Generics and typicality[1]

Robert VAN ROOIJ — *Institute for Logic, Language and Computation, Amsterdam*

Abstract. Cimpian et al. (2010) observed that we *accept* generic statements of the form '*G*s are *f*' on relatively weak evidence, but that if we are unfamiliar with group *G* and we learn a generic statement about it, we still *interpret* it in a much stronger way: (almost) all *G*s are *f*. This paper makes use of notions like 'representativeness' and 'contingency' from (associative learning) psychology to provide a semantics of generics that explains why people *accept* generics based on weak evidence. We make use of the Heuristics and Biases approach of Tversky and Kahneman (1974) and the Associative Theory of Probability Judgements to explain pragmatically why people *interpret* generic statements in a much stronger way. The spirit of the approach has much in common with Leslie's (2008) cognition-based ideas about generics, but the semantics is grounded on Cohen's (1999) relative readings of generic sentences. The basic intuition is that a generic of the form '*G*s are *f*' is true, not because most *G*s are (or tend to have) *f*, but because *f* is typical for *G*, which means that *f* is valuably associated with *G*.

Keywords: generics, association, probabilities, pragmatics.

1 Introduction

Although generics are studied mostly in formal semantics and philosophy, they have recently attracted the attention of cognitive psychologists as well. The reason is that generics play a core role in the way we *learn*, *represent* and *reason* about groups in the world (cf. Leslie (2008)). Indeed, generic statements express very basic kinds of inductive generalizations, learned during the process of categorization. A central hypothesis of this paper is that the way we learn new categories is, and remains, of crucial importance for judgements involving those categories. I will argue that generic statements about categories, or groups, express typical information about these groups, and that the way we learn about a group is of crucial importance for what is typical about this group. The notion of *contingency* from associative learning psychology plays an important role in learning, and I will argue that a slight generalization of it is crucial for typicality as well, and thus for the analysis of generics.

After providing a biased overview of some semantic theories of generics (concentrating in particular on Cohen's proposal) in section 2 and theories of categorization in section 3, I will discuss my own semantic account of generics in section 4. This semantic analysis will give rather weak truth conditions to generic sentences. After that we will provide in section 5

---

[1]The ideas behind this paper were first presented in a conference on lying in Leiden, January 2017. Further presentations on the topic were given in Stockholm, Amsterdam and Zurich, before it was presented at the Sinn und Bedeutung conference in Berlin. I profited from the comments given after these talks, and in particular from those of Nicholas Asher, Nils Franzen, Manfred Krifka, Hannes Leitgeb, and Stephanie Solt. The comments on the abstract were also very useful. I discussed many ideas of this paper with Katrin Schulz. I like to thank all of these for their valuable comments. Finally, I would like to thank Jonathan Pesetsky for correcting my English.

a pragmatic explanation of why generics are normally interpreted in a much stronger way, making use of insights of Tversky and Kahneman (1974) Heuristics and Biases program, and of the Associative Theory of Probability Judgements.

2   Some Semantic Theories of Generics

Generic sentences are sentences that, by their very nature, express useful generalizations. The main question addressed in the literature is about the type of generalization. First, generic sentences are clearly not **universally quantified** sentences: although not all birds fly (penguins don't), (1) is a good generic sentence that most people consider true.

(1)   Birds fly.

Indeed, this is one of the most typical features of generic sentences: they express generalizations that allow for exceptions. But it also need not be the case that almost all, or **most** Gs have feature f in order for the generic 'Gs are f' to be true:

(2)   Birds lay eggs.

Although (2) is true, it is not the case that the majority of birds lay eggs; only the adult female birds do! Moreover, even if most Gs are (or are taken to be) f, the corresponding generic sentence still doesn't have to be true, as exemplified by sentences such as the following:

(3)   Germans are right handed.

Finally, such an analysis is extensional, and that is taken to give rise to problems exemplified by the following, much discussed, generic (4):

(4)   Mail from Antartica is handled by Tanja.

This generic can be true, even though we've never gotten any mail from Antartica. It is normally argued that what such an example points to is a demand for an *intensional* treatment of generics. Arguably, however, (4) is a *normative* generic, and normative generics cannot be given a purely extensional treatment anyway. But, of course, there is a much better reason why generics should not depend on certain actually observed extensionally given sets: if our theory claimed this, we could not account for their inductive, or unbounded, character.

According to the **modal nonmonotonic** approach of Asher and Morreau (1995), Pelletier and Asher (1997) and others, 'Gs are f' is true if and only if for any entity d and all worlds in which d is a normal G, d has feature f. Such theories want to account for a type of **default instantiation**, that is, for the fact that if all we know is that the sentences 'Gs are f' and 'x is a G' are true, we can normally, or by default, conclude by instantiation that x has feature f. Proponents of nonmonotonic logic typically argue that what is normal need not have anything

to do with proportions. But without a independent characterization of what normal is, such a theory is not very predictive (Krifka et al., 1995; Pearl, 1988).[2] Moreover, whatever 'normality' is taken to mean, any analysis that wants to account for 'default instantiation' will have problems accounting for the intuition that the following generics are both true.

(5)   a.   Ducks lay eggs.
      b.   Ducks have colourful feathers.

In order to predict that (5a) is true, it must be the female ducks that should be relevant or normal, while it is the opposite sex that is relevant or normal for (5b).

There is another typical kind of example that is problematic for the analyses discussed so far. Consider the following (seemingly true) generics:

(6)   a.   Ticks carry the Lyme disease.
      b.   Sharks attack people.

Sentences (6a)-(6b) are examples of what Leslie (2008) calls 'striking generics'. A generic sentence is striking if only very few of the Gs need to have feature f for the generic sentence 'Gs are f' to be true. According to Leslie, striking often means 'horrific or appalling'. However, the truth conditions suggested above also seem to hold for familiar examples like (7a)-(7b) that are intuitively not 'horrific'.

(7)   a.   Frenchmen eat horsemeat.   from Schubert & Pelletier (1987)
      b.   Dutchmen are good sailors.

Intuitively, a sentence like (7a) is true not because most Frenchmen eat horse meat, but because *relatively many* Frenchmen do so. These types of examples motivated Cohen (1999) to claim that generic sentences are in fact *ambiguous*. They can both have an *absolute* and a *relative* reading. On the absolute reading, 'Gs are f' is true iff most Gs are f, while on the relative reading it is true iff relatively more Gs are f, than $\neg G$s are, where $\neg G$s are the relevant alternatives to G. Cohen proposes that generic sentences are standardly interpreted in the absolute way, but that sentences that are problematic for many other treatments of generics like (7a)-(7b) should be interpreted in the relative way, just as examples like (5a) 'Ducks lay eggs' and 'Lions have manes'. Presumably, the same is the case for the striking generics (6a)-(6b).

Where Cohen (1999) gives two separate treatments of absolute and relative generics, in Cohen (2001) he provides yet another analysis of **non-descriptive generics** like the following:

(8)   a.   Bishops move diagonally.
      b.   The Speaker of the House succeeds the Vice-President.

---

[2]Of course, this is reminiscent of a similar criticism one might give to semantics of counterfactuals that make crucial use of a primitive notion of 'similarity'.

Cohen (2001) assumes that non-descriptive generics have an underlying logical form that differs radically from their surface form, and that they also have a very different interpretation.

Although I undoubtedly take Cohen's analysis to be a major step forward compared to other analyses of generics, it is certainly not without problems. The **first problem** for Cohen's analysis I take to be the claimed **ambiguity**. Can it really be the case that there is no common core between all types of generics? Should absolute generics really be given a separate treatment from relative and non-descriptive ones? The proposal, for instance, that non-descriptive generics have an underlying logical form that differs radically from their surface form, and that they also have a very different interpretation from descriptive generics is prime facie, at least, problematic. It would be more natural, intuitively, to give all types of generics the same logical form, and have an interpretation of generics that works similarly for them all.

A **related problem** is noted by Leslie et al. (2011), who observe that Cohen's analysis of relative generics predicts that an example like (9) comes out true.

(9)   Dogs have three legs.

Example (9) is clearly predicted to be false on its absolute reading. It is predicted to be true on the relative reading, however. The reason is that dogs have a higher probability of staying alive after losing a limb than wolves, foxes, hyenas, et cetera, because three-legged dogs will be taken care of by their owners. Furthermore, dogs seem to have a higher probability of losing a limb then say hamsters, rabbits, miniature donkeys, and parakeets. Hence, the generic (9) is true relative to both the alternative set Pets and the alternative set Dog-like animals.

Of course, Cohen could simply claim that (9) only has an absolute reading, and on that reading the sentence is correctly predicted to be false. But this move only brings us back to the first problem: how should we determine which reading each generic sentence should have?

What (9) also illustrates is that Cohen's analysis of relative readings is **too weak**. This is not only the case empirically, but also conceptually: they seem to be too weak to be of any use.

Although various aspects of Cohen's analysis of generics have come under attack, in this paper I will build on Cohen's analysis, in particular on his relative readings. Let me here already point to one, I feel, under-appreciated aspect of such readings. Consider generic sentences that express *comparative relations*, like (10):

(10)   Boys are taller than girls.

Although other analyses of generics might be able to account for such readings as well, an analysis in terms of relative readings is almost immediate. The reason is that relative readings are (perhaps implicitly) already treated as comparatives! Let us assume, just for simplicity, an analysis of comparatives as given in Klein (1980): 'John is taller than Sue' is true iff there is a comparison class including (perhaps only) John and Sue such that John is tall with respect to this comparison class, while Sue is not. Similarly, (10) will be true if there is a comparison

class c including (perhaps only) boys and the girls such that with respect to this comparison class boys are tall and girls are not. Notice that according to the relative reading of the generic sentence 'Boys are tall', the sentence is true iff $P(tall(c)/boys) > P(tall(c)/\neg boys)$. If c consists of only the boys and girls this reduces to $P(tall(c)/boys) > P(tall(c)/girls)$. Similarly, the generic sentence 'Girls are not tall' is true on its relative reading in this context iff $P(\neg tall(c)/girls) > P(\neg tall(c)/boys)$, i.e., iff $1 - P(tall(c)/girls) > 1 - P(tall(c)/boys)$ iff $P(tall(c)/girls) < P(tall(c)/boys)$. As a result, sentence (10) is predicted to be true iff $P(tall(c)/boys) > P(tall(c)/girls)$, which indeed seems to be the correct result.

Leslie (2008) observes that although generics are extremely hard to analyse truth-conditionally, we are able to understand and use them successfully with relative ease. She suggests that this is so because generics are the expression of a very primitive default mode of generalizing, which picks up on significant or striking properties and links them to psychologically salient kinds. I completely agree with her **cognitive approach**, and that the analysis of generics should be closely tied to the way we categorize and make inductive generalizations. I would like to focus on this insight as well. As we will see though, this doesn't necessarily mean that truth-conditional approaches like those of Cohen (1999) are wrong headed. Moreover, or so I will argue, this doesn't mean that generics are as (at least 5-way) ambiguous as Leslie et al. (2011) suggest. Perhaps it is possible to give a more uniform semantic analysis of all types of generic sentences, once we know more about typicality and how we learn inductive generalizations.

3   Typicality and associative learning

People have the natural tendency to classify the objects around them in terms of categories. Objects are grouped together to form a category if they have characteristics in common or are roughly similar to one another. Our thinking in terms of categories reduces the complexity of the world around us considerably. Categorization is one of the most common and most important things we do all the time and crucially influences our behavior. One of the most important functions of categories is that they allow us to make use of induction and generalization. Indeed, the process of categorization itself is perhaps the most basic type of generalization we make. It is only natural to assume with Leslie (2008) that generic sentences about categories express these basic generalizations. This suggests that to figure out why we accept certain generic sentences but not others, it is crucial to understand this basic process of categorization.

One of the main claims of this paper is that a generic of the form 'Gs are f' is true if f is a typical feature of Gs, or if typical members of the category G have feature f. Typicality is well-studied in cognitive psychology. According to prototype theory, groups (or categories) are represented by typical members, rather than by all of them and only them, or by typical features, rather than by necessary and sufficient features, because agents have limited attention and limited recall of examples. But what are a group's typical members or features? According to Rosch (1973), it is the *central*, or *average* members of the group, or the features *most* members have. Centrality is determined in terms of a notion of similarity. Barsalou (1985) experimentally showed on the basis of a thorough correlational analysis, however, that at least

for goal-derived artificial categories, the typical members are instead the category's *ideal* members; those that *best* satisfy the goal. For example, the ideal of the category 'things to eat on a diet' presumably is 'zero calories,' which clearly is not a common, but rather an extreme value for members of the category. Idealness can be defined as the extent to which a certain object displays a quality that is directly related to the goal. More recent empirical findings (Ameel and Storms, 2006; Burnett et al., 2005) show that extreme members of a group are also considered typical for many, if not most, other types of categories, namely if categorization is performed in a *contrastive* way. Typical members of a category have features that *distinguish* them from members of other categories; as such, they highlight, but also exaggerate, real differences between groups.

Typical features for a group are taken to be features that are **representative** for the group. This is important for our analysis of generics. Furthermore, I claim that the way we **learn** categories is, and remains, of crucial importance for *judgements* involving those categories.

A popular way to approach the learning of categories involves associative learning based on frequencies and correlations. Much of that psychological research was done before the cognitive revolution in psychology, in classical conditioning. In classical conditioning, what is learned is an association between a cue and an outcome. The cue, *C*, such as the sound of a bell, or a tuning fork, can become associated with an outcome, *O*, which can be thought of either as something like the taste of food, or a shock, or an unlearned reflex response to that, like salivation, or high blood pressure indicating fear. Pavlov hypothesized that the strength of association between cue and outcome depends on the number of times the two are paired.

Subsequent research has revealed, however, that for prediction it is not exactly the number of pairings between cue and outcome that is crucial. In a classic study, Rescorla (1968) showed that rats learn a tone (C) $\rightarrow$ shock (O) association if the frequency of shocks immediately after the tone is higher than the frequency of shocks undergone otherwise. Within associative learning psychology, this difference in frequency is known as the **contingency** of the shock on the tone. The central finding of Rescorla (1968) was that the higher the contingency of shock on the occurrence of the tone, the more the rats anticipated the fear of a shock. Thus, the higher the contingency, the more useful the tone is as a predictor of the shock. Of crucial importance for our paper is that these experiments show that rats will develop a tone $\rightarrow$ shock association even if shocks occur only in, say, 12% of the trials in which a tone is present, as long as the frequency of the shocks experienced otherwise is (significantly) lower. Formally, this contingency, or strength of association, between C (e.g. tone) and O (e.g. shocks) is measured by $P(O/C) - P(O/\neg C)$, abbreviated by ΔP_C^O, where P measures frequencies during the learning phase.[3]

Other experiments in the aversive (i.e. fear) and appetitive conditioning paradigms (Thomas and LaBar, 2008) show that the speed of acquisition increases with the *intensity* of the shock. More generally, stronger emotions promote faster learning, more enduring memories, and stronger associations (Chatlosh et al., 1985). One could say that for trained rats, tones play

---

[3]For a counterfactual analogue of contingency, see Pearl (2000). He shows that under some conditions (exogeneity and monotonicity), his counterfactual notion comes down to the standard statistical notion.

an important role in their categorization of shocks: the tone is a useful predictor and thus provides valuable information to the rat on how to prepare for the future. Moreover, this role of the tone in categorization becomes more entrenched with increased intensity of the shock.

Whereas early work in classical conditioning mostly involved animals, more recent work shows that humans learn associations between the representations of certain cues (properties or features) and outcomes (typically another property or a category prediction) in a very similar way (Gluck and Bower, 1988; Schanks, 1995). On the basis of these findings, on my *preliminary* proposal, I measure the **representativenes** of feature f for category or group G as the contingency between f and G, ΔP_G^f, $P(f/G) - P(f/\neg G)$, where $\neg G$ abbreviates $\bigcup Alt(G)$ (and $G \notin Alt(G)$). Then I will say that a feature f is representative for a group G in a particular context iff there is no relevant alternative feature g with a (significantly) higher contingency with G than f, i.e., $\neg \exists h \in Alt(f) : \Delta P_G^h > \Delta P_G^f$. Notice that a representative feature for group G doesn't have to be one that most, or even many, members of the group have. Instead, a representative feature is one that *distinguishes* group G from its alternative(s) (for simplicity taken to be $\neg G$), which is exactly in line with the view on typicality discussed above: those features are representative for a group that highlight, or exaggerate, differences with other groups. Similarly, even though two features f and h are mutually incompatible for members of a certain group (e.g., no peacock both lays eggs and has fantastic blue-green tails), they can still both be representative, because they are distinguishable, for that same group.

Contingency is important for learning associations. Thus, it is the frequencies that animals and people were exposed to in the learning phase that count. But in many cases people are not exposed to the *actual* frequencies of cues (properties or features) with outcomes (typically another property or a category prediction), but rather with a *distorted* picture of it. Distortion is especially likely to happen when we learn associations through the (social) media. For instance, Kahneman (2011) notes that he had a long-held impression that adultery is more common among politicians than among physicians or lawyers. Only later he realized that this associative belief was probably caused by the fact that the extramarital relations of politicians are much more likely to be reported in the media than the affairs of lawyers and doctors. Still, it is only natural to assume that people will pick up associations from news items in a very similar way that people learn associations through actual exposure. This suggests that learning associations between cues with outcomes from the media also goes via contingency, our ΔP_G^f, but now the frequencies measure not the actual frequencies, but a distorted picture of them via media coverage which is strongly biased towards novelty and poignancy (cf. Kahneman (2011)).

Slovic et al. (2004), among others, argue that there exists a deeper link between representativeness of events or features and our emotional reactions to them. Events which give rise to fear and danger come easy to mind not only because of higher media coverage, but also simply because they give rise to strong emotional reactions. We have seen above that humans are, in this sense, not so different from the animals used in classical conditioning experiments: strong emotions like fear promote faster learning and more enduring memories. The empirical success of reinforcement learning in humans, again, only corroborates this idea.

To incorporate the insight of Slovic et al. (2004) and of fear-conditioning, I will extend our

earlier proposal and will define representativeness in a more general way. I will measure the representativeness of f for G by $P(f/G) \times Intens(f) - P(f/\neg G) \times Intens(f)$, or equivalently $\Delta P_G^f \times Intens(f)$, where $Intens(f)$ measures the *intensity* of f. I will abbreviate this measure by ∇P_G^g. Next I will say in our *final proposal* that f is among the most representative features of G iff there is no contextually salient alternative feature h which has a (significantly) higher measure of representability for G, which I will say holds if $\neg \exists h \in Alt(f) : \nabla P_G^h \gg \nabla P_G^f$. I normally assume that all relevant features have the same intensity, i.e. $\forall f, h \in Alt(f) : Intens(f) = Intens(h)$. This means that under normal circumstances our notion of representability reduces to contingency, $\nabla P_G^f = \Delta P_G^f$.

4   Weak Semantics: Generics state typicalities

In this section I will claim that a generic of the form 'Gs are f' is true if f is a representative feature for G. Therefore I make the following **semantic claim**:

- 'Gs are f' is true iff there is no $h \in Alt(f) : \nabla P_G^h \gg \nabla P_G^f$,
  where '$\gg$' means '*significantly* larger'.

Observe that under any circumstance the generic sentence can only be true if $\Delta P_G^f > 0$, i.e., if the generic is true on Cohen's relative reading.[4] I claim that this general definition can account for the generics I discussed so far, due to the context dependence of various notions involved.[5] To make that clear, let us first make some general observations concerning some special cases:

1. If $Alt(f) = \{f, \neg f\}$ and $Intens(f) = Intens(\neg f)$, the generic '$G$s are f' is true just in case $\Delta P_G^f \gg 0$, i.e., $P(f/G) \gg P(f/\neg G)$, i.e., Cohen's *relative reading*.[6]

---

[4]Instead of using standard contingency, it is tempting to make use of *weighed* contingency or representativeness instead. Let us just consider weighed contingency, defined as follows: $\Delta^\alpha P_G^f := \alpha P(f/G) - (1 - \alpha) P(f/\neg G)$, with $\alpha \in [\frac{1}{2}, 1]$. Now one can define the (simplified) truth conditions of generics for which intensity is irrelevant as follows: 'Gs are f' is true iff $\neg \exists h \in Alt(f) : \Delta^\alpha P_G^h > \Delta^\alpha P_G^f$. One can show easily that in case $\alpha = 1$, Cohen's (1999) absolute reading follows, while if $\alpha = \frac{1}{2}$, the result is Cohen's relative reading. This is certainly an appealing result. However, I won't go for this proposal because (i) in contrast to our definition, if $\alpha \neq \frac{1}{2}$ positive contingency, $\Delta P_G^f > 0$, is not required for generics to be true (which I think is undesirable), (ii) the use of the extra parameter α only adds more context-dependence, and (iii) we can derive the relative reading without making use of α and we can account for the generics that Cohen treats as absolute ones as well.

[5]Sterken (2015) has recently argued that generics are more context dependent than is generally assumed: not only the domain of quantification is context dependent, but also the required force of quantification. Notice that on our analysis the required force of 'quantification' depends on context as well. How high $P(f/G)$ must be in order for the generic 'Gs are f' to be true depends on what $Alt(G)$, and thus $P(f/\neg G)$ is, what $Intens(f)$ is and what $Alt(f)$ is. Thus, I agree with Sterken that the required quantificational force for a generic to be true depends on context, but given our analysis of generics that is not something that should be build on top of the analysis: it just follows from the context dependence required anyway.

[6]Proof: Under these circumstances the sentence is true iff $\Delta P_G^f > \Delta P_G^{\neg f}$. For $\Delta P_G^f > \Delta P_G^{\neg f}$ to be the case it has to hold that $P(f/G) - P(f/\neg G) > P(\neg f/G) - P(\neg f/\neg G)$ iff $P(f/G) - P(f/\neg G) > 1 - P(f/G) - (1 - P(f/\neg G))$ iff $P(f/G) - P(f/\neg G) > P(f/\neg G) - P(f/G)$ iff $P(f/G) > P(f/\neg G)$.

2. If $\forall h \in Alt(f) : Intens(h) = Intens(f)$ and $P(f/G)$ is not high, 'Gs are f' is true just in case $P(f/\neg G)$ is very low, and thus f is very *distinctive* for Gs.

3. If $\forall h \in Alt(f) : Intens(h) = Intens(f)$ and $P(h) \approx P(f)$, 'Gs are f' is true only if $\forall h \in Alt(f) : P(f/G) \gg P(h/G)$, or if ΔP_G^f is only somewhat above 0 and $\forall h \in Alt(f)$, $P(h)$ is not low, $P(f/G)$ has to be (very) high ('*standard*' generics)[7]

4. If ΔP_G^f is only somewhat above 0, and $P(f/G)$ is not high, $Intens(f)$ has to be high for 'Gs are f' to be true. (*striking* generics)

5. If $Alt(f) = \{f, \neg f\}$, $Intens$ is irrelevant and $\bigcup Alt(G) \cap f = \emptyset$, then '$G$s are f' is true just in case $P(f/G) > 0$, i.e., the *existential* reading.[8]

Let us now look at some examples with the above cases in mind.

(1). If $Alt(f) = \{f, \neg f\}$ and $Intens(f) = Intens(\neg f)$, the generic '$G$s are f' is true just in case $\Delta P_G^f \gg 0$. Notice that this is already stronger than Cohen (1999)'s relative reading for which he argues to deal with sentences like 'Dutchmen are good sailors'. However, I think even for these cases that the reading should be stronger. This is what we predict, especially if we assume that $Alt(f)$ can contain many other alternative features than just $\neg f$ (if it contains $\neg f$ at all). That is, the generic is true iff there is no relevantly salient h that is a more distinguishing feature for being a G than f is. I claim that this is exactly not the case for examples like (9) 'Dogs are 3-legged' which indicated that Cohen (1999)'s relative reading of generics is too weak. Indeed, intuitively, one does not distinguish dogs from other pets by looking at whether or not they have three legs; checking whether they bark makes much more sense.

(2). A generic sentence 'Gs are f' is true if f is *very distinctive* for Gs. I claim that a generic like 'Tigers have stripes' is considered true because 'having stripes' is (among the relative alternative features $Alt(f)$) among the most distinctive features of tigers. A generic sentence like 'Germans are right handed', on the other hand, is not predicted to be true simply because 'being right handed' does not distinguish Germans in any significant way from, say, other European citizens.

---

[7]To show this, recall that $P(f/G) > P(f/\neg G)$ just in case $P(f/G) > P(f)$. It follows that if we only take features like h into account such that $P(h) \approx P(f)$, $\Delta P_G^f > \Delta P_G^h$ just in case $P(f/G) \gg P(h/G)$. The same holds if we assume alternatively, and perhaps more naturally, that $\forall h \in Alt(f) : P(h/\bigcup Alt(G)) \approx P(f/\bigcup Alt(G))$. In both cases it means that $P(f/G)$ must be high.

[8]For simplicity I will take $Alt(G) = \{\neg f\}$. Assuming that $Intens$ is irrelevant, we have to check whether $[P(f/G) - P(f/\neg f)] > [P(\neg f/G) - P(\neg f/\neg f)]$. By taking $P(f/G) = p$, what follows is that $[p - 0] > [(1 - p) - 1]$. This reduces to $p > -p$, which holds exactly if $p > 0$, i.e., if $P(f/G) > 0$: the existential reading. Recall that $P(f/G) > -P(f/G)$ iff $P(f/G) + P(f/G) > 0$. This suggests that generics with frequency adverbs like 'Mammals seldom fly' can be interpreted in terms of contingency as well with the same choice of $Alt(f) = \{f, \neg f\}$ and $Alt(G) = \{\neg f\}$. For the above generic to be true we demand that $\frac{1}{2}(\Delta P_G^f - \Delta P_G^{\neg f}) \ll 1$, which comes down to $\frac{1}{2}(P(f/G) + P(f/G)) \ll 1$ and reduces to $P(f/G) \ll 1$, which is Cohen's (1999) analysis. Alternatively, we can simply demand that $\Delta P_G^f \ll 1$, with $Alt(G) = \{\neg f\}$, although this would complicate a compositional analysis.

Our analysis accounts for the intuition that generics like 'Birds fly' and 'Birds lay eggs' are acceptable and true. The reason is that 'flying' and 'laying eggs' are indeed among the most distinguishable features for birds (compared to alternative middle sized animals). Our semantic analysis of generics also explains the following example that is paradoxical to many other theories: although only (adult) male lions have manes, (11a) is an accepted generic, but (11b) is not.[9]

(11)   a.   Lions have manes.
        b.   Lions are male.

The reason is that compared to lions, relatively few other animals have manes, but it is not the case that compared to other animals relatively many lions are male. Our analysis thus correctly predicts that 'Gs are f' can be true and 'Gs are h' false, although $P(h/G) > P(f/G) < \frac{1}{2}$.

I have proposed that generic sentences should be analyzed in terms of representativeness, and that the representativeness of feature f for group G should be measured by ∇P_G^f. I have noted before that this reduces to contingency, ΔP_G^f, if $\forall f, g \in Alt(f) : Intens(f) = Intens(g)$. Notice, now, that ΔP_G^f behaves very similar to two other interesting measures, $\frac{P(f/G)}{P(f)}$ and $\frac{P(f/G)}{P(f/\neg G)}$.[10] It is remarkable that $\frac{P(f/G)}{P(f)}$ and $\frac{P(f/G)}{P(f/\neg G)}$ have been proposed as measures of **stereotypicality** of f for G within social psychology (McCauley et al., 1980; Schneider, 2004). Indeed, just like ΔP_G^f, also $\frac{P(f/G)}{P(f)}$ and $\frac{P(f/G)}{P(f/\neg G)}$ give those features a high value that are distinctive for group G, and thus highlight or exaggerate differences between groups. From this proposal, together with our own, it naturally follows that in case ΔP_G^f is high, we could say that f is a stereotypical feature of G. Is that already enough evidence to propose that 'Gs are f' is a good and true generic?

Indeed, a number of authors, including Declerck (1996) and Geurts (1985), have proposed that generics are about stereotypical properties. This account has been criticized by Krifka et al. (1995), however, and is largely abandoned in the literature. A first argument used by Krifka et al. (1995) is that stereotypes are just widely acknowledged ideas within a speech community, while the truth of a generic depends on actual facts: even if uttered in a culture where everybody believes that cows are a special kind of horse, or that snakes are slimy, they argue, cows are not horses, and snakes are not slimy. This argument is obviously invalid with respect to our analysis of stereotypes, however, if I base my analysis not on a subjective probability function, but on objective frequencies, or propensities. The truth of a generic is then predicted to depend on actual facts. A second counterargument of Krifka et al. (1995) is that stereotypes are tied to well-known groups or situations, while generics are often not about any of those things. But, again, I don't see why this could be problematic for our analysis. A third counterargument is

---

[9]This example is strikingly similar to the famous *conjunction fallacy* of Kahneman and Tversky (1972). In the next section I will show that our analysis accounts for the two 'paradoxes' in the very same way, in terms of our implementation of Tversky and Kahneman (1974)'s *representativeness*-heuristics making use of contingency.

[10]Contingency increases with increasing $P(f/G)$ and decreasing $P(f/\neg G)$ just as $\frac{P(f/G)}{P(f/\neg G)}$. Next, one can show that contingency, ΔP_G^f, behaves monotone increasingly with respect to the standard notion of relevance, $P(f/G) - P(f)$. Obviously, $\frac{P(f/G)}{P(f)}$ increases with increasing $P(f/G)$ and decreasing $P(f)$, just as $P(f/G) - P(f)$.

the fact that although the stereotype states that Hindus don't eat meat, a generic like 'Hindus eat meat' can be true in certain contexts, e.g. as a rejection of the claim that no Hindu eats meat. I agree with the linguistic intuition, but along with Cohen (2004) I think that the truth of this use of the generic is peculiar to its use of rejections. I will come back to this problem later in this section. For now, I conclude that the standard arguments against an analysis of generics in terms of stereotypicality are not valid on our implementation of the latter notion.

Our analysis of generics is based on typicality and as such is very similar to an analysis based on **prototypicality** as well. But the linguistic literature has not been friendly to such an approach. How could I still defend it? Let's see whether we can rebut the troubles typically discussed for such an analysis. First, this approach is criticized for simply passing on the problem of generics to a new problem of what it means to be prototypical. But this can't be a serious problem anymore, given our very explicit proposal, based on psychological research, for what it means to be (proto)typical. A second problem normally discussed is that this approach cannot deal with the fact that the following two sentences both seem to be true:

(12)   a.   Peacocks have fantastic blue-green tails.
         b.   Peacocks lay eggs.

The reason why this example is taken to be a problem is that the proposal to handle generics in terms of prototypicality is mostly taken to be that the sentence '*G*s are *f*' is true just in case the prototypical *G*s have feature *f*. Hence: 'Tigers have stripes' is true if and only if all (proto)typical tigers have stripes. Natural as such an analysis might be, it falsely predicts that (12a) and (12b) cannot both be true, because it is not the case that the typical peacock both has a blue-green tail and lays eggs, simply because there is no peacock that is male *and* female. Fortunately, my analysis differs from the one that is criticized. According to my analysis it is possible that '*G*s are *f*' and '*G*s are *h*' are true, even though *f* and *h* are, in fact, incompatible. It only has to be the case that $\neg \exists g \in Alt(f) : \Delta P_G^g > \Delta P_G^f$ and $\neg \exists g \in Alt(h) : \Delta P_G^g > \Delta P_G^h$, where the sets of alternative features $Alt(f)$ and $Alt(h)$ could be, though need not be, the same. This, obviously, is very well possible: relative to other animals (in general), many peacocks have beautiful blue-green tails and many peacocks lay eggs.[11] What is predicted not to be possible is that both '*G*s are *f*' and '*G*s are ¬*f*' are true (if ¬*f* ∈ *Alt(f)* and *f* ∈ *Alt(¬f)*), which is as it should be according to Hoeltje (2017).

(3). Recall that on our analysis *f* is a representative feature of *G* if *f* is very distinctive for *G*. However, it seems that some generics of the form '*G*s are *f*' are true, without *f* being a very distinctive feature. This holds, arguably, for (13a)-(13d), which are all undoubtedly good generics:

---

[11]Leslie (2008) provides a stronger argument against the prototype theory. She argues that not only (7a) and (7b) are true, but also 'Peacocks have fantastic blue-green tails and lay eggs'. Perhaps a more convincing example is given by Nickel (2010): 'Elephants live in Africa and Asia'. Note that on our analysis it might well be possible that for two mutually incompatible features like *f* and *h* it could be that $\neg \exists g \in Alt(f) : \Delta P_G^g > \Delta P_G^f$ and $\neg \exists g \in Alt(h) : \Delta P_G^g > \Delta P_G^h$, even if $Alt(f) = Alt(h)$. What is obviously not possible on our analysis is that for the *conjoined* feature $f \wedge h$ it holds that $\neg \exists g \in Alt(f) : \Delta P_G^g > \Delta P_G^{f \wedge h}$. Thus, for such cases, '∧' must have wide scope.

(13)   a.   Humans are mortal.
        b.   Birds are warm blooded.
        c.   Dogs are 4-legged.
        d.   Lions are mammals.

Intuitively, these generics are true simply because the vast majority of the mentioned animals have the relevant features. Our analysis can account for such cases as well. Notice, first, that although in all the above cases having the feature f hardly distinguishes the animals involved, Gs, from their alternatives, $\bigcup Alt(G)$, it is still the case that $P(f/G) > P(f/\bigcup Alt(G))$ (taking some immortals into account for (13a)), and thus the feature is predicted to be associated with G, even if not in a maximal way.[12] Second, in examples (13a) and (13b) it is only reasonable to assume that $Alt(f) = \{f, \neg f\}$, and thus $P(f/G) > P(f/\bigcup Alt(G))$ is already enough to make the sentences true. If $P(f/\bigcup Alt(G))$ is high, it just means that $P(f/G)$ has to be *very* high, which indeed is the case. Third, the features involved in (13c) and (13d) (and in (13a) and (13b)) are rather common among all animals. If we only take other such features into account as well (0-legged, 2-legged and birds, fish, reptiles, amphibians), it is demanded on our analysis for 'G is f' to be true that $P(f/G)$ is *very* high (assuming *Intens* to be irrelevant).

Still, it seems that the analysis as it stands is not quite appropriate for examples like (13a)-(13d). For other generics our analysis required that the measure of representativeness, $\nabla P_G^f = \Delta P_G^f \times Intens(f)$ is high. For the above examples, however, that doesn't hold: *Intens* seems irrelevant, and the features are not really distinctive, meaning that ΔP_G^f is low (though positive). Fortunately, one can define a measure closely related to ΔP_G^f (adopted from Shep (1958)), called 'relative difference' and denoted by $\Delta^* P_G^f$, which will have the result that the resulting ∇P_G^f (where in its definition ΔP_G^f is replaced by $\Delta^* P_G^f$) will be high:

- $\Delta^* P_G^f \quad \overset{df}{=} \quad \dfrac{\Delta P_G^f}{1 - P(f/\bigcup Alt(G))}$

Replacing ΔP_G^f in the definition of ∇P_G^f by $\Delta^* P_G^f$ will mean that contingency, and thus distinctness, still plays a major role: for $\Delta^* P_G^f > 0$ it is required that $\Delta P_G^f > 0$, and high ΔP_G^f still results in high $\Delta^* P_G^f$. However, it has the extra effect that $\Delta^* P_G^f$ *increases*, if $P(f|\bigcup Alt(G))$ increases. For instance, if $P(f|\bigcup Alt(G)) = 0.9$, $\Delta^* P_G^f$ will be **ten times** as high as ΔP_G^f (if $\Delta P_G^f > 0$)! Thus, for relatively common features (as in examples (13a)-(13d)) it has the effect that $\Delta^* P_G^f$ will be high, even though ΔP_G^f is relatively low. More intuitively, the use of $\Delta^* P_G^f$ instead of ΔP_G^f has the consequence that for representativeness of f for G, the value $P(f/G)$ is more important than $P(f/\neg G)$.[13]

(4). Next, if $\Delta P_G^f > 0$ but small, and $P(f/G)$ is not high, *Intens*(f) has to be high for 'Gs are f' to be true. Recall that *Intens* was brought in to take over some insights from fear-conditioning.

---

[12]This distinguishes these examples from a sentence like 'Germans are right handed'.

[13]Although the general approach should be stated in terms of $\Delta^* P_G^f$, for simplicity I won't make use of in the rest of this paper, because nothing in the further discussion of this paper relies on it.

I claim that it is exactly this that makes our analysis immediately account for **striking generics** like (6a) 'Ticks carry the Lyme disease' which are problematic for default-based approaches (e.g. Asher and Morreau (1995)). Indeed, Leslie (2008)) notes that 'striking' often means 'horrific or appalling', which means having a high *Intens*.

A feature can be striking also just because it is very peculiar, i.e., uncommon. Learning that members of a group have this peculiar feature more than on average can be very interesting. Making use of Shannon's Information Theory, I will say that A is peculiar exactly if A's informativity, $\inf(A)$, is high. The latter notion is defined as $\log_2 \frac{1}{P(A)} = -\log_2 P(A)$. According to this definition, $\inf(A)$ receives a high value exactly if $P(A)$ is small. ∇P_G^f now comes down to $[P(f/G) - P(f/\neg G)] \times \inf(f)$.[14] If f is a very common feature, $\inf(f)$ will be small and ∇P_G^f can be high only if ΔP_G^f is high. For very uncommon features for members of G and their alternatives, however, ∇P_G^f can be high even if ΔP_G^f is low (but > 0). I claim that this is going on for (at least some) *relative generics* like (7a) 'Frenchmen eat horsemeat'.

(5). Earlier in this section I mentioned an example of a generic statement like (14) that is, intuitively, interpreted **existentially**, and noted that according to Krifka et al. (1995) this shows a problem for any analysis of generics based on stereotypicality. Existential generics like (14) (from (Cohen, 1999)), however, seem to pose a problem for nearly any analysis of generics.

(14)   A. No Indian eats beef.
         B. No! Indians [do]$_F$ eat beef.

Cohen (2004), however, is able to account for existential readings of generics by assuming that these are interpreted on his absolute reading with $Alt(f) = \{f\}$. Although formally appealing, the proposal looks conceptually artificial. For one thing, the focal stress on the verb *do* suggests that $\neg f$ should be an element of $Alt(f)$ as well. What is clear, though, is that for the interpretation, only Indians count, which seems to suggest that our *contrastive* analysis is not well suited to the situation. As shown in a previous footnote, however, I can account for existential readings formally by assuming that *Intens* is irrelevant, $Alt(f) = \{f, \neg f\}$ and $Alt(G)$ is such that $\bigcup Alt(G) \cap f = \emptyset$. Intuitively this seems correct, because the natural way to think about $\bigcup Alt(G)$ is as the set of Indians that verify what is said by A: the Indians that don't eat beef. The result is that the generic in (14) is interpreted as saying that more Indians eat beef than expected.

Let us finally consider **non-descriptive generics** like (8a) 'Bishops move diagonally'. At least since Kripke we know that identity statements can be used in two different ways: (i) to state the identity of meaning (intension) of the two terms, or (ii) to fix the meaning of one term in terms of the meaning of the other. Kripke explains the *a priori* character of a sentence like 'Stick *S* is one meter long' when talking about the ideal stick, or standard meter, preserved in Paris ever since the French Revolution by the second use of identity statements. Generic sentences are much like identity statements and can be used in those two similar ways. On a definitional use

---

[14]What $\inf(f)$ is meant to measure is the informativeness that an arbitrary x has feature f.

of generics of the form 'Gs are f', this means that having f is a *necessary* condition for being a G. If the definition is complete, f is also a *sufficient* condition. Note that if 'Gs are f' is a complete definition of Gs, f is a feature that all and only all Gs must have. But this means that ΔP_G^f will have the maximal value, because in that case $P(f/G) = 1$ (the maximal score) and $P(f/\neg G) = 0$ (the minimal score). This suggests that for definitional, or constitutive, uses of generics, just like for descriptive generics, contingency plays a crucial role!

Not all non-descriptive generics are constitutive in an equally strong manner as (8a) and 'Keys open doors', where without such generics bishops and keys would not even exist. Consider examples like the following:

(15) a. Boys don't cry.
  b. Gentlemen open doors for ladies.

Leslie (2015) argues that (15a), at least on its normative reading, is used to tell a boy that crying is not an appropriate behavior for him and that he should hold back his tears: the sentence says something about what it takes, according to the speaker, to be an *ideal* boy. This picture is reminiscent of the way we categorize goal-derived categories according to Barsalou (1985), and we have seen that what are typical features for such a category are features that do a good job of distinguishing members from this category from members from others. But obviously, this is exactly what our analysis of generics is meant to do in the first place. This suggests that, just like for descriptive generics, contingency $\Delta P_G^f = P(f/G) - P(f/\neg G)$ should play an important role for normative generics like (15a) and (15b) as well.

But what, then, does distinguish normative from descriptive generics? What distinguishes the two types of generics, I would like to claim, is exactly the fact that the generics of the former type are not descriptiv.! More precisely, the probabilities involved in normative generics measure *normative expectations* rather than actual propensities. If so, it immediately follows that ΔP_G^f is high for such a generic of the form 'Gs are f', and thus that except for the interpretation of P, we could interpret non-descriptive generics in exactly the same way as descriptive generics.

Intuitively, normative expectations do not only involve beliefs, i.e., expectations measured by probabilities, but also desires. Indeed, I would think it is natural to claim that (15a) is really about *desirable* features for boys. With a slightly more general definition of ∇P_G^f we can account for this. Let us redefine ∇P_G^f in a slightly more general way than we have so far as $(P(f/G) \times Value(f/G)) - (P(f/\neg G) \times Value(f/\neg G))$. A user of (15a) values it highly if boys don't cry, i.e., if $Value(\neg cry/Boys)$ is high, although for this speaker $Value(\neg cry/Girls)$ might as well be low. Assuming that $P(\neg cry/Boys) \approx P(\neg cry/Girls)$ with P just measuring expectations, it still follows that ∇P_{Boys}^{cry} is high, and the speaker signals that (s)he wants boys not to cry.

5   Strong Pragmatics: from biases to probabilities

On the basis of experimental evidence, Cimpian et al. (2010) concluded that to *accept* a generic about a group we are familiar with, relatively weak conditions have to be fulfilled. At the same time, Cimpian et al. (2010) have observed that hearers *interpret* generics in a much stronger way: (almost) all *G*s are *f*. This holds especially if this generic is about a relatively unknown group. What could explain this strong interpretation?

Our **proposal** is that this is due to the fact that people generally **confuse** representativeness (or stereotypicality) with probability (or prototypicality). This idea might seem ad hoc, but it is in fact at the heart of the whole Biases and Heuristics program of Tversky and Kahneman (1974), and the confusion between contingency and probability is explicitly argued for in the more recent Associative Theories of Probability Judgement. The first program started with Tversky and Kahneman showing that our intuitions involving probability judgements are not in accordance with the norms given by Bayesian probability theory.

Bayesian probability theory is a *prescriptive* theory. Unfortunately, it doesn't seem *descriptively* adequate. The conjunction fallacy of Kahneman and Tversky (1972) shows that in some situations people assign greater probability to a conjunction than to one of its conjuncts, i.e., $P(B \wedge F) > P(B)$, although this is impossible according to the normative Bayesian theory. For example, a woman (Linda) with liberal political views was judged by most participants to be more likely a feminist bank teller than a bank teller. According to their Biases and Heuristics program (Tversky and Kahneman (1974)), to reach a probability judgement, we often do not reason according to Bayesian probability theory, but use simplifying or shortcut heuristics. These heuristics are mostly approximately correct, but also give rise to systematic biases in certain contexts.

We have seen above that the contingency, or associative strength, between cue *C* and outcome *O* is measured by $\Delta P_C^O = P(O/C) - P(O/\neg C)$. The Associative Theories of Probability Judgements (Gluck and Bower, 1988; Lagnado and Shanks, 2002) now make the **further claim** that the subsequent probability judgments are then based on these associations.[15] Although the contingency between *C* and *O* might be very different from the conditional probability of *O* given *C*, association theories of probability judgements claim that when people are asked a question about probability, they readily substitute this with the closely related question about evidential support, or contingency, which sometimes gives rise to an incorrect response.[16] Lagnado and Shanks (2002) show that the Associative Theory of Probability Judgements can account for the conjunction fallacy. In a similar way, associative theories of probability judgments can explain

---

[15]Of course, ΔP_C^O can be anywhere between -1 and 1, whereas probabilities need to be between 0 and 1. How can $\Delta P_C^O \times P(O/C)$ be turned into a probability? For such cases, normally a logistic function is used, a transformation function that turns measures from $[-\infty, \infty]$ into $[0, 1]$ such that everything below 0 goes to somewhere below $\frac{1}{2}$ and analogously to everything above 0.

[16]To give a telling example from Newel et al. (2007), suppose that a football team is as likely to win as to lose when Johan plays, but that the team much more likely loses when Johan is not playing. In that case, although $P(win/Johan\ plays) = P(\neg win/Johan\ plays)$, still people typically will believe that the team will win if Johan is playing.

other ways people deviate from the normative Bayesian theory, such as the fact that people tend to neglect base rates.

The problem I wanted to account for in this section is to explain why people generally interpret generics of the form 'Gs are f' as holding that $P(f/G)$ is high. Our analysis merely predicts that the sentence is true iff ∇P_G^f is high, which means that $[P(f/G) - P(f/\neg G)] \times Intens(f)$ is high. Given the Associative Theory of Probability Judgements, the gap beween the two can easily be bridged in case all features have the same value. Recall that if the value of the features is irrelevant, high ∇P_G^f reduces to a high contingency. By the Associative Theory of Probability Judgements, however, this is confused with high $P(f/G)$, which explains the common intuition under these circumstances.

What if the intensities of the features might be important, i.e., if the relevant features are 'horrific', 'appalling', or 'peculiar'? According to Tversky and Kahneman (1974) *availability heuristics*, people assess the probability of an event by the ease with which instances or occurrences can be brought to mind. Usually this heuristic works quite well; all things being equal, common events are easier to remember or imagine than uncommon events. Unfortunately, sometimes the general rule of thumb doesn't do its job and leads to systematic biases. Some events are more available than others not because they tend to occur frequently or with high probability, but simply because they are inherently easier to think about. Emotionally-charged events, or horrific or appalling features, can be brought to mind easily. Presumably, the same will hold for other newsworthy events or features, in particular for the peculiar ones. Thus, by the availability heuristics, $P(f/G)$ and $P(f/\neg G)$ are considered higher than they actually are, if $Intens(f)$ is high, and the same will thus be the case for the contingency ΔP_G^f. But this means via the associative theory of probability judgement again that $P(f/G)$ is considered higher than it actually is, perhaps close to 1.

Thus, I have argued that hearers that are unfamiliar with group G interpret a generic of the form 'Gs are f' in a much stronger way than one would expect according to the semantic analysis I defended in the previous section, because they confuse representative features, features f where $P(f/G) - P(f/\neg G)) \times Intens(g)]$ is high, with probable feature, features f where $P^*(f/G)$ is high. In the last formula, P^* models subjective belief rather than objective frequencies or propensities. Obviously, if $P^*(f/G)$ is high, we have explained why the generic 'Gs are f' is interpreted as meaning that almost all Gs are f.

There is yet another reason as to why a high representativeness, or contingency, between G and f increases the felt conditional probability $P(f/G)$, i.e., $P^*(f/G)$. Note, first, that it is, or at least was, quite common among psychologists and philosophers of science to use ΔP_G^f to measure the causal strength of G for f. Second, Tversky and Kahneman (1980) show that if we see a correlation, we tend to interpret it in the preferred (strongest) way: as **causal**. Moreover, they show that an event is seen as **more likely** than it actually is, if it can be understood: if it can be causally explained. This, then, is the causality bias: high representiveness of a feature for G as we measured it, leads to higher subjective probability given G than is justified, because we believe that there exists a causal relation between G and f.

This last reason why high representability leads to high conditional probability is closely related with proposals of Barth (1971), Leslie (2015) and Haslanger (2010), Haslanger (2014) why we (and children in particular) *interpret* generics typically in such a strong way. Haslanger (2014) argues that if we use a generic like 'Women are more nurturing than men', we (wrongly) implicate that there is something about what it is to be a woman and about what it is to be a man that explains their supposed differential capacities to nurture. From this she concludes that the utterance of a generic of the form '*G*s are *f*' will normally add to the common ground of the conversation a claim about *f*'s naturalness for the group, or kind, *G*. The generic is (wrongly) taken to be true because of *G*s essence. Haslanger (2014) and Leslie (2015) argue that this is why it is dangerous to claim 'Muslims are terrorists' but not 'Ticks carry the Lyme disease': while for the latter essentialists belief might be true, this is certainly not the case for the former. Although I agree with Barth (1971), Haslanger (2014) and Leslie (2015) that essentialist beliefs play a pragmatically significant role in why we interpret generic statements in such a strong way, I don't think this is the whole reason: it is only one of the biases singled out by Tversky and Kahneman (1974) that are crucial.

6 Conclusion and outlook

In this paper I have based my analysis of generic sentences primarily on an intuition that some authors over the years have claimed would be natural for at least some examples (e.g. Krifka et al. (1995)): a generic of the form '*G*s are *f*' is true iff *f* is a typical, distinguishing, feature for *G*s. Many analyses of generics have been proposed over the years, and none has come out as the clear winner. This is partly due, I suspect, to the vagueness and context-dependence of what is meant by a generic. I have little doubt that my proposal won't meet universal acceptance either. Still, I hope that this paper at least shows that an analysis in terms of typicality can be pushed much further than is generally assumed. I also argued that such a semantic analysis is naturally extended by pragmatic strengthening, making use of insights from Tversky & Kahneman's Heuristics and Biases approach. This popular approach within social and cognitive psychology (as measured by the selling rates of Kahneman (2011)), has, to the best of my knowledge, never been used so far in pragmatics. I think this is a shame, and I hope this paper will help to change things accordingly.

References

Ameel, E. and G. Storms (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language 55*, 402–421.

Asher, N. and M. Morreau (1995). What some generic sentences mean. In Carlson and Pelletier (Eds.), *The Generic Book*, pp. 300–339. University of Chicago Press.

Barsalou, L. (1985). Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory and Cognition 11*, 629–654.

Barth, E. (1971). *De Logica van de Lidwoorden in de Traditionele Filosofie*. Ph. D. thesis, Leiden University.

Burnett, R., C. Medin, Ross, and S. Blok (2005). Ideal is typical. *Canadian Journal of Experimental Psychology 59*, 3–10.

Chatlosh, D., D. Neunaber, and E. Wasserman (1985). Response-outcome contingency: behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation 16*, 1–34.

Cimpian, A., A. Brandone, and S. Gelman (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science 34*, 1452–1482.

Cohen, A. (1999). *Think Generic! The Meaning and Use of Generic Sentences*. Stanford: CSLI Publications.

Cohen, A. (2001). On the generic use of indefinite singulars. *Journal of Semantics 18*, 183–209.

Cohen, A. (2004). Existential generics. *Linguistics and Philosophy 27*, 137–168.

Declerck, R. (1996). The manifold interpretation of generic sentences. *Lingua 68*, 149–188.

Geurts, B. (1985). Generics. *Journal of Semantics 4*, 247–255.

Gluck, M. and G. Bower (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General 117*, 227–247.

Haslanger, S. (2010). *Ideology, generics, and common ground'*, Volume Feminist Metaphysics: Essays on the Ontology of Sex, Gender and the Self, pp. 179–207. Dordrecht: Springer.

Haslanger, S. (2014). The normal, the natural and the good: Generics and ideology. *Politica and Societa 3*, 365–392.

Hoeltje, M. (2017). Generics and ways of being normal. *Linguistics and Philosophy 40*, 101–118.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3*, 430–454.

Klein, E. (1980). The semantics of positive and comparative adjectives. *Linguistics and Philosophy 4*, 1–45.

Krifka, M., F. Pelletier, G. Carlson, A. ter Meulen, G. Chierchia, and G. Link (1995). Genericity: An introduction. In G. Carlson and G. Pelletier (Eds.), *THe Generic Book*, pp. 1–124. Chicago: University of Chicago Press.

Lagnado, D. and D. Shanks (2002). Probability judgment in hierarchical learning: A conflict between predictiveness and coherence. *Cognition 83*, 81–112.

Leslie, S. (2008). Generics: cognition and acquisition. *The Philosophical Review 117*, 1–47.

Leslie, S. (2015). 'Hillary clinton is the only man in the Obama administration?' Dual character concepts, generics, and gender. *Analytic Philosophy 56*, 111–141.

Leslie, S., S. Khemlani, and S. Glucksberg (2011). Do all ducks lay eggs? the generic overgeneralization effect. *Journal of Memory and Language 65*, 15–31.

McCauley, C., C. L. Stitt, and M. Segal (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin 87*, 195–208.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, California: Morgan Kaufman Publishers.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Pelletier, F. and N. Asher (1997). Generics and defaults. In J. van Benthem and A. ter Meulen

(Eds.), *Handbook of Logic and Language*, pp. 1125–1177. Amsterdam: North Holland.

Rescorla, R. (1968). Probability of shock in the presence and absence of cs in fear conditioning. *Journal of Comparative and Physiological Psychology 66*, 1–5.

Rosch, E. (1973). Natural categories. *Cognitive Psychology 4*, 328–350.

Schanks, D. (1995). *The Psychology of Associative Learning*. Cambridge: Cambridge University Press.

Schneider, D. (2004). *The Psychology of Stereotyping*. New York: The Guilford.

Shep, M. (1958). Shall we count the living or the dead? *New England Journal of Medicine 259*, 1210–1214.

Slovic, P., M. Finucane, E. Peters, and D. G. MacGregor (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis 24*, 1–12.

Sterken, R. (2015). Generics in context. *Philosopher's Imprint 15*, 1–30.

Thomas, L. and K. LaBar (2008). Fear relevancy, strategy use, and probabilistic learning of cue-outcome associations. *Learning and Memory 15*, 777–784.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*, 1124–1131.

Tversky, A. and D. Kahneman (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*, pp. 49–72. Erlbaum: Hilsdale.