Journal of Memory and Language 65 (2011) 15–31

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Do all ducks lay eggs? The generic overgeneralization effect

Sarah-Jane Leslie [a,*], Sangeet Khemlani [b], Sam Glucksberg [b]

[a] Department of Philosophy, Princeton University, Princeton, NJ 08544, USA
[b] Department of Psychology, Princeton University, Princeton, NJ 08544, USA

ARTICLE INFO

Article history:
Received 19 November 2009
revision received 20 December 2010
Available online 24 February 2011

Keywords:
Generics
Semantics
Linguistics
Concepts
Philosophy
Quantifiers

ABSTRACT

Generics are statements such as "tigers are striped" and "ducks lay eggs". They express general, though not universal or exceptionless, claims about kinds (Carlson & Pelletier, 1995). For example, the generic "ducks lay eggs" seems true even though many ducks (e.g. the males) do not lay eggs. The universally quantified version of the statement should be rejected, however: it is incorrect to say "*all* ducks lay eggs", since many ducks do not lay eggs. We found that adults nonetheless often judged such universal statements true, despite knowing that only one gender had the relevant property (Experiment 1). The effect was not due to participants interpreting the universals as quantifying over subkinds, or as applying to only a subset of the kind (e.g. only the females) (Experiment 2), and it persisted even when people judged that male ducks did not lay eggs only moments before (Experiment 3). It also persisted when people were presented with correct alternatives such as "some ducks do not lay eggs" (Experiment 4). Our findings reveal a robust generic overgeneralization effect, predicted by the hypothesis that generics express primitive, default generalizations.

Introduction

Statements such as "tigers are striped", "ravens are black", "ducks lay eggs" and "ticks carry Lyme disease" are known as *generics*. Insofar as they are not used to convey information about a particular individual but rather information about a kind, these statements express generalizations (Carlson, 1977; Carlson & Pelletier, 1995). Generics have been widely studied by linguists and philosophers, and have recently attracted the attention of psychologists (Chambers, Graham, & Turner, 2008; Cimpian, Brandone, & Gelman, 2010; Cimpian, Gelman, & Brandone, 2010; Gelman, 2003; Gelman & Bloom, 2007; Gelman, Goetz, Sarnecka, & Flukes, 2008; Gelman & Raman, 2003; Gelman, Star, & Flukes, 2002; Gelman & Tardif, 1998; Goldin-Meadow, Gelman, & Mylander, 2005; Hollander, Gelman, & Star, 2002; Khemlani, Leslie, & Glucksberg, 2009; Khemlani, Leslie, Glucksberg, & Rubio-Fernandez, 2007; Prasada & Dillingham, 2006; Prasada & Dillingham, 2009). This paper presents four studies designed to investigate the relationship between people's interpretations of generics and their interpretations of the quantifiers "all" and "some".

From a logical point of view, statements like "all tigers are striped", which we will refer to as *universally quantified* statements, are only true if every single tiger is striped. The existence of a single stripeless tiger is enough for the universal statement "all tigers are striped" to be false. Unlike universally quantified statements, generics such as "tigers are striped" can be true even if there are some stripeless tigers (Carlson, 1977; Gelman, 2003; Krifka et al., 1995; Lawler, 1973). Further, some generic statements are judged true even though a large percentage of the kind lack the property in question. For example "ducks lay eggs" and "mosquitoes carry the West Nile virus" are true, but male and immature ducks never lay eggs, and over ninety-nine percent of mosquitoes do not carry the West Nile virus (Carlson, 1977; Cohen, 1996; Leslie, 2007; Leslie, 2008; see also Cimpian, Brandone, et al., 2010; Cimpian, Gelman,

* Corresponding author. Address: Department of Philosophy, 1879 Hall, Princeton University, Princeton, NJ 08544, USA.
E-mail address: sjleslie@princeton.edu (S.-J. Leslie).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.12.005

et al., 2010). Recent empirical work confirms that people often judge that these generics are true even when they know that a large percentage of the kind lacks the predicated property (Cimpian, Brandone, et al., 2010; Cimpian, Gelman, et al., 2010; Khemlani et al., 2007; Khemlani et al., 2009).

Sentences involving the quantifier "some" (henceforth, *existentially quantified statements*), e.g. "some tigers are albinos" or "some dogs have only three legs," are true so long as there is at least one albino tiger or one three-legged dog. The corresponding generic statements "tigers are albinos" and "dogs have only three legs", however, are rejected (Carlson, 1977; Carlson & Pelletier, 1995; Khemlani et al., 2009). Some generics are also rejected even though *most* members of the kind have the property – for example, people reject "books are paperbacks" and "Canadians are right-handed", despite knowing that more than fifty percent of the kind has the property (Carlson, 1977; Carlson & Pelletier, 1995; Khemlani et al., 2009). For these reasons, semanticists distinguish the meanings of generic statements from the meanings of both universally and existentially quantified statements (e.g., Carlson, 1977; Krifka et al., 1995; Lawler, 1973).

Unlike quantifiers such as "all", "most", or "some", generics are difficult to analyze from the semantic perspective since they cannot be described in set-theoretic terms (Leslie, 2007). Semantic analyses thus suggest that generics should be much more difficult to acquire and process than quantifiers, due to their greater logical complexity (Leslie, 2008; see also Carlson, 1977; Cohen, 1996; Krifka et al., 1995; Pelletier & Asher, 1997). However, recent developmental findings suggest that generics may be as easy as quantifiers for young children to acquire and process, and in some cases even easier (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998; Gelman & Tardif, 1998; Gelman, Hollander, Star, & Heyman, 2000; Hollander et al., 2002; Gelman, 2003; Gelman & Raman, 2003; Goldin-Meadow et al., 2005; Gelman et al., 2008; Hollander, Gelman, & Raman, 2009; Graham, Nayer, & Gelman, in press; Pappas & Gelman, 1998; Papafragou & Schwarz, 2005/2006; Tardif, Gelman, Fu, & Zhu, in press).

The generics-as-defaults hypothesis

In light of these considerations and others, several theorists have proposed that generics may express default generalizations (Csibra & Gergely, 2009; Gelman, 2010; Gelman & Brandone, 2010; Hollander et al., 2009; Leslie, 2007; Leslie, 2008). That is, the cognitive system may have an automatic, early-developing way of generalizing information from individuals to kinds (Baldwin, Markman, & Melartin, 1993; Graham, Kilbreath, & Welder, 2001; Keates & Graham, 2008; Leslie, 2008). These primitive kind-based generalizations are, according to this hypothesis, later articulated in language as generics. If correct, this hypothesis would explain why generics are understood and produced by young children, despite the semantic complexity that linguists have claimed generics exhibit (Carlson & Pelletier, 1995; Cohen, 1996; Gelman, 2003; Leslie, 2007; Leslie, 2008). The generalizations expressed by quantified statements, in contrast, represent more sophisticated generalizations – not the primitive default ones expressed by generics (Leslie, 2008).

Leslie (2008) notes that such a hypothesis would explain otherwise puzzling cross-linguistic data: generic interpretations are always associated with less marked syntactic forms than quantified statements (Dahl, 1985; Krifka et al., 1995). For example, in English, one uses the words "all" or "every" to mark a universal generalization, and the word "some" to mark an existential generalizations. However, there is no word "gen" that is used to mark a generic generalization – English speakers do not say "gen tigers are striped" like they say "all tigers are striped". Instead, the generic interpretation is associated with the *absence* of a quantifier word: "tigers are striped". Similar patterns are found cross-linguistically – no known language contains a word "gen" that exclusively marks a generic generalization. Rather, like English, generics are signaled in part by the absence of quantifier terms (Carlson & Pelletier, 1995; Dahl, 1985). Default interpretations are often associated with less-marked surface forms (Chomsky, 2000). Thus this cross-linguistic pattern can be explained if generics express default generalizations. Quantifier words such as "all" and "some" signal that the cognitive system must generalize in a non-default manner, whereas the unmarked generic form allows the cognitive system to rely on its default way of generalizing.

The hypothesis that generics, unlike quantifiers, express default generalizations generates a number of empirical predictions (Gelman, 2010; Leslie, 2008). For example, if understanding quantified statements requires deviating from the default mode of generalization, then both children and adults should sometimes fail to execute this deviation, and so should incorrectly treat quantified statements as generics. This tendency might be more pronounced in young children, but if the generics-as-default hypothesis is correct, adults should also be prone to making these errors, at least under some circumstances. Prior research conducted by Hollander et al. (2002) and Tardif et al. (in press) found evidence that young children may indeed treat quantified assertions as generics. However they did not find this with their adult participants.

In their study, Hollander et al. (2002) investigated the extent to which children and adults differentiated generics from universal and existential claims by asking them a variety of yes/no questions, each of which appeared either in universal form (e.g. "do all shoes have laces?"), generic form (e.g. "do shoes have laces?"), or existential form (e.g. "do some shoes have laces?"). They found that four-year-olds and adults successfully differentiated between all three types of questions in their answers, but three-year-olds did not. Instead, the three-year-olds gave the same responses regardless of whether the question was in universal, generic, or existential form. The difference between the age groups was due entirely to differences in their responses to the two quantifiers – Hollander et al. found no developmental differences in the responses to the generic questions across these three age groups. The three-year-olds responded as the adults did to the generic questions – but then *also* responded in that same way to the universally and existentially quantified questions. The three-year-olds apparently handled the generic questions like the older children and the adults did, but unlike the older age groups, the three-year-olds seemed to treat the quantified questions as though they were in fact generic questions (Hollander et al., 2002). Tardif et al. (in press) conducted a similar study of Mandarin Chinese speakers, and found the same results with three-year-olds. Mandarin speaking four-year-olds also showed the effect; it was not until age five that Mandarin speaking children differentiated their responses to quantified questions from their responses to generic questions.

Such findings are predicted by the generics-as-default hypothesis: when confronted with a quantified statement that applies to an entire kind, the young preschoolers failed to inhibit their default tendency toward generic generalizations, and so they treated the quantified assertions as though they were generic. Thus instead of considering whether, e.g., *all* shoes have laces, the three-year-olds instead evaluated the generic "shoes have laces", and responded accordingly. The English speaking three-year-olds were, however, able to respond accurately to quantified questions in a post-test session where all the questions concerned a small discrete set of items (e.g. "are all the crayons in the box?") rather than category-wide generalizations, as was the case in the main experiment (Hollander et al., 2002), and the Mandarin speaking three- and four-year-olds were able to distinguish "all" from "some" in this context as well (Tardif et al., in press). The results of the main experiment were thus not due to a basic lack of competence with the quantifiers; they seemed instead to do with the difficulty of processing category-wide quantified statements. When confronted with a quantified claim about an entire category (as opposed to a specific subset), young preschoolers appear to rely on their interpretation of the corresponding generic.

In Hollander et al.'s (2002) and Tardif et al.'s (in press) experiments, adults did not make the error of treating quantified statements as though they were generics. The generics-as-defaults hypothesis would predict that this would not always be the case, however. We wondered if Hollander et al.'s and Tardif et al.'s adult findings would hold up across a broader range of test items. Would adult participants have any inclination to treat category-wide quantified statements as though they were in fact generics? Work by Jönsson and Hampton (2006), Osherson, Smith, Wilkie, Lopez, and Shafir (1990) and Sloman (1993), Sloman (1998) suggests that this may occur in the case of the quantifier "all". For example, Jönsson and Hampton found that, in a wide range of circumstances, adults judged that it was more likely, e.g., that all ravens are black, than that all young jungle ravens are black. Clearly, such judgments are erroneous from the point of view of logic, because *all ravens are black* entails that *all young jungle ravens are black*. However, if the participants interpreted the universals as generics, then their responses would be much more reasonable, since the generic *ravens are black* could be true even if jungle ravens were, say, light brown when they are young (Jönsson & Hampton, 2006; Jönsson, personal communication). Indeed, Connolly, Fodor, Gleitman, and Gleitman (2007) confirmed that people routinely reason this way with generics (see also Cimpian, Brandone, et al., 2010; Cimpian, Gelman, et al., 2010). Their focus was not on the interpretation of generic statements per se, but rather on prototypical properties (Rosch, 1978) – properties that generics are well suited to express (Declerk, 1986; Geurts, 1985). While the generics-as-defaults hypothesis would explain Jönsson and Hampton's findings, their data constitute only indirect support for the hypothesis. Is it possible to find more direct evidence of adults interpreting quantified statements as generics?

Data from a memory study by Leslie and Gelman (submitted for publication) supports the hypothesis that adults, as well as children, may sometimes treat quantified statements as generics. Leslie and Gelman's data suggest that adults and young children both frequently misremember category-wide statements quantified with "all" and "most" as generics. This is consistent with the idea that adults may at times treat universally quantified statements as generics. In light of these findings, we wondered whether adults would be inclined to agree to some false universal statements when the corresponding generic is true, despite knowing about the falsifying counterinstances. Preliminary data reported by Khemlani et al. (2007) suggested that this was indeed the case – participants agreed to statements such as "all ducks lay eggs" and "all horses give live birth" approximately half the time. We term this the *generic overgeneralization* (GOG) effect, since it involves overgeneralizing from the truth of a generic to the truth of the corresponding universal statement, as the generics-as-default hypothesis would predict. In Experiment 1, we replicated Khemlani et al. (2007) with a larger range of items. We predicted that the findings from Khemlani et al. (2007) would hold up, and that people would often judge statements like "all ducks lay eggs" to be true.

While Hollander et al. (2002) and Tardif et al. (in press) found evidence that preschoolers treat statements quantified with "some" as generics, we reasoned that it was unlikely that adults would make such errors in this experimental context. This is because generics are rarely accepted if *no* members of the kind have the property, i.e., if the corresponding existential statement is false (Carlson & Pelletier, 1995). Thus, it would not be feasible to find cases where an existential statement was false, but the corresponding generic was true. The only possible GOG errors involving "some," then, would involve people incorrectly judging a true existential statement to be false because the corresponding generic was false, e.g. judging "some Canadians are right-handed" to be false because the generic "Canadians are right-handed" is false. However, it is extremely easy to determine that such an existential is true: one only needs to think of a single right-handed Canadian in order to accept it. In this way, true existential statements are easy to evaluate, and so it is unlikely that adults would incorrectly rely on the default generic in such cases. It may be harder, though, for people to confirm that a universal statement is true, since they are only true if every single member of the kind has the property in question. Adults may thus save cognitive effort by relying on the generic to evaluate the universal. If true existential statements are easier to evaluate than universals, however, it is likely that no GOG effect will be found for statements quantified with "some." If the statement is easy to evaluate, participants will be unlikely to rely on the generic default.

The scope of the overgeneralization effect

The generics-as-default hypothesis predicts that adults will tend to incorrectly endorse false universal statements if the corresponding generic is true. A further question is whether this effect would be found for all types of generics, or whether it is more limited in scope. Generics are often used to express *essential* properties of a kind (Gelman, 2003; Gelman, 2010; Gelman & Brandone, 2010), or properties that bear a *principled* connection to the kind (Prasada & Dillingham, 2006; Prasada & Dillingham, 2009); for example "tigers are striped", "horses have hooves", "doctors heal people". Prasada and Dillingham (2006), Prasada and Dillingham (2009) define the notion of a principled connection so as to include only properties that are expected to be highly prevalent among members of the kind. However, generics such as "ducks lay eggs" and "lions have manes" seem to express essential properties in the sense of Medin and Ortony (1989) and Gelman (2003), and otherwise resemble principled properties, despite predicating properties that are only true of mature members of one gender of the kind (Cimpian, Brandone, et al., 2010; Cimpian, Gelman, et al., 2010; Leslie, 2007, 2008; Leslie, Khemlani, Prasada, & Glucksberg, 2009). A discussion of the relationship between principled and essential properties is beyond the scope of this paper (but see Prasada & Dillingham, 2009), so here we adopt the term *characteristic property* to indicate a property that bears a deep causal and explanatory relation to the kind in question – a property that results from the nature of that kind. Characteristic properties can, in some cases, occur in only a minority of the members of the kind; for example, *laying eggs* would seem to stem from the nature of ducks, even though only the mature fertile female ducks possess the property. Leslie (2007), Leslie (2008) terms generics such as "ducks lay eggs" *minority characteristic generics*, since they express properties that are characteristic of the kind, but are only possessed by a minority of its members.

Regardless of prevalence, not all generics express characteristic properties: for example "cars have radios", "pigeons sit on statues", and "sharks attack swimmers" (Leslie, 2007; Leslie, 2008; Prasada & Dillingham, 2006; Prasada & Dillingham, 2009). Would all these generics support a GOG effect? One hypothesis is that all generics produce GOG effects, regardless of whether the property in question is characteristic of the kind or not. On this hypothesis, people should accept universals like "all cars have radios" and "all sharks attack swimmers" as frequently as they accept minority characteristic universals such as "all ducks lay eggs" and "all lions have manes". There are, however, some theoretical and empirical reasons to suspect that the GOG effect might be most pronounced in the case of characteristic properties (Khemlani et al., 2007). Characteristic properties are usually possessed by all or almost all members of a kind; in most cases, the only members of a kind that lack a characteristic (or essential) property are in some respect abnormal, e.g., a stripeless albino tiger or a three-legged dog (Gelman, 2003). Thus, while the corresponding universal statements may be strictly speaking incorrect – since there are *some* stripeless tigers and three-legged dogs – assertions such as "all tigers are striped" and "all dogs have four legs" may be close enough to being correct for practical purposes. This is not so for non-characteristic properties: there is nothing abnormal about the many sharks that never attack swimmers, and so "all sharks attack swimmers" is patently incorrect. In this way, the tendency to substitute a judgment of the generic for the universal will be generally more successful when the property in question is a characteristic one, and so adult participants may be less likely to show a GOG effect if the property is not characteristic of the kind.

However, there are some characteristic properties that do not occur in all or almost all members of the kind, namely minority characteristic ones, as in "ducks lay eggs" and "lions have manes" (Leslie, 2008). If this hypothesis is correct, then the GOG effect should be observed for minority characteristic predications, such as "all ducks lay eggs" and "all lions have manes". Data from Khemlani et al. (2007) favored the hypothesis that the GOG effect was primarily to be found for minority characteristic generics. The limited range of test items prevented any firm conclusions being drawn, and so Experiment 1 used an expanded set of items to test the GOG effect. A full taxonomy of the types of predications used in these studies may be found in Table 1 below.

If adults do indeed erroneously accept statements like "all ducks lay eggs", does this necessarily support the generics-as-default hypothesis? One alternative explanation is that people may interpret the universal quantifier "all" as applying to subkinds of ducks, rather than to individual ducks. That is, people may understand "all ducks lay eggs" to mean *all kinds of ducks lay eggs* – e.g., Mallard ducks lay eggs, Muscovy ducks lay eggs, and so on. On this interpretation, it would be correct to accept "all ducks lay eggs", and so this would not constitute evidence that participants were defaulting to their judgment of a generic in place of a universal. An alternative explanation is that people may interpret "all ducks" to apply to only a subset of ducks, namely the fertile female ducks. That is, the context may restrict the scope of the quantifier, so that people do not interpret "all ducks" to apply to every single duck in the world, but rather to a subset of those ducks (Stanley, 2007; Stanley & Szabo, 2000). If this explanation is correct, then people's acceptance of "all ducks lay eggs" would again not constitute evidence for the generics-as-defaults hypothesis. Experiment 2 sought to address these issues. In Experiment 2a, participants were given population information (e.g. "there are 431 million ducks in the world") so as to prime them to interpret the subsequent statement ("all ducks lay eggs") as quantifying over (a) individual ducks, and (b) every single duck in the world. If the GOG effect is driven by quantification over subtypes, or by contextual quantifier domain restriction, then it should disappear in the context of population information. Further, in Experiment 2b, participants performed a paraphrase task in which they were asked to paraphrase the sentence they evaluated. The paraphrases were then coded for any indications that participants had generalized over subkinds or interpreted the scope of the quantifier to be restricted to a subset of the kind. If the GOG effect was simply due to generalization over subkinds or restricted quantifier scope, then there should be evidence of this in participants' paraphrases. In particular, they should be more likely to provide a paraphrase that either refers to subkinds or reflects quantifier domain restriction if they accepted the universal than if they rejected it.

Another alternative explanation for the effect would be that participants were simply ignorant of the biological facts. Experiment 3 addressed whether the participants who accepted statements such as "all ducks lay eggs" did indeed know that male ducks do not lay eggs. It is possible – though perhaps unlikely – that people accepted these false universal statements simply out of ignorance. The experiment used a blocked design; one block replicated Experiment 1 and the other asked participants to evaluate statements involving counterexamples such as "male ducks lay eggs". In addition to controlling for background knowledge, we were interested in whether the GOG effect would persist after participants had been asked to think about these counterexamples. Adults take counterexamples into consideration when evaluating universals, so if participants had processed these statements as universals but were somehow unable to bring counterexamples to mind, then the GOG effect should disappear, since counterexamples should have been readily called to mind. Conversely, if participants processed the universal statements as though they were generics, as we predicted, then they should have accepted "all ducks lay eggs" to some extent, even after rejecting the statement "male ducks lay eggs". After all, the truth of the generic "ducks lay eggs" is perfectly compatible with male ducks' failure to do so.

Finally, Experiment 4 tested whether the GOG effect persisted when people were asked to choose between two statements such as "all ducks lay eggs" and "some ducks do not lay eggs". The experiment offered people a correct alternative, and thus made salient the potential error that would be involved in accepting the universal. If the GOG effect is due to participants' relying on their judgment of the generic instead of evaluating the universal, then they may continue to erroneously prefer the universal statement.

Experiment 1: the generic overgeneralization effect

Preliminary work suggests that the GOG effect may occur most clearly for predications that express gender-specific properties that are characteristic of the kind in question, e.g., how the members of the kind reproduce (Khemlani et al., 2007). Can any GOG effect be found for non-characteristic properties, or is it limited to characteristic properties in the case of adults? Prasada and Dillingham (2006), Prasada and Dillingham (2009) and Leslie (2007), Leslie (2008) have identified various other types of generics, so we included each identified type in our experiments, plus a category of statements that are false in generic form. Examples are given in Table 1.

Our first and last types of predications, quasi-definitional and false generalizations, were included for comparison purposes. False generalizations tend to be rejected in generic form despite predicating prevalent properties of the kind (Khemlani et al., 2009). The inclusion of this category allowed us to determine which, if any, of the other types were more likely to be accepted in universal form. If other types of statements were accepted in universal form more frequently than false generalizations, then that would suggest a GOG effect for them. At the other end of the spectrum are the quasi-definitional predications, which are genuinely true in universal form. Acceptance in universal form of any predication types other than quasi-definitional ones constitutes an error, since there are always some members of the kind that lack the property. Majority characteristic predications involve characteristic properties that are true of all the *normal* members of the

Table 1
Various types of generic generalizations used in Experiments 1, 2, and 3.

Predication type	Example	Truth value of the generic	Description
Quasi-definitional	Triangles have three sides	True	Property must be universally true of all the members of the kind; no exceptions
Majority characteristic	Tigers have stripes	True	Property must be central, principled or essential (Gelman, 2003; Medin & Ortony, 1989) – namely, it must be directly related to the nature of the kind in question. It must also be prevalent though not universally had among members of the kind; while some exceptional members (e.g. albino tigers) fail to possess it, all the normal members of the kind must possess it
Minority characteristic	Lions have manes	True	Property must be central, principled or essential (Gelman, 2003; Medin & Ortony, 1989) – namely, it must be directly related to the nature of the kind in question. However, it must only be held by a minority of the kind. For our purposes we restricted these items to methods of gestation, methods of nourishing the very young, and characteristic physical traits had only by one gender
Majority	Cars have radios	True	Property must be prevalent among members of the kind, and must not be a principled connection (Prasada & Dillingham, 2006, 2009).
Striking	Pit bulls maul children	True	Property must only be had by a small minority of the kind, and must signify something dangerous and to be avoided
False generalization	Canadians are right-handed	False	Property must be prevalent among members of the kind and there must be a sufficiently salient alternative property (e.g. being left-handed), so that the generic form of the predication sounds false or mistaken

kind (they map onto Prasada & Dillingham's, 2006; Prasada & Dillingham's, 2009 notion of principled connections), and so it would not be surprising if people accepted those statements in universal form. We were particularly interested in minority characteristic, majority, and striking predications throughout our studies. Recent work suggests that these different types of generic predications can be distinguished along various dimensions (Cimpian, Brandone, et al., 2010; Cimpian, Gelman, et al., 2010; Khemlani et al., 2009; Prasada & Dillingham, 2006; Prasada & Dillingham, 2009; Leslie et al., 2009), and so they may be treated in different ways on our tasks. Thus our inclusion of minority characteristic, majority and striking items allowed us to better test the extent and limits of the GOG effect.

Method

Participants and procedure

56 volunteers participated in the experiment over the Internet. Participants were chosen from Amazon's Mechanical Turk system for human interface tasks, an online forum through which individuals can participate in tasks for monetary compensation. All spoke English as their first language and none had participated in experiments concerning generics before. They were asked to judge their agreement to each item, one at a time, and registered their responses by selecting buttons marked 'yes' and 'no' over the Internet through an interface written in Ajax.

Materials and design

Participants were presented with 72 experimental items randomly embedded in a list of 60 fillers. The 72 experimental items consisted of six types of predications, and examples are given in Table 1. A full listing of the materials is provided in the Appendix.

Each predicate type appeared in one of three statement types: existential (e.g., "some ducks lay eggs"), generic (e.g., "ducks lay eggs"), or universal (e.g., "all ducks lay eggs"). This generated a 6 (predicate type) × 3 (statement type) repeated measures factorial design. Statements were counterbalanced such that each participant saw any given predicate in only one statement type. We also included 30 false and 30 true filler items. Filler items were non-quantified factual assertions, such as "Paris is the capital of France" (true) and "New York City is in England" (false).

Results and discussion

Fig. 1 shows the proportions of 'yes' responses to each of the eighteen types of assertions. Existentially quantified

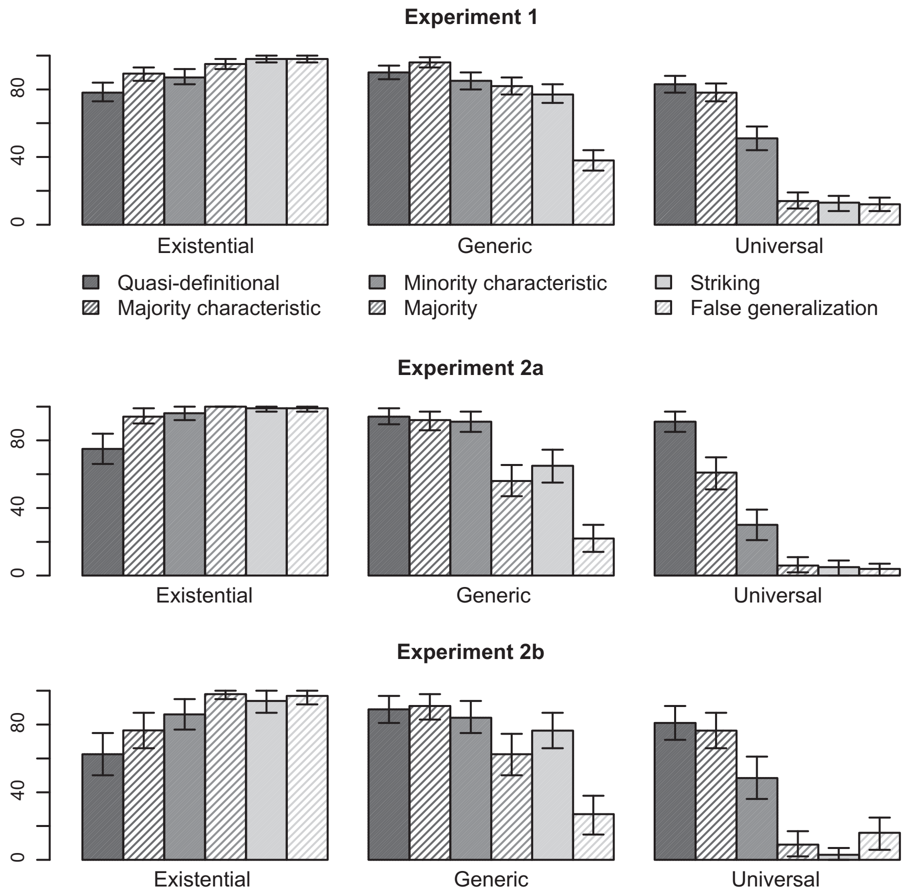

Fig. 1. Proportions of agreement for Experiment 1, 2a, and 2b as a function of statement type and predicate type; 95% confidence intervals shown.

and generic statements were deemed true most of the time, while agreement to universal statements depended on the type of predication expressed. In generic form, quasi-definitional and majority characteristic statements were judged true 90% and 96% of the time, respectively. Minority characteristic predications in generic form were judged true 85% of the time, and majority and striking were judged true in generic form 82% and 77% of the time, respectively. The various assertions in existential form were each judged true over 90% of the time, with the exception of quasi-definitional existentials, which had a 78% acceptance rate. This relatively low acceptance rate can be attributed to the scalar implicature effect, a well-known phenomenon in the language comprehension literature. Given, say, that all triangles have three sides, it is pragmatically odd to assert that *some* triangles have three sides; people sometimes interpret *some* as implying *not all*, and so are reluctant to agree to universally true assertions quantified by "some" (Bott & Noveck, 2004).

False generalizations were rejected most of the time when they appeared in generic form – the agreement rate for them was only 38%. Of more interest are the universally quantified predications. Participants were prone to the GOG effect for minority characteristic and majority characteristic predications: universally quantified minority characteristic predications were judged true 51% of the time and principled predications were judged true 78% of the time when, in fact, all of those statements were false. Further, 40 out of 56 participants judged universal minority characteristic predications true over 30% of the time (binomial test, $p < .005$), and 10 out of 12 items were judged true over 30% of the time (binomial test, $p < .05$), where 30% reflects a proportion high enough to indicate a GOG effect. In contrast, universally quantified majority and striking predications were incorrectly judged true only 13% and 14% of the time, respectively.

The proportions of agreement data were subjected to a 6×3 within-subjects ANOVA. As is conventional, we use F_1 and F_2 to report participant and item analyses respectively, as well as min F' (Clark, 1973). There was a main effect of predicate type, $F_1(5, 275) = 95.21$, $p < .0001$, $F_2(5, 65) = 30.50$, $p < .0001$, min $F'(5, 111) = 23.10$, $p < .0001$, reflecting the higher agreement rate to quasi-definitional, principled, and minority characteristic statements than majority, striking, or false generalization statements. There was also a main effect of statement type, $F_1(2, 110) = 260.33$, $p < .0001$, $F_2(2, 130) = 338.03$, $p < .0001$, min $F'(2, 229) = 147.07$, $p < .0001$, reflecting the higher agreement for 'some' statements (91%) than for generic statements (78%), with the lowest agreement for 'all' statements (42%). Finally, there was a reliable interaction between predicate type and statement type, $F_1(10, 550) = 72.19$, $p < .0001$, $F_2(10, 130) = 42.48$, $p < .0001$, min $F'(10, 301) = 26.55$, $p < .0001$. The interaction can be interpreted in terms of the differences in agreement rate for statement type as a function of predicate type. Agreement rates for existentially quantified assertions were comparable across predicate types. In contrast, quasi-definitional and majority characteristic predicates yielded robust agreement rates in the universal condition, minority characteristic predicates tended towards agreement, and majority, striking, and false generalizations were usually rejected. Finally, generic statements differed as a function of predicate type, with false generalizations eliciting lower agreement than any of the other predications.

Planned comparisons revealed that quasi-definitional, majority characteristic, and minority characteristic assertions were each accepted in universal form significantly more often than false generalizations, $ts(110) > 8.01$, $ps < .0001$, $ds > 1.53$. Neither majority, $t(110) = .57$, $p = .57$, $d = .11$, nor striking assertions were accepted more often than false generalizations, $t(110) = .10$, $p = .92$, $d = .02$. These results support the hypothesis that the GOG effect only occurs for predications that express characteristic properties of the kind.

Experiment 1 found that the GOG effect occurs on approximately half the trials when the property is characteristic of the kind but only occurs in one gender. No evidence was found of the GOG effect occurring when the property in question is not characteristic of the kind. Thus Experiment 1 provides evidence that the GOG effect is limited to characteristic properties, at least for adults. It also provides evidence for the generics-as-default hypothesis more generally, since, as predicted, adults exhibited a tendency to accept false universals with true corresponding generics. However, one alternative explanation of these data is that people interpret the quantifier "all" as applying to subkinds instead of individual members of the kind, in which case Experiment 1 would not in fact constitute evidence for generics being defaults. Another explanation is that people interpret the quantifier "all" as applying only to a subset of the kind – in particular, to only one gender. That is, perhaps they contextually restricted the scope of the universal quantifier so that it does not apply to the entire kind, but only to one gender of the kind – just as if someone asserts "every student passed the exam", she would naturally be interpreted as talking about *the students in her class*, not about every single student in the world (Stanley, 2007; Stanley & Szabo, 2000). Experiment 2 was designed to address these alternative deflationary explanations.

Experiment 2: is the GOG effect due to quantification over subkinds or domain restriction?

To determine whether the GOG effect can be explained by supposing that participants generalize over subkinds or restrict the scope of the universal quantifier, Experiment 2a tested whether the GOG effect persisted in a context in which participants were directed to consider every single individual member of the kind. Experiment 2b solicited paraphrases, to test whether they indicate reliance on a subkind interpretation or quantifier domain restriction.

Experiment 2a: priming individual-based interpretations with population information

Method

Participants and procedure

Twenty-seven volunteers from the same population as Experiment 1 participated over the Internet for pay. All were native English speakers and none had participated in experiments on generics in the past. The study used the same procedure as in Experiment 1, except that on each trial, to prime them to consider each and every individual member of a kind in the world, instead of subkinds or a subset of those individuals, participants were provided with additional information about population estimates for the kind in question. For example, participants were visually presented with the following: ''Suppose the following is true: there are 431 million ducks in the world. Do you agree with the following: all ducks lay eggs.'' The numerical information was artificial, but was designed to convey plausible population estimates for each kind.

Materials and design

Experiment 2a used the same set of items used in Experiment 1, except that it dropped three items: ''rectangles are geometric figures'', ''US Presidents are over 35'', and ''strokes cause paralysis''. These items were dropped because it was not possible to construct meaningful population estimates for each item, e.g., it is nonsensical to say that there are 45 million rectangles in the world, because the set of rectangles is impossible to count. Filler statements were dropped from the study. In total, participants evaluated 69 items. Each item appeared in existential, generic, or universal form. The study thus generated a 6 (predicate type) $\times$ 3 (statement type) repeated measures factorial design.

Results and discussion

Fig. 1 provides the proportions of 'yes' responses to the different assertions used in Experiment 2a. Participants accepted minority characteristic universal assertions 30% of the time, whereas they accepted majority and striking assertions only 6% and 5% of the time respectively. Despite a reduction of the GOG effect, the overall pattern results of the present study replicated those of Experiment 1. The proportions of agreement data were subjected to a 6 $\times$ 3 within-subjects ANOVA, which revealed a main effect of predicate type, $F_1(5, 130) = 100.18$, $p < .0001$, $F_2(5, 62) = 43.02$, $p < .0001$, $\min F'(5, 116) = 30.09$, $p < .0001$, a main effect of statement type, $F_1(2, 52) = 192.50$, $p < .0001$, $F_2(2, 124) = 315.39$, $p < .0001$, $\min F'(2, 117) = 119.54$, $p < .0001$, and a reliable interaction between predicate type and statement type, $F_1(10, 260) = 54.85$, $p < .0001$, $F_2(10, 124) = 31.77$, $p < .0001$, $\min F'(10, 267) = 20.12$, $p < .0001$.

Planned comparisons showed that, as in Experiment 1, minority characteristic assertions were accepted in universal form significantly more often than false generalizations, $t(52) = 5.64$, $p < .0001$, $d = 1.54$. As before, majority $t(52) = .50$, $p = .48$, $d = .19$, and striking assertions were not accepted reliably more often than false generalizations, $t(52) = .57$, $p = .57$, $d = .15$.

Our findings demonstrate that the GOG effect occurred even in a context in which people were asked to consider a population estimate designed to focus attention on every individual member of the kind in the world, rather than on subkinds or on a subset of individuals. This additional information may have reduced the overgeneralization effect; people accepted minority characteristic assertions 30% of the time in this study as compared to 51% of the time in Experiment 1. Thus some of the GOG effect found in Experiment 1 may have been due to some participants employing a subkinds interpretation or a restricted interpretation. However, the results of Experiment 2a indicate that the effect cannot be entirely explained in this way, since the effect remained in a context that encouraged generalization across each and every individual in the world. It is also possible that the GOG effect was reduced here because the context primed participants to employ a more analytic mode of thinking, due to considering numerical information about populations (see Alter, Oppenheimer, Epley, & Eyre, 2007). As a result, the participants may have answered more carefully, and therefore they may have been less likely substitute a default generic judgment for a universal one. The reduction in the GOG effect may thus be due in part to the elimination of subkind and restricted interpretations. Experiment 2b tested the deflationary interpretations further by eliciting participants' paraphrases of quantified and generic assertions.

Experiment 2b: paraphrases of existential, generic, and universal assertions

Method

Participants and procedure

Sixteen volunteers from the same population as Experiment 1 participated over the Internet for pay. All were native English speakers and none had participated in experiments on generics in the past. The experiment consisted of two phases. Participants first performed an evaluation phase in which they carried out the same agreement task as in Experiment 1. They then carried out a second phase, in which they were given the same assertions as in the first phase and were asked to type out a paraphrase of each assertion. Participants were specifically asked to produce paraphrases that retained as much of the original meaning of the statement as possible. The paraphrase task examined whether individuals used language that suggested a subkind interpretation or a restricted interpretation, e.g., whether a participant paraphrased the universal ''all ducks lay eggs'' as ''all types (/kinds/sorts/species) of ducks lay eggs,'' or as ''all female ducks lay eggs.''

Materials and design

Experiment 2b used the same set of items used in Experiment 1, except that all the filler statements were omitted since we were not interested in how the fillers were paraphrased. In total, participants evaluated 72 items. Each item appeared in existential, generic, or universal form. The study thus generated a 6 (predicate type) $\times$ 3 (statement type) repeated measures factorial design.

Results and discussion

The results of the first phase of Experiment 2b replicated those of Experiment 1, and are shown in Fig. 1. Participants accepted minority characteristic universal assertions 48% of the time. The proportions of agreement data were subjected to a 6×3 within-subjects ANOVA. There was a main effect of predicate type, $F_1(5, 75) = 21.46$, $p < .0001$, $F_2(5, 65) = 19.83$, $p < .0001$, min $F'(5, 138) = 10.31$, $p < .0001$, a main effect of statement type, $F_1(2, 30) = 87.23$, $p < .0001$, $F_2(2, 130) = 125.75$, $p < .0001$, min $F'(2, 76) = 50.8$, $p < .0001$, and a reliable interaction between predicate type and statement type, $F_1(10, 150) = 24.89$, $p < .0001$, $F_2(10, 130) = 25.24$, $p < .0001$, min $F'(10, 279) = 12.53$, $p < .0001$. This pattern of results is comparable to those of the previous studies.

In the second phase of the experiment, participants produced paraphrases of the statements that they had evaluated in the first phase. The authors coded the paraphrases on three separate dimensions. First, the paraphrases were coded based on what kind of statement was produced (existential, generic, universal, or other). Coding was based on the semantic significance of the paraphrase, so paraphrases that used the quantifiers "none" or "no", e.g., "no moose are without antlers," were coded as universal quantifiers. Likewise, assertions that made use of the quantifiers "always" and "never" were coded as universals. Definite and indefinite singular assertions, e.g., "a paperback is a book", were coded as generics, as were qualified generics, e.g., "female ducks lay eggs". Paraphrases that made use of the quantifiers "there are" and "a few" were coded as existentials, as were negated universals, e.g., "not all cheetahs are fast runners." Finally, everything that did not fall into these categories was coded as 'other'. Paraphrases were also coded on (a) whether they mentioned subtypes or not, e.g., "some types of pencils are made of wood", and (b) whether they made reference to a restricted subset of the kind, e.g., "female ducks lay eggs". The authors agreed on 79.8% of responses (Cohen's kappa = .73, $z = 43.7$, $p < .0001$). Differences between the coding were reconciled on a case-by-case basis.

Table 2 shows the form of participants' paraphrases as a function of the statement that they were asked to paraphrase. As the table shows, they paraphrased existential assertions as existentials 62% of the time, generics as generics 59% of the time, and universal as universals 51% of the time. Across all responses, only 16% of the paraphrases were classified as 'other'. For those paraphrases that were different from the original assertions, participants paraphrased universals as generics on 36% of trials, which happened significantly more often than when they paraphrased existentials, generics, or universals as any other type of statement (Mann–Whitney tests, all $zs > 3.59$, all $ps < .0001$). (That is, universal-as-generic paraphrases were provided reliably more often than universal-as-existential paraphrases, than generic-as-existential paraphrases, than existential-as-generic paraphrases, and so on for all the other possible combinations.) This finding also is consistent with the generics-as-default hypothesis, as it reveals a tendency to paraphrase universals as generics.

Table 2
Percentages of statements that were paraphrased as an existential, as a generic, as a universal, or in some other form as a function of the statement type of the original assertion in Experiment 2b.

Paraphrased as	Original statement type		
	Existential	Generic	Universal
Existential	62	8	4
Generic	12	59	36
Universal	3	18	51
Other	24	16	9
	100	100	100

Only 1% of paraphrases included subtyping language, and 1.6% of them included a subset restriction across the experiment as a whole. Both proportions were significantly greater than 0 (permutation tests, $ps < .001$). However, the proportion of paraphrases with subtyping language did not correlate with the proportion of GOG errors (Kendall's tau = −.02, $z = .58$, $p = .56$), and neither did the proportion of paraphrases that were restricted to a relevant subset of the kind (Kendall's tau = .02, $z = .76$, $p = .45$). If the GOG effect was driven by these interpretations, then one would expect to see a higher proportion of such paraphrases when the universal was accepted vs. when it was rejected. That is, if the GOG effect was due to participants taking the minority characteristic universals to quantify over subkinds or a restricted subset, then participants who agreed to the universals should be more likely to produce paraphrases that make reference to subtypes or subsets. No such effect was observed, however. In light of this finding, plus the overall low rate of subtyping and restricting paraphrases, Experiment 2b suggests that the GOG effect was not due to subkind or restricted interpretations.

An alternative way of assessing whether participants' answers were due to subkind or restricted interpretations could have been to point out participants' mistakes to them and then see if and how they attempted to defend their original responses. However, people's tendency to produce post hoc rationalizations for phenomena that are driven by wholly unrelated causes is well documented (e.g., Nisbett & Wilson, 1977). Thus, participants may *claim* that in fact they were employing a subkind interpretation in order to justify an otherwise incorrect response that was driven by a default to the generic interpretation – just as Nisbett and Wilson's participants claimed to prefer the detergent Tide because, e.g. their mother used Tide, when in fact their preference was driven by priming effects. For these reasons, we elicited our paraphrases in a neutral context (i.e. without confronting participants about their errors), in order to minimize the chances of post hoc rationalizations.

Experiments 1, 2a, and 2b reveal a GOG effect that cannot be explained by a subkind interpretation of the assertions, or by a domain-restricted interpretation. However, the results may be trivially attributed to a lack of knowledge of the gender discrepancies in the properties expressed by the minority characteristic predications. Perhaps participants were unaware of the gender discrepancies in the minority characteristic statements. To investigate this issue, Experiment 3 tested participants' knowledge of gender discrepancies for the minority characteristic items – e.g., whether people know that male horses do not give live birth. Experiment 3 also tested whether the GOG effect would be eliminated if people were forced to consider relevant knowledge.

Experiment 3: minimizing the generic overgeneralization effect

Participants completed the same evaluation task as in Experiment 1, but were also given a knowledge test in which they judged the truth values of false gendered minority characteristic statements such as "male ducks lay eggs" or "female lions have manes." The study manipulated whether the evaluation test was presented before or after the knowledge task. When the evaluation task was presented first, we expected to replicate the results of Experiment 1, with the additional manipulation check that participants were expected to reject statements such as "male ducks lay eggs" in the subsequent knowledge test. Otherwise, we could attribute performance in Experiment 1 to participants' lack of knowledge of the gender discrepancies in question. When the knowledge test was presented before the evaluation task, then participants may no longer agree with statements such as "all ducks lay eggs" because they would have recently rejected statements such as "male ducks lay eggs", and so would have counterexamples to the universal statements in mind (Johnson-Laird & Hasson, 2003). If participants demonstrated knowledge of the gender discrepancies but nevertheless showed the GOG effect, then this would suggest that the effect is robust and difficult to eliminate.

For the evaluation task, we asked participants from the same population as in Experiment 1 to evaluate the truth values of the eighteen different kinds of assertions. We employed a 6 (predicate type) × 3 (statement type) × 2 (task order) design, and we presented a total of 72 counterbalanced target stimuli. These same participants were given a knowledge test, in which they were given counterexamples to test their knowledge of the relevant gender discrepancies, either before or after the evaluation task.

Method

Participants and procedure

40 volunteers from the same population as Experiment 1 participated over the Internet for pay. All were native English speakers and none had participated in experiments on generics in the past. Participants judged the truth of each item, one at a time, following the same procedure as in Experiment 1.

Materials and design

Experiment 3 used the same set of 72 items used in Experiment 1, examples of which are given in Table 1. We used a blocked design. In one block, participants were presented with a knowledge test of 55 statements, 12 of which were gendered versions of the minority characteristic generic statements used in Experiment 1. The statements were false, and participants were expected to reject them. Table 3 lists the gendered minority characteristic statements used in the study. The rest of the 55 statements were designed to ensure that there were approximately equal numbers of true and false items, and more importantly that participants did not just employ the strategy of rejecting all the gendered statements out of hand. For example, these items included true gendered statements such as "female zebras have stripes".

In the other block, participants evaluated the items as they had in Experiment 1. Half of the participants received the knowledge test before the evaluation task and the other half received it after. The filler statements that were used in Experiment 1 were omitted so as to maximize the likelihood that participants would remember their judgments about the gender specificity of the properties when presented with the knowledge test first. Dropping the filler items thus maximized the likelihood that participants would remember that male ducks do not lay eggs, and would reject "all ducks lay eggs".

Results and discussion

Participants were aware of the relevant gender discrepancies in minority characteristic predications, except for one item: 73% of participants believed that female cardinals are red. Accordingly, this item was dropped from the analyses of all four experiments. The remaining false gendered minority characteristic items were rejected 84% of the time, and levels of rejection did not vary by whether the knowledge task was presented before or after the evaluation task, $t(38) = 1.00$, $p = .32$, $d = .32$. Proportions of agreement to each of the false gendered minority characteristic statements are given in Table 3. We also included

Table 3
Percentage agreement to false gendered minority characteristic statements used in Experiment 3.

Predication type	Example	Proportion of agreement	Agreement rate
Minority characteristic	Male sheep produce milk	3	Low
Minority characteristic	Male snakes lay eggs	3	Low
Minority characteristic	Male horses give live birth	8	Low
Minority characteristic	Male ducks lay eggs	8	Low
Minority characteristic	Male pigs suckle their young	8	Low
Minority characteristic	Male insects lay eggs	18	High
Minority characteristic	Female lions have manes	20	High
Minority characteristic	Female deer have antlers	23	High
Minority characteristic	Female moose have antlers	25	High
Minority characteristic	Female goats have horns	33	High
Minority characteristic	Male kangaroos have pouches	33	High
Minority characteristic	Female cardinals are red[a]	73	–

[a] *Note*: participants erroneously agreed with this item 73% of the time, and so the corresponding ungendered minority characteristic statement, "cardinals are red", was dropped from the analyses of Experiments 1, 2a, 2b, 3, and 4.

items that participants could have deemed gender specific (e.g., "female zebras have stripes"), but we found that they were aware of the lack of gender specificity in those items. These data suggest that the results in Experiment 1 cannot be attributed to mere ignorance on the part of our participants.

The overall agreement rates for the evaluation task were comparable to those found in Experiment 1. Existentially quantified statements were judged as true most of the time (91%). Just as in Experiment 1, quasi-definitional existentials exhibited a scalar implicature effect (76%), i.e., a relatively low agreement level because of the implication that "some" is taken to mean "not all" (Bott & Noveck, 2004). Statements in generic form were judged true 77% of the time, and false generalizations were accepted only 33% of time. Universal statements were accepted 40% of the time. We replicated the GOG effect: minority characteristic universals were judged true 41% of the time. Participants agreed with minority characteristic universals 50% of the time when they were presented before the knowledge test, and this agreement rate dropped to 32% when presented after the knowledge test. Even when participants had recently rejected statements such as "males ducks lay eggs", they still showed the overgeneralization effect – by agreeing with statements such as "all ducks lay eggs" – almost a third of the time. The effect was reduced but not eliminated.

The proportion of agreement data were subjected to a $6 \times 3 \times 2$ mixed measures ANOVA. As in Experiment 1, we found a main effect of predicate type, $F_1(5,190) = 104.02$, $p < .0001$, $F_2(5,65) = 60.75$, $p < .0001$, min $F'(5,146) = 38.35$, $p < .0001$, reflecting the higher agreement rate to quasi-definitional (86%), majority characteristic (88%), and minority characteristic statements (76%) than to majority (61%), striking (58%), and false generalization statements (46%). There was also a main effect of statement type, $F_1(2,76) = 218.19$, $p < .0001$, $F_2(2,130) = 367.03$, $p < .0001$, min $F'(2,160) = 136.84$, $p < .0001$, reflecting the higher agreement for 'some' statements (91%) than for generic statements (77%), with the lowest agreement for 'all' statements (40%), and a significant interaction between predicate type and statement type, $F_1(10,380) = 59.89$, $p < .0001$, $F_2(10,130) = 50.75$, $p < .0001$, min $F'(10,356) = 27.47$, $p < .0001$, again replicating the results in Experiment 1. Despite the reduction of the overgeneralization effect when universal statements were presented after the knowledge test, the ANOVA did not yield a significant effect of task order. Finally, there was no significant three-way interaction, $F_1(10,380) = 1.71$, $p = .08$, $F_2(10,130) = 2.29$, $p < .05$, min-$F'(10,441) = .98$, $p = .46$.

Fig. 2 displays performance for universal statements as a function of predicate type and task order. Planned comparisons showed that indeed, minority characteristic universals were accepted less often when evaluated after the knowledge test as compared to before (32% vs. 50%), $t(38) = -2.05$, $p < .05$, $d = -.65$.

Conversely, majority characteristic universals were accepted more often when evaluated after (90%) vs. before (70%) the knowledge test, $t(38) = 2.95$, $p < .01$, $d = .93$. We interpret this result as a sensitizing effect where the knowledge test, in combination with the subsequent evaluation task, alerted participants to the difference between minority characteristic and majority characteristic items. Majority characteristic universals may have been accepted more often to potentially compensate for the reduction in agreement to minority characteristic universals. No other simple effects comparisons between task orders were reliable.

Planned comparisons between predicate types (collapsed over the two test orders) replicated previous findings, and revealed that minority characteristic assertions were accepted in universal form more than false generalizations, $t(78) = 6.31$, $p < .0001$, $d = 1.41$, whereas neither majority $t(78) = .95$, $p = .34$, $d = .21$, nor striking assertions were accepted reliably more than false generalizations, $t(78) = .48$, $p = .63$, $d = .11$. Importantly, even when the knowledge phase was presented first, minority characteristic predictions were accepted more frequently in universal form than false generalizations, $t(38) = 3.26$, $p < .005$, $d = 1.03$.

As Table 3 shows, false gendered minority characteristic items were rejected most of the time. However, some items were accepted as much as 33% of the time, and so

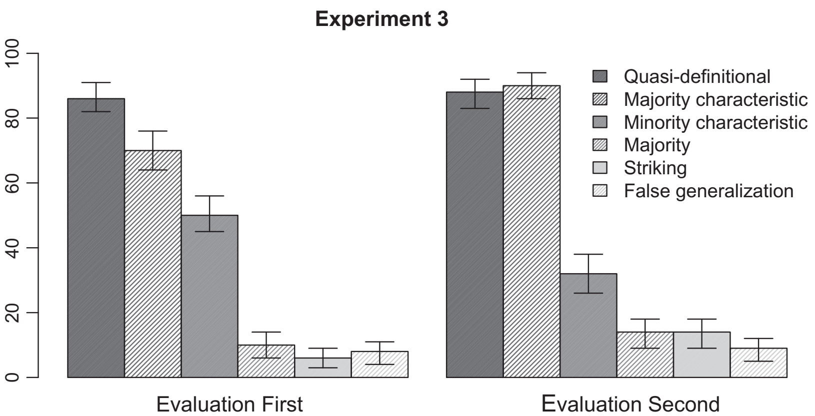

Fig. 2. Proportions of agreement to universals for Experiment 3 as a function of predicate type, separated by whether the evaluation task came first (before the knowledge test) or second (after the knowledge test); 95% confidence intervals shown.

we examined whether the GOG effect was reduced as a result of participants' lack of knowledge about specific items. Five of the 11 remaining minority characteristic items had ignorance rates of less than 10% (with an average of 6%) – i.e. the false gendered items (e.g. "male ducks lay eggs") were accepted by fewer than 10% of participants. The other six items had ignorance rates greater than 10%, with an average of 25% of participants failing to reject the false gendered item. If the GOG effect had been driven by ignorance, then the effect would have been much greater for the second set of items, since participants were less likely to know about the gender differentiations. However, the items for which the ignorance rate was less than 10% yielded a substantial GOG effect (44% of trials), as did those with an ignorance rate greater than 10% (38% of trials). The GOG effects for the two subsets of items did not differ, $t(78) = .70$, $p = .49$, $d = .16$.

Between the knowledge test and the evaluation of minority characteristic universals, there are four possible patterns of response that a given participant could provide for a given minority characteristic item:

(1) A GOG effect response, in which the participant correctly rejected the false gendered minority characteristic assertion, but nonetheless accepted the corresponding minority characteristic universal (e.g. s/he rejected "male ducks lay eggs" but accepts "all ducks lay eggs").
(2) A correct response, in which the participant rejected both the false gendered minority characteristic assertion as well as the minority characteristic universal (e.g. s/he rejected both "male ducks lay eggs" and "all ducks lay eggs").
(3) An 'ignorant' response, in which the participant accepted the false gendered minority characteristic assertion and also accepted the minority characteristic universal (e.g. s/he accepted both "male ducks lay eggs" and "all ducks lay eggs").
(4) A 'noise' response, in which the participant accepted the false gendered minority characteristic but rejected the universal (e.g. s/he accepted "male ducks lay eggs" but rejected "all ducks lay eggs").

Table 4 shows the distribution of the four types of responses in Experiment 3 as a function of the task order. When the evaluation task came before the knowledge test, participants were equally likely to provide GOG responses (40% of trials) as they were to produce correct responses (44% of trials), $t(38) = .49$, $p = .63$, $d = .15$. Participants were significantly more likely to produce GOG responses than ignorant responses (10% of trials), $t(38) = 3.36$, $p < .005$, $d = 1.09$, or noisy responses (6% of trials), $t(38) = 4.53$, $p < .0001$, $d = 1.43$. Further, 17 out of the 20 participants produced more GOG responses than ignorant or noisy responses (binomial test, $p < .0001$, assuming a prior probability of 1/3).

Table 4
Percentages of the four possible types of responses in Experiment 3 as a function of the task order.

Task order	Type of response			
	GOG effect	Correct	Ignorance	Noise
Evaluation first	40	44	10	6
Evaluation second	19	65	13	3

When the evaluation task came after the knowledge task, participants made more correct responses (65%) than GOG responses (19%), $t(38) = 5.61$, $p < .0001$, $d = 1.77$. These two types of responses (GOG vs. correct) were entered into a Chi-square test by task order, and the test revealed a significant interaction, $\chi^2 = 11.52$, $p < .001$, reflecting a higher rate of correct responses when the evaluation phase was proceeded by the knowledge test. However, despite the reduced proportion of GOG responses when the evaluation task came second, participants still made more GOG responses than noisy responses (3% of trials), $\chi^2 = 11.63$, $p < .0001$, suggesting that the GOG effect was not due to random responding, even when participants completed the knowledge test first.

Experiment 3 showed that the GOG effect is not entirely due to ignorance. Participants were generally aware of the relevant gender distinctions, which eliminates the possibility that overgeneralization effects arise simply out of ignorance. When the evaluation block came first, participants produced GOG responses on 40% of trials – that is, they accepted "all ducks lay eggs" despite knowing that male ducks do not lay eggs. Thus, the effect found in Experiment 1 was not due to ignorance.

Participants also sometimes committed the error even when primed with information about the members of a kind who do not possess the predicate in question ('male ducks' and 'female lions' for the examples above). Participants who rejected false gendered statements such as "male ducks lay eggs" nevertheless subsequently agreed with the minority characteristic universal statement "all ducks lay eggs" on 19% of trials. The reduced GOG effect reflects that participants recalled the information from the first half of the task. It is likely that participants were made more sensitive to whether information was gender specific, since the knowledge portion highlighted this. They may have proceeded more cautiously in the second half of the task, especially when it came to the minority characteristic statements. It is thus not surprising that the GOG effect was reduced; heightened sensitivity to gender-specific properties likely lessened the extent to which participants relied on their judgments of generics to evaluate universals, since they would have been more alert to the possibility of error. What is more remarkable is that the GOG effect persisted nonetheless on almost one fifth of trials. This experiment thus suggests that the GOG effect is difficult to eliminate entirely.

The overgeneralization effect persisted in Experiment 3 despite the availability of information that, in principle, should prevent the error. Would the effect still persist even when a correct alternative is presented to participants? People might have agreed with statements such as "all ducks lay eggs" because such statements were presented in isolation. Experiment 4 investigated whether participants would still agree with the above statement when they were explicitly provided with correct alternatives such as "only some ducks lay eggs" or "some ducks don't lay eggs."

Experiment 4: the availability of correct alternatives

Would people continue to show the overgeneralization error in spite of being presented with correct alternatives? Experiment 4 addressed this question by allowing participants to judge two competing statements on whether they agreed with one over the other. The pairs of statements differed only in the quantifiers used to express them. Half the participants were asked to compare universals (e.g., "all Xs are Ys") with affirmative existentials (e.g., "only some Xs are Ys"), and other half compared universals with negative existentials (e.g., "some Xs are not Ys"). Both formulations of the existential provided correct alternatives to the false universals, and so at least in the case of majority characteristic, minority characteristic, majority, striking, and false generalization predications, participants could choose a correct answer by selecting the existential. Participants who express a preference for universals over affirmative existentials (e.g., "only some ducks lay eggs") or negative existentials (e.g., "some ducks don't lay eggs") would show the overgeneralization effect. Given the results of our previous experiments, we expected participants to do so for majority characteristic and minority characteristic predications.

Method

Participants

Forty volunteers from the same population as Experiments 1 and 2 participated over the Internet for pay. All were native English speakers and none had participated in experiments on generics in the past.

Materials and design

Experiment 4 used the same set of 72 items used in Experiments 1 and 2, examples of which are given in Table 1. On each trial, participants saw two versions of each item: a universal version and an existential version. Half the participants received an affirmative existential form ("only some Xs are Ys") and the other half received a negative existential form ("some Xs are not Ys"). This generated a 6 (predicate type) $\times$ 2 (existential polarity) design, where the polarity of the existential was a between-participants manipulation.

Procedure

Participants were asked to evaluate each pair of items over the Internet through an interface written in Ajax. The interface consisted of two locations on a horizontal plane, which were labeled 'A' and 'B'. The pairs of target stimuli appeared in these locations, and the order of their appearance was randomized such that half the time the universals appeared in slot 'A' and half the time they appeared in slot 'B'. Thus, the interface layout appeared as follows:

All Xs are Ys	Only some Xs are Ys
A	B

Participants indicated whether they agreed with one of the items over the other by responding on a 6-point Likert scale ranging from 6 ("I agree with A more") to 1 ("I agree with B more").

Results and discussion

Responses on the Likert scale were coded so that a mean score of 6 indicates that participants agreed with universals over existentials, a score of 3.5 indicates no preference, and a score of 1 indicates that participants agreed with existentials over universals for that particular type of predication. Quasi-definitional statements were given a mean rating of 5.01, majority characteristic 4.61, minority characteristic 3.56, majority 1.94, striking 1.98, and false generalizations 1.74. The means of participants' responses, separated by the polarity of the existential, are given in Table 5.

These differences were assessed via a 6 $\times$ 2 mixed ANOVA, which yielded a significant main effect of predication type, $F_1(5,190) = 141.64$, $p < .0001$, $F_2(5, 65) = 62.50$, $p < .0001$, min $F'(5, 127) = 43.36$, $p < .0001$, but no main effect of polarity and no interaction.

Planned comparisons showed that minority characteristic universals ($M = 3.56$) were preferred more often than majority ($M = 1.94$, $t(78) = 8.6$, $p < .0001$, $d = 1.92$), striking ($M = 1.98$, $t(78) = 8.57$, $p < .0001$, $d = 1.92$), and false generalization universal assertions ($M = 1.74$, $t(78) = 10.18$, $p < .0001$, $d = 2.28$). These data indicate a definite tendency to prefer minority characteristic universals to minority characteristic existentials, as compared to striking, majority, or false generalization universal–existential pairs. The tendency is a dramatic illustration of the GOG effect, as participants need only recognize that, for instance, the population of male ducks satisfies the existential statement "only some ducks lay eggs" and "some ducks don't lay eggs" and thus falsifies the universal statement "all ducks lay eggs."

Use of a Likert scale in this experiment also allowed us to test how confident participants were in their judgments when they displayed a GOG effect. The previous studies used a forced-choice paradigm, so it is possible that a GOG rate of approximately 50% 'yes' responses indicated that participants were uncertain and responding at chance. However, had they relied on their judgment of the generic to evaluate the universal, as per the generics-as-default hypothesis, they should have been confident in their answers. Although the mean preference for minority characteristic universals was 3.56 in this study, participants' confidence ratings were significantly bimodal (Hartigan's dip test, $p < .001$), i.e., their responses fell nearer to the end-points of the scale rather than the mid-point. In other words, participants who preferred the minority characteristic universal did so with confidence. Experiment 4 demonstrated that even when correct

Table 5
Mean ratings of agreement (and standard deviations) to the universal (6) or the existential (1) formulations of assertions as a function of predication type and polarity of the existential in Experiment 4.

Predication type	Polarity of the existential statements	
	Affirmative	Negative
Quasi-definitional	4.95 (1.92)	5.16 (1.61)
Majority characteristic	4.70 (1.93)	4.51 (1.87)
Minority characteristic	3.77 (2.23)	3.35 (2.07)
Majority	2.03 (1.62)	1.85 (1.54)
Striking	1.93 (1.54)	2.03 (1.59)
False generalizations	1.83 (1.63)	1.65 (2.12)

alternatives were available at the time of choice, participants showed the GOG effect.

General discussion

The generic overgeneralization (GOG) effect is a robust phenomenon. Participants agreed to statements such as "all ducks lay eggs" a substantial portion of the time (Experiment 1), despite knowing that male ducks do not lay eggs (Experiment 3). Participants endorsed these universals despite judging only minutes earlier that only half the kind has the property (Experiment 3). Participants also frequently failed to reject these false universal statements when they were explicitly provided with true alternatives such as "some ducks do not lay eggs" (Experiment 4). The effect is not likely to be attributable to participants' employing a subtyping interpretation on which "all ducks lay eggs" would be understood to mean *all kinds of ducks lay eggs*, nor is it likely attributable to participants' contextually restricting the scope of the universal quantifier so that it only applies to, e.g., female ducks. Experiment 2a showed a GOG effect after participants were primed with population information ("there are 431 million ducks in the world") designed to encourage an individual-based interpretation of the statements, and to encourage participants to understand the quantifier to range over every single duck in the world (Stanley & Szabo, 2000). While the GOG effect occurred at a lower rate in this context, it still occurred on a substantial portion of trials. The lower rate may have been due to eliminating any subkind and restricted interpretations, or alternatively the numerical population information may have put participants in a more careful and analytic frame of mind (Alter et al., 2007). Further, Experiment 2b found very little evidence of subtyping interpretations or of domain-restricted interpretations when participants provided paraphrases of the statements they had evaluated. Importantly, there was no reliable correlation between subtyping or restricted paraphrases and acceptance of universal statements, which would be expected if the GOG effect were due to participants employing subtyping interpretations. The findings reported here are instead better explained by the hypothesis that the GOG effect – or at least a substantial portion of the observed effect – was due to participants relying on their judgments of the corresponding (true) generic, and so failing to properly evaluate the universal in question.

If this is so, then the effect that Hollander et al. (2002) and Tardif et al. (in press) identified in preschoolers persists into adulthood, as the generics-as-default hypothesis would predict. The hypothesis that generics express primitive, default generalizations explains a range of findings including acquisition data, cross-linguistic syntactic patterns, and errors in reasoning tasks. For example, if generics express children's default generalizations, then this would explain why generics are acquired early despite their semantic complexity (Carlson & Pelletier, 1995; Gelman, 2003; Gelman, 2010; Gelman & Brandone, 2010; Gelman & Raman, 2003; Graham et al., in press; Leslie, 2007; Leslie, 2008). It would also explain why generic forms are always less syntactically marked than quantificational ones (Chomsky, 2000; Dahl, 1985; Krifka et al., 1995; Leslie, 2008).

Importantly, the generics-as-default hypothesis also predicts that quantified statements will sometimes be incorrectly interpreted as generics. This prediction, if correct, would explain Jönsson and Hampton's (2006) inverse conjunction fallacy, Leslie and Gelman's (submitted for publication) finding that quantified statements are frequently recalled as generics, and Hollander et al.'s (2002) and Tardif et al.'s (in press) finding concerning preschoolers' interpretations of "all" and "some." The studies presented in this paper represent a further test of the predictions of the generics-as-default hypothesis. Consistent with the hypothesis, we found that adults show a robust tendency to evaluate some universal statements as though they were generics. Participants did so across a range of contexts, including one in which they were presented with population information concerning the kind (Experiment 2), one in which they were first primed to consider the counterexamples (Experiment 3) and one in which they were presented with correct alternatives to the universal (Experiment 4).

Interestingly, our findings appear to be limited to predications that involve properties that are characteristic of the kind. Our focus in these experiments was on minority characteristic universals such as "all ducks lay eggs," since these universals offered the clearest illustration of the GOG effect. However, participants also accepted majority characteristic universals such as "all tigers have stripes", despite the fact that these predications are not strictly true in universal form. After all, there are stripeless tigers (as Siegfried and Roy's performances attest). Our experiments did not control for whether or not people are aware that some tigers lack stripes, so we cannot be sure that these statements were not accepted out of simple ignorance. However, Jönsson and Hampton's (2006) studies suggests that something more interesting than ignorance may be driving these results. Forty-six percent of the time their participants judged that a more inclusive universal such as "all ravens are black" was *more* likely to be true than a narrower one such as "all young jungle ravens are black" – even when the two statements were presented side by side for comparison (Jönsson & Hampton, 2006, Study 5). Either this constitutes a clear failure to appreciate the logical features of universal statements, or else their participants were interpreting these universals as generics (Jönsson, personal communication). It should be further noted here than nearly all of Jönsson and Hampton's items would count as majority characteristic predications according to our criteria, and so our participants' willingness to accept majority characteristic universals could potentially be due to their interpreting them as generics also. Similarly, Leslie and Gelman's (submitted for publication) memory task – in which adults often recalled universal statements as generics – exclusively used items that would count as majority characteristic by our criteria.

Jönsson and Hampton's results, together with Leslie and Gelman's, suggest that majority characteristic universals may be interpreted as generics, and our results suggest that this is true of minority characteristic universals as well. We did not find that majority or striking universals were accepted, however – in our studies, the GOG effect was limited to minority characteristic and possibly majority characteristic items. These findings support the hypothesis that the GOG effect is limited to cases where a characteristic property is predicated of the kind.

Part of the explanation here may be that characteristic properties are usually held by all the normal members of a kind, and so GOG effects concerning such properties may often constitute relatively minor errors. Thus, adults could save cognitive effort by relying on the default generic interpretation to evaluate the universal, without risking significant error. The error only becomes significant in the case of these gender-specific properties. For non-characteristic properties, however, the risk of error is more prominent, and so adults may have learned not to rely on their judgments of generics for such items. One outstanding question is whether there are developmental differences in the extent of the GOG effect – e.g. might young children show GOG effects even if the property is not characteristic? Hollander et al.'s (2002) results suggest that the GOG effect may potentially be broader in scope for young children. However, they did not use a wide enough range of items to support definite conclusions. Future research is needed to address this question, and thereby shed more light on why adults only exhibit this effect when the property is characteristic.

In sum, the generic overgeneralization effect – the tendency to judge universally quantified statements to be true when the corresponding generic is true – occurs for minority characteristic statements like "all ducks lay eggs". These universally quantified statements were accepted by participants about half the time, despite their knowing that only female ducks lay eggs. These findings are plausibly due to participants having a tendency to treat universal statements as though they were generics, as one would predict on Leslie's (2008) and Gelman's (2010) hypothesis that generics express primitive, default generalizations (see also Csibra & Gergely, 2009; Gelman & Brandone, 2010; Hollander et al., 2009). This explanation, if correct, has the potential to apply to a range of disparate phenomena, such as Hollander et al.'s (2002) developmental findings, Jönsson and Hampton's (2006) Inverse Conjunction Fallacy, and Leslie and Gelman's (submitted for publication) memory task. These diverse findings may have a common underlying explanation: people have a tendency to treat universally quantified statements as though they were generics. This tendency is what one would expect if generics involve more basic generalizations, while universally quantified statements involve more sophisticated, non-default ones.

Acknowledgments

We are grateful to Paula Rubio Fernandez, Axit Fumero, Susan Gelman, Geoffrey Goodwin, Phil Johnson-Laird, Mark Johnston, Alan M. Leslie, Louis Lee, Ray Nickerson, Sandeep Prasada, and Carlos Santamaria for their helpful comments. These studies were supported in part by a NSF Graduate Research Fellowship awarded to the second author.

Appendix

Materials for Experiments 1, 2, and 3 in generic form.

Predication type	Item (in generic form)	Predication type	Item (in generic form)
Quasi-definitional	Ants are insects	Majority	Barns are red
Quasi-definitional	Bachelors are unmarried	Majority	Bars are noisy
Quasi-definitional	Cats are animals	Majority	Bikers have tattoos
Quasi-definitional	Dogs are mammals	Majority	Cars have radios
Quasi-definitional	Elms are trees	Majority	Clocks are round
Quasi-definitional	Mushrooms are fungi	Majority	Italians eat spaghetti
Quasi-definitional	Poodles are dogs	Majority	Jackets have zippers
Quasi-definitional	Preschoolers cannot vote	Majority	Pigeons sit on statues
Quasi-definitional	Rectangles are geometric figures	Majority	Pigs are kept in pens
Quasi-definitional	Sows are pigs	Majority	Shoes have laces
Quasi-definitional	US Presidents are over 35	Majority	Squirrels live in parks
Quasi-definitional	Vixens are foxes	Majority	Subways are crowded
Majority characteristic	Bees have wings	Striking	Birds carry avian flu
Majority characteristic	Cats have whiskers	Striking	Coyotes kill pets
Majority characteristic	Cheetahs run fast	Striking	Insects carry viruses
Majority characteristic	Cows eat grass	Striking	Lions eat people
Majority characteristic	Dogs have tails	Striking	Mosquitoes carry malaria
Majority characteristic	Horses have four legs	Striking	Pit bulls maul people
Majority characteristic	Kangaroos hop	Striking	Rats carry disease
Majority characteristic	Leopards have spots	Striking	Rottweilers maul children
Majority characteristic	Pencils are wooden	Striking	Sharks attack swimmers
Majority characteristic	Ravens are black	Striking	Strokes cause paralysis
Majority characteristic	Sparrows have wings	Striking	Ticks carry Lyme disease
Majority characteristic	Tigers have stripes	Striking	Tigers attack people
Minority characteristic	Cardinals are red[a]	False generalization	Animals are mammals
Minority characteristic	Deer have antlers	False generalization	Athletes are students

(*continued on next page*)

Appendix (*continued*)

Predication type	Item (in generic form)	Predication type	Item (in generic form)
Minority characteristic	Ducks lay eggs	False generalization	Bees are worker bees
Minority characteristic	Goats have horns	False generalization	Books are paperbacks
Minority characteristic	Horses give live birth	False generalization	Canadians are right-handed
Minority characteristic	Insects lay eggs	False generalization	Computers are pcs
Minority characteristic	Kangaroos have pouches	False generalization	Ducks are female
Minority characteristic	Lions have manes	False generalization	Engineers are male
Minority characteristic	Moose have antlers	False generalization	Humans are over three years old
Minority characteristic	Pigs suckle their young	False generalization	Lions are male
Minority characteristic	Sheep produce milk	False generalization	Teachers are female
Minority characteristic	Snakes lay eggs	False generalization	Trees are deciduous trees

[a] This item was dropped from the analyses reported in the paper (see Table 3).

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*, 569–576.

Baldwin, D. A., Markman, E., & Melartin, R. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development, 64*, 711–728.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437–457.

Carlson, G. (1977). Reference to kinds in English. PhD dissertation, University of Massachusetts, Amherst.

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

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

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University 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.

Cimpian, A., Gelman, S. A., & Brandone, A. C. (2010). Theory-based considerations influence the interpretation of generic sentences. *Language and Cognitive Processes, 25*(2), 261–276.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Cohen, A. (1996). Think generic: The meaning and use of generic sentences. PhD dissertation, Carnegie Mellon University.

Connolly, A., Fodor, J., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition, 103*, 1–22.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.

Dahl, O. (1985). *Tense and aspect systems*. Oxford: Blackwell.

Declerk, R. (1986). The manifold interpretations of generic sentences. *Lingua, 68*, 149–188.

Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press.

Gelman, S. A. (2010). Generics as a window onto young children's concepts. In F. J. Pelletier (Ed.), *Kinds, things, and stuff: The cognitive side of generics and mass terms. New directions in cognitive science* (Vol. 12). New York: Oxford University Press.

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

Gelman, S. A., & Brandone, A. C. (2010). Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development, 6*, 223–240.

Gelman, S. A., Coley, J., Rosengren, K., Hartman, E., & Pappas, A. (1998). *Beyond labeling: The role of parental input in the acquisition of richly structured categories* (Vol. 63, No. 1). Monographs of the Society for Research in Child Development (SRCD), Series no 235. Ann Arbor, MI: SRCD.

Gelman, S. A., Goetz, P. J., Sarnecka, B. S., & Flukes, J. (2008). Generic language in parent–child conversations. *Language Learning and Development, 4*, 1–31.

Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. In D. Medin (Ed.). *Psychology of learning and motivation* (Vol. 39, pp. 201–263). New York: Academic Press.

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

Gelman, S. A., Star, J. R., & Flukes, J. E. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development, 3*, 179–199.

Gelman, S. A., & Tardif, T. (1998). A cross-linguistic comparison of generic noun phrases in English and Mandarin. *Cognition, 66*, 215–248.

Geurts, B. (1985). Generics. *Journal of Semantics, 4*.

Goldin-Meadow, S., Gelman, S. A., & Mylander, C. (2005). Expressing generic concepts without a language model. *Cognition, 96*, 109–126.

Graham, S. A., Nayer, S. L., & Gelman, S. A. (in press). Two-year-olds use the generic/non-generic distinction to guide their inferences about novel kinds. *Child Development.*

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2001). Words and shape similarity guide 13-month-olds' inferences about nonobvious object properties. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty third annual conference of the Cognitive Science Society* (pp. 352–357). Hillsdale, NJ: Erlbaum.

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

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology, 38*, 6.

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory &Cognition, 31*, 1105–1113.

Jönsson, M. L., & Hampton, J. A. (2006). The inverse conjunction fallacy. *Journal of Memory and Language, 55*, 317–334.

Keates, J., & Graham, S. A. (2008). Category labels or attributes: Why do labels guide infants' inductive inferences? *Psychological Science, 19*, 1287–1293.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (2009). Generics, prevalence, and default inferences. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S., Leslie, S. J., Glucksberg, S., & Rubio-Fernandez, P. (2007). Do ducks lay eggs? How humans interpret generic assertions. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society*. Nashville, TN: Cognitive Science Society.

Krifka, M., Pelletier, F. J., Carlson, G., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: An introduction. In G. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 1–125). Chicago: Chicago University Press.

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

Leslie, S. J. (2007). Generics and the structure of the mind. *Philosophical Perspectives, 21*, 375–405.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *The Philosophical Review, 117*(1), 1–49.

Leslie, S. J. & Gelman, S. A. (submitted for publication). Quantified statements are recalled as generics: Evidence from preschool children and adults.

Leslie, S. J., Khemlani, S., Prasada, S., & Glucksberg, S. (2009). Conceptual and linguistic distinctions between singular and plural generics. *Proceedings of the 31st annual cognitive science society*. Amsterdam: Cognitive Science Society.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.

Papafragou, A., & Schwarz, N. (2005/2006). Most wanted. *Language Acquisition, 13*, 207–251.

Pappas, A., & Gelman, S. A. (1998). Generic noun phrases in mother–child conversations. *Journal of Child Language, 25*, 19–33.

Pelletier, F. J., & Asher, N. (1997). Generics and defaults. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 1125–1179). Cambridge, MA: MIT Press.

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

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense perception. *Cognition, 33*(3), 401–448.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale: Lawrence Erlbaum Associates.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35*, 1–33.

Stanley, J. (2007). *Language in context: Selected essays*. Oxford: Oxford University Press.

Stanley, J., & Szabo, Z. (2000). On quantifier domain restriction. *Mind and Language, 15*(2 and 3), 219–261.

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (in press). Acquisition of generic noun phrases in Chinese: Learning about lions without an "-s". *Journal of Child Language.*