Genericity, exceptions and domain restriction: experimental evidence from comparison with universals

Dimitra LAZARIDOU-CHATZIGOGA – Queen Mary, University of London
Linnaea STOCKALL – Queen Mary, University of London

Abstract. Generic statements are characteristically associated with two features that distinguish them from universally quantified statements: first, they are tolerant of exceptions, and, second, they are not associated with any overt quantifier or determiner. We present data from a timed Truth Value Judgement Task (TVJT) that investigates the consequences of these two features for processing. We discuss these results in the context of recent proposals that generic interpretations are a more 'default' or 'basic' kind of interpretation than universal quantification and argue that our results do not support these proposals.

Keywords: domain restriction, exceptions, genericity, universal quantification

1. Introduction

While linguists have studied genericity for decades, it is still not clear what the precise meanings of generic statements (henceforth GSs) are, nor how those meanings are generated. GSs are characteristic for making reference to kinds rather than individuals and for being tolerant of exceptions (Carlson 1977, Gelman 2003, Krifka et al. 1995 and Lawler 1973), in contrast to universally quantified statements (UQSs). Take for example the GSs below:

(1) Tigers have stripes.
(2) Cats have whiskers.

Both statements can be truthfully uttered even in the face of exceptions, such as the existence of stripeless tigers or cats with no whiskers. If we compare these with the UQSs in (3) and (4)(4) we see that the situation changes and that we cannot claim that the UQSs are true if there is even one tiger that doesn't have stripes or one cat that doesn't have whiskers.

(3) All tigers have stripes.
(4) All cats have whiskers.

The tolerance of exceptions by GSs is quite complicated, as we see by comparing (5) & (6). (5) seems to be true even though as few as 1% of mosquitoes actually carry the virus (Hayes et al. 2005), while (6) is not, even though considerably more than half of all books published are indeed paperback (Shaffer, 2012):

(5) Mosquitoes carry the West Nile virus.
(6) Books are paperbacks.

Clearly, GSs do not have straightforward truth and licensing conditions. Furthermore, if we compare GSs with quantified statements, we see that although the latter involve overt operators like the universal quantifier *all* in (3) and (4) or the existential quantifier *some* in (7) and the proportional quantifier *most* in (8) below, GSs are not tied to any overt operator.

(7) Some tigers have stripes.
(8) Most cats have whiskers.

In (1) and (2) a Bare Plural (BP from now on) is used to make a generic statement and, thus, we may treat this as a statement with no determiner. As we will discuss though later on, this is not the only option to express genericity.

Two issues seem to arise then, which form the generics puzzle:

a) GSs are robust to (some) exceptions, whereas UGSs are not.
b) The absence of an overt quantifier in the case of GS raises the question of where the generic interpretation comes from in the first place.

The issue of exceptions seems to be closely related to another characteristic of generics, which is the fact that generics resist contextual restriction. According to the received view (Krifka et al. 1995: 8) generics cannot be contextually restricted, while universals like *every* can:

(9) Context: There are lions and tigers in this cage.
a. Every lion is dangerous. (*Can* mean 'Every lion in this cage is dangerous')
b. Lions are dangerous. (*Cannot* mean 'Lions in this cage are dangerous')

Thus, while generics cannot undergo domain restriction, the domains of universals like *every* can be claimed to be contextually restricted by covert domain variables at LF in the sense of Stanley and Szabó (2000). According to this proposal, the nominal argument of *every*, in this case *lion,* would have a domain variable C, which would refer to a contextually salient property. The nominal argument of *every* would be thus a contextually salient set of lions, i.e. the lions in the cage referred to[1]. This is not a possible interpretation for the generic BP *lions* above.

Furthermore, generics are statements that express non-accidental properties (Dahl 1975), or *essential* properties of a kind (Gelman 2003) or properties that bear a *principled connection* to the kind (Prasada & Dillingham 2006). Thus, they are treated as 'eternal' truths in the sense that they do not depend on the specific context, in which they are uttered, but rather on the properties that determine a kind. On the contrary, from a logical point of view, the truth or falsity of a quantified statement has been linked to the set quantified over. In order to determine the truth/falsity of a quantified statement, one needs to look at the extension of the kind in question. The result can be then described in set-theoretic terms.

The truth- and licensing-conditions of generic interpretations are hotly debated in the formal

[1] This is only one possible way to derive domain restriction, but it suffices for the purposes of this discussion.

semantics literature: a spate of theories postulate modal operators in possible-world semantics (Pelletier & Asher 1997), non-monotonic inferences (Asher & Morreau 1995), relevant quantification (Schubert & Pelletier 1987), prototypicality (Nunberg & Pan 1975), stereotypicality (Geurts 1985) and probability of the information conveyed (Cohen 1996) as relevant factors. It is important to highlight that the theoretical impasse is at least partly due to the lack of reliable data, since usually the issues have been addressed through the researcher(s)' reflective intuition, which prohibits consensus even about the most fundamental facts about genericity. In the last few years some questions about genericity have been addressed with experimental paradigms drawn from experimental and developmental psychology (Leslie 2008, Leslie et al. 2011, Gelman 2010). This literature has focused on a fundamental puzzle: how do children acquire the meaning of GSs in the absence of dedicated words or morphemes that encode genericity?

In this paper we review the theoretical and experimental literature investigating these two issues and present data from a timed Truth Value Judgement Task (TVJT) experiment that addresses both features outlined above.

The structure of the paper is as follows: In section 2 we review the theoretical literature on generics and exceptionality mentioned in (a) above. In section 3 we discuss the learnability and processing issues that arise given (b) above and the recent work by Leslie and colleagues suggesting a solution. In section 4 we present our own experiment, which addresses weaknesses in the Leslie et al. 2011 experiments and incorporates the issue of exceptionality. In section 5 we present and discuss the results. In section 6 we provide a summary and conclude.

2. Generic interpretation and exceptions

In a recent article, Pelletier (2010) poses the question: "How many exceptions can a generic statement tolerate and still be true?" Based on different types of generics, we see that the percentage of exceptions might be bigger than just some *abnormal, weird* cases, ranging from individual, abnormal cases (10) to around 50% (12) and (13) and even to 99% (16):

(10) Tigers have tails.
(11) Telephone books are thick.
(12) Ducks lay eggs.
(13) Guppies give live birth.
(14) Italians are good skiers.
(15) Turtles live to an old age.
(16) Mosquitoes carry the West Nile virus.

Greenberg (2007) sheds some light on how to investigate this issue further. Greenberg (2007) claims that it is important to differentiate two types of exceptions: exceptional individuals and contextually irrelevant individuals. On the one hand, exceptional individuals would be non standard or abnormal with respect to some relevant aspect, i.e. legitimate exceptions to *dogs* *have four legs* are dogs that in addition to not having four legs have mutations, have undergone an accident, etc. On the other hand, contextually irrelevant individuals depend on the utterance context or are contributed by presuppositions, implicatures or real-world knowledge of the VP. For instance, in considering ‘Snakes lay eggs’, male individual snakes are considered irrelevant because the VP (‘lay eggs’) presupposes giving birth, which is only possible for females. This differentiation provides a promising way to look into the issue of exceptions, which is exploited in our experiment.

Even though the examples above involve Bare Blurals (BPs), GS are not expressed only with BPs. As Krifka et al. (1995: 8) argue, GS put no limitations on what kinds of NPs may occur in them. We can find different kinds of NPs in them, as we see below:

(17)
a. John drinks coffee.
b. My brother drinks coffee.
c. A teacher drinks coffee.
d. Every teacher drinks coffee.
e. Coffee is healthy.

The important observation here is that genericity is not encoded in a unique and unambiguous way by the use of exclusively generic forms (e.g. by a generic determiner or qunaitifer). This is not a particular characteristic of English. Generic meaning is not encoded as a ‘GEN’ quantifier systematically, either within one language or cross-linguistically. Rather, generic interpretation typically results from the interaction of a number of variable factors: the lexical semantics of the constituting elements, pragmatic knowledge, discourse situation, grammatical marking of determination and quantification on the NPs and syntactic position of the NPs. Thus, although in English genericity is typically expressed via a BP, in several other European languages such as Spanish or Greek, a definite plural NP is by far the most frequently used type of NP used in GS. In Greek (see Marmaridou-Protopapa 1984) a typical GS is the following:

(18) I rinokeri ine thilastika.
the rhinos are mammals.
‘Rhinos are mammals.’

Krifka et al (1995) propose a compositional semantic analysis of generic interpretations that posits a GEN operator, a dyadic modal operator which composes with a restrictor and a matrix as in (19), similarly to quantificational adverbs like *always* or *usually* (Lewis 1975). This generic operator has no phonological exponent, which provides a simple account for why a whole range of different sentence types can be associated with generic interpretations.

(19) Tigers have stripes.
Gen_x [tiger (x)] [have.stripes (x)]
restrictor matrix

However, this analysis does not address the issue of how listeners know that there is a generic operator in a sentence, or, especially, how children learning a language would come to posit such an operator. Leslie and colleagues, in a recent series of papers, have proposed an answer to this question that does not assume a generic operator.

They argue instead that the generic interpretation is the default interpretation, and is essentially available for free. Several theorists have proposed that generics express default generalizations (Csibra & Gergely 2009, Gelman 2010, Gelman & Brandone 2010, Hollander et al. 2002 and Leslie 2007, 2008) mainly based on developmental studies. Generics are said to be available first to children, while universal quantification is more complex, and acquired later. The empirical predictions these theorists make are that children and adults should sometimes incorrectly treat quantified statements as generics. Hollander et al. (2002) and Tardif et al. (2012) claim to confirm this prediction for children. This idea seems to also find support in a view of cognition set forth by Kahneman and colleagues (2002), according to which there are two systems of cognition: System 1, which is a fast, automatic, effortless lower-level system and System 2, a slower, more effortful higher-level system. Generics could be then part of System 1, as Leslie (2007) argues.

Leslie et al. (2011) present the following arguments for their generics-as-defaults hypothesis:
(a) There is no dedicated marker for genericity across languages.
(b) Adults incorrectly treat quantified statements as generics, as we discuss in detail below.
(c) A memory study by Leslie and Gelman (2012) claims that adults and preschoolers alike tend to recall quantified statements as generics.

In our experiment, we decided to address first the issues with adults. This may give us clues about the acquisition of generics by children as well.

3. Generics and universals: review of the experimental literature

Empirical data can shed some light into the hard-to-draw distinctions between universal quantification and genericity that we discussed in the introduction above. The few experiments that address the issue of the differences between UQSs and GSs focus on the difference between *all* and *generic* Bare Plurals:

(20) All tigers are striped.
⇒ true if every single tiger is striped, false in the face of exceptions
(21) Tigers are striped.
⇒ true despite possible exceptions

Leslie et al. (2011) compare generics with universals. If understanding quantified statements requires deviating from the default mode of generalization, then their prediction is that adults should sometimes fail to execute this deviation and so should incorrectly treat quantified statements as generics, when the corresponding generic is true. They then report experimental evidence that adults do sometimes judge universal statements as true, despite knowing that they are truth-conditionally false which they call the Generic Over Generalization (GOG) effect.

In their first experiment participants judged either generic or universal statements that involved the following types of properties: quasi-definitional (*triangles have three sides*), majority characteristic (*tigers have stripes*), minority characteristic (*ducks lay eggs*), majority non-characteristic (*cars have radios*), striking (*pit bulls maul children*), and false generalizations (*Canadians are right-handed*). They claim that the GOG effect occurs in approximately half the trials when the statement involves characteristic properties, that is, properties that bear a deep causal and explanatory relation to the kind in question, which may occur either in the majority of the members of the kind or only in a minority. The GOG effect is then attributed to the fact that generics constitute the default interpretation. Other possible explanations for the GOG effect that are considered, but later on discarded by Leslie et al., are the following:

(a) Subkind interpretation.
(b) Ignorance of the facts.
(c) Domain restriction.

They discarded (a) above after doing a paraphrase task, where the participants had to paraphrase the statements they had just read (their experiment 2b) and no evidence of subkind interpretation was found. They discarded (b) by distributing a knowledge test, which did not show ignorance of the relevant facts (their experiment 3). Our focus is the third possible explanation, which they addressed in experiment 2a. In order to check for the possibility of domain restriction in the sense of Stanley and Szabó (2000), as discussed above, they provided the participants with information about population estimates for the kind in question in the following form:

(22) ‘‘Suppose the following is true: there are 431 million ducks in the world. Do you agree with the following: all ducks lay eggs?’’

This information was supposed to prime interpretation over individual ducks and not over different kinds. The authors claim that the GOG effect still occurred on a substantial portion of trials, reporting an acceptance rate for *all* statements at 55%[2] for majority characteristic statements and 30% for minority characteristic statements.

Leslie et al.’s experiment 2a and their interpretation of the results are challenged by the following observations. In the first place, the contexts Leslie et al. use to induce specific/individual interpretations do not make salient the exceptions that would make the universal quantification over individuals interpretation untrue. Second, claims about universal quantifiers in general ought to be further refined. Languages have different types of universal quantifiers, so the question then arises: if we accept that *all* allows the GOG effect, do all different universal quantifiers allow it? *All* may or may not be interpreted with respect to the context given. A closer look at the linguistics literature reveals proposals for the non typical

2 Leslie et al. (2011) do not report the exact number, but this is an approximate value based on the graph they present.

behaviour of *all* leading to claims that *all* is not a quantifier (Brisson 2003) or that it has a maximizing effect acting as a slack regulator (Lasersohn 1999), both of which would predict that (22) could be judged as true. And, finally, 'all ducks lay eggs' *has* a generic interpretation in any theory - there is no 'overgeneralization' required to interpret this sentence as quantifying over duck kinds/species, merely a failure to use the context in the way Leslie et al. intended.

So our experiment addressed all these three issues, and furthermore collected timing data in order to develop a more nuanced perspective on the process of evaluating generic and quantified statements and making Truth Value Judgments.

4. The experiment

We addressed the issues of the hypothesized GOG effect and of the systematic exceptions in minority characteristic statements. Following Leslie et al. we used BPs as our generics. We added two additional types of UQSs, since we wanted to test whether the GOG effect occurred only with *all* or with universal quantifiers in general. The different types of universal quantifiers tested were *all, all the* and *every*[3]. Our predictions were that 'all the ducks lay eggs' would be judged as false, since it cannot be interpreted generically anyway, and that 'every duck lays eggs' would be interpreted as contextually restricted or as generic, but that the generic interpretation should be allowed much less than with 'all ducks lay eggs'. We designed a different way to prime domain restriction by presenting a specific context rather than providing information about the kind in general. We used contexts that make the systematic exceptions to minority characteristic generic statements salient and focused on sex differences in animals.

4.1 Method: participants and procedure

Twenty native-English speaking Queen Mary, University of London students participated for credit for a linguistics course. Participants sat before a computer screen. The software used was DMDX on a PC running Windows. Participants read a background context on three consecutive screens and then a statement on the next screen. A final screen at the end asked them whether the statement they read was true or not. Responses were recorded by button press. Participants were instructed to judge the statement with respect to the background they had just read. Statements involved either a minority characteristic or a majority characteristic property and one of the four determiner types.

4.2 Materials and design

The experimental items used form a subset of the items used by Leslie et al. (2011) in experiment 2a. Norming of the items through questionnaires distributed to a similar population

[3] We didn't include *each* for two reasons: a) in order for the set of variables to be easier to handle, given time and analysis restriction and b) because it has been argued that *each* cannot receive generic interpretation, while *every* is compatible with generic interpretation.

to the one tested in the experiment lead us to exclude 8 items (e.g. 'Cardinals are red' and 'Sheep produce milk') because our participants did not seem to be sure of the facts. Only items that were consistently judged to be true statements by 80% of our norming participants were included.

The two conditions we manipulated were:

a) the type of the characteristic statement: majority or minority characteristic statement
b) the determiner type: *generic, all, all the, every*

We used 8 statements of characteristic type, which gave us 64 unique trials. Each participant saw 32 items in total, 16 test items and 16 fillers. Within the 16 test items were 4 *generic N* (i.e. bare plurals), 4 *all N*, 4 *all the N* and 4 *every N*. Each participant saw 4 trials of each determiner type = 8 trials of each statement type = 2 trials of each determiner+statement type.

The two dependent measures were acceptance (Truth Value Judgment) and reaction time. Here are samples of the items used, a GS with a majority characteristic property (*tigers have stripes*) and a UQS with a minority characteristic property (*all ducks lay eggs*):

(23) SCREEN 1: [appears line by line]
BACKGROUND:
Linton zoo's viewing windows in some of the enclosures make
it possible to get as close as a whisker. It currently hosts
12 female tigers among many other animals.

SCREEN 2: [appears line by line]
STATEMENT:
Tigers have stripes.

SCREEN 3:
TRUE?
Press F for 'yes' and N for 'no'[4]

(24) SCREEN 1: [appears line by line]
BACKGROUND:
Clissold Park is a big community park in Stoke Newington. It
has two artificial lakes, which are managed as wildlife
refuges. There are 80 male ducks in both lakes.

SCREEN 2: [appears line by line]
STATEMENT:
All ducks lay eggs.

4 These buttons were selected due to their position on the keyboard.

SCREEN 3:
TRUE?
Press F for 'yes' and N for 'no'

Table 1 provides examples of all the statement types used in the experiment:

	GS: BARE PLURAL	UQS 1: ALL	UQS 2: ALL THE	UQS 3: EVERY
MAJORITY CHAR.	*Tigers have stripes.*	*All tigers have stripes.*	*All the tigers have stripes.*	*Every tiger has stripes.*
MINORITY CHAR.	*Ducks lay eggs.*	*All ducks lay eggs.*	*All the ducks lay eggs.*	*Every duck lays eggs.*

Table 1: Examples of statement types

The full set of statements used can be found in table 2. Note that the majority and minority characteristic lists are not list or pairwise matched for length, lexical content, etc, following Leslie et al (2011):

MAJORITY CHARACTERISTIC	**MINORITY CHARACTERISTIC**
Cats have whiskers.	Deer have antlers.
Cheetahs run fast.	Ducks lay eggs.
Cows eat grass.	Goats have horns.
Horses have four legs.	Insects lay eggs.
Kangaroos hop.	Kangaroos have pouches.
Leopards have spots.	Lions have manes.
Sparrows have wings.	Moose have antlers.
Tigers have stripes.	Snakes lay eggs.

Table 2: Set of items used[5]

5. Results and discussion

5.1 Acceptance rates

Table 3 summarizes the number of yes responses to the TVJ question in each condition.

	DETERMINER TYPE			
CHARACTERISTIC TYPE	*gen*	*all*	*all the*	*every*
majority	40	0	0	4
minority	39	0	0	0

Table 3: The number of yes responses to the TVJ

The difference between the generic statements and the other three statement types is very

5 The two columns are not matched.

striking: all but one generic statement was judged to be true, and all universally quantified statements were judged to be false except for 4/40 sentences with ‘every’. Clearly no inferential statistical tests are required to be confident that generic statements were associated with a different pattern of responses than universally quantified ones.

Leslie et al (2011) discuss their TVJ data in terms of the proportion of yes responses, rather than the actual number of responses. For the purposes of easy comparison, we can present our data in this way as well, in Figure 1.

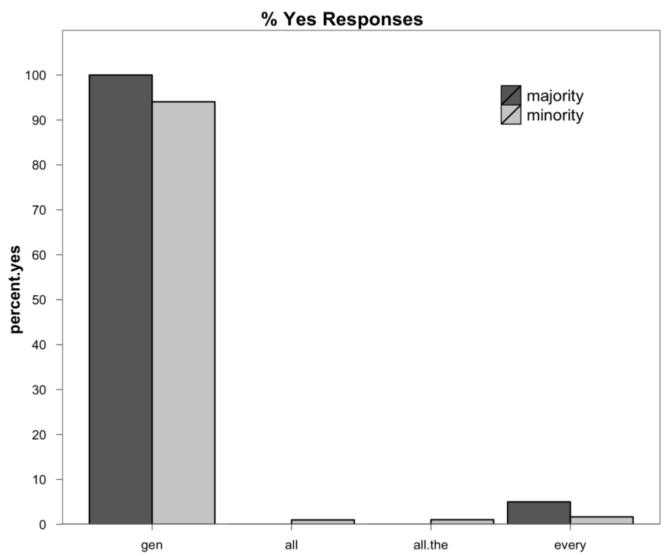

Figure 1: % of YES responses

As table 3 & figure 1 make very clear, we did not replicate the Leslie et al. finding for *all*. There is no hint that anyone in our experiment judged sentences quantified with ‘all’ as generic - even in the majority characteristic trials where the exceptions to the generalization are not made salient by the context. Leslie et al.’s results with context were 55% yes responses for majority characteristic sentences and 30% for minority, while ours are 0% for both cases. Not one single person said that ‘all tigers have stripes’ is true, while every single participant said that 'tigers have stripes' is true.

Not only are our rejection rates for universally quantified sentences much, much lower than Leslie et al's, but our acceptance rates for generic sentences are also higher. Generic statements involving a minority characteristic were judged to be true by 97.78% of our participants, and generic statements involving a majority characteristic were judged true by 100% of participants, as compared to about 85% for both conditions in Leslie et al's experiment. We speculate that this difference may be attributable to the fact that our experiment was conducted in a formal psycholinguistics laboratory setting, and all our participants were university students, while Leslie et al. collected their data using Mechanical Turk, and thus had a more variable environment and set of participants. Mechanical Turk, and other web based data collection methods, are becoming increasingly commonly used as tools to conduct cognitive science experiments (Buhrmester et al. 2011), and the larger and more diverse participant pools that result from this technique are clearly an advantage in many ways, but the difference we find here suggests that the results of lab based and online data collection methods are not always straightforwardly comparable.

Because we conducted our experiment in the lab, we were able to collect timing data as well as record the Truth Value Judgments of participants.

5.2 Reaction Time: Question Answering Response Time

We measured the time it took participants to make their TVJ decision: that is, to judge the statement as true or false. We hypothesized that if the participants had to decide between a kind level and an individual level interpretation, as is potentially the case with sentences like ‘all ducks lay eggs’ or ‘every duck lays eggs’, coming to a decision about the truth of the statement should be more difficult than in the case of sentences with no such ambiguity like ‘all the ducks lay eggs’. The extra difficulty associated with deciding between competing interpretations should be associated with slower decisions, and we expected that the slow down should be most pronounced for minority characteristic statements, since those are the cases where the kind level interpretation is true, while the individual level interpretation is quite saliently false.

The predictions for generic statements were less straightforward. First, it was not clear whether to expect that generic statements would be judged more quickly or more slowly than universally quantified statements. If generic interpretations are, indeed, a default, Kahneman System 1 type of interpretation, as argued by Leslie (2007), then accessing and making use of these interpretations should be a fast, effortless task. However, given that generic interpretations are not signaled with any overt cue in the sentence, computing and evaluating a generic interpretation could be a slow, effortful task. Secondly, it was not clear whether the majority vs. minority characteristic distinction should affect processing difficulty and speed. We might expect that since generics are tolerant of exceptions, whether there is a context making those exceptions salient or not should be irrelevant, so there should be no difference between majority and minority characteristic conditions. But given that the background context is only potentially relevant in the minority characteristic cases, we might expect that trials in which the exceptions to the generalization are made salient would be more effortful to evaluate since participants might actually have to deliberate. Given a situation in which the existence of male ducks has just been made very salient, participants might consider the possibility that ‘ducks lay eggs’ is false, even though they ultimately conclude it is true. For the majority characteristics, the exceptions to the generalization are not made salient by the context and thus these trials may be less difficult for participants.

Table 4 summarizes the mean question answering response times by condition, and the difference between characteristic types, and Figure 2 plots these results graphically (with the same data plotted in two graphs to aid visual comparison).

DETERMINER TYPE	MAJORITY CHARACTERISTIC	MINORITY CHARACTERISTIC	DIFFERENCE
gen	1397.45 *(1054.00)*	1925.08 *(1903.69)*	-527.63
all	1644.67 *(1214.82)*	2224.89 *(1802.82)*	-580.22
all the	1719.02 *(1002.09)*	1497.92 *(950.76)*	221.10
every	1454.75 *(1083.05)*	1474.32 *(894.64)*	-19.57

Table 4: Question Answering Mean Response Times and *(Standard Deviations)*

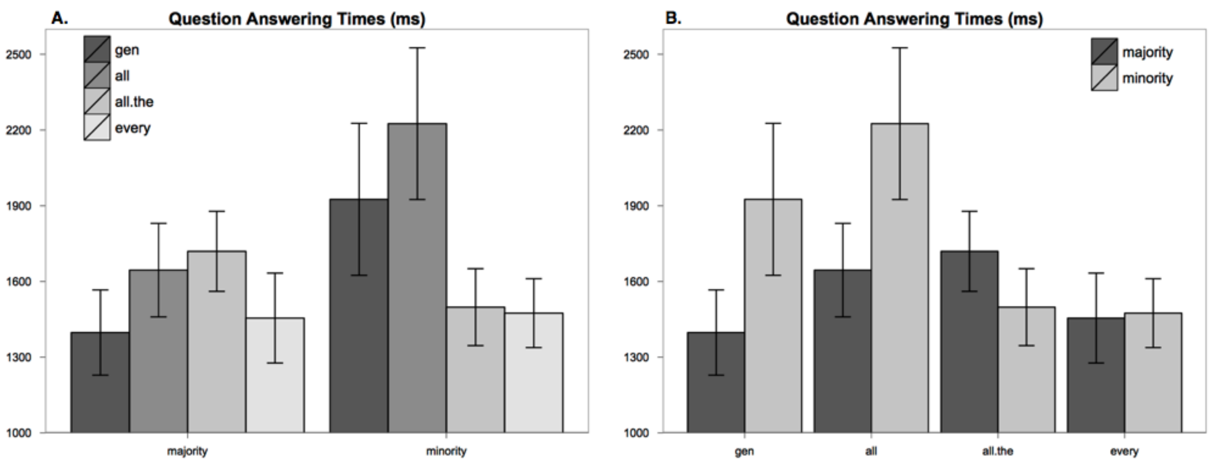

Figure 2: Question Answering Response Times. Error bars plot standard error.

Figure 2(a) highlights the contrast between the majority characteristic statements, where response times do not vary much between the four determiner types, and the minority characteristic sentences, where generics and ‘all’ sentences were judged much more slowly than the other two sentence types. Figure 2(b) plots the same data, but this time in a way that emphasizes the differences between the majority and minority characteristic statements for generics and ‘all’ sentences, but not for the other two types. We analyzed response times with a linear mixed effects model with characteristic type and determiner type as crossed fixed effects and subject and item as random effects. The model statistics are in table 5.

	ESTIMATE	STD. ERROR	T VALUE	PMCMC
INTERCEPT (MAJORITY, GENERIC)	7.08	0.12	60.08	0
CHAR (MINORITY)	0.11	0.08	1.48	0.16
DET (ALL)	0.16	0.1	1.61	0.1
DET (EVERY)	-0.02	0.1	-0.25	0.83
CHAR X DET (MINORITY, ALL)	-0.27	0.14	-1.87	0.06
CHAR X DET (MINORITY, ALL THE)	0.03	0.15	0.23	0.85
CHAR X DET (MINORITY, EVERY)	0.16	0.14	1.12	0.29

Table 5: Estimates, standard errors, *t* values, and *p* values of the linear-mixed effects model for

the Question Answering Response Time analysis.

Only the interaction between majority generic statements and minority ‘all’ statements emerges as statistically significant given alpha < 0.05, due to the relatively small number of observations per condition, however the numerical differences are striking and worth considering carefully.

The pattern of results in Figure 2 does not follow clearly from any of our predictions. We do see the hypothesized slow down for ‘all ducks lay eggs’ as compared to either ‘all the ducks lay eggs’ or ‘all tigers have stripes''. Sentences that involve both a minority characteristic property and ‘all’ quantification are correctly judged to be false, but making this judgment takes participants longer than sentences with ‘all’ + a majority characteristic (where the kind vs. individual interpretations are not made saliently distinct by the context) or than sentences involving a minority characteristic property, but ‘all the’ quantification (where there is no kind level interpretation available). However, we predicted the same pattern for sentences with the quantifier ‘every’, and we do not find it. Numerically, sentences with ‘every’ are judged false (correctly) faster than any of the other universally quantified sentences, and there are no differences between the two characteristic types.

Also unexpectedly, generic statements are associated with the same pattern as ‘all’ sentences: a much slower response in the minority characteristic condition than in the majority characteristic condition, or than in any of the other conditions in the experiment except minority characteristic + ‘all’. This similarity is particularly striking given that the actual responses participants made were categorically different for generic vs. ‘all’ statements.

The response time slow down for both ‘ducks lay eggs’ and ‘all ducks lay eggs’ in contexts where non-egg-laying ducks (namely males) have just been made salient suggests that although generics are still ultimately judged to be true in the face of exceptions, those exceptions are not entirely irrelevant to the process of coming to that judgment.

5.3 Reaction Time: Statement Reading Time

To further investigate the processing of generic and universally quantified interpretations, we also analyzed the time it took participants to read the critical statements themselves. Making a TRUE/FALSE judgment about a sentence is a metalinguistic task, which surely involves all kinds of non-linguistic and task dependent factors, thus the statement reading time may reveal something about the initial steps involved in interpreting these kinds of statements that the TVJT did not.

Table 6 reports the statement reading times by condition and the differences between characteristic types. Figure 3 plots these data graphically (again with the same data plotted in two graphs to aid visual comparison)

DETERMINER	MAJORITY CHARACTERISTIC	MINORITY CHARACTERISTIC	DIFFERENCE

gen	2204.48 *(1126.82)*	3070.91 *(1638.68)*	-866.43
all	2640.16 *(1213.09)*	2898.25 *(1487.01)*	-258.09
all the	2842.18 *(1509.39)*	3425.47 *(1255.58)*	-583.29
every	3133.3 *(1446.59)*	3197.68 *(1357.47)*	-64.38

Table 6: Statement Reading Times and *(Standard Deviations)*

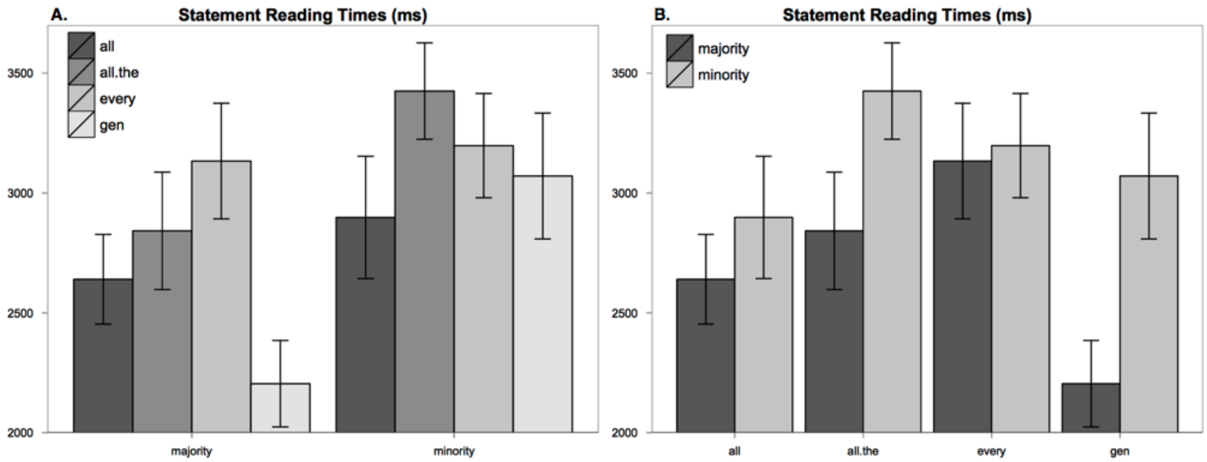

Figure 3: Statement Reading Times. Error bars plot standard error.

Statement reading times were analyzed in a linear mixed effects model as above. The model statistics are in Table 7.

	ESTIMATE	STD. ERROR	T VALUE	PMCMC
INTERCEPT (MAJORITY, GENERIC)	7.78	0.08	102.52	0
CHAR (MINORITY)	0.15	0.05	2.94	0.01
DET (ALL)	0.25	0.07	3.46	0
DET (ALL THE)	-0.01	0.07	-0.20	0.87
DET (EVERY)	0	0.07	-0.02	0.94
CHAR X DET (MINORITY, ALL)	-0.21	0.10	-1.99	0.04
CHAR X DET (MINORITY, ALL THE)	-0.03	0.10	-0.32	0.69
CHAR X DET (MINORITY, EVERY)	-0.11	0.10	-1.11	0.30

Table 7: Estimates, standard errors, *t* values, and *p* values of the linear- mixed effects model for the Statement Reading Time analysis.

In both the numerical patterns and the statistical analysis, the most striking feature is the reading times for the generic + majority characteristic condition, which are significantly faster than any other condition[6].

[6] In order to rule out any possibility that the generic reading times were faster due to the shorter length of these sentences, which, of course, lack an overt determiner, a model including sentence length in characters was conducted. Model comparison revealed no significant difference between the fit of the two models - adding length

We can also note that across the board, all four quantification types are read more slowly when they involve a minority characteristic than when they involve a majority characteristic, with the largest effects for generics and 'all the' sentences. Recall that the minority characteristic sentences are all read following a background statement which makes the exceptions to the generality of the characteristic highly salient, while the background statement does not make the exceptions to the majority characteristic statements salient. So it is not possible to determine whether this difference in reading times is due to the difference between majority vs. minority characteristic properties, or to the difference in the degree of relevance of the background context.

6. Conclusions

In this paper we presented the results of a Truth Value Judgement Task that addressed the basic question of how people generate and evaluate generic interpretations. This is a challenging question because generic statements are typically morpho-syntactically unmarked (there is no consistent overt cue to trigger the generic interpretation), and are famously robust to even very systematic exceptions, and yet not to all exceptions (recall the example of paperback books). Even individually, these features would seem to pose challenges for children acquiring a system, or for fluent speakers parsing sentences.

The basic goal of our research was to determine whether, in fact, fluent adult speakers of English do indeed find generic interpretations difficult to generate or evaluate.

Recent research by Leslie and her colleagues (Leslie et al, 2007; 2011) seems to suggest that they do not. Instead, Leslie et al claim to have discovered a Generic Over Generalisation effect, whereby speakers interpret universally quantified statements such as 'all ducks lay eggs' as if they were an equivalent generic statement (eg. 'ducks lay eggs'). They observe this GOG effect both in truth value judgment tasks, and in paraphrase tasks. Leslie et al argue that this evidence that speakers seem to find generic interpretations to be easier to generate, evaluate and recall than universally quantified interpretation is consistent with a hypothesis that generic interpretations are simpler, more basic, and more automatic than quantified interpretations.

This proposal is appealing, since it proposes that generic interpretations do not involve quantification, and are available more or less for free. Children must acquire the syntax and semantics of the natural language quantifiers and determiners that have more restrictive interpretations, while generic interpretations are the starting point.

However, Leslie et al only compared bare plural generic statements ('ducks lay eggs') with 'all' quantified statements ('all ducks lay eggs'), and only collected data on the outcome of the

did not improve the fit of the model.

interpretation of these statements (the TRUE/FALSE response)[7]. Drawing conclusions about universal quantification in general, or about the ease and automaticity with which generic interpretations are processed requires more subtle manipulations.

Thus our experiment compared bare plural generics to three types of universally quantified statements, systematically manipulated whether the context made possible exceptions to the generic/universal statement salient, and measured statement reading times, and question answering times, as well as the value of the T/F judgment.

We failed to replicate the GOG effect. There is no hint that anyone in our experiment judged sentences quantified with 'all' as generic. Prima facie, this at least fails to support the generics as default analysis, although of course it is not evidence against this hypothesis. It's possible that by prefacing all our statements with background information that makes individual members of the species of animal the statement is about, we made the universal quantification over individuals interpretation very salient, and thus easy to access, cancelling out the generics advantage seen in Leslie et al's experiments.

Since our acceptance rates for generic sentences were also higher than in the Leslie et al experiments, it's also possible that the very different results are attributable to the differences in the setting of the experiment and the set of participants.

Intriguingly, when we look at the time it took participants to make their judgments, we do find that generic and 'all' sentences are associated with similar responses, just as Leslie et al do, however where Leslie et al find responses they argue are evidence that generic interpretations are easy, we find the opposite. While generic and 'all' statements that involve properties that are characteristic of the majority of the relevant species (such as having a tail) are judged just as quickly as statements with 'all the' and 'every' quantification, generic and 'all' statements involving minority characteristic properties are judged significantly more slowly than all other statement types. Thus just in the two cases where the background context makes the systematic exceptions to the statement salient, participants are much slower to decide whether the statement is true or false. They ultimately judge correctly 99% of the time, but the judgment is clearly more difficult in these cases. This is the first systematic demonstration we are aware of that although the ultimate truth of a generic interpretation is not sensitive to context, assessing a generic interpretation is. Note that we only find context sensitivity for the statements that are evaluated against the background of what Greenberg (2007) calls contextually irrelevant individuals – exactly the opposite of what we might have predicted. However, we confounded characteristic type and background relevance in this experiment (only minority characteristic trials included a background that made salient the exceptions to the generalization), so further research is clearly required to better understand these results.

[7] Recall also our concerns about the extent to which Leslie et al (2011) successfully controlled for kind level interpretations for their 'all' sentences.

Finally, we found that the reading times for the generic + majority characteristic statements were significantly faster than any other condition. Given the fact that our majority and minority characteristic items were not pairwise matched for length, lexical frequencies, etc, the statement reading time results must be interpreted cautiously. However, they may be taken as support for the generics as default hypothesis after all. At least in the case where the systematic exceptions to the generic statement were not made salient by the context, we find that generic statements are much faster to read than universally quantified statements. This is precisely what is predicted by the hypothesis that generic interpretations are simple, basic interpretations that are more easily generated and parsed than quantified interpretations.

In the end then, our results muddy the waters more than they clear them. On the one hand, we find no evidence for a GOG effect, whereby universally quantified statements are interpreted as generics, and we find evidence that may suggest that generics are actually more sensitive to contextual restriction than is generally assumed. Both these findings are consistent with analyses of generics as involving operators of some kind, and being at least as complex as other forms of quantification. On the other hand, the statement reading time results suggest that (at least some) generic statements are, in fact, much easier to process than their corresponding quantified statements, which is inconsistent with these accounts. Further research to more systematically investigate these findings, and better understand just how we generate generic interpretations to begin with is clearly motivated.

Acknowledgements

The first author would like to thank Napoleon Katsos for insightful discussion.

References

Asher, N. and Morreau M. (1995). What Some Generic Sentences Mean. In G. Carlson & F. J. Pelletier (Eds.) *The Generic Book*. Chicago. Chicago University Press, 300–339.

Brisson, C. (2003). Plurals, All, and the Nonuniformity of Collective Predication. *Linguistics and Philosophy 26*(2), 129-184.

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3-5.

Carlson, G. N. (1977). *Reference to Kinds in English*. Ph.D. thesis, University of Massachusetts at Amherst.

Carlson G. (1999). Evaluating Generics. In P. Lasersohn (Ed.), *Illinois Studies in the Linguistic Sciences*, 29:1, 1-11.

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.

Cohen, A. (1996). *Think generic: The meaning and use of generic sentences*. Ph.D. thesis, Carnegie Mellon University.

Condoravdi, C. (1992). Strong and weak novelty and familiarity. In *Proceedings of the Second Conference on Semantics and linguistic theory*, 17–37.

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

Dahl, 0. (1985). *Tense and aspect systems*. Oxford: Blackwell.

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), 100-123. New York: Oxford University Press.

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

Geurts, Barts (1985). Generics. *Journal of Semantics 4*(3), 247-255.

Greenberg, Y. (2007). Exceptions To Generics: Where Vagueness, Context Dependence And Modality Interact. *Journal of Semantics 24*(2), 131-167.

Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, Campbell GL (2005). *Epidemiology and transmission dynamics of West Nile virus disease*. Emerging Infect. Dis. 11 (8), 1167–73.

Hollander, M. A., Gelman, S. A., and Star, J. (2002). Children's Interpretation of Generic Noun Phrases. *Developmental Psychology 36*(6), 883–894.

Kahneman, D. and Frederick, S. (2002). "Representativeness revisited: Attribute substitution in intuitive judgment." In T. Gilovich, D. Griffin, and D. Kahneman (Eds.) Heuristics & Biases: The Psychology of Intuitive Judgment. Cambridge University Press. New York, 49–81.

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. (1987). An Outline of Genericity, partly in collaboration with Claudia. Gerstner. SNS-Bericht 87-23, University of Tübingen.

Krifka, M., F. Pelletier, G. Carlson, A. ter Meulen, G. Chierchia, and G. Link. (1995). Genericity: An Introduction. In G. Carlson & F. J. Pelletier (Eds.) *The Generic Book*. Chicago. Chicago University Press, 1–125.

Lasersohn, P. (1999). Pragmatic Halo. *Language 75(3),* 522-551.

Lawler, J. (1973). Studies in English generics. *University of Michigan Papers in Linguistics 1*(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., S. Khemlani, & S. Glucksberg. (2011). All Ducks Lay Eggs: The Generic Overgeneralization Effect. *Journal of Memory and Language 65*(1), 15-31.

Leslie, S. J. & S. Gelman. (2012). Quantified statements are recalled as generics: evidence from preschool children and adults. *Cognitive Psychology 64*(3), 186-214.

Lewis, D. (1975). Adverbs of Quantification. In E.Keenan (Ed.) *Formal Semantics in Natural Languages*. Cambridge University Press, 3-15.

Marmaridou-Protopapa, S.A.A (1984). *The study of reference, attribution and genericness in the context of English and their grammaticalization in M. Greek noun phrases*. Unpublished Ph.D. thesis. Darwin College. Cambridge.

Nunberg G. and C. Pan (1975). Inferring Quantification in generic sentences. *Papers from the XIth Regional Meeting of CLS*, 412-28.

Pelletier, F.J. (2010). Are All Generics Created Equal?. In F.J. Pelletier (Ed.) *Kinds, Things and Stuff*, Oxford University Press, 60-79.

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. Schubert, L. K. and F. J. Pelletier (1987) "Problems in the Representation of the Logical Form of Generics, Plurals, and Mass Nouns", in E. LePore (ed.), New Directions in Semantics, Academic Press, London, 385-451.

Prasada S, Dillingham EM. (2006). Principled and statistical connections in common sense conception. *Cognition. Feb 99*(1), 73-112.

Shaffer, A. (2012). *How Paperbacks Transformed the Way Americans Read*. Mental Floss. Aug, 14.

Stanley, J. and Z.G. Szabó (2000). On Quantifier Domain Restriction. *Mind and Language 15* (2-3): 219-61.

Tardif, T., Gelman, S.A., Fu, X., and Zhu, L. (2012). Acquisition of generic noun phrases in Chinese: learning about lions without an '-s'. Journal of Child Language, 39, 130-161.