Title
Generic and Universal Generalisations:Contextualising the ‘Generic Overgeneralisation’ Effect

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Lazaridou-Chatzigoga, Dimitra
Katsos, Napoleon
Stockall, Linnaea

Publication Date
2017

Peer reviewed

Generic and Universal Generalisations: Contextualising the 'Generic Overgeneralisation' Effect

Dimitra Lazaridou-Chatzigoga (dl518@cam.ac.uk)
Institut für Anglistik und Amerikanistik, Humboldt-Universität zu Berlin/University of Cambridge
Unter den Linden 6, 10099, Berlin, Germany

Napoleon Katsos (nk248@cam.ac.uk)
Department of Theoretical and Applied Linguistics, University of Cambridge
9 West Road, Cambridge, CB3 9DP, UK

Linnaea Stockall (l.stockall@qmul.ac.uk)
School of Languages, Linguistics and Film, Queen Mary, University of London
Mile End Road, London, E1 4NS, UK

Abstract

In this study, we focused on the Generic Overgeneralisation (GOG) effect (Leslie, Khemlani, and Glucksberg 2011) and tested the relevance of context and an explanation based on quantifier domain restriction for the pattern of judgement data observed. Participants judged generic majority characteristic statements like *tigers have stripes* or statements with universal quantifiers that have different sensitivity to context ('all', 'all the', 'each') preceded by one of three levels of context: a) neutral, where the information in the context does not interact with the truth value of the critical statement, b) contradictory, where it presents an exception which should rule out a universally quantified statement, and c) supportive. Our results suggest that proponents of the generics-as-default view ruled out context prematurely and that in fact context is a viable alternative explanation for much of the so-called GOG effect.

Keywords: context; generalisation; genericity; quantification; quantifier domain restriction

Introduction

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):

(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 Dahl 1975; Carlson 1977), how to characterise their semantic interpretation and how to model their truth conditions remains a controversial topic (see discussion in 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, 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):

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

Generic generalisations can be made using a wide range of different grammatical means, including definite and indefinite singulars and bare plurals in English, but no language has a unique, unambiguous marker of genericity 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 a recent psychological approach to generics, capitalises upon.

This 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, 2008, 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 2-system view of cognition 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. 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.

The GaD approach argues that the fact that there is no overt generic operator in any known language is because generics are the unmarked, System 1, case. On this view, only effortful, non-default quantificational generalisations require overt linguistic exponence. 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. Leslie refers to Chomsky (2000), but she gives no definition of markedness. Intuitively, it seems that what is at stake is surface level overt realisation (third notion of markedness of Haspelmath 2006).

A challenge for both types of approach is to determine which properties can be used in generic statements. Generic generalisations range from definitional statements that do not tolerate any exceptions (*triangles have three sides)* to statements that involve characteristic properties and are true of the majority of the kind with only a few exceptions (*tigers have stripes*) - which are called 'majority characteristic' by Leslie et al. (2011) - through statements that are true of a minority of the kind yet are characteristic *(ducks lay eggs*) - which are called 'minority characteristic'- to statements that have low prevalence but involve a property that is noteworthy in some way (*sharks attack people*) - which are called 'striking'. A further complication is that statistical prevalence is neither a necessary nor a sufficient condition to license generic generalisations, as statements like *books are paperbacks* may be true of the majority of the kind, but are typically judged as false and thus fail as generic generalisations ('false generalisations').

In Lazaridou-Chatzigoga, Katsos and Stockall (2015) we juxtaposed the linguistic approach to genericity and the experimental research investigating the GaD hypothesis and we highlighted the significant challenges for each approach. We concluded that interdisciplinary work, integrating the tools and perspectives of both strands of investigation, is needed in order to advance our understanding of genericity.

In Lazaridou-Chatzigoga, Stockall and Katsos (2017) we focused on the effect called 'Generic Overgeneralisation' (GOG) (Leslie et al. 2011), which has been used to support the GaD view on generics. The Generic Overgeneralisation (henceforth 'GOG') effect is "the tendency to overgeneralise the truth of a generic to the truth of the corresponding universal statement" (Leslie et al. 2011:17). In that paper, we discuss a set of four non-mutually exclusive explanations for the GOG effect: a) ignorance of the relevant facts, b) subkind (taxonomic) interpretation, c) the atypical behaviour of *all* and d) Quantifier Domain Restriction (QDR), which will be explained in more detail in the next section. We proposed 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 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. The overall aim in that paper was 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.

In this paper, we address the effect of context on generic and universally quantified generalisations empirically.

The GOG effect

The first detailed investigation of the GOG effect is found in Leslie et al. (2011). In their experiment 1, participants performed a truth-value judgement task on sentences that were presented in one of three forms: generic, universal (*all*), or existential (*some*). The statements involved different kinds 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 generalisations (*Canadians are right-handed*). The authors report that adults sometimes judge universal statements as true, despite knowing that they are truth-conditionally false. For example, participants judged a quantified statement like *all tigers have stripes* as true, even though they know it is false given that there are albino tigers. The authors claim that the participants made this 'error' because they relied on the corresponding generic statement, which is true. They find that the GOG effect is restricted to characteristic properties and that it occurs in more than half the trials: 78% for majority characteristic and 51% for minority characteristic statements.

The authors argue that these elevated acceptance rates 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. Before reaching that conclusion, the authors acknowledge three alternative explanations, which they argue are ruled out with subsequent experiments: a) *ignorance of the relevant facts*, namely, that participants do not know that male ducks do not lay eggs, which they ruled out by administering a knowledge test that confirmed knowledge of the relevant facts (their experiment 3), b) *subkind interpretation,* namely, that participants interpret *all ducks lay eggs* as 'all kinds/types of ducks lay eggs', which is a true statement, which they discarded through a paraphrase task (their experiment 2b), where only 1% of the paraphrases explicitly mentioned subkinds, and c) Quantifier Domain Restriction (henceforth QDR), to which we turn in the next paragraph.

Under an explanation based on *QDR* participants might interpret a statement like *all ducks lay eggs* as applying only to a relevant subset of ducks, namely the mature fertile female ducks. Quantified statements are interpreted within a context, which may restrict the scope of the quantifier (see Stanley and Szabó 2000). Thus, the reason why people accept the *all* statement is because (they believe) it is true once the quantifier has been restricted to the relevant subset of ducks. The authors addressed this alternative explanation in their 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:

(6) ''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 make it difficult/impossible to interpret *all* as restricted to only the 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 of the kind above. However, the GOG effect still occurred on a substantial portion of trials for statements with *all*, with a 60% acceptance rate 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 percentage. The authors thus concluded that domain restriction could not be the sole explanation for the GOG effect.

On the above grounds, Leslie et al. (2011) rejected all three alternative explanations and argued they had found evidence for a strong generic bias, according to which people sometimes treat universally quantified statements as if they were generic.

Overview of the present study

In the present study, we addressed QDR as an explanation for the GOG effect building on a design used by Lazaridou-Chatzigoga and Stockall (2013). We chose to focus on the relevance of context and QDR given that the latter 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 universal quantifier can be restricted in the following sense: in a discourse like *'There was rhubarb pie for dessert, Everyone developed a rash'* (example modified from von Fintel 1994) a quantifier like *everyone* does not quantify over all the individuals in the world, but rather over the contextually restricted set of individuals. 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 *everyone* with respect to the available set of individuals is not only plausible but also charitable to the speaker.

We hypothesised that 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 the observed tendency to accept universally quantified statements as true can be explained through independently motivated mechanisms and that there is no need to appeal to GOG.

Rather than the population statistics contexts used by Leslie et al. (2011), which only had a moderate effect on participant behaviour, we decided to use three different types of contexts. Furthermore, because of the design we adopted, we decided to focus only on majority characteristic statements ('tigers have stripes') leaving minority characteristic statements ('ducks lay eggs') for future investigation. We varied the context 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 exceptions which should rule out a universally quantified statement are made salient, and c) supportive, where a paraphrase of the critical property is given, which makes its generality salient. Two of the three context types (contradictory and supportive) made the relevant domain for QDR salient, while the neutral context served as a baseline measure. The contradictory and supportive contexts turned the implicit restriction to 'all normal' individuals to an explicit one by either highlighting some *abnormal* individuals (contradictory) or by stating that the relevant individuals had the relevant property, i.e. they were *normal* individuals (supportive).

In addition to manipulating context, however, a compelling test of the QDR view also requires testing whether the GOG effect is observed only with *all* or whether different universal quantifiers would show such an effect. There are reasons to believe that *all* should not be treated as a representative universal quantifier. It has been argued to a) participate in fallacious reasoning (Jönsson and Hampton 2006), b) be prone to hyperbolic/loose use similar to 'almost all' (Claridge 2011), and c) be ambiguous between distributive and collective interpretation (Beghelli and Stowell 1997). Thus, using different types of universal quantifiers is essential to test the scope of the GOG effect. Furthermore, a study that specifically addresses the relevance of QDR should include universal quantifiers with different sensitivity to QDR. More specifically, QDR is less likely if the universal quantifier used does not require linking with a set under discussion, as is the case with *all*, compared to *each* and *all the,* which have to be interpreted as D(iscourse)-linked (Partee 1995, Pesetsky 1987).

To recapitulate, (a) the contextual manipulations used were expected to make the implicit domain restriction explicit and salient to the participants and (b) this manipulation was expected to influence truth-value judgements by showing a decrease in acceptance rates in the contradictory condition and an increase in acceptance rates in the supportive condition.

Method

Participants and procedure

120 volunteers (49 male, 70 female, 1 other; aged 19-67; mean age 37.28; SD 13.06) participated in the experiment

over the Internet. Participants were recruited through Amazon's Mechanical Turk system for human interface tasks. All spoke English as their first language and lived in the United States. The study was presented in the online platform Qualtrics. Each trial consisted of three displays. In the first display participants read a background context, in the second display they read a statement and in the third display they were asked to judge whether they agreed with the statement they just read. Their response was recorded by selecting keyboard keys ('A' for yes and 'K' for no).

Materials and design

Participants were presented with 84 statements, including 48 fillers presented in a randomised order. The 12 test items consisted of majority characteristic statements like *tigers have stripes* and *horses have four legs*. We included 26 control items, 12 definitional statements like *ants are insects* and 12 false generalisations like *books are paperbacks* to get baseline measures and to (semi)-counterbalance the percentage of expected True/False responses. All the contexts and items were normed beforehand by English native speakers, who did not take part in the experiment. This was done to make sure that the context manipulations worked as intended. Most of the experimental items used are a subset of the items used by Leslie et al. (2011). The two conditions we manipulated for the majority characteristic items were: a) determiner type: bare plural generic/*all/all the/each*, and b) context type: neutral/contradictory/supportive. The statements were presented in one of the determiner forms and were preceded by one of three types of context: a) neutral, b) contradictory, or c) supportive, examples of which are given below:

(7)

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 colour.

c. **supportive context:** Linton Zoo is home to three tigers, Tibor, Baginda and Kaytlin, whose black and orange coats visitors love to photograph.

Given the 4 determiners (generic/*all/all the/each*) we created 4 lists with 3 sublists each that varied with respect to the pairing of the items with context type, which gave us 12 lists in total. There were 10 participants in each sublist, who were assigned randomly. Here is a sample of a trial of a statement with *all* after a neutral context:

(8)

DISPLAY 1:
Background:
Linton Zoo is home to three tigers, Tibor, Baginda and Kaytlin, whose playful games visitors love to watch and photograph.

DISPLAY 2:
Statement: All tigers have stripes.
DISPLAY 3:
Do you agree with the statement?
o Yes (A) o No (K)

The definitional and false generalisations were in the generic form in all lists. Fillers served to ensure the percentage of expected True/False responses was similar. The definitional and false generalisations, as well as the fillers, were preceded by a context that did not vary across conditions. The materials can be viewed at http://www.dimitra-lazaridou-chatzigoga.com/cogsci-paper/

Results and discussion

The final model used included 116 out of the 120 participants. 4 participants were excluded based on their responses to definitional statements; we excluded subjects that responded correctly at fewer than 10 out of 12 items.

Acceptance rates

Table 1 summarises the proportion of 'TRUE' responses to the TVJ question for the test items (majority characteristic statements) in each condition. We report proportion of 'TRUE' responses rather than the actual number of responses to facilitate comparison with Leslie et al.'s (2011) results.

Table 1: Mean Proportion (SE) of 'TRUE' responses as a function of context and determiner type.

	Context		
	neutral	**contradictory**	**supportive**
determiner			
GEN (ø)	99.14 (3.12)	87.07 (0.86)	100 (0)
all	80.56 (3.82)	48.15 (4.83)	87.96 (3.14)
all the	78.33 (3.78)	37.50 (4.43)	90 (2.76)
each	79.17 (3.72)	30.83 (4.23)	85.83 (3.2)

As we see above, generics were accepted at higher rates overall than universals, as expected, given that we had chosen items that were true in generic form. Both in the neutral and the supportive condition acceptance rates for generics were at ceiling (99% and 100% respectively) and were only lower in the contradictory condition (87%). With universals, the picture is more complicated. All three universals (*all, all the, each*) were accepted at similar rates in both the neutral and the supportive condition, showing only a small increase in the supportive condition. In the neutral condition, *all*-statements were accepted 81% of the time, *all the*-statements 78% of the time and *each*-statements 79% of the time. In the supportive condition, *all*-statements were accepted 88% of the time, *all the*-statements 90% of the time and *each*-statements 86% of the time. Universals after a contradictory context yielded fewer acceptances overall, as expected: *all*-statements were accepted 48% of the time, *all the*-statements 38% of the time and *each*-statements 31% of the time. Acceptance rates for both generics and universals differed significantly between the neutral and the contradictory condition.

On the surface, we do get many 'TRUE' responses to universal quantifiers, as in the GaD literature, which might look like a GOG effect. We predicted that contradictory context should decrease acceptances across the board, while supportive context should increase them. We expected a smaller effect for generics because the generic statements were constructed so as to be true and because of their resistance to contextual restriction (i.e. we expected ceiling effects) and we predicted differences between the universal Qs depending on their sensitivity to QDR/D-linking. Nevertheless, we had specific predictions about the *relative* rates depending on the level of context, which according to Leslie et al. should not differ. In order to appreciate the relative effect of context on acceptance rates, we subtracted the average means of the contradictory condition from the average means of the neutral condition, as well as the average means of the supportive condition from the average means of the neutral. We took acceptance rates in the neutral condition as our baseline plotted at 0. Negative values mean fewer acceptances and positive values mean more acceptances. We interpreted the rates obtained as the relative effect of context on acceptance rates plotted in figure 1 below.

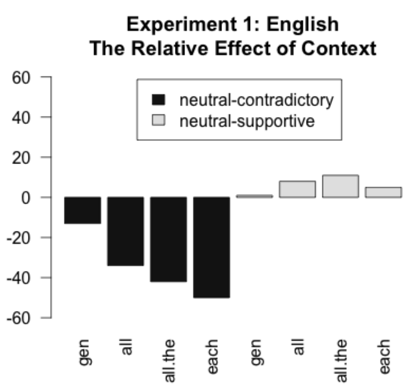

Figure 1: The relative effect of context.

We get the effect for the contradictory context exactly as predicted for the universal Qs. The relative effect is bigger for those quantifiers that require QDR because of their semantics (*all the, each*) than for the one that allows but does not require QDR to the relevant subset (*all*). We also find that context does affect GEN in the contradictory context.

The prediction about the supportive context was not borne out because of ceiling effects in the neutral condition. Adding explicit information supporting the statement hardly mattered, as acceptance rates did not rise significantly. The ceiling effect might be due to participants being charitable and/or exceptions not being immediately salient.

We used R (R Core Team, 2012) and the lme4 package (Bates et al. 2015) to perform a generalised mixed-effects linear analysis of the relationship between determiner and context, specifying a binomial family. Responses were treated as a dummy coded categorical variable and were modelled with glmer. First, we fitted a full model with det.type and context.type as fixed effects (with an interaction term) and with random intercepts for subjects and items. We performed a likelihood ratio test of the full model with an interaction term against a model without the interaction term and the comparison proved non-significant ($\chi^2(6) = 8.3455$, $p = 0.2139$). Thus, including an interaction term did not improve model fit, so we used the model without the interaction term for all subsequent analyses/comparisons.

We then fitted versions of the full model, in which a single effect was removed and we compared the reduced model to the model without interaction. To test the main effect of context, we removed context. A likelihood ratio test of the model without interaction against the model without context proved highly significant ($\chi^2(2) = 311.81$, $p < 2.2e\text{-}16$). Thus, we concluded that there was a main effect of context. To test the main effect of determiner, we removed determiner. A likelihood ratio test of the model without interaction against the model without determiner proved highly significant ($\chi^2(3) = 58.183$, $p = 1.436e\text{-}12$). Thus, we concluded that there was a main effect of determiner.

General Discussion

We set out to explore one of the alternative explanations for the judgement data that concern universally quantified statements, which have been used as evidence of a GOG effect. The study presented here provides experimental evidence for the relevance of a QDR-based explanation of the purported GOG effect. In our study, context did not only affect acceptance rates for *all* and other universal quantifiers (*all the, each*), but it further predicted the levels of QDR depending on the level of context. The effect of context was greater for *all the* and *each,* two quantifiers that require QDR, while it was smaller for *all,* whose domain is only optionally restricted. Leslie et al. had ruled out the relevance of context and hence predicted no differences in acceptances across contexts for *all.* Furthermore, even though they only discuss *all* they make general claims about (universal) quantification being prone to the GOG effect. We argue that drawing conclusions about universal quantification (and by extension about genericity) requires more subtle manipulations. The differences we found between the different universal quantifiers are predicted according to the QDR view advanced here, but ought to be inconsistent with the GaD view, were they to discuss them.

We also find that context matters for generics too, a fact that bears further investigation, but is in line with recent work that claims that generics display some context sensitivity (Sterken 2015). This might be more consistent with an analysis in which GEN also involves (some form of) quantification rather than one that treats GEN as categorically/ontologically different from universal quantifiers.

Overall, we argue that there exist alternative explanations for big portions of the supposed GOG effect. The study discussed here did not address all the alternatives, but so far in the literature it has been shown that at least pure error, ignorance and now context all significantly affect acceptance rates. In work in progress, we address cross-linguistic variation in the realisation of generic and universal generalisations. The general thrust of this work is that, rather than being under the influence of a default bias, adults are simply sensitive to the subtle interplay of quantifier semantics and pragmatics on the one hand, and context on the other. This approach has the advantage of accounting for data without postulating ad-hoc mechanisms such as GOG just for generics.

Acknowledgments

This work was supported by a British Academy/Leverhulme Trust Small Research Grant (SG-132271). We are grateful to Faidra Faitaki and Valentina Hu for their help with experimental materials and data entering, to Anne Beshears, Patrick Elliott, Shivonne Gates, David Hall, Jessica Soltys, Tom Stanton, Nikola Vukovic and Elspeth Wilson for their help with norming and the experimental software, and to Stephen Politzer-Ahles for advice on the statistics.

References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Beghelli, F., & Stowell, T. (1997). Distributivity and negation. In Ways of Scope Taking, edited by A. Szabolcsi, 71-107. Dordrecht: Kluwer.

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

Chomsky, N. (2000). New Horizons in the Study of Language and Mind. Cambridge: CUP.

Claridge, C. (2011). Hyperbole in English: A corpus-based study of exaggeration. Cambridge, UK: CUP.

von Fintel, K. (1994). Restrictions of quantifier domains. PhD dissertation: University of Massachusetts, Amherst.

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, 100-123. New York: OUP.

Grice, H.P. (1975). "Method in Philosophical Psychology: From the Banal to the Bizarre", *Proceedings and Addresses of the American Philosophical Association* (1975), 23-53.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics* 42, 25-70.

Heim, I. (1991). Artikel und Definitheit. In A. von Stechow and D. Wunderlich (eds.), Semantik: Ein internationales Handbuch der zeitgenössischen Forschung, 487-535. Berlin: de Gruyter.

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

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

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

Lazaridou-Chatzigoga, D., and Stockall, L. (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, D., Katsos, K. and Stockall, L. (2015). Genericity is easy? Formal and experimental perspectives. *RATIO* 28(4), 470-494.

Lazaridou-Chatzigoga, D., Stockall, L. and Katsos, N. (2017). A new look at the ‘Generic Overgeneralisation’ effect. *Inquiry.* Advance online publication. doi: http://dx.doi.org/10.1080/0020174X.2017.1285993

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., Khemlani, S. and Glucksberg, S. (2011). All Ducks Lay Eggs: The Generic Overgeneralization Effect. *Journal of Memory and Language 65*(1), 15-31.

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

Mari, A., Beyssade, C. and del Prete, F. (2013). Genericity. Oxford: OUP.

Partee, B. (1995). Quantificational structures and compositionality. In E. Bach, E. Jelinek, A. Kratzer and B. Partee (eds.). Quantification in Natural Languages. Dordrecht: Kluwer, 541–601.

Pesetsky, D. (1987). Wh-in-Situ: Movement and Unselective Binding. In the Representation of (In)definiteness, Eric J. Reuland & Alice G. B. ter Meulen, eds. MIT Press: Cambridge, Mass.

R CoreTeam (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

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

Sterken, R. (2015). Leslie on Generics. *Philosophical Studies*, 172(9), 2493-2512.