described generalizable properties (e.g., eating fruits and vegetables), and half described idiosyncratic properties (e.g., getting mud in their hair). The generalizable and idiosyncratic properties were matched in average length (both *M*s = 4.75 words). Moreover, in a separate norming study, we asked participants (N = 43) to judge how many members of a kind were likely to possess these properties (e.g., “If you had to guess, what percentage of stups get mud in their hair?”), given that at least one member of the kind had the property. The results confirmed that all of the generalizable properties were indeed judged to be more generalizable (range = 75.1% to 85.7% of category members have the property) than all of the idiosyncratic properties (range = 14.0% to 55.4%); this difference was significant, $M_{generalizable}$ = 80.9% versus $M_{idiosyncratic}$ = 42.2%, $t(42)$ = 9.56, $p < .001$.

The 16 facts were presented in one of three random orders, each of which had two versions. The three random orders were generated with the constraint that no more than three facts of the same form (i.e., “all” or generic) or containing the same type of property (i.e., generalizable or idiosyncratic) should occur in a row. The two versions of each order were identical except with respect to the generic/universal form of each fact: If a fact was generic in one version, it was universally quantified in the other version, and vice versa.

Procedure and Design. Testing occurred in small groups of up to six participants. To avoid overwhelming participants' memory capacity, the 16 facts were split into two blocks of eight, each of which contained four facts in generic form and four in "all" form, as well as four generalizable and four idiosyncratic properties. The same three-phase procedure, described next, was followed for both blocks.

1) *Learning phase*. In the No Load condition, the experimenter asked participants to pay close attention to the sentences because they would be asked to recall them in a later test. Then, she read aloud the eight facts from the first block. As the participants listened to the facts, they followed along in a booklet that contained only line drawings of the novel animals referenced in these facts (and not the facts themselves). The procedure for the Cognitive Load condition was identical, except that participants were also asked to rehearse a string of six digits while listening to the facts and following along in their booklets. Immediately after listening to the facts, participants were asked to recall the digits in the order in which they were presented (for similar methods of inducing cognitive demands, see Baddeley & Hitch, 1974; Gilbert & Hixon, 1991).

2) *Distractor phase*. All participants were then asked to complete a four-minute distractor task in which they completed a series of multi-digit multiplication problems.

3) *Recall phase*. Next, participants received a second booklet with the same drawings as the booklet from the learning phase. Participants were asked to go through the booklet and write what they remembered of the sentences that the experimenter had read for each page. They were asked to write in full sentences. Because our main interest was in participants' memory for the scope of the facts (generic vs. "all") rather than in their memory for the content of these facts, we provided two strong clues to the content of each sentence: the bare singular form of the relevant novel noun (e.g., zorb) and an additional noun from the fact (e.g., vegetable; see Table 1 for full list of clues).

Once participants finished writing down their recall responses for the first block, the three phases (learning, distractor, and recall) were repeated for the second block of eight sentences.[2]

The design of our study can be summarized as follows: 2 (fact form: generic vs. "all"; within subject) × 2 (property type: generalizable vs. idiosyncratic; within subject) × 2 (cognitive load: load vs. no load; between subjects).

Coding. One researcher coded participants' recalled sentences into one of three mutually exclusive categories depending on their scope: generic, "all," and "other" (which also included failures to recall anything). If a sentence was about a kind as a whole (e.g., "Zorbs like to eat vegetables"), it was coded as generic.[3] If a sentence was about all members of a kind (e.g., "All zorbs eat fruits and vegetables"), it was coded as "all." If a sentence was about a single instance of a kind or had indeterminate scope (e.g., "Eat fruits and vegetables"), it was coded as "other." These three categories accounted for 49.7%, 31.0%, and 19.4% of participants' responses, respectively.[4] A second researcher, blind to the load condition and the original form of the fact, coded the responses of 167 of the 187 subjects (20 subjects were used for training). Cohen's

---

[2] Approximately half of the participants received a slight variant of this procedure, in which (1) the booklets used during the learning phase listed the same two clues as those in the recall phase (rather than no clues), and (2) the Cognitive Load condition involved rehearsing a string of eight digits (rather than six). Because this procedural variant did not interact significantly with either of the variables of interest (property type and cognitive load), we report the data collapsed across it.

[3] The vast majority of sentences coded as generic had either bare plural noun phrases (e.g., "Reesles like to swim in the ocean") or indefinite singular noun phrases (e.g., "A dax stores its food in its cheeks") in the subject position. However, we also coded as generic a number of sentences with (what appeared to be) bare singular noun phrases in the subject position (e.g., "Glippet keep their nests on mountain peaks"). This coding decision was based on the assumption that some participants may have been unsure of how to pluralize the novel nouns provided (e.g., some may have thought that the plural of *glippet* may be *glippet*, on analogy with *sheep* or *deer*). This latter type of generic accounted for only 5.3% of responses coded as generic. Moreover, when the data were analyzed without these generics, the results replicated those reported in the main text. Also note that we did not code definite singular noun phrases as generic. Even though such noun phrases can in principle refer to a kind, their generic use is rare (e.g., Gelman, Coley, Rosengren, Hartman, & Pappas, 1998). Moreover, since one of the clues provided to participants was a picture of a *single* exemplar from the relevant category, the task context made it very likely that subjects' definite singular nouns (e.g., "the oller") were referring to the exemplars on the page in front of them.

[4] The 19.4% "other" responses consisted of 11.8% responses about single instances ($M_{generalizable}$ = 10.6%; $M_{idiosyncratic}$ = 13.0%) and 7.6% responses that had indeterminate scope (the latter percentage includes skipped responses).

kappas for the generic, "all," and "other" coding categories were .97, 1.0, and .95, respectively, and disagreements were resolved through discussion.

We also coded Cognitive Load participants' memory for the digits they were asked to rehearse while they were listening to the facts. Two researchers independently rated each participant's recalled digit strings on a scale from 1 [completely wrong or missing] to 5 [completely correct].[5] Inter-rater agreement was high, $r = .93$. Each subject's final rating was the average of the two researchers' ratings, except in cases where their scores differed by more than one point. In such cases, the researchers discussed the disagreement and reached a mutually agreeable rating.

Dependent Measure. In light of the prior evidence for the efficiency of kind-based computations (e.g., Hollander et al., 2002; Leslie & Gelman, 2012), it is likely that participants will, on the whole, be more likely to mistakenly recall "all" statements as generic rather than the reverse. Our proposal of a bias to generalize to categories makes two additional predictions: (1) the magnitude of this asymmetry in memory errors (i.e., more "all"-to-generic than generic-to-"all" conversion errors) should be greater when the facts concern generalizable properties than when they concern idiosyncratic properties, and (2) this property effect should be observed even when participants have few cognitive resources available.

To test these predictions, we calculated the difference score between "all"-to-generic and generic-to-"all" memory conversions, separately for the generalizable and the idiosyncratic properties. This calculation proceeded in two steps. First, we calculated the percentage of statements that were originally presented in "all" form that were instead recalled in generic form, and the percentage of statements originally presented in generic form that were instead recalled

---

[5] The intermediary scale points were labeled as follows: 2 [very few correct numbers in the correct order], 3 [some of the correct numbers in the correct order], and 4 [minor errors (e.g., one number missing or out of order)].

in "all" form. These conversion scores were calculated for each individual participant, separately for the generalizable and idiosyncratic properties. Second, we took each participant's percentages of "all"-to-generic conversions and subtracted from them the participant's percentages of generic-to-"all" conversions, again separately for the generalizable and idiosyncratic properties. Thus, each participant received two final difference scores (one for each property type), which we will refer to as *generalization-bias scores* from here on.

Our predictions can be assessed by testing, first, whether participants' generalization-bias scores are higher for facts that describe generalizable properties than for facts that describe idiosyncratic properties, and second, whether this difference is present both when cognitive resources are intact and when they are taxed.

Results

Data Analysis Strategy. Participants' generalization-bias scores clustered in the upper half of the range and were thus non-normally distributed (Shapiro-Wilk test, $p < .001$). Because of this violation of parametric assumptions, we analyzed the data using ordinal logistic regressions (OLRs) computed using the Generalized Estimating Equations command in SPSS. Cognitive load was a between-subjects factor in this analysis, and property type was a within-subject factor.[6]

Cognitive Load Manipulation Checks. If participants in the Cognitive Load condition complied with our instructions to rehearse the string of digits provided by the experimenter, then they should have reasonably accurate memory for these digits. Very poor digit recall is most likely a sign that the subjects did not rehearse the digits and were not actually under a cognitive load (see Gilbert & Hixon, 1991, for a similar argument). Thus, we excluded from the analyses any subjects whose average digit memory scores were below 2 on the 1–5 scale described in the

[6] Despite the assumption violations, the results of the OLRs were replicated with analyses of variance (ANOVAs).

Method (n = 13; $M_{digit\ memory}$ = 1.37). This left 81 subjects in the Cognitive Load condition. (Note that all of the significant results reported below remain significant even if these subjects are not excluded.)

As an additional check that the 81 remaining Cognitive Load participants were indeed under a load, we tested their accuracy on the primary task (fact recall) relative to the participants in the No Load condition. If the cognitive load imposed a burden on working memory resources, Cognitive Load participants should have less accurate memory compared to participants in the No Load condition. Consistent with this prediction, Cognitive Load participants were significantly less likely than No Load participants to recall the facts in the correct form (*M*s = 38.0% and 52.2% of responses were recalled in the correct form, respectively), Wald $\chi^2(1)$ = 28.89, $p < .001$, d = 0.75 (see also Table 2 and the Appendix). Thus, it seems reasonable to assume that our cognitive load manipulation was successful in inducing different cognitive demands on the two groups of participants.

First Prediction: A Main Effect of Property Type. To reiterate, we proposed that people are biased to spontaneously generalize to kinds. In the context of our task, this bias might prompt spontaneous generalizations to the kind level especially when the evidence warrants such generalizations. Thus, when a novel property is generalizable—the sort of property that is typically true of kinds—participants may be particularly likely to use the quantified evidence at hand (e.g., “All zorbs eat fruits and vegetables”) to implicitly infer something about the kind itself (e.g., zorbs, as a kind, have this sort of diet). These generalizations, should they occur, would lead to higher generalization-bias scores for generalizable properties than for idiosyncratic properties. In line with this prediction, the OLR revealed a significant main effect of property type, such that participants had higher generalization-bias scores for facts about generalizable properties (M = 24.1% more "all"-to-generic than generic-to-"all" conversions) than for facts about idiosyncratic properties (M = 16.0%), Wald $\chi^2(1) = 12.11$, $p = .001$, $d = .13$.

As a reminder, participants' generalization-bias scores are calculated as the difference between their "all"-to-generic and generic-to-"all" conversions. However, it is the "all"-to-generic conversions that are of most interest to us here because they are the direct byproducts of the hypothesized bias to generalize to kinds. Therefore, we also tested whether these key "all"-to-generic conversions were significantly more common for the generalizable than the idiosyncratic properties. Indeed, as predicted, participants were significantly more likely to misrecall "all" facts as generic when the facts described generalizable properties (M = 48.0% of "all" facts) than when they described idiosyncratic properties (M = 41.7%), Wilcoxon $Z = 3.62$, $p < .001$.

We also explored whether the effect of property type held up at the level of individual participants. Specifically, we compared the number of participants who had higher generalization-bias scores for the generalizable than for the idiosyncratic properties with the number of participants who had the opposite pattern (higher generalization-bias scores for idiosyncratic properties). Consistent with our prediction, there were significantly more participants with higher generalization-bias scores for the generalizable properties (37.9% of participants) than participants with higher generalization-bias scores for the idiosyncratic properties (17.8% of participants), $p < .001$ by a sign test.

Finally, participants' generalization-bias scores were significantly greater than zero, both for the generalizable properties (one-sample Wilcoxon test, $Z = 4.92$, $p < .001$) and for the idiosyncratic properties (one-sample Wilcoxon test, $Z = 3.13$, $p = .002$), indicating that there were significantly more "all"-to-generic than generic-to-"all" conversions for both of these types of facts. These differences are consistent with prior arguments that suggest generic facts impose a lower processing burden relative to quantified facts (e.g., Leslie & Gelman, 2012).

Second Prediction: An Effect of Property Type in Both the Cognitive Load and the No Load Conditions. We also hypothesized that the bias to generalize to kinds operates without much cognitive effort. Thus, our second prediction was that this bias should influence participants' memory, regardless of whether or not they are asked to perform another task while listening to the experimenter's facts. In other words, we predicted that there would be a statistically significant effect of property type in both the No Load and the Cognitive Load conditions.

Consistent with our prediction, participants' generalization-bias scores were higher for statements that described generalizable properties than for statements that described idiosyncratic properties both in the No Load condition, Wald $\chi^2(1) = 6.63$, $p = .010$, $d = 0.15$, and in the Cognitive Load condition, Wald $\chi^2(1) = 5.41$, $p = .020$, $d = 0.12$. Additionally, there was no significant difference between the magnitude of the property type effect in each load condition, as the OLR revealed no trace of an interaction between property type and cognitive load, Wald $\chi^2(1) = 0.08$, $p = .78$ (see Table 2 and the Appendix for means).[7]

Individual participants' response patterns pointed to the same conclusion: There were significantly more participants who had higher generalization-bias scores for the generalizable (vs. the idiosyncratic) properties than participants who had higher scores for the idiosyncratic (vs. the generalizable) properties in both the No Load and the Cognitive Load conditions, *p*s = .005 and .053, respectively, by sign tests.

Discussion

[7] The main effect of cognitive load was not significant either, $M_{No\ Load} = 18.3\%$ vs. $M_{Cognitive\ Load} = 22.1\%$ more "all"-to-generic than generic-to-"all" conversions, Wald $\chi^2(1) = 0.15$, $p = .70$, $d = 0.06$.

To summarize, we found that participants were more likely to misremember quantified facts as generic (rather than vice-versa) when these facts were about properties that are typically generalizable to a kind (e.g., “All zorbs eat fruits and vegetables”) than when they were about properties that are typically more idiosyncratic (e.g., “All stups get mud in their hair”). This finding is in line with our main proposal that people are biased to make spontaneous kind generalizations and are therefore likely to generalize quantified evidence to kinds even in circumstances where such generalizations interfere with correct performance.

There is, however, an alternative explanation for these findings. Perhaps people don’t generalize the quantified evidence about generalizable properties to the relevant kinds, as we hypothesized. Rather, universally quantified and generic statements may simply be *closer in their meaning*—and thus *more confusable*—when they describe generalizable properties than when they describe idiosyncratic properties. For instance, hearing that “zorbs eat fruits and vegetables” might lead people to expect that the vast majority of zorbs do so (Cimpian et al., 2010), which would then make this statement similar in meaning with a statement such as “all zorbs eat fruits and vegetables.” If generic and “all” statements are seen as meaning roughly the same thing in this particular case, then people might just produce the shorter of the two statements at recall, leading to an increased rate of “all”-to-generic conversions compared to generic-to-“all” conversions. (Again, no kind generalizations of the sort we hypothesize are invoked by this alternative account.) In contrast, generic statements about idiosyncratic properties (e.g., “Stups get mud in their hair”) may *not* be seen as being similar/confusable in meaning with the corresponding universally quantified statements (e.g., “All stups get mud in their hair”), in part because generics about such properties suggest relatively low prevalence levels (Cimpian et al., 2010, Experiment 3). For idiosyncratic properties, then, people may be *less* inclined to use the shorter generic statements as stand-ins for universally quantified statements (because their meanings are not seen as interchangeable), leading to lower, and more symmetrical, numbers of "all"-to-generic and generic-to-"all" errors.

Although intuitively plausible, this alternative cannot explain other aspects of the results obtained in Experiment 1. For example, if generic and "all" statements are more confusable for generalizable properties, then it is hard to explain why the percentage of correct responses for these properties (M = 46.5%) was nearly identical to the percentage of correct responses for the supposedly less-confusable generic and "all" statements about idiosyncratic properties (M = 44.8%; Wilcoxon Z = 1.23, p = .22). The confusability alternative straightforwardly predicts that people should be correct *less often* for the generalizable properties. This result, however, is suggestive but not conclusive, so we conducted Experiment 2 in order to provide a more definitive test of this alternative explanation. Specifically, we measured and statistically adjusted for the perceived similarity/confusability of "all" and generic statements about generalizable and idiosyncratic properties. If the confusability alternative is correct, then taking participants' judgments of meaning similarity into account when testing for an effect of property type should eliminate the difference in generalization-bias scores observed in Experiment 1. In contrast, we predicted that the effect of property type on generalization-bias scores would replicate even when controlling for this measure of similarity/confusability.

Experiment 2

Method

Participants. Eighty-six participants were recruited using Amazon's Mechanical Turk platform and completed the study online. All were native English speakers residing in the US. The reward for participation was $0.75.

Items. We used the same items as in Experiment 1.

Procedure. The procedure was similar to the No Load condition of Experiment 1, with a few methodological changes necessitated by (1) the switch from in-lab to online testing, and (2) the addition of the key control variable of this experiment (namely, similarity/confusability).

The changes were as follows. First, instead of listening to the sentences while looking at a booklet, participants read the sentences on their computer screens, each on a separate page. Each page was programmed to automatically advance after 15 seconds in order to equate encoding time across sentences. To ensure that the participants attended to the stimulus sentences, we also required them to type out these sentences in a text box on the page on which they were displayed. Second, the distractor phase was shortened to two minutes rather than four. A two-minute delay is more in line with the brevity of typical studies on Mechanical Turk, and yet it is still long enough to ensure that participants had to rely on long-term memory at recall (e.g., Peterson & Peterson, 1959). Third, the distractor task consisted of arithmetic problems that could be solved without needing a pen and paper, unlike the multi-digit multiplication problems used in Experiment 1. Fourth, to assess the similarity/confusability of "all" and generic statements describing generalizable and idiosyncratic properties, we presented participants with all 16 pairs of generic and "all" statements (e.g., "Stups get mud in their hair" and "All stups get mud in their hair"; see Table 1) and asked them to rate how similar in meaning these pairs were ("How similar are the meanings of these two sentences?"). Participants marked their answers on a 10-point scale (from 1 = "very dissimilar" to 10 = "very similar"). Each pair was presented on a different screen. The order of the 16 pairs was randomized for each subject. Also, the order in which the generic and "all" sentences were displayed within the pairs was randomized across participants. These rating questions were always presented after the recall phase so as to not interfere with the memory task.

Because online testing makes cheating on the memory task a possibility, at the very end of the session we asked participants to report whether they had written down, copied, or used any other external sources to help them remember the sentences. To encourage truthful responses, we made it very clear to participants that they would receive payment regardless of how they answered this question. Two participants reported cheating and were excluded, leaving 84 participants in our sample.

Coding. The coding scheme was the same as in Experiment 1. Inter-coder agreement was calculated over all 84 transcripts and was again excellent: Cohen's kappas for the generic, "all," and "other" coding categories were .97, .99, and .94, respectively. Disagreements were resolved through discussion.

Dependent Measures. Generalization-bias scores were calculated just as in Experiment 1, by taking the difference between "all"-to-generic and generic-to-"all" memory conversions separately for the generalizable and the idiosyncratic properties. Participants' similarity scores were also calculated separately for the two property types by averaging the similarity ratings for the eight generalizable items and the eight idiosyncratic items.

Results and Discussion

Our claim is that people make more asymmetric memory errors (more "all"-to-generic than generic-to-"all" conversions) for generalizable than for idiosyncratic properties because of an implicit bias to generalize to kinds, and not because of low-level factors having to do with the confusability of generic and universally quantified statements describing these types of properties. Thus, we predicted that generalization-bias scores would be significantly higher for the generalizable than for the idiosyncratic properties *even when adjusting for any differences* *between the similarity/confusability of "all" and generic statements about the two property types*.

To begin, we examined whether such similarity/confusability differences exist in the first place. Participants did in fact rate universally quantified and generic statements as being more similar when they described generalizable properties (M = 8.34 on a 1–10 scale, SD = 1.37) than when they described idiosyncratic properties (M = 8.05, SD = 1.56), Wilcoxon Z = 3.63, p < .001. Importantly, however, this difference did not account for the difference found between generalization-bias scores for generalizable and idiosyncratic facts. We submitted participants' generalization-bias scores to an OLR with property type (generalizable vs. idiosyncratic) as a predictor and similarity/confusability scores as a covariate. As predicted, the main effect of property type was replicated even when controlling for the similarity variable: Participants had significantly higher generalization-bias scores for the generalizable properties (M = 34.5% more "all"-to-generic than generic-to-"all" conversions) than for the idiosyncratic properties (M = 25.3%), Wald $\chi^2(1) = 7.37$, p = .007, d = .15. (We report unadjusted means here and in Table 2 and the Appendix.) Moreover, the similarity/confusability covariate was not a significant predictor of generalization-bias scores, Wald $\chi^2(1) = 1.11$, p = .29.[8,9]

When we considered only participants' "all"-to-generic conversions, we found that these conversions (which are most pertinent to our argument of a bias to generalize to kinds) were more likely for facts that described generalizable properties (M = 56.8% of "all" facts) than for facts that described idiosyncratic properties (M = 47.6%), Wilcoxon Z = 3.31, p = .001.

[8] The same results were found with a repeated-measures ANOVA with similarity/confusability as a covariate.
[9] As an additional means of investigating whether the difference in similarity/confusability was related to the difference in generalization-bias scores between generalizable and idiosyncratic properties, we calculated the correlation between (1) a generalization-bias difference score (generalization-bias score for generalizable properties minus generalization-bias score for idiosyncratic properties) and (2) a similarity/confusability difference score (similarity score for generalizable properties minus similarity score for idiosyncratic properties). This correlation was non-significant, $r(82) = -.08$, p = .45, which speaks against the alternative hypothesis tested here and is consistent with the non-significant covariate effect reported in the main text.

Finally, as in Experiment 1, the main effect of property type was replicated at the level of individual participants. There were significantly more participants with generalization-bias scores in the predicted direction (generalizable > idiosyncratic; 33.3%) than participants with generalization-bias scores in the unpredicted direction (idiosyncratic > generalizable; 14.3%), $p = .017$ by a sign test.

These findings speak against the possibility that participants' responses in our task are due to a similarity/confusability confound. Instead, it seems more likely that participants are spontaneously generalizing the provided quantified information to the relevant kinds whenever appropriate, revealing an implicit bias to generalize to kinds.

Experiment 3

In Experiment 3, we addressed an additional alternative explanation for our findings. According to this alternative, participants' responses in Experiments 1 and 2 could have been driven entirely by the clues given during the recall phase. To elaborate, the clue words provided for the generalizable properties in those experiments were slightly different in content from the clue words provided for the idiosyncratic properties (see Table 1): More of the clues for generalizable properties referred to typical animal habitats (e.g., ocean, mountain) and diets (e.g., vegetables). Given that elements such as habitats and diets often figure in kind-wide properties, participants who saw these clues during the recall phase may have been artificially induced to generate statements about kinds *even if they hadn't drawn any kind inferences from the original statements* (or, for that matter, even if they didn't remember anything about the original statements). It is possible, then, that this asymmetry in the content of the memory clues used in the first two studies might explain the greater number of "all"-to-generic (vs. generic-to-"all") errors for generalizable properties than for idiosyncratic properties. Experiment 3 was conducted to investigate this alternative explanation. Specifically, we presented a new group of participants with the clues (both pictures and words) given during the recall phase of Experiments 1 and 2 and asked them to generate either generic or universally quantified statements using these clues. If the clues alone were driving our effect, we should find more generic responses for the clues provided for generalizable properties in the first two studies compared to clues provided for the idiosyncratic properties. In contrast, we predict that there will be no difference in the number of generic responses between the two sets of clues.

Method

Participants. Eighty-seven participants were recruited using Amazon's Mechanical Turk platform and completed the study online. All were native English speakers residing in the US. The reward for participation was $0.75.

Procedure. We asked participants to *create* sentences using exactly the same clues that were shown in the recall phase of Experiments 1 and 2. We also provided a few rules to guide participants' sentence creation; these rules were meant to ensure that their expectations about the types of sentences they were supposed to generate were similar to those of participants who had previously gone through a learning phase (Experiments 1 and 2). Specifically, we asked participants to (1) generate full sentences, to (2) begin their sentences with either bare plural nouns (e.g., "zorbs") or universally quantified nouns (e.g., "all zorbs"), and to (3) avoid using the same type of beginning across all 16 sentences they would be asked to create. Participants were also told that the second clue could be used in another form than it was given ("For example, if you are given the word 'eye,' you can use the word 'eyes' in your sentence instead"). These rules were reiterated on each trial, so that participants didn't have to remember them. Across participants, we randomized the order in which the two types of noun phrases (bare plural and universally quantified) were mentioned in the rules, so as to avoid biasing participants toward using one or the other.

Coding. The same coding criteria were used as in Experiments 1 and 2. However, because the instructions provided in this study explicitly limited participants' responses to sentences beginning with bare plural nouns or universally quantified nouns, there were almost no responses that were ambiguous in scope.[10] As a result, we did not ask a second researcher to code participants' responses in this study.

Results and Discussion

There was no difference in the number of generic sentences participants generated for clues previously provided for facts describing generalizable properties (M = 60.1% of all responses) and for clues previously provided for facts describing idiosyncratic properties (M = 61.4%), Wilcoxon $Z = 0.72$, $p = .47$.[11] This result suggests that our findings in Experiments 1 and 2—namely, the higher generalization-bias scores for generalizable properties than for idiosyncratic properties—were not an artifact of the clues provided during the recall phase. Instead, it is more likely that participants were actually *misremembering* the quantified generalizable facts as being about the relevant categories, arguably because of the hypothesized bias to make kind generalizations.

Experiment 4

We are proposing that humans are biased to generalize information about quantified sets to the level of entire kinds. So far, however, the quantified evidence we have provided to participants was extremely strong: namely, that *all* members of a kind exhibit a certain feature.

[10] Although instructing participants to begin their sentence with a bare plural does not guarantee that they will produce a generic statement, in fact in this study they almost always did.

[11] Participants produced more generic responses than would be expected by chance (50%) for both sets of clues, one-sample Wilcoxon Zs = 4.27 and 4.79, ps <. 001, for the generalizable and the idiosyncratic property clues, respectively.

Kind generalizations based on universally quantified evidence are, of course, easy to make. For a stronger test of our proposal, in Experiment 4 we asked whether people would still make spontaneous kind generalizations (and the errors caused by these generalizations) if the quantified information were weaker. Specifically, we tested whether people would spontaneously generalize information about *many* members of a kind to the kind itself.[12]

Method

Participants. One hundred and two participants were recruited using Amazon's Mechanical Turk platform and completed the study online. All were native English speakers residing in the US. The reward for participation was $0.75. Three additional participants were tested but excluded from the sample because (1) they reported cheating (n = 2), or (2) answered "1" to all 30 arithmetic questions during a distractor block, which suggested they did not take the survey seriously (n = 1).

Materials and procedure. We used the same items as in Experiments 1 and 2, except the sentences quantified with "all" were replaced with sentences quantified with "many." The procedure was the same as that in Experiment 2 (that is, online testing with a similarity control).

Coding. The coding was the same as in Experiments 1 and 2, except the "all" coding category was replaced with an analogous "many" category (e.g., "Many stupps get something in their hair"). A second researcher coded the responses of approximately 25% of the participants (n = 25) to assess reliability. Cohen's kappas were .94, .97, and .92, for the generic, "many", and "other" categories, respectively, indicating excellent agreement.

[12] In principle, our argument applies to quantified evidence that is even weaker. In the context of our paradigm, however, scalar implicature may block kind generalizations from evidence conveyed via statements with weaker quantifiers. For example, if subjects heard that "*some* zorbs eat fruits and vegetables," they might infer that most actually do not, which might block the relevant generalization to zorbs as a kind. This is a problem specifically when the evidence is conveyed linguistically via quantified sentences: Actually *witnessing* some zorbs eating fruits and vegetables would be perfectly compatible with a kind inference.

Results and Discussion

Even though the quantified evidence was weaker in this experiment, we predicted that participants would still spontaneously generalize it (when appropriate) to the relevant kinds, which would lead to higher generalization-bias scores for the generalizable properties relative to the idiosyncratic properties. The results supported this prediction. Participants' generalization-bias scores were significantly higher for facts that described generalizable properties (M = 57.4% more "many"-to-generic than generic-to-"many" conversions) than for facts that described idiosyncratic properties (M = 50.0%), even when controlling for the perceived similarity in meaning of the "many" and generic facts, Wald $\chi^2(1) = 5.11$, $p = .024$, $d = .17$. [13, 14] Also, as in Experiment 2, participants' perceived similarity between the meaning of the "many" and generic statements did not predict their generalization-bias scores, Wald $\chi^2(1) = 0.06$, $p = .80$. When we considered only the percentage of "many"-to-generic conversions (rather than the generalization-bias scores as a whole), we again found more of these key conversions for generalizable properties (M = 63.5% of "many" facts) than for idiosyncratic properties (M = 57.1%), Wilcoxon $Z = 2.22$, $p = .026$.

An analysis of individual participants' response patterns revealed that, as in Experiments 1 and 2, there were more participants who showed the predicted pattern (higher generalization-bias scores for the generalizable than the idiosyncratic properties) than participants who showed

[13] The average similarity rating for "many" and generic facts that described generalizable properties (M = 7.70) was not statistically different from the average rating for facts that describe idiosyncratic properties (M = 7.63), Wilcoxon Z =1.50, p = .14.

[14] The generalization-bias scores are generally higher here than in Experiment 2 (which used the quantifier "all"). This difference seems to be driven in part by the fact that generic-to-"many" conversions were less frequent in this experiment than generic-to-"all" conversions had been in Experiment 2 (see Table 2). One possible reason for this difference might be that "all" statements were seen as more similar in meaning to generic statements than "many" statements were, as revealed by participants' similarity ratings. Thus, participants may have been more likely to accidentally misremember generic statements as being about all rather than many members of a kind. However, this greater similarity of "all" and generic statements did not similarly boost "all"-to-generic conversions, which were actually slightly lower than "many"-to-generic conversions. This further speaks against the possibility that confusability of meaning was responsible for the crucial quantified-to-generic conversions in our studies (see the discussion of this alternative explanation in Experiment 2).

the opposite pattern (36.3% vs. 23.5% of participants, respectively). However, this difference did not reach statistical significance in this study, $p = .12$ by a sign test.

These findings provide additional support for the proposed bias to generalize to kinds. Even though the quantified statements in this experiment provided weaker evidence than those in Experiments 1 and 2, people nevertheless spontaneously generalized the information they conveyed to the corresponding kinds, which in turn interfered with participants' ability to perform well on the memory task.

General Discussion

We proposed that humans exhibit a powerful bias to spontaneously generalize to kinds. To test this claim, we used a task that—unlike most other research on kind generalizations—provided no prompts to generalize. In fact, this task (modified from Leslie & Gelman, 2012) arguably discouraged participants from making generalizations because the goal of the task was simply to *memorize* a series of generic and quantified statements. We found that participants made the precise pattern of errors one would expect to see if they were implicitly, spontaneously formulating category generalizations based on the quantified statements: That is, participants were more likely to misremember quantified statements as generic (rather than the reverse) when those statements described generalizable properties as opposed to when they described idiosyncratic properties. We also found this increased rate of memory errors for the generalizable properties when participants were under additional cognitive demands (Experiment 1), suggesting that the hypothesized generalization bias can operate "under the radar," without taking up much cognitive capacity. Finally, our findings ruled out two alternative explanations. In Experiments 2 and 4, we found that the predicted difference in generalization-bias scores between facts that describe generalizable and idiosyncratic properties persisted when controlling for the perceived similarity/confusability of the quantified and generic forms of these facts. Moreover, in Experiment 3 we demonstrated that the clues provided in the recall phase of the memory experiments could not have been responsible for the differences observed in generalization-bias scores. Together, these studies support our proposal that people are biased to make kind generalizations—to spontaneously generalize novel information to kinds whenever such generalizations are justifiable.

To highlight the striking nature of these results, we should point out that there are valid considerations that could have prompted participants to convert the quantified statements about *idiosyncratic* properties to generic form, which would have led to a pattern opposite of what we actually observed. Typically, idiosyncratic properties of the sort used in our study apply to fewer category members than generalizable properties do (e.g., the norming study in the Method of Experiment 1; Cimpian et al., 2010). As a result, quantified statements that imply idiosyncratic properties are highly prevalent (e.g., "*All/Many* stups get mud in their hair") are less plausible than analogous statements about generalizable properties (e.g., "*All/Many* zorbs eat fruits and vegetables"). In principle, then, participants could have preferred to convert the quantified statements about idiosyncratic properties to generic form because generic statements can plausibly be true even if there is little statistical evidence to support them (e.g., Cimpian et al., 2010; Leslie, 2008). The fact that participants converted instead more of the quantified statements about *generalizable* properties to generic form, despite the fact that they were plausible as is, strengthens our claim that participants' behavior in our task was driven by a bias to generalize to kinds.

It is also important to note that the present results cannot be fully explained by, and thus go beyond, an ease-of-processing bias of the sort previously proposed in the literature (e.g., Hampton, 2012; Leslie & Gelman, 2012). If the only factor driving participants' responses in our task had been a relative difference in the difficulty of processing quantified and generic statements (with quantified statements being more effort-intensive to understand, evaluate, store, etc., than generic statements), then participants should have converted an equal number of quantified statements to generic form across the two property types. The ease-of-processing account alone does not predict that quantified statements vary in their computational complexity depending on whether they refer to, say, eating fruits and vegetables or getting mud in one's hair. Thus, the greater asymmetry between the number of quantified-to-generic and generic-to-quantified memory conversions we observed for generalizable relative to idiosyncratic properties is consistent with, and was predicted a priori by, our argument that human cognition is biased to generalize evidence about quantified samples to the entire relevant kinds whenever such generalizations are warranted. From a broader perspective, the presence of such a bias is likely to have a powerful impact on the development of our conceptual knowledge, facilitating the acquisition of a tremendous amount of category knowledge from experience with particular samples.

We should also clarify that the present evidence for a bias to generalize to kinds does not speak against the ease-of-processing account. In fact, in our experiments we consistently found more conversions from quantified facts to generic facts than conversions in the opposite direction, for both the generalizable and the idiosyncratic properties. This overall conversion asymmetry replicates Leslie and Gelman's (2012) findings and is in line with their arguments that reasoning about generic facts is less cognitively demanding than reasoning about quantified facts. Therefore, the generalization bias and the ease-of-processing bias are likely to operate in tandem to influence how people learn about the world.

Returning to the present studies, our results also suggest that the bias to draw kind inferences might operate without making many demands on cognitive resources. We base this conclusion on the fact that the effect of property type (i.e., higher generalization-bias scores for generalizable than for idiosyncratic properties) was as strong for participants who were under a cognitive load as it was for those who were not. Thus, the implicit bias to generalize to kinds may function not just when people have the luxury of explicitly focusing on learning about categories. Rather, this fundamental bias probably operates under most everyday circumstances, even when we are engaged in other activities and not deliberately trying to acquire generic knowledge.

Our proposal of a bias to generalize to kinds is compatible in spirit with previous hypotheses and evidence from the developmental literature that highlight the privileged status of kind representations in human cognition. Consider, for example, the recent claims of an early—perhaps even innate—sensitivity to social cues that signal the transmission of generic knowledge (Csibra & Gergely, 2006, 2009). To illustrate, communicative cues such as eye contact and pointing to an object lead 9-month-olds to encode and remember kind-relevant properties of that object (e.g., shape, color) rather than kind-irrelevant ones (e.g., location; Yoon, Johnson, & Csibra, 2008; see also Futó, Téglás, Csibra, & Gergely, 2010; Butler & Markman, 2012). Such ostensive/pedagogical contexts have also been shown to elicit higher rates of generic language (Gelman, Ware, Manczak, & Graham, 2013), which provides a very effective means of conveying generic knowledge and is understood by children as young as 2 and 3 (e.g., Cimpian & Markman, 2008; Cimpian, Meltzer, & Markman, 2011; Gelman & Raman, 2003; Graham, Nayer, & Gelman, 2011). Also related to our present argument, Cimpian and Erickson (2012) demonstrated that preschool-age children are better able to recall information that pertains to a kind compared to identical information about an individual. Children's ability to retain kind knowledge faithfully in long-term memory dovetails nicely with the proposed bias to generalize to kinds and thereby acquire such kind knowledge.

In conclusion, the current findings suggest that the human mind may be biased to make spontaneous kind generalizations whenever the evidence at hand allows such generalizations. This bias is likely to exert a powerful—yet often unnoticed—influence on our learning, guiding us towards knowledge at the level of abstract kinds.

References

Baddeley, A. D. & Hitch, G. J. (1974). Working Memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory, 8,* 47-90, New York: Academic Press. doi:10.1016/S0079-7421(08)60452-1

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*(4), 1416-1428. doi:10.1111/j.1467-8624.2012.01775.x

Carlson, G. N. & Pelletier, F. J. (Eds.). (1995). *The generic book.* Chicago, IL: 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*(8), 1452-1482. doi:10.1111/j.1551-6709.2010.01126.x

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: new evidence that categories are privileged in children's thinking. *Cognitive Psychology*, *64*(3), 161-185. doi:10.1016/j.cogpsych.2011.11.002

Cimpian, A., & Markman, E. M. (2008). Preschool children's use of cues to generic meaning. *Cognition*, *107*(1), 19-53. doi:10.1016/j.cognition.2007.07.008

Cimpian, A., Meltzer, T. J., & Markman, E. M. (2011). Preschoolers' use of morphosyntactic cues to identify generic sentences: Indefinite singular noun phrases, tense, and aspect. *Child Development, 82*(5), 1561-1578. doi:10.1111/j.1467-8624.2011.01615.x

Cimpian, A., & Park, J. J. (2014). Tell me about *pangolins*! Evidence that children are motivated to learn about kinds. *Journal of Experimental Psychology: General, 143*(1), 46-55.

doi:10.1037/a0031687

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 249-274). Oxford, UK: Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148-153. doi:10.1016/j.tics.2009.01.005

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science, 21*(12), 1871-1877. doi: 10.1177/0956797610388810

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition, 117*(1), 1-8. doi:10.1016/j.cognition.2010.06.003

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95. doi:10.1016/0010-0285(88)90025-4

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

Gelman, S. A. (2010). Generics as a window onto young children's concepts. In: *Kinds, things, and stuff: Mass terms and generics,* ed. J. F. Pelletier, pp. 100-120. Oxford University Press.

Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., & Pappas, A. (1998). Beyond labeling: The role of parental input in the acquisition of richly-structured categories.

Monographs of the Society for Research in Child Development, 63(1), 1-157. doi:10.2307/1166211

Gelman, S. A, & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183-209. doi:10.1016/0010-0277(86)90034-X

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*(6), 1532-1541. doi:10.1111/j.1467-8624.1987.tb03864.x

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development, 74*, 308-325. doi:10.1111/1467-8624.00537

Gelman, S. A., Ware, E. A., Manczak, E. M., & Graham, S. A. (2013). Children's sensitivity to the knowledge expressed in pedagogical and nonpedagogical contexts. *Developmental Psychology, 49*(3), 491-504. doi:10.1037/a0027901

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*(4), 509-517. doi:10.1037/0022-3514.60.4.509

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*(2), 409-427. doi:10.1111/j.1467-8624.2004.00683.x

Graham, S. A., Nayer, S. L., & Gelman, S. A. (2011). Two-year-olds use the generic/nongeneric distinction to guide their inferences about novel kinds. *Child Development, 82*(2), 493-507. doi:10.1111/j.1467-8624.2010.01572.x

Hampton, J. A. (2012). Thinking intuitively: The rich (and at times illogical) world of concepts.

Current Directions in Psychological Science, 21(6), 398-402. doi:10.1177/0963721412457364

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, *38*(6), 883-894. doi:10.1037//0012-1649.38.6.883

Jönsson, M. L. & Hampton, J. A. (2006). The inverse conjunction fallacy. *Journal of Memory and Language, 55*, 317-334. doi:10.1177/0963721412457364.

Keates, J., & Graham, S. A. (2008). Category markers or attributes: Why do labels guide infants' inductive inferences? *Psychological Science*, *19*(12), 1287-1293. doi:10.1111/j.1467-9280.2008.02237.x

Lawler, J. M. (1973). Studies in English generics. *University of Michigan Papers in Linguistics, 1*(1), 1-184.

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

Leslie, S. J. (2012). Generics articulate default generalizations. *Recherches Linguistiques de Vincennes*, *41*, 25-45. doi:10.4000/rlv.2048

Leslie, S. J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, *64*(3), 186-214. doi:10.1016/j.cogpsych.2011.12.001

Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, *65*(1), 15-31. doi:10.1016/j.jml.2010.12.005

Mannheim, B., Gelman, S. A., Escalante, C., Huayhua, M., & Puma, R. (2011). A developmental analysis of generic nouns in Southern Peruvian Quechua. *Language Learning and Development*, *7*(1), 1-23. doi:10.1080/15475441003635620

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Meyer, M., Gelman, S. A., & Stilwell, S. M. (2011). Generics are a cognitive default: Evidence from sentence processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Boston, MA: Cognitive Science Society.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198.

Prasada, S., Khemlani, S., Leslie, S. J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition, 126*(3), 405-422. doi:10.1016/j.cognition.2012.11.010

Shipley, E. F. (1993). Categories, hierarchies, and induction. *The Psychology of Learning and Motivation, 30,* 265-301. doi:10.1016/S0079-7421(08)60299-6

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Sutherland, S. L., & Friedman, O. (2012). Preschoolers acquire general knowledge by sharing in pretense. *Child Development*, *83*(3), 1064-1071. doi:10.1111/j.1467-8624.2012.01748.x

Sutherland, S. L. & Friedman, O. (2013). Just pretending can be really learning: Children use pretend-play as a source for acquiring generic knowledge. *Developmental Psychology, 49*(9), 1660-1668. doi:10.1037/a0030788

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2012). Acquisition of generic noun phrases in

Chinese: Learning about lions without an “-s”. *Journal of Child Language*, *39*(1), 130-161. doi:10.1017/S0305000910000735

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America, 105*(36), 13690-13695. doi:10.1073/pnas.0804388105

Table 1

The 16 items, in generic and universally quantified format

	Fact Format	
Property Type	**Generic**	**Universally Quantified**
Generalizable[a]	Cheebas sleep through the winter	All cheebas sleep through the winter
	Daxes keep food in their cheeks	All daxes keep food in their cheeks
	Reesles like to swim in the ocean	All reesles like to swim in the ocean
	Blins sweat through their paws	All blins sweat through their paws
	Mooks shed their skin every year	All mooks shed their skin every year
	Zorbs eat fruits and vegetables	All zorbs eat fruits and vegetables
	Lorches taste with their feet	All lorches taste with their feet
	Glippets build their nests on mountain peaks	All glippets build their nests on mountain peaks
Idiosyncratic[b]	Stups get mud in their hair	All stups get mud in their hair
	Ollers have broken legs	All ollers have broken legs
	Ackles get fungus infections in their ears	All ackles get fungus infections in their ears
	Kweps chip their teeth on nuts	All kweps chip their teeth on nuts
	Zoovs fall out of trees while sleeping	All zoovs fall out of trees while sleeping
	Kazzes trip over logs and rocks	All kazzes trip over logs and rocks
	Sapers twist their ankles	All sapers twist their ankles
	Flooms get dust on their faces	All flooms get dust on their faces

[a]The memory clues for the *generalizable* properties were as follows: “cheeba” and “winter,” “dax” and “cheek,” “reesle” and “ocean,” “blin” and “paw,” “mook” and “skin,” “zorb” and “vegetable,” “lorch” and “foot,” “glippet” and “mountain.”
[b]The memory clues for the *idiosyncratic* properties were as follows: “stup” and “hair,” “oller” and “leg,” “ackle” and “ear,” “kwep” and “tooth,” “zoov” and “tree,” “kazz” and “rock,” “saper” and “ankle,” “floom” and “face.”

Table 2

Average percentages for various measures in Experiments 1, 2, and 4, by property type and cognitive load (standard deviations in parentheses below)

	Generalizable Properties				**Idiosyncratic Properties**			
Measure	**Exp. 1 (CL)**	**Exp. 1 (NL)**	**Exp. 2 (“all”)**	**Exp. 4 (“many”)**	**Exp. 1 (CL)**	**Exp. 1 (NL)**	**Exp. 2 (“all”)**	**Exp. 4 (“many”)**
Generalization-bias score (quantified-to-generic minus generic-to-quantified conversions)	25.6 (62.6)	22.8 (61.6)	34.5 (60.1)	57.4 (42.6)	18.5 (60.7)	13.7 (63.6)	25.3 (59.0)	50.0 (41.9)
Quantified-to-generic conversions (originally presented “all”/“many” statements recalled as generic)	47.8 (40.9)	48.1 (37.1)	56.8 (37.3)	63.5 (35.8)	39.8 (37.9)	43.3 (34.2)	47.6 (36.9)	57.1 (35.8)
Quantified – correct (originally presented “all” or “many” statements recalled as “all” or “many,” respectively)	27.8 (33.8)	41.4 (37.1)	29.8 (34.8)	14.7 (23.4)	28.4 (33.7)	44.6 (37.0)	37.8 (35.5)	16.4 (22.4)
Quantified – other (originally presented “all”/“many” statements recalled as neither “all”/“many” nor generic)	24.4 (31.6)	10.5 (24.3)	13.4 (28.9)	21.8 (31.4)	31.8 (35.1)	12.1 (23.2)	14.6 (26.9)	26.5 (35.5)
Generic-to-quantified conversions (originally presented generic statements recalled as “all”/“many”)	22.2 (29.6)	25.3 (32.0)	22.3 (31.3)	6.1 (15.9)	21.3 (30.9)	29.6 (35.9)	22.3 (30.3)	7.1 (16.3)
Generic – correct (originally presented generic statements recalled as generic)	50.3 (38.8)	64.5 (35.4)	64.6 (35.8)	71.8 (33.3)	45.7 (39.7)	58.3 (36.9)	61.9 (34.4)	68.1 (33.8)
Generic – other (originally presented generic statements recalled as neither generic nor “all”/“many”)	27.5 (33.7)	10.2 (23.1)	13.1 (27.0)	22.1 (30.1)	33.0 (34.4)	12.1 (25.7)	15.8 (29.8)	24.8 (34.0)

Note. CL = Cognitive Load. NL = No Load. The three quantified rows add up to 100% (within each column), as do the three generic rows.

Appendix

Table A1

Raw averages for various measures in Experiments 1, 2, and 4, by property type and cognitive load (standard deviations in parentheses below)

	Generalizable Properties				Idiosyncratic Properties			
Measure	**Exp. 1 (CL)**	**Exp. 1 (NL)**	**Exp. 2 ("all")**	**Exp. 4 ("many")**	**Exp. 1 (CL)**	**Exp. 1 (NL)**	**Exp. 2 ("all")**	**Exp. 4 ("many")**
Generalization-bias score (quantified-to-generic minus generic-to-quantified conversions)	1.02 (2.50)	0.91 (2.47)	1.38 (2.40)	2.29 (1.70)	0.74 (2.43)	0.55 (2.54)	1.01 (2.36)	2.00 (1.68)
Quantified-to-generic conversions (originally presented "all"/"many" statements recalled as generic)	1.91 (1.64)	1.92 (1.48)	2.27 (1.49)	2.54 (1.43)	1.59 (1.51)	1.73 (1.37)	1.90 (1.48)	2.28 (1.43)
Quantified – correct (originally presented "all" or "many" statements recalled as "all" or "many," respectively)	1.11 (1.35)	1.66 (1.49)	1.19 (1.39)	0.59 (0.94)	1.14 (1.35)	1.78 (1.48)	1.51 (1.42)	0.66 (0.90)
Quantified – other (originally presented "all"/"many" statements recalled as neither "all"/"many" nor generic)	0.98 (1.26)	0.42 (0.97)	0.54 (1.16)	0.87 (1.26)	1.27 (1.41)	0.48 (0.93)	0.58 (1.08)	1.06 (1.42)
Generic-to-quantified conversions (originally presented generic statements recalled as "all"/"many")	0.89 (1.18)	1.01 (1.28)	0.89 (1.25)	0.25 (0.64)	0.85 (1.24)	1.18 (1.44)	0.89 (1.21)	0.28 (0.65)
Generic – correct (originally presented generic statements recalled as generic)	2.01 (1.55)	2.58 (1.42)	2.58 (1.43)	2.87 (1.33)	1.83 (1.59)	2.33 (1.48)	2.48 (1.38)	2.73 (1.35)
Generic – other (originally presented generic statements recalled as neither generic nor "all"/"many")	1.10 (1.35)	0.41 (0.92)	0.52 (1.08)	0.88 (1.20)	1.32 (1.38)	0.48 (1.03)	0.63 (1.19)	0.99 (1.36)

Note. CL = Cognitive Load. NL = No Load. The three quantified rows add up to 4 (within each column), as do the three generic rows.