**A New Look at the 'Generic Overgeneralisation' Effect**

ABSTRACT:
*While generic generalisations have been studied by linguists and philosophers for decades, they have only recently become the focus of concentrated interest by cognitive and developmental psychologists, who propose the generics-as-default view. In this paper we focus on the 'Generic Overgeneralisation' (GOG) effect (Leslie, Khemlani, and Glucksberg 2011) and the native speaker judgments that have been used to support it, and by extension, the generics-as-default view. We take a step back to look at the history of the GOG effect in order to contextualise it. We review existing experimental evidence and discuss four non-mutually exclusive explanations for the GOG effect: ignorance, subkind interpretation, atypical behaviour of* all and *quantifier domain restriction. We conclude that a closer look at the semantics and pragmatics of generics and universal quantifiers may provide a more nuanced explanation for the pattern of judgment data than that proposed by the generics-as-default view.*

## 1 Introduction

In this paper, and in our broader research program, we are investigating the similarities and differences between different ways of expressing generalisations in natural language. Quantificational generalisations, as in (1), are expressed in quantitative, statistical terms, while generic generalisations, as in (2)-(3), make general claims about kinds of entities and refer to a property that is characteristic of the kind in question, but not necessarily statistically prevalent, as in (3), (only adult, male lions have manes).

(1) Some lions live in cages.
(2) Lions roar.
(3) Lions have manes.

Generic generalisations have long been studied in formal semantics, within which genericity is frequently viewed as a species of quantification. Even though generics have been studied since the seventies (see Lawler 1972, 1973; Dahl 1975; Carlson 1977) they still remain a rather controversial topic when it comes to deciding how to characterise their semantic interpretation and how to model their truth conditions (see recent discussion in Carlson 2011; Mari, Beyssade, and del Prete 2013). Within formal semantics, modal logic and probabilistic approaches are most prominent, both of which treat genericity as akin to quantification. According to the modal approach, which is the most widely adopted formal analysis of genericity (see Mari, Beyssade, and del Prete, 2013, 43), generic meaning is obtained as the effect of an underlying operator or quantifier dubbed 'GEN', which is not phonologically realised but which is active in the composition of the sentence meaning and is an unselective variable binding operator similar to adverbs of quantification like *usually, typically, always*, as analysed in Lewis (1975). This operator is sentential and is represented by a tripartite structure as in (4) (Krifka et al. 1995). Thus, the logical form of (2) may be given as follows in (5):[1]

(4) GEN [restrictor] [matrix]
(5) $GEN_x$ [Lions (x)] [Roar (x)]

A competing probabilistic account has been developed by Cohen (1999, 2004), according to which the important factor is the probability of the individual having the property in question. This probabilistic account also relies on a covert generic quantifier.[2, 3]

Carlson (2011, 1172) argues that "the GEN analysis is both rich and complex, interacting with the context, information structure, and subtleties of the syntax in a variety of ways. While the details of various analyses that have employed it may be called into question, that there is *some* kind of operator akin to GEN in generics is a reasonably secure claim at this point; this, despite the fact that it does not have direct and fully consistent morphological/phonetic realization in English or any languages that have been studied extensively to date".

This quote brings our attention to the fact that no language seems to have an overt realisation of the 'GEN' operator. Generic generalisations can be made using a wide range of different grammatical means, both within, as illustrated for English in (6), and across languages (see Behrens [2000] for typological comparisons and discussion[4]), but no language has a unique, unambiguous marker of genericity

---

[1] The tripartite structure was introduced by Heim (1982) and Farkas and Sugioka (1983) as a major novelty against Carlson's (1977) unitary operator *Gn.* A major motivation for the tripartite structure implicit in quantification is that it readily accommodates intuitions of ambiguity, such as those associated with sentences like "typhoons arise in this part of the Pacific", which can be interpreted as either "typhoons in general have a common origin in this part of the Pacific", or as "there arise typhoons in this part of the Pacific". See Krifka et al. (1995, 23) for discussion.

[2] Cohen's probabilistic account proposes that it is probability rather than modality that forms the basis of the semantics of generics, at least for 'absolute generics' such as *ravens are black*. This account holds that *As are B* is true just in case the probability of an arbitrary *A* being a *B* is greater than 0.5, that is, greater than chance. Cohen (2004, 531) introduces a homogeneity condition, according to which "the generic **gen**(ψ,φ) presupposes that its domain, ψ, is homogeneous, in the following sense: for any psychologically salient criterion by which ψ may be partitioned into subsets, the conditional probability of φ ought to be roughly the same given every such subset of ψ." Salient partitions are e.g. space, numerical scales, gender, subject matter and abstract domains. As Mari, Beyssade, and del Prete (2013; 84) illustrate, in Cohen (1999), he proposes that there is a covert generic quantifier GEN, which gives rise to the following representation:

i. Birds fly.
   GEN (bird(x), fly(x)) P(fly | bird) > 0.5 (the probability of an object flying given that the object is a bird is greater than 0.5)

[3] Within formal semantics, the view that generics are not quantificational is either considered to have been surpassed by the modal account (cf. Carlson 1977) or has not yet been shaped into a fully spelled out account. See though Deo and Maniman (2015) for an account based on gradability and stochastic comparison, as well as discussion in Mari et al. (2013) and further references therein, where they state that an account based on noun ambiguity à la Dayal (2004) could provide an account for genericity without assuming a generic operator. Within philosophy, the view that generics are not quantificational is a common feature of the following proposals: (a) generics are simple subject/predicate sentences that predicate properties of kinds (Liebesman 2011), (b) genericity has a psychological, rather than a linguistic, basis (Collins 2015) and (c) a sophisticated kind-predicate view à la Carlson (1977) needs to be revisited (Teichman 2015).

[4] In languages without articles, such as Finnish, which morphologically conflates referential marking and role marking, the morphological case of a phrase might be a relevant feature in generic marking. Korean and Tagalog employ topic-marking elements, while in Vietnamese some types of generics

equivalent to a quantifier or determiner. It is important to note that none of the analyses that posit a 'GEN' operator offer an explanation for this, a point that the *Generics-as-Default* view (see next paragraph) capitalises upon.

(6) a. Echidnas have claws. *bare plural*
b. An echidna has claws. *indefinite singular*
c. The echidna has claws. *definite singular*

In contrast to the quantificational analysis of generics, a growing body of experimental and developmental psychological work on the topic proposes that genericity is categorically different from (and significantly simpler than) quantification (Leslie 2007; Gelman 2010). This latter hypothesis, called the *Generics-as-Default* view (GaD view henceforth) treats generics as an innate and default mode of thinking. This idea is linked to the view of cognition that assumes two different systems, argued for by Kahneman and Frederick (2002) among others, which includes a distinction between System 1, a fast, automatic, effortless lower-level system, and System 2, a slower, more effortful higher-level rule-governed system.[5] According to this view, the fact that no language has a dedicated overt 'GEN' operator does not come as a surprise: given that generics are the most primitive default generalisations, children do not need to learn anything in order to acquire them. Thus, generics come essentially for free.

Determining which properties or attributes can be generically predicated has proven very challenging for both types of approach. Generic generalisations can range from exceptionless definitional statements such as *triangles have three sides*, or *the walrus is a mammal*, through what Leslie, Khemlani, and Glucksberg (2011) call 'majority characteristic' statements such as *dogs have four legs* or *tigers have stripes* (which are true of the overwhelming majority of instances, with only a few exceptional individuals), through 'minority characteristic' statements like *ducks lay eggs* or *lions have manes* (which all involve primary or secondary sexual characteristics of animals, and are thus only true of no more than 50% of individuals), to 'striking' generalisations like *sharks attack swimmers*, or *mosquitoes carry malaria* (which are true of only a tiny fraction of individuals, but involve properties which are noteworthy in some way).

To make matters more complicated, not only is statistical prevalence not necessary to licence generic generalisations, it is not sufficient either. Statements like *books are paperbacks* or *Canadians are right-handed* may be true of 80% or more of individuals, and yet are not typically judged as true, and thus fail as generic statements.

The two approaches to genericity have their own different challenges. Within the formal semantics approach, a critical challenge has been to understand and model the ways in which generic generalisations are licensed, and, as just discussed, to account for the lack of any overt realisation of 'GEN' in any known language.

---

contain classifiers. For further discussion of the typological parameters of genericity see Behrens (2000).

[5] The idea that there are two distinct systems of cognition, one of which is intuitive in nature, and one of which is reflective has been common across scientists who study human reasoning in different domains, i.e. in conditional and probabilistic reasoning (Evans and Over 1996; Sloman 2002; Stanovich 1999), decision making (Kahneman and Frederick 2002) and social cognition of various sorts (Petty and Cacioppo 1986; Chaiken, Liberman, and Eagly 1989).

In the psychological GaD approach, because generic generalisations are understood as a basic mode of thinking, some of the specific challenges for the quantificational analysis are avoided. For instance, according to this view, there is no overt generic operator in any known language because generics are the unmarked, System 1, case. On this view, only effortful, non-default quantificational generalisations require overt linguistic exponence. Moreover, the GaD approach fits naturally with findings that generics are frequent in children's natural speech from a very early age (Gelman et al. 2008). However, while assigning generics to a more basic, unmarked System 1, mode of thinking may sound intuitive at some level, it rests on a vague and undefined notion of markedness.[6] Moreover, it is not obvious that this approach fares any better in offering a principled explanation for why *Italians are good skiers* is typically judged as true (and thus an acceptable generic statement), while *Canadians are right-handed* is not.

In Lazaridou-Chatzigoga, Katsos and Stockall (2015), we juxtapose the formal semantic and the psychological approaches and in doing so we highlight some of the significant challenges for each of them. We argue that the evidence for the GaD proposal is significantly weakened by a lack of cross-linguistic considerations and serious engagement with the formal semantics of quantification and specificity. We also question the appeal to effortful processing much of the evidence for the proposal rests on. On the other hand, the formal semantics models do not offer any explanation for the robust findings from early child language competence, namely that generic utterances and generic interpretations are present in children as young as 2 years old, despite not being associated with any overt morpho-syntactic marker in any known language.

In this paper we will focus on the effect called 'Generic Overgeneralisation' (GOG) (Leslie, Khemlani, and Glucksberg 2011), which has been used to support the GaD view on generics. We will take a step back to look at the history of the GOG effect in order to understand it better. We will use conceptual arguments, as well as discuss a set of four alternative explanations for the data that seem to support GOG: a) ignorance of the relevant facts, b) subkind (taxonomic) interpretation, c) the atypical behaviour of *all* and d) Quantifier Domain Restriction (QDR). We will propose that all these factors play a role in explaining the attested behaviour by adults. These factors are independently attested and known to interact with the interpretation of generic and quantified statements. We will propose that they likely suffice to explain the data that have been used to justify the existence of the GOG effect. We suggest that even the name of the GOG effect might be misleading. The effect mainly tries to capture the behaviour observed with the quantifier *all,* which supposedly gets a generic interpretation as a result of an overgeneralisation bias. Thus, perhaps a better name for that effect would be 'Quantifier Reanalysis' effect, because this term would direct the focus where we believe it belongs: on the interpretation of *all*, or more generally of quantifiers, rather than the interpretation of generic statements. The overall aim is to showcase the role of linguistic factors (both semantic and pragmatic) in the interpretation of generic and quantified statements, and to underscore the relevance of linguistically-motivated explanations.

---

[6] While it may be intuitive to say that *ducks lay eggs* is simpler, and therefore less marked, than *all ducks lay eggs*, no such easy comparison can be made between the generic *an echidna lays eggs* and the specific indefinite *an echidna lives here*. The appeal to markedness in the GaD literature seems to rely heavily on the fact that many generic statements in English involve a bare plural subject combined with a simple present verb form. A significantly more sophisticated notion of markedness will be required to extend the proposal to the other ways to form a generic in English and to other languages.

## 2 A history of the Generic Overgeneralisation effect

In this section we will review the Generic Overgeneralisation effect, given that it has been adduced as one of the main predictions and, subsequently, as one of the main pieces of evidence in support of the GaD view.

Leslie, Khemlani, and Glucksberg (2011) use GOG to refer to "the tendency to overgeneralise the truth of a generic to the truth of the corresponding universal statement" (17). In order to understand the importance of this effect, we propose to look at the reasoning behind it: if we assume that generics invoke the cognitive system's default mode of generalisation, Leslie (2008, 23) claims that "we might further conjecture that the comprehension of non-default quantifiers requires the conceptual system to override or inhibit its default operations". If generics are the default interpretation of any generalisation, then understanding quantified statements requires deviating from the default interpretation. Thus, it is more effortful. Both children and adults might sometimes fail to perform this deviation and thus, despite knowing how to correctly interpret quantified statements in most cases, they would occasionally "treat them as if they were generics" (Leslie 2007, 398).

Leslie (2008) argues that generalising with generics follows a default process, whereas generalising with quantifiers follows a non-default process. She compares these processes to the default processes that a chicken's spinal column implements, which triggers basic locomotion (evident in headless chickens that are nevertheless able to run), vs. the non-default processes of its brain that inhibit/manage such behaviour in the usual circumstances. Other examples of (non)default processes which Leslie (2008) points out as models for the GaD can be found in areas such as cellular processes, neural networks, glandular regulatory systems, visual cognition, action-planning systems, and even at the conscious level of temptation resistance.

Leslie's view of the difference between generics vs. quantifiers is argued to be very similar to Kahnemanian System 1 vs. System 2 distinctions. One piece of evidence for the existence of two systems is the fact that they can lead to conflicting judgments. Conflicts can arise between what people judge on an intuitive basis and what people judge on a reflective basis. When there is a conflict, this might result in fast, automatic System 1 responses, when slower, more effortful System 2 responses are required. The distinction is thus between 'intuitive System 1' and 'reflective System 2'.

Research has demonstrated experimentally that in situations where intuition and reflection diverge, participants arrive at conflicting judgments, one issued by System 1 and the other by System 2 (see Frederick [2005]). This can occur when people are given a task that requires a System 2 operation and response, but instead they give a System 1 response that is *easier, quicker and more automatic*. Thus, System 1 is erroneously *over*-used.

Leslie (2007, 395) cites Frederick's (2005) "cognitive reflection test", to illustrate the idea that the two systems might arrive at conflicting judgments: "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" Most people report an initial inclination to answer "10 cents". System 1 supplies this first fast, but erroneous response. The correct response of "5 cents" requires algebraic reasoning, part of the slower System 2.

Thus, the idea is that we might find experimental situations where we would see evidence of a similar type of error for making and understanding generalisations, which emerges when people rely on the default process instead of the non-default. Leslie claims that these situations might be sufficiently demanding cognitive tasks.

Errors then would arise when "people interpret quantified statements as though they were generics" (Leslie 2007, 398).

If we accept that generics reflect this default mode of generalisation while other methods require inhibitory processing, one expects some empirical consequences, Leslie argues. The list of the ones envisaged (Leslie 2008, 24-25) includes:

a) children are expected to produce and comprehend generics with greater ease than quantifiers, and at earlier ages
b) generics are expected to show up more frequently in maternal speech than quantifiers
c) if faced with a "sufficiently demanding cognitive task, young children might even erroneously interpret sentences containing explicit quantifiers as though they were generics"
d) we might also expect adults to be susceptible to the same type of error and treat nongeneric generalisations, as if they were generics, i.e. "nongeneric generalisations would, from time to time, inappropriately exhibit some characteristics of generics, especially if the information-processing demands were made great enough"

We will discuss here the third and fourth predictions, but see Lazaridou-Chatzigoga, Katsos and Stockall (2015) for a review of the developmental studies concerning points (a) and (b) above.[7]

Research by Hollander, Gelman and Star (2002) has been argued to show this effect for children, thus providing evidence for prediction (c). In experiment 1 they asked 3- and 4-year-old children and adults to answer questions like *Are {fires/all fires/some fires} hot?* They found that while both 3- and 4-year-olds gave adult-like responses to generic questions, only the 4-year-olds were adult-like with *all* and *some* questions. The 3-year-olds answered all three question types following the same pattern. Crucially, this response pattern was the one observed for generics for both 4-year-olds and adults. Thus, given that 3-year-olds gave to *all* and *some* the same response pattern as to generics, the authors conclude that they treated quantified sentences as if they were generic, as predicted by the GaD hypothesis. This is not due though to children's poor understanding of quantifiers in general, they argue, as 3-year-olds understand quantifiers in a less demanding task (experiment 2). Experiment 1 was more demanding because it asked children to think about abstract generalities rather than about a restricted set that was readily available.

Leslie (2008) argues that, even though Hollander, Gelman and Star (2002) did not find a similar error with their adult participants, adults may still be susceptible to similar errors under certain circumstances. Thus, in order to test prediction (d) above, Khemlani, Leslie, Glucksberg, and Rubio-Fernández (2007) tested a larger set of data than the one used in Hollander, Gelman and Star (2002). More specifically, participants were presented with minority characteristic statements such as *(all) ducks lay eggs* and *(all) horses give live birth* and were asked whether they agreed with the

[7] One reviewer points out the "natural pedagogy" theory proposed by Csibra and Gergely's (2009) as potentially relevant for the data in question. This theory suggests that infants fast-learn generic knowledge from adults positing a "genericity bias". We will leave this discussion for future work given that (a) the acquisition of generics falls outside the main focus of the present paper, and (b) due to space limitations, as the "natural pedagogy" theory is not purely linguistic, but rather aims at providing a framework for cultural transmission and communication more broadly, the relevance of which would need to be properly addressed at length.

statement or not. According to the authors, participants should reject the *all*-statements, because they are false. Nevertheless, participants agreed to the *all*-statements approximately half of the time (46%), which the authors take to provide preliminary evidence for a GOG effect in adults.

The first detailed investigation of the GOG effect is found in Leslie, Khemlani, and Glucksberg (2011). In their experiment 1, participants judged the truth or falsity of a list of generic and *all*-quantified statements that were presented one after the other without any background context. In more than half of the trials when the *all*-statements involved characteristic properties, participants judged these statements to be true: 78% for majority characteristic such as *all tigers have stripes* and 51% for minority characteristic statements such as *all ducks lay eggs*. By contrast, *all*-statements, which did not involve characteristic properties, such as *all cars have radios*, were only judged true 13% of the time. The authors argue that these high acceptance rates for the characteristic-property *all*-statements are due to participants interpreting the 'false' universally quantified statements as if they were their 'true' generic counterparts, and are thus a clear case of GOG.

As the authors acknowledge, however, these elevated acceptance rates might be due to alternative explanations, which they sought to address in subsequent experiments. Leslie, Khemlani, and Glucksberg (2011) considered three possible explanations for why their participants were so prone to accept statements like *all ducks lay eggs* as true: a) ignorance of the relevant facts, b) a subkind (taxonomic) interpretation of *all*, and c) quantifier domain restriction.

a. Ignorance of the relevant facts

The simplest explanation for the behaviour of the participants in Leslie, Khemlani, and Glucksberg's (2011) experiment 1 is that they simply didn't know some of the relevant facts. All the items in the majority and minority characteristic conditions involved properties of animals, and many of these properties involved primary or secondary sexual characteristics relevant to mating and reproduction. If participants were ignorant, they might have accepted *all ducks lay eggs* because they actually thought that all ducks, including males, do so.

The authors addressed this possibility with experiment 3, which included two different tasks: (a) the truth value judgment task as in experiment 1 and (b) a knowledge task, in which they asked participants to judge the truth or falsity of false gendered minority characteristic statements such as *male ducks lay eggs* or *female lions have manes*. The authors manipulated whether participants did the knowledge task before or after the truth value judgment task. They found that participants were generally aware of the relevant facts (rejecting the false statements 84% of the time). When the knowledge task followed the truth value judgment task, the results from experiment 1 were replicated: both minority and majority characteristic *all*-statements were judged true more than 50% of the time (70% for majority and 50% for minority). The knowledge task results show that even participants who know the facts still produced the GOG effect. And even when the knowledge task preceded the truth value judgment task, which ensured that the relevant facts were salient, participants still accepted majority characteristic statements such as *all tigers have stripes* 90% of the time, and minority characteristic items (*all lions have manes*) 32% of the time. They conclude that the GOG effect cannot be attributed to ignorance, and appears to be robust even in the face of the exceptions to the generalisation being made salient.

b. Subkind (taxonomic) interpretation

If we take sentences like *all tigers have stripes* or *all ducks lay eggs* to quantify over tiger or duck subkinds, *all tigers have stripes* could be interpreted as 'all kinds/types of tigers have stripes', including the Bengal tiger, the Sumatran tiger, the Siberian tiger, etc. and *all ducks lay eggs* could be interpreted as 'all kinds/types of ducks lay eggs', including the Mallard, the Eider, the Goldeneye, etc. Under such a subkind or taxonomic interpretation, the correct response to these statements is to accept them. Thus participants in the truth value judgment task in experiment 1 may have been generating subkind interpretations on the trials they judged to be true.[8]

In order to address this alternative explanation, the authors ran experiment 2b, where participants followed the truth value judgment task with an additional second phase, in which they were asked to provide paraphrases of each statement they had judged in the first phase. The authors then coded the responses and looked for mentions of subtypes, finding that only 1% of the paraphrases included subtyping language. These results lead them to reject the possibility that subkind interpretation was driving the participants' behaviour.

c. Quantifier Domain Restriction (QDR)

A sentence like *all ducks lay eggs* might be judged as true, if people interpret it as applying to only a relevant subset of ducks, namely to the mature fertile female ducks. This explanation takes quantified statements to be interpreted within a context, which may restrict the scope of the quantifier (as per Stanley and Szabó 2000, Stanley 2002). Thus, the reason why people accept the above statement is because (they believe) it is true once you have restricted the quantifier to the relevant subset of ducks.

In order to address this alternative explanation the authors ran experiment 2a, where they provided the participants with a background context, which was presented before each statement. These contexts included artificial population estimates of the following form:

(7) ''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 quantification over every individual duck in the world, and thereby to make it difficult/impossible to interpret *all* as restricted to only the (female) ducks that are presupposed by *lay eggs*. If acceptance of *all ducks lay eggs* in the first experiment was driven by QDR, the authors predicted that it would disappear in the context of population information.

Nevertheless, the GOG effect still occurred on a substantial portion of trials, with a 60% acceptance rate for *all* statements for majority characteristic statements and 30% for minority characteristic statements - less than when the statements appeared with no preceding context (78% and 51% respectively), but still a high

[8] In further support for this explanation note that the generalisations used in the majority non-characteristic, and striking conditions, which were almost always rejected in their *all*-statement form, are less susceptible to a subkind or taxonomic interpretation. A 'striking' statement such as *all pit bulls maul children* does not easily have the interpretation 'all kinds/subtypes of pit bulls maul children', nor does a majority non-characteristic statement such as *all clocks are round* ('all subtypes/kinds of clocks are round'), since there are no distinct (sub)species of pit bull or clock.

percentage. The authors thus concluded that quantifier domain restriction could not be the sole explanation for the GOG effect.

On the basis of the results of the additional experiments that addressed the three alternative explanations for the acceptance of *all*-statements with majority and minority characteristic statements, Leslie, Khemlani, and Glucksberg (2011) rejected all of them and argued to have found support for a strong generic bias, according to which people sometimes treat universally quantified statements as if they were generic.

## 3 Alternative explanations for the GOG effect

In this section we articulate the reasons we believe that the rejection of the alternative explanations by Leslie, Khemlani, and Glucksberg (2011) was premature, and that these options, together with a fourth option they do not consider (atypical behaviour of *all*), are plausible explanations for the putative GOG effect. We will argue, then, that the GOG effect is not due to participant error based on the erroneous overgeneralisation of generic readings to universally quantified statements, but rather a perfectly reasonable response given the semantics and pragmatics of quantification.

We briefly revisit the ignorance explanation, and then discuss subkind interpretations and the atypical behaviour of *all*. We will discuss the fourth possible alternative explanation, quantifier domain restriction, in its own section, to do proper justice to the nature of the explanation.

### a. Ignorance/Failure to recall the relevant facts

We are persuaded by Leslie, Khemlani, and Glucksberg's (2011) experiment 3 that their participants know the relevant facts about the minority characteristic items (*all ducks lay eggs, all deer have antlers*), and we accept that the 30% acceptance rates for such sentences even when the knowledge probing task came first means that general or temporary 'ignorance' (failure to remember) cannot entirely explain the GOG effect. However, making the exceptions salient does reduce acceptance rates for these items by 18%, suggesting failure to recall the facts does play some role in generating 'true' responses to these items.

Furthermore, the knowledge task and task order manipulation (knowledge task before or after the truth value judgment task) produce very puzzling results for the majority characteristic items (*tigers have stripes*). The knowledge task included only false minority characteristic statements like *male ducks lay eggs* and thus it failed to test whether participants are aware that majority characteristic generalisations also have exceptions (participants were not asked to judge whether *albino tigers have stripes* or *amputated horses have four legs*). Doing the knowledge task (which as we noted included only minority characteristic statements) before the critical truth value judgment task had an effect on majority characteristic statements too. In fact, it **increased** acceptance rates for majority characteristic *all*-statements by 20% compared to doing this task after the judgment (acceptance increased from 70% to 90%). This is counter-intuitive, because for minority characteristic statements the knowledge task had the effect of decreasing acceptances (and hence decreasing the putative GOG effect). Leslie and colleagues appeal to a strategic task effect to explain this strange result, but they do not provide convincing evidence that participants in the experiments knew that *all tigers have stripes* is not true.

Thus, experiment 3 showed that participants accepted minority characteristic items 18% less of the time when they did the knowledge task first (acceptance dropped from 50% to 32%). Therefore, it could be argued that 18% of the GOG effect is explained by ignorance/failure to recall but a remaining 32% of acceptances should be safely attributed to GOG. However, the fact that acceptance rates went up for majority characteristic statements when the knowledge task preceded the experiment suggests that something else is going on apart from ignorance/failure to recall. We believe that this is the atypical behaviour of *all*, which we outline in section (c) below.

b. Subkind (taxonomic) interpretation

The paraphrase task used by Leslie, Khemlani, and Glucksberg (2011) does not provide conclusive evidence to exclude the subkind interpretation explanation. The distinction between implicit knowledge and explicit knowledge is one of the most fundamental distinctions in cognitive science (see Dienes and Perner [2002] for an overview of cases where participants are employing rules in categorisation and judgment tasks that they are not able to report in explicit versions of the task). In our context, participants may not have been sufficiently consciously aware that they were interpreting *all ducks* as 'all kinds of duck' in the main experiment to be able to explicitly report this in a paraphrase task that was administered later on, but this does not guarantee that they did not do so. Thus, we cannot rule out the possibility that at least some of the times when people accept a statement like *all ducks lay eggs*, they are interpreting it as quantifying over different duck subkinds, such as the Mallard, the Eider, the Goldeneye, etc., which renders the universal statements trivially true.

c. Atypical behaviour of *all*

When Leslie, Khemlani, and Glucksberg (2011) discuss the GOG effect, they reason that it would be unlikely that adults would make such errors with *some,* whereas data from Hollander, Gelman and Star (2002) suggest that 3-year-olds make similar errors with both *some* and *all.* The authors hypothesise that adults might be inclined to do so only with universal statements given that "it may be harder, […] to confirm that a universal statement is true" (Leslie, Khemlani, and Glucksberg 2011, 17). Thus, Leslie and colleagues decided to focus on the universal quantifier *all.* In this subsection we will review some factors that suggest that *all* cannot be treated as a representative universal quantifier, couching these within a broader consideration of the behaviour of *all.* It might be then that the GOG effect is due more to the atypical behaviour of *all* rather than to the default nature of generic readings.

First, the behaviour of *all* has been discussed in work that investigates fallacious reasoning (Jönsson and Hampton 2006). *All* has been found to participate in a special phenomenon called the "inverse conjunction fallacy", according to which participants often expressed a belief in *all*-statements with a simple restrictor such as *all <u>ravens</u> are black*, while rejecting a matched statement with a conjoined restrictor such as *all <u>young jungle ravens</u> are black*, or, even when they accepted both statements, they were inclined to give greater confidence to the first one. These judgments are erroneous from the point of view of logic, given that the first statement entails the second. If one accepts the first statement, one should also accept the second, and if one accepts both, it would be inconsistent to express greater confidence only in one of them. Jönsson and Hampton (2006, 323) propose that one of the interpretations of their results is that "people may not interpret the word *all* as mapping onto the universal quantifier in logic". Results such as these suggest that if people are not forced to interpret *all* strictly, they might interpret it loosely or vaguely, akin to 'almost all'.

Second, the behaviour of *all* has been discussed in corpus linguistics, in work that investigates the expression of exaggeration. Claridge (2011) argues that universal quantifiers such as *all* and *every(thing)* are prone to hyperbolic use, with a meaning similar to 'very many' or 'almost all'. To illustrate this use, she quotes the fragment "*I didn't bring none of my clothes back ... I left 'em all down there"* from Labov (1984) and argues that it is clear from the context that the speaker is unlikely to intend the statement in its strict literal sense. *All* in this context gets a more reasonable interpretation, according to which it means 'almost all (my clothes)' plus intensity.

Thus, it is possible that participants in the experiments we are reviewing here might not have a reason to interpret *all tigers have stripes* as if speaking strictly. Furthermore, given that they know that universal quantifiers are hyperbole-prone, they might interpret it as 'almost all/very many tigers have stripes' and thus judge it as true even though they are aware of the existence of albino tigers.

A further issue with *all* is that it is distinct from other English universal quantifiers like *every* and *each*, in terms of the property of 'distributivity'. At first glance, it might seem, for instance, that (8) and (9) mean the same thing. However, while *every* requires a singular nominal restrictor (cf. **Every girls run*), *all* requires a plural (cf. **All girl run*), and in episodic sentences *all* must always appear with a definite plural, as in (11), while *every* again requires a bare singular (12). This morpho-syntactic difference can be explained if we assume that *all* operates over plural sets/collectivities, while *every* operates over atomic individuals/singleton sets. Further support for this distinction comes from the fact that *every* forces a distributive interpretation, as in (13), while *all* allows for a collective interpretation, as in (14). As a consequence, *all* is compatible with collective predicates such as *gather,* as in (15), while *every* is not, as in (16) (see Beghelli and Stowell 1997 for further discussion).

(8) Every girl runs.
(9) All girls run.
(10) Every girl was running.
(11) All *(the) girls were running.
(12) Every (*the) girl(*s) was/(*were) running.
(13) Every girl built a raft.
  (i) different raft for each girl
  (ii) *a single raft built as a team
(14) All the girls built a raft.
  (i) different raft for each girl
  (ii) a single raft built as a team
(15) All the girls gathered in the pool.
(16) *Every girl gathered in the pool.

Thus, *all* is not only potentially ambiguous between logical 'all' and conventional/hyperbolic 'almost all', *all* is also ambiguous between distributive and collective interpretations. A collective interpretation of *all ducks lay eggs* could be that 'ducks, collectively/as a set, lay eggs', which is similar to the kind reading 'ducks, as a kind, lay eggs'. This is true, providing yet another alternative route to an apparent GOG response.

In sum, *all* is not straightforwardly the best choice for a universal quantifier if one wishes to study the GOG effect – there are several reasons to doubt that it is reliably and consistently interpreted as $\forall$, independently of any possible tendency to interpret universals as generics.

It is clear from the above comparison between different universal quantifiers that any general claim about the tendency of universal quantifiers to be interpreted as if they were generic cannot overlook variation in the realisation of universally quantified statements within a language. Thus, experiments that compare generic and different types of universally quantified statements such as *all the, every* or *each* need to be done (see Lazaridou-Chatzigoga and Stockall 2013 for the first experiment that addresses this challenge). Additionally, as will be discussed more in detail in section 4.2, the different sensitivity universal quantifiers show to quantifier domain restriction might prove crucial in order to understand the purported GOG effect.

We have argued that ignorance, subkind interpretations, and the atypical behaviour of *all* may each contribute significantly to explaining why participants judge statements like *all tigers have stripes* to be true in Leslie, Khemlani, and Glucksberg's (2011) experiments, without needing any appeal to a Generics-as-Default explanation. The fourth possible alternative explanation, Quantifier Domain Restriction, is more complicated than the others and merits its own section.

## 4 Quantifier Domain Restriction as an explanation for the GOG effect

### 4.1 What is Quantifier Domain Restriction

In this section we will focus on Quantifier Domain Restriction (QDR henceforth) and whether it can explain these unexpected results about the behaviour of *all* that Leslie, Khemlani, and Glucksberg (2011) have used to support the GaD view on generics. An approach that relies on QDR has the additional advantage of accounting for the psychological data without postulating ad-hoc mechanisms just for generics.

We argue that the rejection of the QDR explanation by Leslie and colleagues was not justified. Specifically, merely providing participants with population estimates is not enough to make QDR to only the relevant (potentially egg-laying) ducks impossible, and may even serve to make this subset interpretation salient by priming a 'biology class' mode of thinking.[9] Thus, we think it is quite plausible that when people accept a statement like *all ducks lay eggs* they interpret it as a claim only about the relevant restricted set of mature fertile female ducks.

QDR is a pervasive phenomenon affecting quantifiers and their interpretation within a context, and is routinely invoked in quantification (Heim 1991). According to QDR, the domain of a quantifier can be restricted in the following sense: *everyone* in (17) does not quantify over all the individuals in the world, but rather over a contextually restricted set of individuals, i.e. the dinner guests who had rhubarb pie for dessert (example modified from von Fintel [1994, 33]). Furthermore, listeners are known to be charitable (Grice 1975). Thus, in a conversation one assumes that speakers take the most sensible positions and make the most plausible assertions. Under this view, interpreting *everyone* as quantifying over all the individuals in the world seems a rather unlikely intended interpretation and moreover one that is not charitable to the speaker because it renders her utterance false, whereas interpreting

[9] Leslie et al. (2011) do not provide a complete set of materials, only the example of a context and generalising statement pair we give in (7) above, so we can only speculate about the extent to which this example is representative.

*everyone* with respect to the available set of individuals is not only plausible but also charitable to the speaker.

(17) There was rhubarb pie for dessert. Everyone developed a rash.

How we encode QDR in the grammar is currently under debate and opinions vary as to whether QDR is part of the syntax/semantics or not, and, if yes, where exactly we should place the covert domain variables, on the nominal (Stanley and Szabó 2000; Stanley 2002) or on the quantifier (von Fintel 1994; Martí 2003; Giannakidou 2004). One execution of the approach that represents contextual variables at LF (Stanley and Szabó 2000) argues, for instance, that the nominal argument of *every* in (18), in this case *lion*, is contextually restricted by an unpronounced domain restriction variable [i] which is present in the logical form. Other approaches propose that unarticulated constituents of this sort can be supplied by the pragmatic process of free enrichment (Recanati 2002; Hall 2008) and need not be postulated in the logical form. Finally, another line of proposals follows a situation-based view of domain restriction (Cooper 1995; Recanati 1996; Schwarz 2012; Kratzer 2004), according to which situation arguments, rather than covert domain variables, are responsible for domain restriction.

(18) *Context: There are lions and tigers in this cage.*
Every lion[i] is dangerous.

Arguably, this is not the place to decide upon this issue (see Kratzer [2004] for an overview). Irrespective of whether one represents QDR as a syntactic, semantic or pragmatic phenomenon, there is an abundance of independent motivation for its existence.

If we take seriously the possibility of QDR as an alternative explanation to GOG, then participants who accept universally quantified statements like *all ducks lay eggs* are not interpreting them as generic. Instead they are simply engaging with the process of QDR, which is naturally implicated in the interpretation of universally quantified statements in any case. The following thought experiment spells out this alternative in detail.

### 4.2 A thought experiment involving QDR

If we could show that the amount of 'GOG' behaviour can be altered by carefully manipulating different levels of contextual information preceding the critical utterance, we would have evidence that what is at play is QDR and not GOG. Including quantifiers with different sensitivity to QDR would further sharpen the predictions and make potential results stronger.

Rather than the population statistics contexts used by Leslie, Khemlani, and Glucksberg (2011), which did not seem to have any (big) effect on participant behaviour, we suggest using contexts that make the relevant domain for QDR salient. To fully test the effect of such contexts, one could vary the type of contextual information preceding the critical utterance as follows: a) neutral, where the information in the context does not interact with the truth value of the critical statement; b) contradictory, where an exception which should rule out a universally quantified statement is made salient, and c) supportive, where the generality of the critical property is made salient.

In addition to manipulating context, however, a compelling test of the QDR view would also require comparing *all* (which allows, but does not require QDR) to other universal quantifiers like *all the* and *each* (which require it because of their semantics).

An illustration of the three levels of context for a majority characteristic property in a universally quantified statement is given below:

(19)
a. neutral context:
Linton Zoo is home to three tigers, Tibor, Baginda and Kaytlin, whose playful games visitors love to watch and photograph.
b. contradictory context:
Linton Zoo is home to three tigers, Tibor, Baginda and Kaytlin, whose fur is all white due to a recessive gene that controls coat color.
c. supportive context:
Linton Zoo is home to three tigers, Tibor, Baginda and Kaytlin, whose black and orange coats visitors love to photograph.

Statement:
All tigers have stripes./All the tigers have stripes./Each tiger has stripes.

On the GaD view, all three contexts should give rise to similar levels of GOG responses, for all three quantifiers. The 'more marked' system 2 quantified statements should be 'erroneously' interpreted as generic statements at similar rates to previous experiments (somewhere between 30% and 90% of the time). The contradictory context may act like the knowledge probe task in Leslie, Khemlani, and Glucksberg's (2011) experiment 3, and reduce acceptance rates for quantified statements by 15%-20%, but proponents of the GaD view would have no reason to expect a difference between the neutral and supportive contexts, or between different quantifiers, since the GaD view is that participants fail to interpret the quantifiers as universal quantifiers, and that the generic statement counterparts (*tigers have stripes*) should be judged as true in all three contexts.

The predictions of the QDR account, though, critically diverge from the GaD view with respect to the different universal quantifiers, since we predict variation in different contexts because of a range of different reasons for participants to offer a 'true' response. We expect that participants' responses are not biased towards a generic interpretation, but are rather dependent on the sensitivity of the quantifier to QDR and the availability of kind level interpretation.[10]

More specifically, QDR is less likely if the universally quantified statement used does not require linking with a set under discussion, as is the case with *all,* compared to *each* and *all the*, which do (Partee 1995). For instance, the QDR view would predict that providing a context in which the existence of tigers that do not have stripes is asserted (the contradictory context) should evoke very low acceptance rates for *all the* and *each* statements, while *all*-statements could still trigger relatively high acceptance rates. Participants are required to interpret the discourse-linked (D-linked, as in Pesetsky [1987]) quantifiers as referring to the stripeless tigers just

[10] Another potential licensor of a 'true' response, which we do not discuss here due to space limitations, is the availability of a prototypical interpretation. According to this interpretation, *all tigers have stripes* might be judged as true, because the prototype of the concept TIGER is strongly associated with the concept STRIPE.

introduced in the context – if we observed GOG-like responses to these items in contradictory contexts, we would have a reason to reject the QDR explanation and accept that participants might be erroneously interpreting quantified statements as generics. With *all*-statements, however, participants could adopt a QDR interpretation, but they are not required to. They could alternatively generate any of the other options discussed above (kind-level interpretation, loose/hyperbolic interpretation), which would licence a 'true' response. They could not, however, be ignorant of the existence of stripeless tigers, so we would still predict that acceptance rates would be lower in the contradictory context than in the neutral or supportive contexts.

Supportive contexts should evoke the mirror pattern for *all the* and *each*: explicitly predicating variegated coloration of the tigers under discussion means that the only licit judgment to these statements would be 'true'. The difference between these quantifiers and the *all*-statements may be minimal in supportive contexts, however, given that the QDR interpretation is easily available, and charitable, and that the non-QDR alternatives also licence a 'true' response.

The neutral context would provide a baseline to observe effects of differences between quantifier types, while keeping the information supplied by QDR constant. To judge whether *all the* or *each* statements are true in the neutral context, participants will be required to access their own knowledge about the kind under discussion; 'true' answers will be due to ignorance of/failure to recall the facts. The *all* statements in the neutral contexts could evoke a 'true' response for any of the reasons discussed above (ignorance, kind interpretation, loose/hyperbolic interpretation).

By using this careful manipulation of context, one could aim to show that people interpret generic and universally quantified statements relying on their knowledge of the semantics and pragmatics of genericity, quantification and on their availability for a contextually restricted interpretation (see Lazaridou-Chatzigoga and Stockall [2013] for preliminary support for such manipulations, as well as Lazaridou-Chatzigoga, Katsos and Stockall [2016] for an implementation of this thought experiment).[11]

We have made a strong case for the view that QDR is one of the four non-mutually exclusive explanations for the behaviour that looks like a 'GOG' effect and we have outlined how one could further investigate such a claim. In the next section we revisit, and question, the argument that the observed GOG effect is what we would expect given the GaD.

## 5 Cognitive load and processing demand

In this section we would like to turn to yet another argument addressing the status of the GOG effect. Given that the GOG effect describes the tendency to interpret false

[11] Manipulating context might additionally illuminate the relevance of QDR for generics themselves, something that seems to go against the received view for generics (Krifka 1987), but is in line with more recent work, according to which generics display some context sensitivity (Sterken [2015], see also discussions in Greenberg [2007] and Carlson [1999]). In Lazaridou-Chatzigoga and Stockall (2013) we found that minority characteristic generics like *ducks lay eggs* preceded by a contradictory context that made specific male ducks salient, pattern with *all*-statements in eliciting very slow truth value judgments, as compared to *all the* or *every* statements. These results suggest that participants might engage in a costly process of domain restriction even with generics. However, this pilot study did not manipulate enough features of the context to be conclusive, so more research is needed in order to establish a possible link between generics and context.

universals as the corresponding true generics, the GOG is treated as an *erroneous* judgment people make. Furthermore, it is argued that the GOG is more likely to be observed in situations of cognitive load and processing demand. As will be shown here, actual evidence that GOG 'errors' are more likely to occur due to higher cognitive demand is rather sparse.

In the case of Hollander, Gelman and Star's (2002) acquisition data, we might accept that generalising about abstract entities is more cognitively demanding than generalising over a specific set that is readily available. However, it is not clear what is so cognitively demanding in the majority of the other experiments reported in the literature that claim to show GOG errors with adults. These data mainly concern adult native speakers who participate in truth value judgment tasks, who are all competent both with generics and quantifiers.

For example, Leslie, Khemlani, and Glucksberg's (2011) experiment 1 was administered via the Amazon Mechanical Turk online data collection platform, which allows participants to complete questionnaires for payment, from the comfort of their own home or office. Participants were required to read simple sentences such as *ducks lay eggs*, and select a button labelled 'yes' or 'no' to indicate whether they agreed with the statement. Participants were not required to respond within any particular time frame, and were not required to engage in any other task at the same time. Participants each judged 132 items in total and they likely spent anywhere from 15-30 minutes depending on how quickly they worked through the task.[12]

These non-obviously demanding tasks resulted in very high rates of apparent GOG: "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" (Leslie, Khemlani, and Glucksberg 2011, 21).

It is not clear to us that a 78% error rate in response to a very simple task is consistent with the hypothesis that the GOG effect should occur "from time to time" (Leslie 2008, 25), especially when this rate varies by condition, while the task remains the same.

But more generally, in order to convincingly claim that a task like this is cognitively demanding and that this is the reason why the processing load is high enough so that GOG errors are more likely to be observed, one needs to actively manipulate this dimension. This has simply not been done in the vast majority of the experiments that claim to show the GOG effect. There is nothing prima facie challenging about the tasks the participants were faced with, and, no evidence that higher than normal cognitive demands played any role in generating the high levels of acceptance of 'false' quantified statements. It is beyond the scope of the current paper to review the enormous literature on cognitive load, working memory, individual differences, manipulated task demands etc. that are relevant to this issue. Interested readers may find Lewis, Vasishth and Van Dyke's (2006) review a useful starting place.

It suffices to mention just one relevant, highly influential result from decades old research on relative clause processing. King and Just (1991) report two experiments comparing how well individuals with high vs. low working memory

[12] Note that this is merely informed speculation. Leslie, Khemlani and Glucksberg (2011) do not provide details about the duration of the experiment, or report any kind of variance in performance between participants.

capacity (as assessed by the Reading Span test) are able to read and comprehend subject vs. object relative clause containing sentences under conditions of varying concurrent working memory load.[13] Using a moving-window self-paced reading paradigm (Just, Carpenter, and Wooley 1982), and a working memory load manipulation in which subjects had to recall 0, 1 or 2 cued words from the sentence(s) they read, King and Just found that all three manipulations (sentence type, number of words to recall, and participant working memory span) affected recall rates, comprehension question accuracy and reading times. It is not necessary for present purposes to discuss the details of all the main effects and interactions, or their interpretations, but, it is worth noting that King and Just found that for High Span participants, comprehension accuracy for object relatives declined 10% as the recall load increased from 0 to 2, but there was no decline in their accuracy for the easier subject relatives.

Thus, in an experiment with a much more demanding basic task (moving-window self-paced reading, followed by complex comprehension questions), and linguistic materials, where even the 'easy' subject relatives condition is considerably more complex than the 3-5 word sentences in Leslie, Khemlani, and Glucksberg's (2011) experiment 1 (compare: *The reporter that attacked the senator admitted the error publicly after the hearing* with *All ducks lay eggs*), King and Just find only a 10% increase in error rates associated with an increasingly demanding concurrent cognitive task, only for object relative clauses. And even then, the overall error rates only reached about 20%.

Seen in this context, Leslie and colleagues' explanation for experiment 1, whereby 78% of the time their participants erroneously judged statements like *all tigers have stripes* to be true because of high cognitive load is untenable.

Furthermore, the only experiment we are aware of that actually manipulates task demands in the processing of generics is Meyer, Gelman and Stilwell (2011). The rationale behind their experiment was that if generics are default, it should be easier to judge that a property is characteristic of a generic (i.e. dogs) than of a quantified set (i.e. all dogs) and they expected this generic advantage when participants were under time pressure. Meyer et al. recorded truth value judgments and judgment times in response to majority characteristic generic vs. *all* quantified statements like *{all} dogs have four legs* and definitional statements like *{all} giraffes have long necks*, and varied whether participants were told to answer as quickly as possible, or to take as much time as they wanted.

The main results were that a) participants in the speeded condition were more accurate at making a truth value judgment (90% for generics vs. 55% for quantified statements) and faster to respond to generics (~900ms) than to universals (~1100ms) and b) for definitional statements, participants were faster to respond to generics than universals in both the speeded (~900ms vs. ~1000ms) and the unspeeded (~1400ms vs. ~1800ms) condition. Both results seem to be consistent with the GaD hypothesis: participants were, indeed, more likely to say 'true' to *all dogs have four legs* in the speeded condition than in the non-speeded condition, and were slower to judge quantified than generic statements, which suggests quantification is effortful, and

[13] Object relative clauses have been long known to be more effortful to process than matched subject relatives (Holmes and O'Regan 1981), and this difference is typically taken to reflect the increased processing demands of retaining the wh-dependency in working memory longer in object relatives (compare: subject relative: The author who [t knew the editor] met her friend. vs. object relative: The author who [the editor knew t] met her friend.)

participants were more susceptible to a GOG error when they were under time pressure.

But these results merit a closer look. In the unspeeded condition, participants were numerically, though not statistically, **less** accurate at judging the truth of generic majority characteristic statements than in the speeded condition (85% vs. 90%), while they were significantly more accurate for quantified statements (80% vs. 55%). The difference between the 'error' rate for the majority characteristic quantified statements between the speeded and unspeeded conditions is about 35% - a rather large effect for a relatively minor cognitive demand increase, but still nowhere near the 78% error rate found by Leslie, Khemlani, and Glucksberg's (2011) experiment 1, which imposed no time pressure or other high-load demands on participants.

These results then, suggest that while it is indeed possible to increase the truth value judgment error rate for at least the majority characteristic quantified statements by increasing task demands, processing load is unlikely to come close to accounting for the GOG effects reported in the literature.

We propose, instead, that most of the time, an apparently GOG response is not, in fact, a processing error at all. We argue that mechanisms for which we have independent motivation can account for these kinds of responses, i.e. Quantifier Domain Restriction, as well as subkind (taxonomic) interpretation and the semantic and pragmatic ambiguity of *all.* And Leslie, Khemlani, and Glucksberg (2011) themselves find that ignorance of, or failure to recall, the facts can account for 18% of GOG responses. Thus, evidence that the GOG responses are erroneous judgments is much less compelling than it might at first appear.

## 6 Conclusion

We observed that there is both a long-established interest in the study of generic statements in the formal semantics literature, and a relatively recent surge in research on genericity from a psychological perspective. These views seem to take categorically different stances on genericity, as the formal semantics view treats generics as quantificational and the GaD view treats generics as categorically different from quantifiers (though exactly what this view entails requires further clarification). Nevertheless, each approach offers an explanation for important phenomena (relating to the interpretation of generics and the pattern of acquisition, among others) and each faces its own challenges. The main case that we focussed upon in this paper is the evidence from adult processing data that has come to light thanks to the intense psycholinguistic interest in genericity. We reviewed existing experimental evidence, outlined possible future experimental directions and concluded that a closer look at the semantics and pragmatics of generic and universally quantified statements, as well as a more careful consideration of the potential effects of processing demands, may provide a more nuanced explanation for the pattern of adult judgments than that proposed by the psychological GaD view.

We suggested that even the name of the GOG effect might be misleading. The effect mainly tries to capture the behaviour observed with the quantifier *all,* which supposedly gets a generic interpretation as a result of an overgeneralisation bias. Thus, perhaps a better name for that effect would be 'Quantifier Reanalysis' effect, because this term would direct the focus where we believe it belongs: on the interpretation of *all,* or more generally of quantifiers, rather than the interpretation of generic statements.

Ultimately, we believe that by integrating the tools and perspectives of both strands of investigation, we can gain a better understanding of what generic generalisations are, how they differ from quantificational generalisations and what the effect of context is, if any, when interpreting these two types of generalisations.

**Acknowledgements**

The authors would like to thank the two anonymous reviewers for their constructive comments, as well as Rachel Sterken and Bernhard Nickel, who organised the workshop *Genericity in Linguistics and Philosophy 'The Generic Workshop'* at Harvard University. We are grateful to all participants for the discussion, especially to John Collins for his insightful commentary on our presentation. The first author would also like to thank the organisers for the financial support received to attend the workshop. We also acknowledge the helpful feedback from audiences, where parts of this research project were presented: Queen Mary Ling Lunch Talk 2012, Sinn und Bedeutung 17 at École Normale Supériore, XPRAG London Master Class at UCL, XPrag 2013 at Utrecht University, London Semantics Day 1 at Queen Mary and Experimental Semantics and Pragmatics Workshop at the University of Cambridge.

**Funding**

The authors were partially supported by a British Academy/Leverhulme Trust Small Research Grant [SG-132271].

**References**

Beghelli, Filippo, and Tim Stowell. 1997. Distributivity and negation. In *Ways of Scope Taking*, edited by Anna Szabolcsi, 71-107. Dordrecht: Kluwer.

Behrens, Leila. 2000. *Typological Parameters of Genericity*. Arbeitspapier 37 (Neue Folge). Institut für Sprachwissenschaft, Universität zu Köln.

Carlson, Greg. 1977. *Reference to kinds in English*. PhD dissertation, University of Massachusetts, Amherst.

Carlson, Greg. 1999. Evaluating Generics. In *Illinois Studies in the Linguistic Sciences*, edited by Peter Lasersohn, 29:1, 1-11.

Carlson, Greg. 2011. Genericity. In *Semantics. An International Handbook of Natural Language Meaning* (HSK 33.2) edited by Klaus von Heusinger, Claudia Maienborn and Paul Portner, 1153-1185. Berlin: Mouton de Gruyter.

Chaiken, Shelly, Akiva Liberman, and Alice Eagly. 1989. Heuristic and systematic processing within and beyond the persuasion context. In *Unintended Thought*, edited by James S. Uleman and John A. Bargh, 212-252. New York: Guilford Press.

Cohen, Ariel. 1999. "Generics, frequency adverbs, and probability." *Linguistics and Philosophy* 22: 221-253.

Cohen, Ariel. 2004. "Generics and Mental Representation." *Linguistics and Philosophy* 27: 529-56.

Collins, John. 2015. Generics sans Gen. ms. Distributed at The Generic Workshop: Genericity in linguistics and philosophy, Harvard University.

Cooper, Robin. 1995. The role of situations in generalized quantifiers. In *Handbook of Contemporary Semantic Theory*, edited by Shalom Lappin, 65-86. Oxford: Blackwell.

Claridge, Claudia. 2011. *Hyperbole in English: A corpus-based study of*

*exaggeration*. Cambridge, UK: Cambridge University Press.
Csibra, Gergely, and Gergely, György. 2009. "Natural pedagogy." *Trends in Cognitive Sciences* 13(4): 148-153.
Dahl, Östen. 1975. On generics In *Formal semantics of natural language*, edited by Edward Keenan. 99-111. Cambridge: Cambridge University Press.
Dayal, Veneeta. 2004. "Number marking and (in)definiteness in kind terms." *Linguistics and Philosophy*, 27, 393–450.
Deo, Ashwini, and Madiman, Mokshay. 2015. "Scale structure and genericity: Understanding generics using stochastic comparison." Available at http://ling.auf.net/lingbuzz/002364
Dienes, Zoltan, and Perner, Josef. 2002. "What sort of representation is conscious?" *Behavioral and Brain Sciences* 25, 336-337.
Evans, Jonathan St. B. T., and David E. Over. 1996. *Rationality and reasoning*. Hove, East Sussex, UK: Psychology Press.
Farkas, Donka, and Yoko Sugioka. 1983. "Restrictive if/when clauses." *Linguistics and Philosophy* 6(2): 225–258.
Frederick, Shane. 2005. "Cognitive reflection and decision making." *Journal of Economic Perspectives*, 19 (4), 25–42.
von Fintel, Kai. 1994. *Restrictions of quantifier domains*. PhD dissertation: University of Massachussets, Amherst.
Gelman, Susan A. 2010. Generics as a window onto young children's concepts. In *Kinds, things, and stuff: The cognitive side of generics and mass terms*, *New directions in cognitive science*, edited by Francis J. Pelletier, 100-123. New York: Oxford University Press.
Gelman, Susan A., Peggy J. Goetz, Barbara W. Sarnecka, and Jonathan Flukes. 2008. "Generic language in parent-child conversations". *Language Learning and Development*, 4, 1–31.
Giannakidou, Anastasia. 2004. Domain restriction and the arguments of quantificational determiners. In *Semantics and Linguistic Theory 14*, edited by Robert B. Young. 110-128. Ithaca, NY: Cornell.
Greenberg, Yael. 2007. "Exceptions To Generics: Where Vagueness, Context Dependence And Modality Interact". *Journal of Semantics* 24(2), 131-167.
Grice, H. Paul. 1975. Method in Philosophical Psychology: From the Banal to the Bizarre. In *Proceedings and Addresses of the American Philosophical Association*, 23-53.
Hall, Alison. 2008. "Free enrichment or hidden indexicals." *Mind and Language* 23(4): 426-456.
Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. New York: Garland.
Heim, Irene. 1991. Artikel und Definitheit. In *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, edited by Arnim von Stechow and Dieter Wunderlich, 487-535. Berlin: de Gruyter.
Hollander, Michelle A., Susan A. Gelman, and Jon Star. 2002. "Children's Interpretation of Generic Noun Phrases." *Developmental Psychology* 36(6), 883-894.
Holmes, V. M., and J. K. O'Regan. 1981. "Eye fixation patterns during the reading of relative clause sentences.". *Journal of Verbal Learning and Verbal Behavior* 20, 417-430.
Jönsson, Martin L., and Hampton, James A. 2006. "The inverse conjunction fallacy." *Journal of Memory and Language* 55, 317–334.

Just, Marcel A., Patricia A. Carpenter, and Jacqueline D. Wooley. 1982. "Paradigms and Processes in Reading Comprehension." *Journal of Experimental Psychology*: General, 111 (2), 228-238.
Kahneman, Daniel, and Frederick, Shane. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman. 49-81, Cambridge University Press. New York.
Khemlani, Sangeet, Leslie, Sarah-Jane, Glucksberg, Sam, and Paula Rubio-Fernández. 2007. "Do ducks lay eggs? How humans interpret generic assertions." In *Proceedings of the 29th annual conference of the Cognitive Science Society*, edited by D. S. McNamara and J. G. Trafton, 395-400. Nashville, TN: Cognitive Science Society.
King, Jonathan, and Marcel A. Just. 1991. "Individual differences in syntactic processing: The role of working memory." *Journal of Memory and Language*, 30: 580–602.
Kratzer, Angelika. 2004. Covert quantifier restrictions in natural languages. Talk given at Palazzo Feltrinelli in Gargnano June 11, 2004.
Krifka, Manfred. 1987. An Outline of Genericity, partly in collaboration with Claudia Gerstner. Technical Report SNS-Bericht 87-25 Seminar für natürliche Systeme, Tübingen University Germany.
Krifka, Manfred, Francis J. Pelletier, Greg Carlson, Alice ter Meulen, Gennaro Chierchia, and Godehard Link. 1995. Genericity: An Introduction. In *The Generic Book*, edited by Greg Carlson and Francis J. Pelletier, 1-125. Chicago. Chicago University Press.
Labov, William. 1984. Intensity. In *Meaning, Form and Use in Context: Linguistic Applications*, edited by Deborah Schiffrin, 43-70. Washington: Georgetown University Press.
Lawler, John. 1972. Generic to a Fault. In *Papers from the 8th Regional Meeting, Chicago Linguistic Society* (CLS 8), 247–258.
Lawler, John. 1973. Studies in English generics. University of Michigan Papers in Linguistics 1(1). University of Michigan Press.
Lazaridou-Chatzigoga, Dimitra, and Linnaea Stockall. 2013. Genericity, exceptions and domain restriction: experimental evidence from comparison with universals. In *Proceedings of Sinn und Bedeutung* 17, edited by E. Chemla, V. Homer, and G. Winterstein, 325-343. École Normale Supérieure, Paris.
Lazaridou-Chatzigoga, Dimitra, Napoleon Katsos, and Linnaea Stockall. 2015. Genericity is easy? Formal and experimental perspectives. In *RATIO* 28(4), Special Issue: Investigating Meaning: Experimental Approaches, edited by Nat Hansen and Emma Borg, 470-494.
Lazaridou-Chatzigoga, Dimitra, Napoleon Katsos, and Linnaea Stockall. 2016. The effect of context on generic and quantificational statements. In *Pre-proceedings of Trends in Experimental Pragmatics*, edited by Fabienne Salfner and Uli Sauerland, 87-94.
Leslie, Sarah-Jane. 2007. "Generics and the structure of the mind." *Philosophical Perspectives* 21, 375-405.
Leslie, Sarah-Jane. 2008. "Generics: Cognition and acquisition." *The Philosophical Review* 117(1), 1-49.
Leslie, Sarah-Jane, Sangeet Khemlani, and Sam Glucksberg. 2011. "All Ducks Lay Eggs: The Generic Overgeneralization Effect." *Journal of Memory and Language* 65(1), 15-31.

Lewis, David. 1975. Adverbs of Quantification. In *Formal Semantics in Natural Languages*, edited by Edward Keenan, 3-15. Cambridge University Press.
Lewis, Richard L., Shravan Vasishth, and Julie A. Van Dyke. 2006. "Computational principles of working memory in sentence comprehension." *Trends in Cognitive Science* 10, 447-454.
Liebesman, David. 2011. "Simple generics." *Noûs* 45: 409-442.
Mari, Alda, Claire Beyssade, and Fabio del Prete. 2013. Genericity. Oxford: Oxford University Press.
Martí, Luisa. 2003. *Contextual variables*. PhD dissertation, University of Connecticut.
Meyer, Meredith, Susan A. Gelman, and Sarah M. Stilwell. 2011. Generics are a cognitive default: Evidence from sentence processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Boston, MA: Cognitive Science Society.
Partee, Barbara. 1995. Quantificational structures and compositionality. In Quantification in Natural Languages, edited by Emmon Bach, Eloise Jelinek, Angelika Kratzer and Barbara Partee, 541–601. Dordrecht: Kluwer.
Petty, Richard E., and John T. Cacioppo. 1986. *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.
Pesetsky, David. 1987. Wh-in-Situ: Movement and Unselective Binding. In *The Representation of (In)definiteness*, edited by Eric J. Reuland and Alice G. B. ter Meulen, 98-129. MIT Press: Cambridge, Mass.
Recanati, François. 1996. Domains of discourse. *Linguistics and Philosophy* 19: 445-475.
Recanati, François. 2002. Unarticulated constituents. *Linguistics and Philosophy* 25: 299-345.
Schwarz, Florian. 2012. Situation pronouns in Determiner Phrases. *Natural Language Semantics*, December 2012, Volume 20, Issue 4: 431-475.
Sloman, Steven A. 2002. Two systems of reasoning. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman. 379-396, Cambridge University Press. New York.
Stanley, Jason. 2002. Nominal restriction. In *Logical Form and Language*, edited by Gerhard Preyer and Georg Peters, 365-388. Oxford University Press.
Stanley, Jason and Zoltán G. Szabó. 2000. "On Quantifier Domain Restriction." *Mind and Language* 15 (2-3), 219-261.
Stanovich, Keith E. 1999. *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
Sterken, Rachel. 2015. "Leslie on Generics." *Philosophical Studies*, 172(9), 2493-2512.
Teichman, Matt. 2015. Genericity and quantification. ms. Presented at The Generic Workshop: Genericity in linguistics and philosophy, Harvard University.