Normal Objects, Normal Worlds and the Meaning of Generic Sentences

REGINE ECKARDT
University of Konstanz

Abstract

It has sometimes been proposed that generic sentences make statements about prototypic members of a category. In this paper I will elaborate this view and develop an account where generic sentences express quantification about the normal exemplars in a category—here and in counterfactual worlds sufficiently similar to our own.

Comparing the account to the currently most widespread analysis which views generic sentences as universal quantifications in carefully chosen best-possible worlds, we find that an analysis that is based on the choice of normal objects does better justice to the data in question than an analysis that relies on a choice of normal worlds alone.

A further conceptual advantage of an explicit separation of (a) a choice of best exemplars and (b) a modal component of generic quantification consists in the fact that it highlights that different generic sentences can rely on different kinds of choice of best exemplar. Comparing their logical behaviour, I will demonstrate that we should at least distinguish between *normal*-generic sentences and *ideal*-generic sentences.

Finally, the paper proves that the account I propose is a modal variant of some recent purely extensional default logics, developed in AI.

I INTRODUCTION

The main aim of my paper is to propose a semantics for the generic operator GEN, widely used in the semantic investigation of generic sentences, a semantic account which moreover does justice to certain well known empirical observations about generic quantification:

(1.1) Generic sentences allow for exceptions:
    Dogs bite postmen
    Dogs don't bite postman Otto.
(1.2) Generic sentences involve a modal component:
    Rose handles mail from Antarctica.
(1.3) Generic sentences can be embedded in other modal constructions:
    If every postman got dog training, then dogs wouldn't bite postmen.
(1.4) Nested generic quantification:
    Dogs bite men who are afraid of dogs.

For a more general overview over the discussion of genericity, the reader is referred to the comprehensive survey article of Krifka, Pelletier, Carlson, ter

Meulen, Chierchia and Link (Krifka *et al.* 1995) as well as Pelletier and Asher (Pelletier & Asher 1997). Note that in the light of these surveys, the enterprise of the present paper is restricted to a sub-question in the investigation of genericity, the question about the nature of the GEN operator. Even if I might occasionally talk about the 'investigation of generic sentences' in the following, this is always meant in that limited sense.

The account I propose will be based on two main ingredients. First, I will introduce a family of operators that select for each domain *P* the subset of normal individuals in *P*. Second, I will use a notion of accessible counterfactual worlds, the dispositional orbit of the world of evaluation, which will do justice to the modal nature of generic statements. Finally, the discussion will reveal the fact that we have to distinguish two types of generic sentences, normal-generic statements and ideal-generic statements.

In proposing a new account, it is furthermore of interest to see in what respect this account can do better than previous theories. I will mainly concentrate on a comparison to a family of theories for GEN that I will call 'Best World Theories'. I will raise two main criticisms. First, these accounts only allow for one-dimensional quantification, which causes problems in certain cases. Second, Best World Theories try to capture several interacting factors in generic quantification by one un-analysed operator, and I will argue that this obscures the facts under consideration.

The paper is organized in the following way. In sections 2 and 3, Best World Theories of generic sentences and the Normality Based Theory of generic sentences will be introduced. Section 4 will examine the treatment of exceptions in either account. It will turn out that Best World Theories are forced to follow a strategy of 'stepwise quantification', which will be shown to be inappropriate upon closer investigation in section 5. The Normality Based Theory can face exceptions without problems. Section 6 investigates genericity in complex modal constructions. The remainder of the paper will argue in favour of a distinction between normal-generic sentences and ideal-generic sentences. In section 7, I will take the example of the 100-year-old turtle as an intuitive starting point to introduce this distinction. Section 8 will elaborate on this point, arguing that different kinds of axiomatic restrictions are appropriate for either kind of statement. While 'normality' should be more statistical in spirit, 'ideality' refers more to a conceptual ideal. Section 9 addresses the question of a sound conceptual basis for theories of GEN. An Appendix relates the approach to theories of default reasoning in AI literature. It turns out that the account can integrate widely used systems of nonmonotonic reasoning in a modal framework—a combination that already bore fruit in section 6.

2 BEST WORLD THEORIES

In this section, some reasons to prefer a modal account for GEN over other proposals will briefly be reviewed, and one most recent such account, the theory of Pelletier & Asher (Pelletier & Asher 1997), will be introduced. This theory will provide a concrete object of comparison to the account to be developed in this paper.

It has often been observed that generic sentences cannot be analysed in a purely extensional manner. The most striking evidence in favour of this observation is generic sentences that express regularities that, in fact, have never so far been instantiated. Examples (2.1) and (2.2), taken from Krifka *et al.* (1995), illustrate this point:

(2.1) Rose handles mail from Antarctica.
(2.2) This machine peels oranges.

(2.1) might be true even if no letter from Antarctica has ever reached our office, and (2.2) can felicitously describe a machine that is brand new and has never before been set to work, and would still be true even if the machine got destroyed by an accident before it had seen its first orange.[1]

The overview article by Krifka *et al.* (1995) sketches a modal treatment of the GEN operator in Section 1.2.6, which is based on the rich and elaborate modal accounts developed in Kratzer (1978) and Heim (1982). Apart from that proposal, Krifka *et al.* (1995) list a number of modal treatments of the GEN operator that I will refer to with the cover term 'Best World Theory'. For concreteness, I compare my approach with the recent account proposed by Pelletier and Asher (Pelletier & Asher 1997) which, however, for the issues under discussion can be taken to represent a much larger class of similar approaches, some of them listed below.

Best World accounts of a generic statement of the form $GEN[x_1, \ldots, x_n](\Phi, \Psi)$ elaborate the following paraphrase: take any a, go to those possible worlds which are most normal, or most undisturbed, or most ideal, with respect to matters in question—that is, the proposition $\Phi(a)$—and see whether Ψ holds true. If yes (for all a, for all normal-$\Phi(a)$-worlds), the generic statement is true. If no, it isn't.

Formally, Pelletier & Asher assume a function * which maps pairs of worlds and propositions onto the set of those worlds where the proposition holds true in the most normal way:

(2.3) $*: W \times P(W) \to P(W)$

---

[1] The sentences can also become *false* under such circumstances. This is the more striking observation, as generic sentences tend to be regarded as universal quantifications of some kind.

$^*(w, p) = p'$

where $p' =$ all the normal p worlds according to w, that is, all the worlds in which p, along with the typical consequences of p holds true.

Next, a conditional operator $>$ is defined on basis of the * function in the following way:

(2.4) $M, w, g \models \Phi > \Psi$ iff $^*(w, \llbracket \Phi \rrbracket^{M,g}) \subseteq \llbracket \Phi \rrbracket^{M,g}$,
  where $\llbracket \Phi \rrbracket^{M,g} = \{w \in W: M, w, g \models \Phi\}$

We can now translate the logical form of generic sentences, something of the form

(2.5) $\text{GEN}[x_1, \ldots, x_n](\Phi, \Psi)$

into the logical expression

(2.6) $\forall x_1, \ldots, x_n \ (\Phi > \Psi)$

and will thus get exactly the semantics for (2.5) that was given in the paraphrase above.

Further restrictions of * are discussed in Pelletier & Asher (1997), like FACTICITY and OR. I will give FACTICITY for illustration, an axiom that ensures that worlds that are normal for p also support p:

(2.7) FACTICITY: $^*(w, p) \subseteq p$

Although such axioms, of course, influence the resulting logic of GEN, they will not be of primary importance in my criticisms of best world theories. I therefore refer the reader to the original source for more details, as well as a treatment of defeasible inferences on the basis of this logic.

With respect to the issues to be discussed below, the theory of Pelletier & Asher can be seen as standing for a range of related accounts, ranging from Morreau (1996), Asher & Morreau (1995), back to classical papers like Delgrande (1987, 1988).

3 NORMAL OBJECTS

This section serves to introduce the basic notions of the account I want to propose as an alternative to Best World Theories. The most important ingredient will be the idea of distinguishing normal from not-so-normal objects in a category.

If one starts out from a world w to look for those worlds w' in $^*(w, p)$ which are normal for p, one will in many cases immediately leave w. Formally speaking, $w \notin {}^*(w, p)$ for most propositions p and worlds w, and this should be so, because otherwise default implication would come too

close to classical implication, as demonstrated in Pelletier & Asher (1997), or Morreau (1996). In face of this (quite informal) observation, one might ask how speakers can acquire this kind of sophisticated knowledge about counterfactual worlds. All they can look at is the real world around them. Generic beliefs should be the result of '[speakers'] desire to understand and characterise the world immediately surrounding them' (Pelletier & Asher 1997: 1129), yet the 'normal case in the world immediately surrounding one' seems to be deeply hidden in Best World Theories.

One could propose to define the set of objects a which are normal Ps, for every property P, in world w:

(3.1) a is normal with respect to P in w iff
$$w \in {}^*(w, {}^\wedge P(a))$$

Evidently, the notion of 'real, normal Ps' is an indirect part of the Best World Theory. Yet it remains unclear how the global function $*$ builds on such real cases which are intuitively fundamental for our generic beliefs about P.

We could also turn things upside-down and introduce an explicit notion of 'normal object in P' in order to analyse generic sentences.[2] The basic idea is to distinguish between *all Ps* and *normal Ps*. In order to do that, we introduce a family of functors

(3.2) $N_n \colon W \times (D_e)^n \rightarrow W \times (D_e)^n$

These functors will map all n-ary properties P on to their normal parts $N(P)$. For all worlds w, $N_n(P)(w)$ is the set of all those tuples $a_1, \ldots, a_n$ which are normal Ps in w. Evidently, these $a_1, \ldots, a_n$ should also be P in w, such that we sensibly require (3.3) to hold true.[3] Assumption (3.3) will be repeated as (N 1) in section 8, where more axiomatic restrictions will be discussed.

(3.3) For all w: $N(P)(w) \subseteq P(w)$

On the basis of these operators, we can present a first try at analysing generic sentences.

(3.4) FIRST VERSION:
$$\text{GEN}x_1, \ldots, x_n (\Phi \; ; \; \Psi) \text{ iff}$$
$$\forall x_1, \ldots, x_n (N_n(\lambda s \lambda x_1, \ldots, x_n \Phi)(x_1, \ldots, x_n) \rightarrow \Psi)$$

---

[2] In the survey part of Pelletier & Asher (1997), this strategy is discussed in section 2.3 on 'Prototypes'. The theory to be developed here can actually be seen as a spell-out of that sketch. A similar sketch can also be found in Krifka *et al.* (1995).

[3] I will sometimes omit the arity index of N in the following when this is possible without causing misunderstandings.

(3.4) reads as follows: 'Form the n-ary property Φ, and use it as an argument of the N functor. That will give you the normal Φs. If all of them show Ψ, the generic statement holds true.'

The analysis proposed in (3.4) captures the intuition nicely that generic statements are just talking about normal objects. However, it can easily be argued that the treatment in (3.4) is still too extensional. Many such universal sentences come out true by accident, or because of the finiteness of our world. Consider the following scenario where no choice, however sophisticated, of normal Ps can rescue the case.

(3.5) • It is true that no pope ever had the name 'Bartholomew'.
   • Normal popes are a subset of the actual popes.
   • Thus, (3.4) renders the generic sentence 'Popes aren't called Bartholomew' true.
   • Actually, the sentence is false.

Similar examples arise whenever the set of Ps in a world w is small enough to fail to exemplify some property that, in and of itself, would be quite natural for some P to have. And interestingly, speakers can have quite clear feelings about whether some property isn't exemplified by accident or by principle. (Compare the generic statement: 'Popes aren't called Goofy'.) Another well-known problem for a purely extensional analysis are generic sentences about something that has not occurred yet. Let me repeat example (2.1) for illustration.

(3.6) Rose handles letters from Antarctica.

According to the translation suggested in (3.4), we could only say that for all normal letters x from Antarctica (arriving in our office), Rose handles x. In the (likely) case that no letter from Antarctica has ever reached us, we would quantify over an empty set, which does no harm in making the universal true—however, more things will spuriously be true, due to the fact that there are no letters form Antarctica: All generics in (3.7), for example, would be predicted to be true:

(3.7) Rose eats letters from Antarctica.
   Rose answers letters from Antarctica rudely.
   Joe handles letters from Antarctica.

This is, of course, unwelcome. We can avoid these kinds of accidental universal truths by evaluating (9) not only in ours, but also in a range of other possible counterfactual worlds. Thus, we arrive at an analysis like in (3.8):

(3.8) SECOND VERSION:

$$w \models \text{GEN} x_1, \ldots, x_n(\Phi \ ; \ \Psi) \text{ iff}$$

$$\forall w'(w \approx w' \rightarrow \forall x_1, \ldots, x_n(N_n(\lambda s \lambda x_1, \ldots, x_n \Phi)(x_1, \ldots, x_n)(w') \rightarrow \Psi(w'))$$

A paraphrase would be 'Go to any world w' that is related by $\approx$ to w. Now, check for all objects $a_1, \ldots, a_n$ that are normal Φ in w' whether they are also Ψ in w'. If this works, for all $a_1, \ldots, a_n$, in all w', the generic statement is true.'

Definition (3.8) makes use of the new binary relation $\approx$ which singles out those counterfactual worlds that are relevant in the evaluation of the generic sentence. Formally, the relation $\approx$ is an accessibility relation between possible worlds. Coming from some world w, all those other worlds are $\approx$-accessible that are like w with respect to causal and statistical dependencies and regularities, but may differ from w in isolated accidental facts. (3.9) is meant to capture the intuitive content of $\approx$:

(3.9) $w \approx w'$ stands for: w' is like w with respect to all dispositions, causal and statistical dependencies and regularities but may differ in other facts.

We will call $\{w' | w \approx w'\}$ the *dispositional orbit* or briefly *DO* of w. Note that there is a crucial difference between the underlying content of world selection in Best World Theory, coded as * function, and accessibility $\approx$ in (3.8). While * selects *better* worlds than ours, $\approx$ accesses those worlds which behave *like* our own. These need not be more normal in any way, but may differ with regard to facts that are relevant for our generic beliefs. In example (3.5) above we might reach worlds w' where some normal pope is called Bartholomew, where some unnormal pope has the name Goofy, where no normal pope is called John, but none where a normal pope is called Goofy. Having strong faith in normality operators N_n, one might even think about not restricting the universal quantification at all. As long as N_n picks out the proper subsets in all words, might not the universal in (3.8) hold in *all* worlds w', not only in some carefully selected ones? It will become clear in sections 6 and 9 that dispositional orbits are indispensable, and their role will become more colourful. Yet we will first look at some quite un-intensional cases in order to see the approach at work.

4 NORMAL OBJECTS AND EXCEPTIONS

In this section, I will show that Best World Theories are not very robust in dealing with exceptions. The only way in which they can account for them will lead to further trouble in section 5. The Normality Based Theory does not face comparable difficulties.

Consider the two sentences in (4.1):

(4.1) a. Dogs bite postmen.
  b. Dogs don't bite the postman Otto.

The two sentences are not contradictory. In fact, they perfectly illustrate that generic sentences allow for exceptions. Surprisingly, Best World Theories face some annoying difficulties when treating exceptions. A first natural representation of (4.1a) in the Best World Theory introduced in section 2 could look like this, where the reasonable reference to occasions s when postmen and dogs meet is added:

(4.2) $\forall x \forall y \forall s (DOG(x)\ \&\ POSTMAN(y)\ \&\ MEET(x, y, s) > BITE(x, y, s))$

Formula (4.2) is true in w_0 iff for all instantiations a, b, s for x, y and s the set of normal 'a is a dog and b is a postman and they meet in s'—worlds is a subset of the set of worlds where a bites b in s.

(4.3) $*(w_0,\ DOG(a)\ \&\ POSTMAN(b)\ \&\ MEET(a, b, s)) \subseteq BITE(a, b, s)$

In particular, this should also hold true if b happens to be **otto** the brave:

(4.4) $*(w_0,\ DOG(a)\ \&\ POSTMAN(otto)\ \&\ MEET(a, otto, s))$
  $\subseteq BITE(a, otto, s)$

Sentence (4.1b), on the other hand, will be represented as in (4.5):

(4.5) $\forall x (DOG(x)\ \&\ POSTMAN(OTTO)\ \&\ MEET(x, OTTO, s)$
  $> \neg BITE(x, OTTO, s))$

Given that the constant name OTTO is interpreted as the individual **otto**, this amounts to the condition in (4.6):

(4.6) For all a, s:
  $*(w_0,\ DOG(a)\ \&\ POSTMAN(otto)\ \&\ MEET(a, b, s))$
  $\subseteq \neg BITE(a, otto, s)$

We may faithfully assume that the set of worlds where a bites **otto** in s and where a does not bite **otto** in s are disjoint. Therefore, the conditions in (4.4) and (4.6) contradict each other. This is an unwelcome result, because the consistency of (4.1) is a paradigm issue in genericity.

The shortcoming was pointed out, among others, by Lehman (1989). Pelletier & Asher answer it by offering a different analysis of (4.1a):

(4.7) $\forall x (DOG(x) > \forall y (POSTMAN(y) > \forall s (MEET(x, y, s) > BITE(x, y, s))))$

Matters become quite intricate at this point, which is why we will take the time to spell out (4.7) in detail:

(4.8) For all **a**:
$$*(w_o, DOG(a)) \subseteq$$
$$\{w'| \text{ for all } \mathbf{b}: *(w', POSTMAN(b)) \subseteq$$
$$\{w'| \text{ for all } \mathbf{s}: *(w''', MEET(a, b, s)) \subseteq BITE(a, b, s))\}\}$$

This reads as follows: we pick an **a**. We proceed, starting from w_o, to all those worlds where **a** is a normal dog. There we are. Now we pick up some **b**. We proceed, starting from w', to those further worlds w'' where **b** is a normal postman. In w'', we pick some **s** and proceed to those worlds w''' where **s** is a normal meeting of **a** and **b**. Now, **a** should bite **b** in **s**.

Example (4.1b) receives the corresponding representation given in (4.9):

(4.9) $\forall x(DOG(x) >$
$$\forall s(POSTMAN(OTTO) \ \& \ MEET(x, OTTO, s) >$$
$$\neg BITE(x, OTTO, s)))$$

In this case, we choose an **a** and proceed from w_o first to all worlds which are normal for **a** being a dog. In those, we choose some **s** and go further to worlds w'' which are normal for **s** being a meeting of **a** and the postman **otto**. As the world need not be normal for Otto being a postman, Pelletier & Asher would argue,[*] the meeting can be such that **otto** does not get bitten, although he would be if we were in one of those worlds where he is a normal postman. So far, so good.

We see that if the Best World Theory wants to account for examples like (4.1), it needs to adopt the strategy of 'stepwise quantification' where only one variable is bound at a time. For instance in (4.7), we have one default implication for variable x, one for y, and one for z. Treating them all in one step, like in formula (4.2), was fatal.

At this place, it is necessary to remember what all that walking around in possible worlds was good for. When we had picked some individual **a**, in (4.8), we moved to worlds w' where **a** was a normal DOG. Next, we picked some **b** and went from w' to somewhere else, w'', where **b** was a normal POSTMAN. The problem seems to be that **a** might cease to be a normal DOG (or worse, cease to be a dog at all) once we have reached w''. After all, the global function * in w' cannot take care of some odd chosen dogs when looking for further worlds where **b** conforms to the typical POSTMAN. Let me put it even more dramatically: if * indeed could care for such things,

---

[*] Actually, I am not sure whether they would defend this claim. In the example they present, they can afford to keep their twin of postman Otto, the zookeeper Joe, out of all antecedents of default implications. We cannot represent Otto in the consequent of >, because Otto is part of the event description in the antecedent. Thus, it might be that Pelletier & Asher fail at this example, but let us act as if they didn't.

then the first step from w_o to w' where a is a normal dog must have already preserved, without us noticing, all those normal cats, normal mice, normal sausages etc. which we might want to use in the next sentence about the normal dog a. Evidently, the star operator $*$ has to perform a much more tedious task than authors in Best World Theories acknowledge.

A normal-operator analysis of (4.1) is unproblematic, because we have the explicit means of specifying which variables refer to normal individuals, and which do not. We chose which things should be normal, we form the property with respect to which they should be normal, we can now apply the normality operator of an appropriate arity, and get the antecedent for the implication as a result. The representations for (4.1a) and (4.1b) are given in (4.10) and (4.11), respectively:

(4.10) $\forall w'(w_o \approx w' \rightarrow$
$\forall xys(N_3(\lambda W \lambda XYS.\mathrm{DOG}_W(\underline{X})$ & $\mathrm{POSTMAN}_W(\underline{Y})$ &
$\mathrm{MEET}_W(\underline{X,Y,S})))(x,y,s)(w')$
$\rightarrow \mathrm{BITE}(x,y,s)(w'))$

(4.11) $\forall w'(w_o \approx w' \rightarrow$
$\forall xs(N_2(\lambda W \lambda XS.(\mathrm{DOG}_W(\underline{X})$ & $\mathrm{POSTMAN}_W(otto)$ &
$\mathrm{MEET}_W(\underline{X},otto,\underline{S})))(x,s)(w')$
$\rightarrow \neg \mathrm{BITE}(x,otto,s)(w'))$

Some remarks about notation: in order to make things more readable, I have underlined the arguments of N_3 and N_2. The possible world parameter of properties is notated, somewhat inconsistently, sometimes as an subscript of the predicate in question $(\mathrm{DOG}_w(x))$, sometimes as an extra argument $(\mathrm{DOG}(x,w))$. I hope that this is not too disturbing. I moreover adopt the convention to use capital letters for the lambda-bound variables X and λX in the argument $\lambda X \Phi(X)$ of N in order to facilitate reading.

Formula (4.10) says that in all w' in the dispositional orbit of w_o, we find that all triples of normal dogs, postmen and their meeting are such that the dog will bite the postman. In contrast, formula (4.11) says that in all worlds in the dispositional orbit of w_o, we find that the normal meetings of normal dogs with postman Otto are such that Otto doesn't get bitten. In particular, Otto can stay an unnormal postman throughout all of the DO of w_o. This is, as we shall see in section 6, one of the explicit functions of the dispositional orbit.

5 NORMAL PAIRS AND PAIRS OF NORMALS

In this section, it will be shown that Best World Theories cannot account for situations where normal pairs over A and B are not the same as pairs of

normal A and normal B. Moreover, other perspectival issues in generic statements will be addressed.

The strategy of stepwise quantification becomes more problematic in cases where normal encounters of A and B are usually not encounters of normal A and normal B. I will use the Wolves-and-Men example for illustration. The following generic sentence reflects traditional wisdom about wolves and men. Picturesque tales to that end can already be found in general survey articles on *lupus lupus* as in Brehm (1876) and, more recently, Grzimek (1987).

(5.1) Wolves kill men.

However, modern descriptions of wolves tend to stress the fact that the generic sentence (5.1) is not, strictly speaking, true. In fact, normal wolves tend to avoid encounters with normal (healthy, adult) humans altogether. In the rare case that a normal wolf meets a normal man, so these treatises report, it is usually easy for the human to frighten the wolf away.

Yet these facts are compatible with the assumption that the generalization expressed in (5.1) is talking about what normally happens under those circumstances when wolf and man meet at all: Such meetings most probably take place at times when the wolves are unusually hungry, and where the human is of a special kind, namely looking helpless and easy-to-attack: a sick or wounded person or a child. Although the generic sentence does not talk about normal wolf and normal man, its content is still important enough for humans. We are equally interested in sick and healthy, young and old, when it comes to the question who should be killed and eaten by wolves: nobody should. Therefore, even regularities about normal encounters of a rare kind are worth reporting, if they have such far-reaching consequences.

Stepwise quantification cannot handle such examples properly. Sentence (5.1) will acquire a representation as in (5.2):

(5.2) $\forall x(\text{WOLF}(x) > \forall y(\text{MAN}(y) > \forall s(\text{MEET}(x, y, s) > \text{KILL}(x, y, s))))$

As in previous examples, we select an object a and go where a is a normal wolf, we pick a man **b** and make it a normal man, and now we look for normal encounters of these. Either we shall find that such encounters usually end with the man successfully chasing away the wolf, and will deny the truth of (5.2) on these grounds. Or we can claim that none of these encounters is like what we had in mind (our expectations being shaped by the things that immediately surround us), such that there are none. In this case, (5.2) becomes true, as do many other generic sentences about wolves and men, because vacuous universal quantification is a powerful truth maker.

We could try and make all three, wolf **a**, man **b** and encounter **s** normal at once: a world that is most normal for the proposition 'a is a wolf and b is a man and s is a meeting of a and b' need not support the normality of either conjunct in isolation. Hungry wolves can meet weak men and kill them in such worlds, without changing our standards of a normal man.

(5.3) $\forall x \forall y \forall s (\text{WOLF}(x) \ \& \ \text{MAN}(y) \ \& \ \text{MEET}(x,y,s) > \text{KILL}(x,y,s))$

However, the structure of this formula corresponds to (4.2). The same argument as above will show that sentence (5.1), in the representation suggested in (5.3), will contradict sentence (5.4) which states that Waldo the Wolf is an exceptional case. (Waldo might be a tame wolf.) Again, this prediction does not fit in with our intuitions:

(5.4) Waldo the Wolf doesn't kill men.

An analysis in terms of normality operators will properly distinguish between (5.1) where we quantify over wolves, men and their meetings, and make a statement about the normal triples of that kind, and example (5.4) where we only quantify over normal encounters of Waldo with men.

(5.5) $\forall w' (w_o \approx w' \rightarrow$
$\forall xys (N_3 (\lambda W \lambda XYS(\text{MAN}_W(X) \ \& \ \text{WOLF}_W(Y) \ \&$
$\underline{\text{MEET}_W(X,Y,S)}))(x,y,s)(w')$
$\rightarrow \text{KILL}(x,y,s)(w'))$

(5.6) $\forall w' (w_o \approx w' \rightarrow$
$\forall xs (N_2 (\lambda W \lambda XS(\text{MAN}_W(X) \ \& \ \text{WOLF}_W(\text{waldo}) \ \&$
$\underline{\text{MEET}_W(X,\text{waldo},S)}))(x,s)(w')$
$\rightarrow \neg \ \text{KILL}(x,\text{waldo},s)(w'))$

Let me briefly discuss a further constellation in which normal encounters with A are not the same as encounters with normal A. These highlight the fact that 'Normality' is a subjective notion.

(5.7) Bees are busy.

The generic statement in (5.7) is a variant of the proverbial 'Bienenfleiß' ('busy as a bee'). However, on closer investigation it was found that the normal bee most of the time is dozing hidden in the hive, and that (5.7) is but an unimportant generalization about the normal-bee-we-see and what it does on normal-occasions-when-we-see-it. When we see a bee, we see it working, but this does not mean that most bees work most of the time. (It means that there are many bees in hives, though.)

Example (5.7) is thus based on some selection of occasions when we meet bees. Similar generalizations arise when we meet only a special subset of a set of A (what would we think about bees if all we knew were the drones?).

Real world examples of that kind, however, already have the nasty smell of a prejudice, which becomes even more stinging if the generic statement reflects what some special person x thinks is true about members of a class A. Yet one should be aware of the fact that 'being a normal A' in many cases amounts to a characterization of 'the kind of A I meet with highest frequency'. If those A that I meet are the same as those everyone else meets, the arising subjective notion of 'normal A' is accepted by the whole community. If we can be sure, moreover, that the 'normal A we meet' is also the most common A there is, our generalizations about 'the normal A' become trustworthy observations of reality.

One can capture these three stages of subjectivity by adding a further hidden parameter to the normality operators: $N(x,P)$ yields normal Ps seen from the perspective of x. $N(A,P)$ yields normal Ps seen from the perspective of the community A, something like a weighted generalization of the single subjective normalities $N(a,P)$ of members a of A. In fact, such $N(A, .)$ is all we ever can get, taking larger and larger communities A, but in some cases one might want to express that we have good reason to expect that these subjective normalities will not be shaken by any further community, in which case we could use an unparametrized notion $N(P)$.

Note that a subjective notion of normality can not help us to safe example (5.1). It is not that we only see a special selection of meetings between wolves and men. And even the wolf might realize that the men he would be apt to meet (in order to kill) are not those that are most common. Sentence (5.1) is a generalization about normal encounters of wolves and men, and not a statement about normal men and normal wolves, from whatever perspective.

6 NORMAL HERE AND NORMAL ELSEWHERE

This section discusses the normality based theory and its predictions in modal embeddings.

Let us come back to example (4.1)a, repeated here as (6.1). The corresponding semantic representation is given in (6.2).

(6.1) Dogs bite postmen.

(6.2) $\forall w'(w_o \approx w' \rightarrow$
$$\forall xys(N_s(\lambda W\lambda XYS(DOG_W(X) \ \& \ POSTMAN_W(Y) \ \&$$
$$\underline{MEET_W(X,Y,S)}))(x,y,s)(w')$$
$$\rightarrow BITE(x,y,s)(w'))$$

What should worlds look like which we look at in the evaluation of (6.2)? Certainly both; dogs and postmen should generally look like they do in our

world. For example, we should not reach a world w'' where postmen generally get special dog training in order to impress dogs. If that was the case, the normal postman in such a world would probably not be bitten by dogs, and thus the universal in (6.2) would become false. Therefore, such a w'' should not be element in the dispositional orbit DO of w_o.

However, it makes sense to argue that, if we look for triples of a dog, a postman and a meeting of these which are like those that we see in our everyday surroundings, then the postman of the triple should be like the postmen here in such triples, and in particular he should not have special dog training. The worlds in DO are there to provide counterexamples to accidental universal truths in our world, and to give examples for concepts which happen to be empty, but they are not there to change dispositions.

Things are different in counterfactual embeddings of generic sentences. Sentence (6.3) provides an example where the antecedent of the counterfactual already carries us into counterfactual worlds which have dispositional orbits quite different from ours. In such a world, different generic statements can hold true. The representation is given in (6.4).[5]

(6.3) If every postman got dog training, then dogs would not bite postmen

(6.4) $w_o \models \lambda w(\forall x(\text{POSTMAN}_w(x) \rightarrow \text{DOGTRAINED}_w(x)) \; \Box\!\!\rightarrow$
$\qquad \lambda w'(w' \approx w \rightarrow \forall y \forall z(\text{N}(\lambda W \lambda Y \lambda Z.\text{DOG}_W(Y) \;\&$
$\qquad \text{POSTMAN}_W(Z))(y, z, w')$
$\qquad \rightarrow \neg\,\text{BITE}(y, z, w')))$

Formula (6.4) can be paraphrased as: 'For all "nearby" worlds w where all postmen get dog training, it holds true that in all worlds w' in the dispositional orbit of w, all pairs of normal dogs and postmen you find are such that the dog will not bite the postman in w'.'[6] The closest worlds where the antecedent is true are beyond the DO of the actual world. In the DOs reached from that remote starting point, all postmen get dog training and dogs remain sane enough not to bite them. That is, the notion of closest worlds in the sense of Lewis is used in counterfactual reasoning with generic sentences to ensure that everything else stays as *normal* as it was in

---

[5] I will suppress reference to meetings between postmen and dogs in (6.4). The arguments in the following will mainly be concerned with the looks of normal dogs and postmen and not with the exact nature of their regular meetings. I use explicit quantification over worlds here, following the analysis of counterfactual statements proposed by Lewis (Lewis 1973a,b). The box-arrow is to be read as qualifying the universal quantification over possible worlds in the sense defined by Lewis.

[6] The shortcut 'nearby' in the paraphrase abbreviates the Lewisian quantification in counterfactual implication:

$w_o \models {}^{\wedge}\phi \; \Box\!\!\rightarrow {}^{\wedge}\psi$ iff

for all worlds w where ϕ is true, there is a world w' which is equally or more similar to w_o such that for all further worlds w'' of equal or greater similarity to w_o where ϕ holds true, ψ will be true also.

the world we started from. In particular, even in worlds where postmen get dog training, normal dogs should remain as they are in the actual world. Dogs should not, for example, change their character and all act like pitbull terriers who would fiercely attack even the best-trained postman. In fact, all schemes of improving the chances of postmen against dogs are based on the expectation that dogs will remain like they are while postmen get better equipped for their job.

Let us also have a look at nested generic quantifications, which turn out to be another case of modality in genericity, for example in relative clause constructions as in (6.5):

(6.5) Dogs bite men who are afraid of dogs.

The representation in (6.6 d) explicates that we make a generic statement about people with certain dispositions, that is for whom other generic statements are true. I give some intermediate steps of the derivation in (6.6a–c). The parts in boldface are those that are new, respectively.

(6.6) a. $\lambda w \lambda y \forall w' (w' \approx w \rightarrow \forall x (N(\lambda W \lambda X.DOG_W(X))(x, w') \rightarrow$
$FEAR(y, x, w'))$
= the property of being someone who is afraid of dogs

(6.6) b. $\lambda w \lambda y (\mathbf{MAN_w(y)} \ \&$
$\forall w' (w' \approx w \rightarrow \forall x (N(\lambda W \lambda X.DOG_W(X))(x, w') \rightarrow FEAR(y, x, w')))$
= be a man who is afraid of dogs

(6.6) c. $N[\lambda W \lambda X \lambda Y (\mathbf{DOG_W(X)} \ \&$
$MAN_W(Y) \ \& \ \forall w' (w' \approx s$
$\rightarrow \forall x (N(\lambda W' \lambda X' DOG_{W'}(X'))(x, w') \rightarrow FEAR(y, x, w')))))]$
= property of being a normal dog–man-pair such that the man is afraid of dogs

(6.6) d. $\forall w^* (w_o \approx w^* \rightarrow$
$\forall x \forall y (N[\lambda W \lambda X \lambda Y (DOG_W(X)$
$\& (MAN_W(Y) \ \& \ \forall w' (w' \approx W \rightarrow \forall x (N(\lambda W' \lambda X'.DOG_{W'}(X'))(x, w')$
$\rightarrow FEAR(y, x, w'))))](x, y, w^*) \rightarrow \mathbf{BITE(x, y, w^*)})$

In (6.6)b, we collect all pairs of men and worlds such that 'man a is afraid of normal dogs, in w'. This might be the case because a is a shy person in w and has had bad experience with dogs, which behave like our dogs do otherwise. However, it also might be the case that normal dogs in w are very fierce animals; they might be 2 metres high, very aggressive, and a needs no special characteristics in order to be afraid of them. It even might be that a is a lunatic and dogs are all very small, peaceful and friendly animals but he fears them nevertheless. (It might even be that *all* men show this kind of lunacy in some world, such that even the normal ones are afraid of friendly dogs.) The operator N in (6.6c) takes an argument that is the

conjunction of DOGS and MEN who are afraid of dogs *in any sense*. Only in (6.6d) do we restrict our attention to those worlds that are in the dispositional orbit of w_o, the world of evaluation. In doing this, we get rid of worlds where a normal peaceful, small, friendly dog meets the normal lunatic who is afraid of dogs—and does not bite him. Formula (6.6d) states that in all worlds that are dispositionally like our own, a normal pair comprising a man who is afraid of a normal dog (in our sense) is such that the dog will bite the man.

7 NORM AND IDEAL: THE 100-YEAR-OLD TURTLE

I propose to distinguish between normal-generic statements and ideal-generic statements. In this section, the distinction will be motivated in terms of plausibility arguments.

The following (true) generic sentence is known from the literature as a puzzling case:[7]

(7.1) Turtles live to be a hundred years or more old.

Sentences like (7.1) were used in the first place to argue that generic quantification does not amount to a quantifier like 'most' or 'the most common'. We know that the vast majority of turtles die very young and only the strongest, happiest exemplars live to a biblical age. It is simply false to claim that (7.1) is true because most turtles live to 100 years.

However, sentence (7.1) is problematic even for more sophisticated theories, as Pelletier & Asher (1997) demonstrate. Those Best World Theories that imply the existence of absolutely best worlds (e.g. Delgrande 1987) will have to assume a world where in fact all turtles do live for 100 years. It is easy to imagine that such a world would not be the absolutely best world in many ecological respects.

Pelletier & Asher represent sentence (7.1) as in (7.2):

(7.2) $\forall x (\text{TURTLE}(x) > \text{LIVE-TO-100-YEARS}(x))$

Thus, we take any **b**, move to a world that is normal for **b** being a turtle, and see **b** become very old. As we can check for each turtle separately, Pelletier & Asher would argue, there is no need for worlds where all turtles get old at once, and ecological disasters are avoided.[8]

---

[7] Quoted from Pelletier & Asher (1997: 1164).

[8] Remember, however, our observation in section 4 that stepwise quantification requires that objects once made normal remain normal in the next step. Therefore the question whether Pelletier & Asher can avoid such ecological disasters is not finally settled yet.

In principle, we can easily mimic the analysis in (7.2) by the representation in (7.3). Yet I will argue that (7.3) offers a hint that there is more to the 100-year-old turtle than that captured in (7.2)/(7.3).

$$(7.3)\ \forall w(w \approx w_o \rightarrow \forall x(N(\lambda W \lambda X.\text{TURTLE}_W(X))\,(x, w) \rightarrow$$
$$\text{LIVE-TO-100-YEARS}(x, w))$$

Taking (7.3) seriously, we have to acknowledge that not many turtles are normal in our world. Even if we ignore the many little baby turtles that get eaten soon after birth and argue on the basis of turtles-we-meet (compare section 5), we have to face the fact that (7.3) requires that the turtle we normally meet is not a normal turtle. In the extreme, it would even be possible that there was *but one known turtle* that was observed to be that old. Given the biological background we have, we would still accept (7.1) to be true. This amounts to the claim that there was but one known normal turtle.

It is certainly obscure to claim that all the generic knowledge we have about turtles has come about from the observation of this one single exemplar. The Methuselah Turtle might have many accidental features. We perhaps acquired our knowledge about interbreeding behaviour from quite different turtles, and gourmets' expectations about the taste of turtle meat were certainly shaped by younger exemplars.

The standard reply to that kind of worry would be that generic generalizations are something more sophisticated than talking about 'the average Q': generic sentences need not always talk about the average Q. Yet there is a certain tension between the standard reply and the equally plausible claim that 'people notice regularities in nature, and form ... folk-laws to codify these regularities and to predict what the future will bring' (Pelletier & Asher 1997: 1129). The quotation offers a sane conceptual basis for generic quantification. The position underlying the standard reply, on the other hand, can only say that there might be some basic cases in which 'generic' means 'normal' but that, in an undescribed process of holistic integration of all kinds of bits of folk-law, the * operator, or N functor, emerges which selects worlds/objects due to laws quite different from those of statistics.

I propose that we should treat (7.1) as a different kind of generic sentence, and we should do that because it actually *is* different from statements like (6.1). (7.1) states something about the potential age of turtles, but matters are such that the normal course of the *world* rather *inhibits* turtles from exemplifying this potential. A turtle that is 'normal' in the sense of (7.1) is one where, quite un-normally, all those incidents that cause an early end for most turtles did not take place. The 'normal' turtle in (7.1)

and the 'normal' course of events are in conflict. Therefore, I propose to analyse (7.1) in terms of *ideal* rather than normal turtles.

Generalizations like (6.1), on the other hand, are such that the normal Q (dogs, postmen, etc.) talked about are those Q which are the product of the normal course of the world. A normal world does not prevent dogs from developing an infelicitous sportive interest in postmen's trousers. And it is not the case that only those postmen who are spared the experience of the everyday postman show dog-incensing timidity: normal postmen are those we find in the normal world. We thus need to distinguish *normal* from *ideal* generic sentences:

- Normal Qs are those produced by the normal course of events. Ideal Qs might be rare, because the normal course of events inhibits them. Thus, we select normal and ideal Qs by different criteria.
- Generic sentences about normal Qs can be falsified by pointing out a large number of counterexamples. Generic sentences about ideal Qs are immune to such an argument.
- Many properties that we observe with 'normal Qs', and that are not supported by any kind of theory, are—perhaps accidentally—never exhibited by one of the rare ideal Qs. If our generic beliefs about Qs were based on ideal exemplars, the corresponding generic sentences should be judged false. This is an intuitively wrong prediction.

Let me give another example. It is well known that books on mushroom with drawn pictures are more reliable guides for the mushroom-hunting gourmet than books with photographs. The reason is that it seems to be almost impossible to take a picture of an exemplar that has escaped all damage to its ideal shape. The painter can develop a picture of an ideal exemplar, relying on other, *normal*-generic sentences: 'Normally, we find a bite like that when a snail has eaten from a mushroom. Thus, a snail has damaged this mushroom. If the mushroom had escaped this damage, the bite would be missing.' We know what an ideal undamaged mushroom looks like because we know what features are the result of damage. And, even if there was not a single undamaged toadstool in the world, we would still be sure that the undamaged ideal fly agaric was poisonous, because we know that all normal ones are.

In formal terms, we will assume a further family of functors $(I_n)_n$ N of the same type as the 'normal' functors we have used so far. We can replace N by I throughout all definitions and will get a semantic account for ideal-generic sentences analogous to the account for normal-generic sentences. For now, the difference between a representation in terms of N and a representation in terms of I is barely a conceptual one. The points at which

this conceptual difference yields quite concrete consequences will be pointed out in the next section.

Clearly, Best World Theory would have the parallel option to add a second * operator which selects for Φ-ideal rather than Φ-normal worlds. Yet it is symptomatic of the conceptual vagueness of the approach that the necessity of such a move for the treatment of the 100-year-old turtle was never noticed.

A first empirical observation in favour of a separation of normal-generic sentences and ideal-generic sentences is that both (7.4a) and (b) apply correctly, in some sense, to some unnamed scientific journals:

(7.4) a. Papers get reviewed in about 8 weeks.
      b. Papers get reviewed in about 8 months.

A more far-reaching observation in favour of separating ideal-generic and normal-generic sentences concerns the desirable link to theories of speakers' reasoning about new objects introduced in discourse. Formal investigations of speakers' understanding of longer texts have revealed that, in order to 'make sense' of a given piece of text, they have to rely substantially on *default inferences* in order to conjecture plausible anaphoric links, interpret definite descriptions and to form hypotheses about temporal and causal relations between the facts and events reported. It is natural to assume that the contents of generic sentences form, so to speak, the *database* in such everyday reasoning and that nonmonotonic logic investigates the *inference patterns* that speakers apply. Finally, speakers can apply generic laws to the special case if they assume that, unless told otherwise, the individuals they talk about are as normal as can be in the categories as members of which they were introduced in the discourse. This is indeed the intuitive core of several attempts to offer a formal spell-out of this link (see Delgrande 1988; Asher & Morreau 1995; Pelletier & Asher 1997).

Once more it turns out that normal-generic sentences play a different role in this kind of reasoning from ideal-generic sentences. Assume that we are engaged in a discourse about the turtle Agatha. While we can safely hypothesize that Agatha is a normal turtle who lives under water, moves slowly, is dark-greenish, etc., we will be more hesitant with respect to the assumption that Agatha will live to be 100 years old. (One might compare one's own expectations about the age one will reach with the generic statement that *humans* live to be 90 years or more.) We have argued that the ideal is not usually what we meet every day. In a normal world, Agatha will most probably *not* be an ideal turtle. Accounts of generic sentences that do not make this distinction will, however, predict that we expect Agatha to become a Methuselah Turtle as naturally as we expect her to live in water.

8 NORM AND IDEAL:
SOME LOGICAL CONSIDERATIONS

In this section, the distinction between normal-generic and ideal-generic sentences will be further defended looking at the different logical consequences speakers draw from either kind of statement.

The present section discusses some axiomatic restrictions on the behaviour of normality and ideality operators. In testing our intuitions about appropriate axiomatic restrictions and their logical consequences, we find further support for the distinction between ideal-generic statements and normal-generic statements. While I will propose axioms for the normality operators N that turn normality into a quasi-statistical notion, the ideality operators I behave differently. The relation between N and certain similar proposals made in the AI literature (which elaborate the statistical origin of the definitions in more depth than I can do here) will be discussed in the Appendix.

The first requirement that is reasonable for both normality and the ideal is that normal Φ's must be Φ's, and equally, ideal Φ's must be Φ's. This is reflected in (8.1) and (8.2). The appropriate first-order axioms are given together with the respective set-theoretic clauses in order to facilitate reading.

(8.1) (N 1): $\forall x(N(\lambda W\lambda X.\Phi_W(X))(x,w) \rightarrow \Phi(x,w))$
(8.1a) $N(A) \subseteq A$
(8.2) (I 1): $\forall x(I(\lambda W\lambda X.\Phi_W(X))(x,w) \rightarrow \Phi(x,w))$
(8.2a) $I(A) \subseteq A$

If something is, for example, a normal cup, it must in particular be a cup, or if someone is an ideal postman, then he must also be a postman. Not only are these assumptions intuitively plausible, but similar requirements are also made in Best World Theories, under the label of FACTICITY (see section 2).

The next point is already more problematic. If there are Ps, should there also be normal/ideal Ps? Intuitions diverge at this point. With respect to normal Ps, it can reasonably be argued that no natural class of objects consists *entirely* of exceptions. Perhaps the class falls into objects of varied shapes and properties which we all encounter with equal frequency, but then this rather means that all of them are equally normal, and not that all of them are equally abnormal. Elaborating this thought ($N(P) = P$) amounts to the prediction that normal-generic statements about the class in question amount to, at least in the world under discussion, universal statements about P. This seems a fairly innocent conclusion.

Matters are evidently different for ideal objects. In stressing that ideal objects in P might arise only under quite non-normal circumstances, we

already allowed for the possibility that some quite normal worlds (among them the real one) could be without any ideal P objects, for some categories P. Even if there are Ps, $I(P)$ might be empty. Therefore, I propose the following non-emptiness condition only for N, but not for I.[9]

(8.3) (N 2): $\exists x \Phi(x, w) \rightarrow \exists x N(\lambda W \lambda X. \Phi_W(X))(x, w)$
(8.3a) $A \neq \varnothing \rightarrow N(A) \neq \varnothing$

The lack of an analogous axiom (I 2) will lead to different predictions about the logical strength of normal-generic and ideal-generic sentences which are empirically justified. We will come to these in a moment.

The final restriction will relate normal/ideal Ps to normal/ideal Ps *that are* Q. Axiom (8.4)/(8.5) are motivated by more intricate considerations than (8.1)–(8.3). Basically, they will allow one to maintain the global perspective that everyday default reasoning *is* reasoning in terms of generic sentences. (8.4) is the weakest way of restricting the range of possible N functors in a way such that N will support the default inference patterns discussed below: RATIONAL MONOTONICITY, CAUTIOUS MONOTONICITY and WEAK CONTRAPOSITION.[10]

(8.4) (N 3): $\exists x(N(\lambda W \lambda X. \Phi_W(X))(x, w) \wedge \Psi(x, w)) \rightarrow$
$\forall x(N(\lambda W \lambda X. \Phi_W(X) \wedge \Psi_W(X))(x, w) \leftrightarrow$
$N(\lambda W \lambda X. \Phi_W(X))(x, w) \wedge \Psi(x, w))$
(8.4a) $N(A) \cap B \neq \varnothing \rightarrow N(A \cap B) = N(A) \cap B$
(8.5) (I 3): $\exists x(I(\lambda W \lambda X. \Phi_W(X))(x, w) \wedge \Psi(x, w)) \rightarrow$
$\forall x(I(\lambda W \lambda X. \Phi_W(X) \wedge \Psi_S(X))(x, w) \rightarrow$
$I(\lambda W \lambda X. \Phi_W(X))(x, w) \wedge \Psi(x, w))$
(8.5a) $I(A) \cap B \neq \varnothing \rightarrow I(A \cap B) \subset I(A) \cap B$

(8.4) requires that if there are normal Ps that are Q at all, then the normal P-*and*-Q's should be those normal Ps which are Q. It predicts, for example, that if we find normal postmen who have a beard then the normal-postmen-who-wear-beards are those normal postmen who wear beards anyway. No different standards of normality should apply as a result of further known circumstances.

In the case of idealness, the weaker requirement (8.5) can be paraphrased as 'if some ideal Ps are Q, then the ideal P-*and*-Q's should at least be a

---

[9] We will see reasons that allow I to be a partial function. We seem to have *statistical folk knowledge* about any kind of category, but not all categories need to have something like a 'prototype'.

[10] The list does not exhaust the range of 'prominent' patterns in nonmonotonic reasoning but will serve to exemplify the usefulness of (8.4)/(8.5). For a fuller discussion of the status of, and logical relations among such inference patterns, see Chellas (1974), Ginsberg (1994), Kraus/Lehmann/Magidor (1990), Rott (1996)/(to appear), to name but a few. The Appendix shows how our approach ties in with this kind of literature.

subset of the ideal *Ps* that are *Q*.' Once more, this reflects the fact that the logic of ideal objects is different. The ideal objects in a large and undifferentiated class can be less specific than the ideal of a more narrow class. Consider the following example:

(8.6) Cups are (ideally) made of porcelain
    Porcelain cups (ideally) have a gold rim
    But: Do cups indeed (ideally) have a gold rim?

Although the notion of a cup evokes the material 'porcelain', the idea we have of a spitting image of a porcelain cup is even more specific than the ordinary cup-of-porcelain. This is captured by (8.5) while retaining the intuition that even ideal *P-that-are-Q* should not lie far beyond the ideal *P*, as long as *Q* in and of itself is not an entirely un-ideal property of *P*.

Note that if the latter should actually be the case, that is, if *Q* is a somehow odd property, normal/ideal *P-and-Q* are free to have any appearance. This is illustrated by the following example:

(8.7) Dogs (normally/ideally) don't suffer from cancer.
    Dogs who suffer from cancer may have quite different properties
        from dogs in general.

Before moving on to such cases where statements about atypical exemplars are made, I shall list some simple consequences of (N1)–(N3) and (I1) and (I3). Unless stated otherwise, the laws hold for both, normal-generic and ideal-generic statement.

RATIONAL MONOTONICITY:

If As are C, and it is not the case that As are non-B, then As that are B are C.

Proof: Assume $N(A) \subseteq C$, and not $N(A) \subseteq \neg B$. Then $N(A) \cap B \neq \emptyset$, and therefore $N(A \cap B) = N(A) \cap B$. Thus $N(A \cap B) \subset N(A) \subseteq C$. This holds in all possible worlds. For *I*, we must argue in some more detail. If $I(A) \neq \emptyset$, the conclusion follows as in the above case. However, we might be in a world *w* where $I(A) = \emptyset$. Yet there must be worlds in the DO of *w* where $I(A) \neq \emptyset$, as otherwise the second, negated ideal-generic sentence would not be true. (If we had no ideal *As* anywhere, anything could be claimed to hold true about ideal *As*.) In all those worlds, $I(A) \neq \emptyset$ and moreover $I(A) \subseteq C$, as otherwise the first ideal-generic sentence would not hold true. As before, we can deduce that $I(A \cap B) \subseteq C$ in all these worlds.

RATIONAL MONOTONICITY is one of the most widely accepted requirements to generic implication, and default reasoning in general (see Pelletier &

Asher 1997, as well as the sources quoted therein). An example is given below:

(8.8) If dogs have hair, and it is not the case that dogs are generally not brown, then brown dogs have hair, too.

CAUTIOUS MONOTONICITY:

If As are normally B, and As are normally C, then As that are B should normally be C.

Suppose that there are ideal As. Then, if As are ideally B, and As are ideally C, then ideal As that are B should ideally be C, too.

Proof: Assume that $N(A) \subseteq B$, and $N(A) \subseteq C$. As $N(A) \cap B$ is thus nonempty, $N(A \cap B) = N(A) \cap B$, and therefore is a subset of $N(A)$. Therefore $N(A \cap B) \subseteq C$.

In the case of ideal-generic statements, we have to make the assumption that there are worlds in the DO of ours such that these contain ideal As. As before, the conclusion follows in these worlds in the same way as in the proof for the normality case. Given that generic sentences make necessity claims about all worlds in DO, the claim about ideal-generic statements follows.

(8.9) If dogs have four legs, and dogs love sausage, then dogs that have four legs love sausage.

Note that Pelletier & Asher doubt the validity of inferences like (8.9). I propose that this is the case because we tend to give causal force to the more narrow description of dogs in the consequent of (8.9): not only do dogs that have four legs love sausage, but they do so somehow *because* they have four legs. The sentence 'dogs love sausage', in contrast, suggests that dogs in general love sausage, not only when or because they have four legs. However, if we strip off these further implicit assumptions of the sentence, the result is a convincing deduction.

(8.10) Dogs have four legs.
Dogs love sausage.
Dogs that have four legs love sausage.

A final property that can be proved for normal-generic sentences, yet not for ideal-generic sentences, is the principle of WEAK CONTRAPOSITION:

WEAK CONTRAPOSITION:

If normal As are B, then

(i) either normal $\neg$ B's are $\neg$ A,
(ii) or else $\neg$ B is an un-normal case altogether.

I will first give examples for both cases of the disjunct. Example (8.11) looks like a case where something like a principle of contraposition would be quite welcome. Example (8.12), in contrast, illustrates a case where general contraposition would yield absurd consequences, and which exactly fits the second case of the above disjunction. This discussion is adopted from Ginsberg (1994).[11]

(8.11) Koala bears live on eucalyptus.
        → Animals who don't live on eucalyptus are normally not Koala bears.
(8.12) Men (normally) do not suffer from diabetes.
        ↛ People who suffer from diabetes are normally female.

Formally, the relation between both pairs of sentences in (8.11) and (8.12) conforms to the pattern known as contraposition. The implication in (8.11) looks sound. Something like it might actually be in use when a biologist, knowing not much more about Koala bears than (8.11), looks out for Koala bears in the jungle: If the putative Koala bear is seen regularly eating coconuts, the biologist will adopt the hypothesis that the animal is probably not a Koala bear after all.

Yet example (8.12) would clearly be an undesired implication. The reason seems to be that suffering from diabetes is uncommon both for men and their complement (in context), women. Therefore, if normal men do not suffer from that disease, normal women may not suffer from it either.

These two cases are exactly what is allowed by WEAK CONTRAPOSITION. The proof that it holds true for N is not so simple as the properties we checked so far. In particular, it relies on (N1) to (N3) in full strength and therefore will not carry over to ideal-generic sentences. However, this might even be desirable.

We argued in section 7 that normal-generic sentences and ideal-generic sentences should be kept separate. Ideal-generic statements say something about how a·member in P will look like under maximally undisturbed circumstances, while normal-generic statements say something about members of P under 'normally disturbed' circumstances. Being confronted with some new, so far unknown, P object in conversation, one will assume that it most likely is a normal P, but probably not an ideal P. Therefore, default conclusions on the basis of normal-generic sentences, but not ideal-generic sentences, will be drawn. Another facet of the same observation is that contrapositions of ideal-generic sentences like the one in (8.13) are

---

[11] Note that we, following Ginsberg and similar discussions by von Fintel (1998), talk as if the entire universe were, for these matters, restricted to animals or humans, respectively.

intuitively wrong. (8.14) offers the general pattern which is used in the discussion below.

(8.13) Turtles live to be 100 years old.
$\not\rightarrow$ Animals who don't get to be 100 years old are not turtles.
(8.14) *As* are ideally *B*.
$\not\rightarrow$ non-*B*'s are non-*A*. (in whatever reading)

In (8.13), it is not the case that the negated consequent $\neg B$, 'not living to be 100 years old', is generally a marked (non-normal or non-ideal) property in the animal kingdom, like 'suffering from diabetes' was for humans. Therefore, the failure of (8.13) would not be captured by something like case (ii) of weak contraposition. The deeper reason is that no version of contraposition makes any sense for ideal-generic sentences. Knowing that the ideal A has property B, we will not conclude anything about $\neg B$'s, normal or ideal, and for various reasons.

If we read (8.14) as talking about normal $\neg B$, we have to take into account that we were talking about ideal *As* and may therefore not be surprised if quite normal, but simply un-ideal *As* already show $\neg B$. Therefore, $\neg B$ in the normal case should not lead us to conclude anything about being *A*. (If an animal dies at the age of 80, it may still be a turtle without being anything extraordinary otherwise.)

If we read (8.14) as taking about ideal $\neg B$, we will very often make the observation that the complement of a category, which is such that we have a notion of an ideal exemplar in that category, is not such that a similar 'prototypical $\neg B$' would exist (see fn. 9). Consider the above example: what should an ideal *animal-which-doesn't-reach-100-years* look like? Would it rather be furry or skinny? Does it have four legs or more? These questions seem hard to answer.

Thus, the implication in (8.14) in such cases does not hold true for reasons beyond those captured in weak contraposition: the second sentence in (8.14) is, first of all, a generic sentence. If we read it as a normal-generic sentence, the implication becomes false because normal $\neg B$s may be *A* in spite of ideal *As* being $\neg B$. If we read it as an ideal-generic sentence, it is probably not even interpretable because the antecedent 'animals of age less than 100 years' is not an appropriate argument of the *I* functor. We will capture this by assuming that *I* are partial functors.[12]

However, sometimes matters are such that the complement $\neg B$ in context is a category that is an eligible argument for *I* (that is, is the kind of

---

[12] An equally satisfying assumption, conceptually, might be that $I(\neg B) = \varnothing$ in such cases. However, this would lead to evident complications, because the respective generic sentences would then be predicted to be, first of all, true (empty universal quantification), and further assumptions would be necessary to explain why that makes them unacceptable.

category where it makes sense to ask about ideal exemplars). One of the most prominent constellations is, of course, that we are in a context where our attention is restricted to humans, such that B and $\neg$B correspond to males and females. I am aware of the difficulty of producing any inoffensive generic statements involving gender, but let me, for this one occasion, nevertheless try.[13]

(8.15) Ideal gynaecologists are women.
Ideal men are not gynaecologists.

Intuitively, the two sentences are logically independent. Or, more bluntly, why should our image of the ideal gynaecologist have any influence on our ideas about ideal men?

To summarize, intuitions with respect to contraposition provide further evidence in favour of a distinction between normal-generic and ideal-generic sentences. It remains to be shown that weak contraposition for normal generic-sentences follows from (N1)–(N3). This will require some more technical considerations.

It can be shown that axioms (N1)–(N3) induce a *global* ordering of all objects in the universe of the model, according to their degree of normality. Formally, we can derive for each possible world w a family of functions $(R_n)_{n \in \omega}$ and ordered sets Ω_n such that

(i) R_n maps D^n into Ω_n.
(ii) for each subset A of D^n which is definable in the logical language L we use $N_n(A) = \{a \in A \mid R_n(a) \text{ is minimal in } A\}$ in w.

Intuitively, the mapping R_n transfers the ordering of Ω_n on to D^n, the set of all n-tuples over the universe D. The ordering of D^n is to be read as a ranking according to degrees of *abnormality*, such that the elements of *lowest* rank are the most *normal* ones while the ones with *higher* ranks are the more and more *unnormal* ones.[14] Figure 1 illustrates this correlation. The proof that this is the case will be given in the Appendix.

The use of such ranking functions, based on ordinal numbers Ω, was proposed by Spohn (1988). Its generalization to sets of individuals was first suggested independently by Brafman (1996) and Weydert (1997). The latter two approaches, however, are explicitly set up in purely extensional terms, and it is not immediately evident how a modal dimension is to be

[13] I can adopt at least the first sentence among my personal ideals—but remember that 'ideal', like 'normal', in the end is a notion that reflects a personal perspective on the world . . .

[14] The direction of this ordering is a heritage of the statistical origins of this kind of modelling. I keep it in order to facilitate the formal comparisons to be made in section 9. I apologize for the extra effort thereby caused for the reader. It takes some time to get accustomed to the inverse correlation that the *more* normal an object, the *lower* its rank.
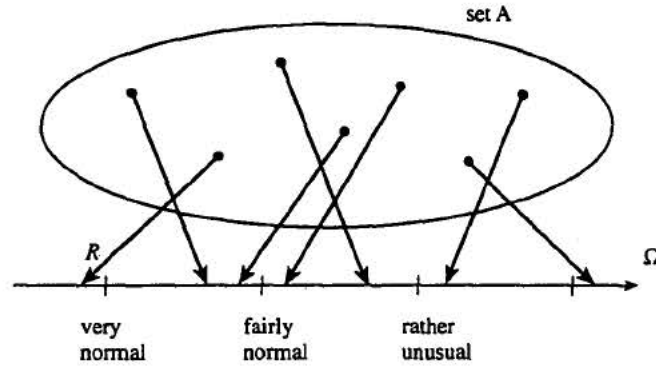
set A

very
normal

fairly
normal

rather
unusual

R

Ω

Figure 1

introduced. We have argued at various places (and follow the general discussion on generic sentences in that respect) that a purely extensional treatment of generic sentences is inadequate. In this sense, our treatment extends these approaches in a way necessary for reasonable application in natural language semantics, even if (as will be shown) our treatment of normality, extensionally speaking, is equivalent to the proposals by Brafman and Weydert.

With this ordering at hand, we can now prove WEAK CONTRAPOSITION for N_n. Let me repeat the claim in a more formal manner:

WEAK CONTRAPOSITION: Let M be a model of a language L that includes a family of normality operators $(N_n)_{n \in \omega}$ as introduced above. Assume that (N 1) to (N 3) hold true in M. Let A and B be L-definable subsets of D^n such that $N(A) \subseteq B$. Then

    (i) $N(\neg B) \subseteq \neg A$ or
    (ii) $N(B \cup \neg B) \subseteq B$

Proof: Assume that (i) does not hold true. This means that $N(\neg B) \cap A \neq \emptyset$, and we now have to show that $N(D) \subseteq B$. If $N(\neg B) \cap A \neq \emptyset$, then we know that an element x in the intersection must have higher rank $R_n(x)$, that is, be *less* normal, than all elements in $N(A)$, because the latter—most normal ones in A—are all in B. The normal elements b in B in turn are either in part among those in $N(A)$, or else are all outside A and have thus even lower rank (= are even more normal). In any case, $R_n(x) \geq R_n(b)$, which means that the most normal objects in D are in B. This is exactly what (ii) states.

Before concluding this section, I will offer some more sample constellations of global ordering according to normality. The first constellation, depicted in Figure 2 is one where we have A and B intersecting, and where some, but not all normal A are normal B.
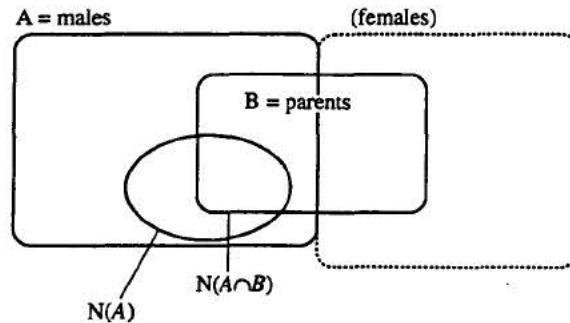
Figure 2

We could, for example, instantiate A as *male* and B as *parent*. The above picture would then say that some, but not all, normal males have children. (Or, in other words, neither sentence 'men have children', nor 'men don't have children', is a true normal-generic sentence in our world—which I take is a valid assumption.) Given this, the global notion of normality encoded by (N1)–(N3) will predict that the normal male parent is simply a normal male who happens to have children. In particular, this has the consequence that all normal-generic sentences about males carry over to males with children, which is a reasonable prediction.

The diagram in Figure 3 offers a graphical summary of the diabetes example. Normal males do not suffer from diabetes. Nor do normal females. Suffering from diabetes is just an extraordinary property.

Finally, Figure 4 illustrates the widely discussed Quaker/Republican example. We can have two properties A and B where $N(A)$ does not intersect with B, and $N(B)$ does not intersect with A either. No further implications follow. This is exemplified (quoting generic statements about
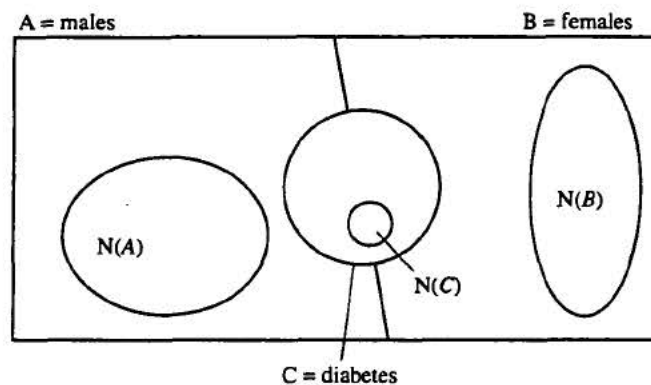

Figure 3
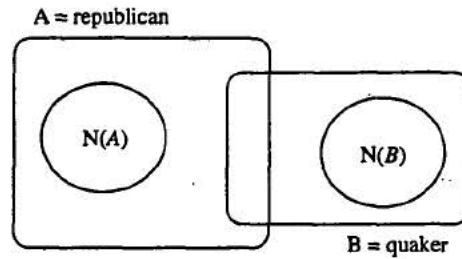
A = republican

N(A)   N(B)

B = quaker

Figure 4

political interests from the literature) by the observation that Republicans normally are not Quakers, and Quakers normally are not Republicans. Logically speaking, this leaves all options open for those persons who are both a Quaker and a Republican. They are just not covered by the normal-generic knowledge expressed in the sentences 'Quakers normally are not Republicans' and 'Republicans normally are not Quakers'.

To summarize, the proposed set of three axioms turn out to restrict the logic of N, the normality operators, in a strong way. I will argue in the next section that such a strong version of normality is necessary for general reasons. Yet notice that one advantage of the account is its flexibility. The precise nature of axiomatic restrictions we want to adopt for normal-generic, or for ideal-generic statements, can easily be adjusted according to further considerations, while the underlying intended content of the operator remains stable. This distinguishes our account from proof theoretic approaches to default reasoning, where new sets of derivation rules lead to entirely new logical systems without making it clear in what sense all are competing spell-outs of a uniform underlying idea.

9 SOME CONSIDERATIONS ABOUT THEORY ARCHITECTURE

In this section, I will discuss potential circularities luring in Best World Theory and Normality Based Theory. I will argue that the distinction between normal-generic and ideal-generic statements is a first step towards avoiding such circularities.

How do generic sentences relate to the more basic facts in the world? It seems reasonable to assume that generic facts, or at least some of them, are something like rough generalizations over simple facts (i.e. facts about single objects and their properties). We have already seen some statements to this end in the literature on Best World Theory.

However, Best World Theory ultimately does not reflect much of this insight. In section 3, it was shown that 'normal objects in P' can be defined only indirectly, and that the *-operator does not systematically build on 'everyday objects in our world'. In section 4, we saw that the choice of worlds in $*(w, p)$ has to respect the properties of previously chosen individuals a which were ideal in w for some property Φ. (Of course, the more choices have to be made in stepwise quantification, the more individuals the *-operator will have to keep in mind.) In section 7 I showed that worlds w in $*(w_o, p)$ are sometimes selected rather for idealness than the normality of the matters in question. Thus, it is not clear whether the 'meaning' of * can be characterized in independent terms, and for now the only safe thing that can be said about * is the somewhat uncharitable (9.1):

(9.1) Let Δ be the set of all desirably true generic sentences $\text{GEN}x(\gamma \; ; \; \delta)$ about γ in w_o.
A world w is normal for proposition $\gamma(a)$ (that is, $w \in *(w_o, \gamma(a))$)
if and only if $w \models \delta(a)$ for all δ such that $\text{GEN}x(\gamma \; ; \; \delta)$ is in Δ.

What (9.1) states is this: you want to know what generically holds true for γ's? Look into all those worlds that are normal for γ. How do I know what worlds are normal for γ? First of all, they have to fulfil all generic implications for γ . . . .[15]

Now, (9.1) would be innocent if any independent criteria were in sight to characterize $*(w_o, \gamma)$. Yet we have seen that the shape of object a is not enough for a world to be normal for $\gamma(a)$, that even the idea that generally everything should be as *normal* as possible for γ was not enough, and that in the end $*(w_o, \gamma)$ was the holistic integration of various complex steps of generalisation. Therefore, (9.1) for now is all Best World Theory can offer when it comes to characterising the star operator *.[16]

Naturally, a different set-up does not automatically prevent such obscurities. An uncharitable characterization (9.2) of normality N can make the same point against Normality Based Theory, as (9.1) does for Best World Theory:

(9.2) An object a is normal in γ, $a \in N(\gamma)$, only if for all generic implications δ we want to hold true for γ's, $\delta(a)$ holds true.

Yet, as normality N is something we first and most simply evaluate in our own real world, it can also be understood differently. We have consistently

---

[15] I omitted universal quantification and instantiation in the paraphrase, for the sake of brevity.
[16] It might not be an accident that Pelletier & Asher even use a similar formulation in their definition that was quoted here in (2.3). They do not seem to find this a problem.

presented normality as a kind of folk-statistical notion. This led to the more specific observations that

- Normal encounters of participants of different kinds need not be the same asencounters of normal participants of these kinds
- Normality involves a perspective
- Normality is not the same as the ideal
- Normality is restricted by axioms (N1)–(N3)

Therefore the first important building block of the theory, normality, can be taken to be an independent notion.

The second building block in the interpretation of (normal-)generic statements was the choice of the dispositional orbit *DO* of the world of evaluation. What is the status of this accessibility relation? In order to answer this question, it is helpful to remember why the dispositional orbit was introduced. The worlds in *DO* are needed to exemplify cases that *by accident* are not instantiated in the real world: the pope called 'Bartholomew', or the first letter from Antarctica. They should *not* provide counterexamples to generalizations that we actually want to become true (like the normal pope called 'Goofy').

On the one hand, such worlds can exemplify cases where a real accident seems to have played a role, and I assume freely at this place that we *have* a notion of 'accident' available. Yet it might well be that such cases are not of overwhelming importance because upon closer investigation it turns out that the notion of an 'accident' is itself based on knowledge of many similar cases which ended differently, and thus knowledge about the general limits to the possibilities available.

This leads us to a second, and perhaps more important, function of the *DO*. There are cases where the *DO* actually *does* reflect modal knowledge that lies at the basis of generic sentences. Consider one last time the example 'Rose handles mail from Antarctica'. This scenario certainly does not say anything about the normal case in a statistical sense, because there is no meaningful statistics over the empty set of events. In contrast, it is natural to assume that the sentence reports some kind of intentional (and thereby intensional) planning in Rose's office. Discussing various counter-factual scenarios of what might happen, the employees of that office *decide* that the counterfactual scenario of mail arriving from Antarctica should end so that *Rose* is the person to take care of it.[17] *DO* is shaped by our plans and intentions. This second function of *DO* is in concord with a distinction drawn by Carlson (1995: 233f.), the distinction between generic sentences

---

[17] Again, under generally unchanged circumstances, which probably include something like the expectation that not much mail from Antarctica will ever arrive. Otherwise the office might decide to employ a further secretary in charge of the fans at Antarctica.

which have an inductive flavour and generic sentences which seem to call for a 'rules and regulations' approach.[18] One function of the *DO* is to take care of the 'rules and regulations' aspect of genericity. Yet the *DO* is combined with the (more) inductive notion of normality, such that the overall theory allows to capture mixed cases, like the pope example: certainly, there are no explicit laws about names for popes, but there are some general expectations and guidelines about how to behave when you are to be God's representative on earth. These will exclude worlds in the *DO* where popes eagerly chose the name 'Goofy', while other papal names are allowed even without an explicit list of eligible names hidden in the safe of the Vatican.

In summary, the analysis of normal-generic sentences is built on the independent notions of normality and dispositional orbit. This is, however, only a partial answer to the question whether the overall theory is free from circular definitions, because I cannot, at present, offer a similar justification for ideal-generic sentences. My hope would be that, as sketched in section 7, at least part of our notion of an ideal is rooted in our knowledge about what usually happens when all disturbing factors can be excluded. It is likely, however, that for instance perceptual notions of ideal Gestalt involve cognitive processes that are beyond the reach of this paper.

Acknowledgements

---

[18] It would probably surprise Carlson to see that an account called 'Normality Based Theory' is one where this distinction can be discussed cleanly. He expresses the hope that an analysis for the 'rules-and-regulations' examples might lie more in the direction of proof theoretic accounts for default reasoning. Pelletier & Asher (1997), however, argue nicely that these accounts are not satisfying on semantic grounds. Another direction for analysing the 'rules-and-regulations' examples, which was pointed out by Carlson, is the Best World Theory proposed in Asher & Morreau (1995). For a more recent proposal, see Cohen (1999).

APPENDIX: RELATIONS TO SOME OTHER NORMALITY-BASED ACCOUNTS

In this appendix, I will prove the fact that (N1)–(N3) turn normality N into a global notion. Our treatment of normal-generic statements on the basis of N, restricted by these axioms, is equivalent to certain statistically committed default logics developed in AI.

I first give some definitions on which the following results are based. Note that, throughout this section, I adopt the convention that boldface variables abbreviate tuples of variables: x stands for $x_1, x_2, \ldots, x_n$. In the same way, boldface letters $m, n \ldots$ will stand for tuples of elements of the model domains under discussion.

Operator-based normality

Let L be a logical language with λ-abstraction for variables that range over individuals. We can thus, even without introducing an entire hierarchy of types, define what it means for a term to be of type $\langle e^n, t \rangle$. Moreover, we will assume that L contains a family of functor symbols $(N_i)_{i \in \omega}$ where for each term ϕ of type $\langle e^i, t \rangle$ the term $N_i(\phi)$ is also of type $\langle e^i, t \rangle$ The functor symbols will be interpreted by functions that map the power set over D_e^n into itself. They are to be thought of as functions that map each relation on to its normal subpart. The interpretation of N_i is restricted by axiom scheme (N1) that states that a normal P should always also be a P, and by (N2) which requires that we should always be able to find some normal P. These axioms correspond to (N 1) and (N 2) in the previous section. The restrictions will come in different shape, depending on whether the language L provides quantification over relations or not. The second-order analogues are listed in (N1^2) and (N2^2). (Second-order versions will generally be marked with a superscript 2 in the following.)

(N1) For all terms P of type $\langle e_n, t \rangle$,
$$\forall x (N_n(P)(x) \rightarrow P(x))$$

(N2) For all terms P of type $\langle e^n, t \rangle$
$$\exists x P(x) \rightarrow \exists x N_n(P)(x)$$

(N1^2) Where P is a variable ranging over properties of arity n, we have
$$\forall P \forall x (N_n(P)(x) \rightarrow P(x))$$

(N2^2) Where P is a variable ranging over properties of arity n, we have
$$\forall P (\exists x P(x) \rightarrow \exists x N_n(P)(x))$$

The difference between the first-order schemata and the second-order axioms is that the latter restrict the behaviour of N_i with respect to all arguments while the former only make claims about definable subsets.

Normality based on ranking

The logic languages that are at the basis of normality with ranking do not refer explicitly to any normal objects. Instead, they provide universal quantification that is restricted to the normal subpart to the respective domain of quantification, in a way much similar to the default implication used in Best World Theories. The following definitions are based on work by Weydert (1997) and Brafman (1996), which is rooted in research in nonmonotonic logic (see Kraus et al. 1990), and also on the work on ordinal conditional functions by Spohn (1988).

Definition: Let L be a first order logical language which is augmented by the following kind of formulae: for all formulae ϕ, ψ in L, and variables $x_1, \ldots, x_k$ the following is also a formula in L:

$$\phi \rightarrow_{x1, \ldots, xk} \psi$$

The language L is interpreted in structures (M, R) where M is an L-model in the usual sense and R is a family of ranking functions of the following shape:

(i) $R = (R_n)_{n \in \omega}$
(ii) For each $n \in \omega$, R_n is a function of D_e^n into an ordered set Ω.
(iii) For all formulae ϕ and ψ

$(M,R) \models \phi \rightarrow_{x_1, \ldots, x_i} \psi$ iff

there is an $m \in \{k \mid (M, R) \models \phi(k)\}$ such that for all

$n \in \{k \mid (M, R) \models \phi(k)\}$ with $R_i(n) \leq R_i(m)$ we find that $(M, R) \models \psi(n)$.

Further possible restrictions on R are discussed both in Brafman and Weydert. While Weydert attempts to fix the ranking so as to come close to a probability measure on the domain D, Brafman keeps his models more flexible. Both authors give a sound and complete axiomatization for their respective versions of L. The following axiom (W1) was suggested by Weydert (p.c.) in order to ensure that each L-definable subset $A \subseteq D_e^n$ contains one or more elements of minimal rank. These models are called 'smooth models' in the literature.

(W1) $\forall y(\phi(x) \rightarrow_x x \neq y) \rightarrow \forall x(\neg \phi(x))$

where y does not share any variable with x.

Intuitively, (W1) asserts that for all sets A, there are $a \in A$ such that for all $c \in A$, $R(c) \geq R(a)$. An analogous effect would be achieved by requiring that the ordered set Ω in fact has to be an ordinal (this approach is explored in Spohn 1988). This even ensures that *each* set contains elements of minimal rank, not only definable subsets. However, the resulting models can no longer be characterized by a recursive set of axioms, which is why I refrain from this move.

Ranked models can be mimicked with normality operators

Let (M, R) be an L-model with ranking. Let us further assume that (M, R) validates axiom (W1), such that all L-definable subsets in D_e^n have minimal elements. We can now construct an L_{norm}-model M' which simulates (M, R):

Let L_{norm} be the language with normality functors which equals L in its standard part. We will augment M to yield an L_{norm}-model M' by the following definition:

$$N_n(A) := \{m \in A \mid \forall n \in A(R_n(m) \leq R_n(n))\}$$

for all definable subsets A of D_e^n and

$$N_n(A) := A \text{ otherwise.}$$

Now define a mapping * of L_{rank}-formulae into L_{norm}-formulae in the obvious way:

For all atomic formulae ϕ, let $\phi^* := \phi$.

For $\phi = \neg \psi, \psi \wedge \theta, \psi \vee \theta$ let $\phi^* = \neg \psi^*, \psi^* \wedge \theta^*, \psi^* \vee \theta^*$.

For $\phi = \exists x \psi$, let $\phi^* := \exists x \psi^*$

For $\phi = \psi \rightarrow_x \theta$, let $\phi^* := \forall x (N(\lambda x.\psi^*)(x) \rightarrow \theta^*)$

Using this definition, we can prove the following equivalence to hold true:

$$(M, R) \models \phi \Leftrightarrow M' \models \phi^*$$

Proof: By induction on the complexity of formulae. The crucial step is to show that:

$$(M, R) \models \psi \rightarrow_x \theta \Leftrightarrow M' \models \forall x (N(\lambda x.\psi^*)(x) \rightarrow \theta^*)$$

This, however, is ensured by the definition of N_n and the obvious observation that all sets involved are definable. □

The observation that ranking should give rise to a notion of normality is not surprising. We can, however, prove the stronger proposition that (N 1)–(N 3), or their corresponding axiom schemes, are sufficient to restrict L_{norm}-models to those which have an equivalent L_{rank}-counterpart. It is this stronger proposition that we used in section 8 when discussing WEAK CONTRAPOSITION.

Certain normality models can be mimicked by ranked models

The central property of ranked models is that intersections of sets inherit their normal parts from the respective supersets if the intersections are nonempty. While we get a strong equivalence between ranking and normality functors if we allow ourselves second-order quantification, we can at least translate the default implication part of the L_{norm}-model with ranking, without second-order quantification. The latter is expressed more precisely in Theorem 1^1, while the former is formulated in Theorem 1^2.

Theorem 1^1: Let L_{norm} be a language with normality operators and L_{rank} be the corresponding language with default implication, as defined above. (That is, both languages share the same classical core: the same constant symbols, relation symbols, functions.) Let M be an L_{norm}-model in which satisfies (N1) and (N2) above, and where moreover (N3) holds true, for all terms A, B of type $\langle e^n, t \rangle$ and arities n.

(N3) $M \models \exists x (N_n(A)(x) \land B(x)) \rightarrow \forall x (N_n(\lambda y.A(y) \land B(y))(x) \leftrightarrow N_n(A)(x) \land B(x))$

Then we can define a mapping from a subset of L_{norm} on to L_{rank} such that the following holds true:

$$M \models \phi \text{ iff } (M, R) \models \phi^{\#}$$

The proof is given below.

Theorem 1^2: Let L_{norm} be a language with normality operators that moreover provides quantification over relations, and L_{rank} again be the corresponding language with default implication. Let M be an L_{norm}-model which satisfies $(N1^2)$ and $(N2^2)$, and in which for variables A, B of type $\langle e^n, t \rangle$ the following $(N3^2)$ hold true (for all arities n):

$(N3^2)$ $M \models \forall A \forall B \exists x (N_n(A)(x) \land (B(x))$
$\rightarrow \forall x (N_n(\lambda y.A(y) \land B(y))(x) \leftrightarrow N_n(A)(x) \land B(x))$

Then M can be turned into a model M' for L_{rank} which remains identical with respect to the classical part of L and where, for each definable set A

$$N(A) = \{m \mid \forall k \in A (R_n(m) \leq R_n(k))\}.$$

We will start with the proof of the weaker Theorem 1^2. The argument corresponds closely to a similar construction in (Spohn 1988). Having accomplished this, we will proceed to prove the more intricate Theorem 1^1.

Proof of 1^2: We define a partition on the Cartesian products D_ϵ^n of the domain D_ϵ of individuals:

$$m_o := N_n (D_\epsilon^n) \text{ and } m_o^* := m_o$$

$$m_1 := N_n (D_\epsilon^n \backslash m_o^*) \text{ and } m_1^* := m_o^* \cup m_1$$

$$\dots$$

$$m_{k+1} := N_n(D_\epsilon^n \backslash m_k^*) \text{ and } m_{k+1}^* := m_k^* \cup m_{k+1} \text{ for successor ordinals}$$

$$m_\lambda := N_n (D_\epsilon^n \backslash \bigcup_{i<\lambda} m_i^* \text{ and } m_\lambda^* := \bigcup_{i<\lambda} m_i^* \cup m_\lambda \text{ for limit ordinals.}$$

Let κ be the smallest ordinal number such that $m_\kappa = \emptyset$. We can now define a mapping R_n from D_ϵ^n into κ, with the intention that $(R_n)_{n \in \omega}$ will be the ranking for L_{rank}. For each m in D_ϵ^n let $(R_n)(m) := \mu$ if and only if $m \in m_\mu$. R_n is thus a mapping of D_ϵ^n into an ordered set, that is, an appropriate ranking function. We interpret the language L_{rank} on the basis of this ranking function. It remains to be shown that for each set $A \subseteq D_\epsilon^n$, $N_n(A) = \{m \mid \forall k \in A(R_n(m) \leq R_n(k))\}$. Take an arbitrary subset A of D_ϵ^n. We can now reason as in step (i) to (iv):

(i) There is a smallest ordinal number μ such that $A \cap m_\mu \neq \emptyset$. According to construction, $\{m \mid \forall k \in A(R_n(m) \leq R_n(k))\} = A \cap m_\mu$.

(ii) According to construction, there is a set $P \subseteq D_\epsilon^n$ such that $m_\mu = N_n(P)$. We can even give P more precisely: $P = D_\epsilon^n \backslash (\bigcup_{i < \mu} m_i)$.

(iii) As m_μ is the first set in the partition to have nonempty intersection with A, we know that $A \subseteq P$, that is, $A \cap P = A$.

(iv) As $A \cap N_n(P) \neq \emptyset$, it follows that

$$M \models \forall x (N_n(\lambda y.P(y) \wedge A(y))(x) \leftrightarrow N_n(P)(x) \wedge A(x))$$

Therefore $N_n(A) =_{(iii)} N_n(A \cap P) =_{(iv)} N_n(P) \cap A =_{(ii)} m_\mu \cap A$

$$=_{(i)} \{m \mid \forall k \in A(R_n(m) \leq R_n(k))\} \text{ as desired. } \square$$

Note that the proof of 1^2 even validates the stronger claim that each model with functor-based normality will be equivalent to a model ranked on the basis of an *ordinal* Ω. This class of ranked models cannot be characterised with a recursive set of first order axioms. We therefore know that the second order quantification which we allowed ourselves in the axiom schemes above were not only a convenient shortcut, but restricted the model class in a nontrivial way. It is mainly for this reason that the more general statement in (1) becomes interesting: We can show that not only a very limited class of operator-based models for normality are equivalent to (a subclass of) ranked models but that the equivalence holds in general. Clearly, we will have to revise the original proof, because the construction made substantial use of the fact that the crucial property (N 3) held for all subsets of D_ϵ^n.

In the following lemmas, we are always using models M of a language L_{norm} without second order quantification. I will use the abbreviation $A \cap B$ for $\lambda x.A(x) \wedge B(x)$. I will also

occasionally omit the arity index n in the function N_n. All models are assumed to validate axiom schemes $(N1)'–(N3)'$.

The overall strategy of the proof is this: we show that N induces a linear ordering on the definable power set over D_e^n. This is done in Lemma 1—9; the linear ordering is given in Definition 2. Using this ordering, we then construct an L_{rank}-model (M, R).

Lemma 1: *Let M be as described. It follows that for all $A, B \subseteq D_e^n$ we have that*

(i) $N(A \cup B) \cap B = \varnothing$ *and* $N(A \cup B) = N(A)$ *or*
(ii) $N(A \cup B) \cap A = \varnothing$ *and* $N(A \cup B) = N(B)$ *or*
(iii) *both intersections are nonempty and* $N(A \ll B) = N(A) \cup N(B)$.

Proof: If $N(A \cup B) \cap B \neq \varnothing$, then $(N3)$ will yield $N((A \cup B) \cap B) = N(B) = N(A \cup B) \cap B$. In the same way, if $N(A \cup B) \cap A \neq \varnothing$, it follows from $(N3)$ that $N((A \cup B) \cap A) = N(A) = N(A \cup B) \cap A$. Finally, we know that $N(A \cup B) \subseteq A \cup B$ and therefore $N(A \cup B) = [N(A \cup B) \cap A] \cup [N(A \cup B) \cap B]$. Thus, the three constellations listed above are the only possible ones. $\square$

Lemma 2: *If $N(B) \subseteq A$ and $N(A \cup B) = N(A) \cup N(B)$ then $N(B) \subseteq N(A)$.*

Proof: We assumed that $N(A \cup B) = N(A) \cup N(B)$ (*). Therefore, $N(A \cup B) \cap A \neq \varnothing$, due to Lemma 1. Thus, $N(A \cup B) \cap A = N(A)$ (again from Lemma 1). Yet, as $N(B) \subseteq A$, we know moreover, due to (*) and $N(B) \subseteq A$, that $N(A \cup B) \cap A = N(A) \cup N(B)$. Thus, $N(A) = N(A) \cup N(B)$ and therefore $N(B) \subseteq N(A)$. $\square$

Lemma 3: *If $N(A \cup B) \subseteq A$, then $N(A \cup B) = N(A)$.*

Proof: In case that $N(A \cup B) \cap B = \varnothing$, the claim follows from Lemma 1. Else $N(A \cup B) \cap B \neq \varnothing$, and therefore $N(B) \subseteq N(A \cup B) \subseteq A$. The claim then follows from Lemma 2. $\square$

Definition 1: *We define $N(A) < N(B)$ to be an abbreviation for*

$$N(A \cup B) \subseteq A \text{ and } N(A \cup B) \cap N(B) = \varnothing$$

Remark: It follows easily that $N(A) < N(B)$ if and only if $N(A \cup B) \cap B = \varnothing$.

Lemma 4: *The relation $<$ is transitive.*

Proof: Assume that for nonempty sets A, B, C,

(1) $N(A \cup B) \subseteq A$ and $N(A \cup B) \cap N(B) = \varnothing$ and
(2) $N(B \cup C) \subseteq B$ and $N(B \cup C) \cap N(C) = \varnothing$.

We have to show that $N(A \cup C) \subseteq A$ and $N(A \cup C) \cap N(C) = \varnothing$. We do this by contraposition: assume that $N(A \cup C) \cap C \neq \varnothing$. We will now compute $N(A \cup B \cup C)$. According to Lemma 1, we know that $N(A \cup B \cup C) = N(A \cup B)$ or $N(A \cup B \cup C) = N(C)$ or $N(A \cup B \cup C) = N(A \cup B) \cup N(C)$. Due to our assumption, we can conclude that in any case

(*) $N(B) \not\subseteq N(A \cup B \cup C)$

We will now distinguish two cases: $N(A) \cap C \neq \varnothing$ or $N(A) \cap C = \varnothing$.

In the case that $N(A) \cap C \neq \emptyset$, we can immediately infer that $N(A \cup B) \cap (B \cup C) \neq \emptyset$, as well. According to Lemma 1, we know for arbitrary definable sets X, Y that $N(X) \cap Y \neq \emptyset$ implies $N(Y) \subseteq N(X \cup Y)$. Take now X to be $A \cup B$ and Y to be $B \cup C$. We thus get that $N(B \cup C) \subseteq N((A \cup B) \cup (B \cup C))$. Due to assumption (2), we can conclude that $N(B) \subseteq N(A \cup B \cup C)$. This contradicts (*).

If it was such that $N(A) \cap C = \emptyset$, we would get that $N(A \cup C) = N(C)$ (by Lemma 1). In that case, however, we can show that neither $N(A)$ nor $N(B)$ nor $N(C)$ are in $N(A \cup B \cup C)$, thus coming to a contradiction: we know that $N(A \cup (B \cup C)) = N(A)$ or $N(A \cup (B \cup C)) = N(B \cup C)$ or $N(A \cup (B \cup C)) = N(A) \cup N(B \cup C)$. Because $N(B \cup C) \cap N(C) = \emptyset$, we know that $N(C) \not\subseteq N(A \cup (B \cup C))$. Analogously, we can argue that $N(B) \not\subseteq N((A \cup B) \cup C)$ and that $N(C) \not\subseteq N((A \cup C) \cup B)$. This means that $N(A \cup B \cup C) = \emptyset$, which is only possible if A, B and C are $= \emptyset$.

We have shown that $N(A \cup C) \cap C = \emptyset$. Thus, $N(A) < N(C)$ which completes the proof. □

Lemma 5: *If $N(A) < N(B)$ and $N(C) \subseteq N(B)$, we infer that $N(A) < N(C)$.*

Lemma 6: If $N(B) \subseteq N(A)$ and $N(B) < N(C)$, we can infer that $N(A) < N(C)$.

In order to prove Lemma 5 and 6, we first have to cover a number of intermediate observations.

Lemma 7: *If $N(B) \subseteq N(A \cup B \cup C)$ and $N(B) \subseteq N(A)$,*

then $N(A) \subseteq N(A \cup B \cup C)$.

Proof: $N(B) \subseteq N(A \cup B \cup C)$ implies that $N(A) \cap N(A \cup B \cup C) \neq \emptyset$. Hence, $A \cap N(A \cup B \cup C) \neq \emptyset$. We infer that $N(A \cup B \cup C) \cap A = N(A)$ and therefore (by Lemma 1) $N(A) \subseteq N(A \cup B \cup C)$.

Lemma 8: *If $N(B) < N(C)$ then $N(C) \cap N(A \cup B \cup C) = \emptyset$.*

Proof: We know that $N(B \cup C) = N(B)$ and $N(B \cup C) \cap N(C) = \emptyset$. Assume that $N(C) \cap N(A \cup B \cup C) \neq \emptyset$. Because we generally know that $N(C) \subseteq C \subseteq (B \cup C)$, we can conclude that $(B \cup C) \cap N(A \cup B \cup C) \neq \emptyset$. If this is the case, we know by (N3) that we can conclude that $N(B \cup C) = N((A \cup B \cup C) \cap (B \cup C)) = N(A \cup B \cup C) \cap (B \cup C)$. We assumed that $N(C) \cap N(A \cup B \cup C) \neq \emptyset$. We therefore know that $N(C) \cap N(A \cup B \cup C) \subseteq (B \cup C) \cap N(A \cup B \cup C)$ where $N(B \cup C) = (B \cup C) \cap N(A \cup B \cup C)$ (because $N(C) \subseteq (B \cup C)$). Therefore $N(C) \cap N(B \cup C) \neq \emptyset$ which contradicts our assumptions. Therefore $N(C) \cap N(A \cup B \cup C) = \emptyset$.

Proof of Lemma 5:

(i) $N(A \cup B \cup C) = N(A \cup B)$ or $N(C)$ or $N(A \cup B) \cup N(C)$. (Lemma 1)
(ii) Due to Lemma 8, we know that $N(B) \cap N(A \cup B \cup C) = \emptyset$, because

      $N(A) < N(B)$.

(iii) Therefore, $N(C) \cap N(A \cup B \cup C) = \emptyset$, as $N(C) \subseteq N(B)$ (due to assumption).
(iv) However, $= N((A \cup C) \cup B) = N(A \cup C)$ or $N(B)$ or $N(A \cup C) \cup N(B)$.
(v)   From (ii), we can infer that $N((A \cup C) \cup B) = N(A \cup C)$.

(vi) Assume now that $N(A \cup C) \cap C \neq \emptyset$. It follows that

$N((A \cup B) \cup C) \cap C \neq \emptyset$ and thus $N((A \cup B) \cup C) \cap C = N(C)$. This means that $N(A \cup B \cup C) \cap N(C) \neq \emptyset$, in contradiction with (iii). Thus, $N(A \cup C) \cap C = \emptyset$, and thus $N(A) < N(C)$, as required.

Proof of Lemma 6:

(i) $N(A \cup (B \cup C)) = N(A)$ or $= N(B \cup C)$ or $= N(A) \cup N(B \cup C)$.

(ii) $N(B) = N(B \cup C)$ and $N(B) \subseteq N(A)$. Therefore, $N(B) \subseteq N(A \cup B \cup C)$.

(iii) Due to assumption and Lemma 7, $N(A) \subseteq N(A \cup B \cup C)$.

(iv) As $N(B) = N(B \cup C)$ and $N(B) \subseteq N(A)$, we know that $N(A \cup B \cup C) = N(A)$.

(v) $N((A \cup C) \cup B) = N(A \cup C)$ or $= N(B)$ or $= N(A \cup C) \cup N(B)$.

(vi) $N(A) \subseteq A \subseteq A \cup C$ and $N(A) \subseteq N(A \cup B \cup C)$, thus, $N(A \cup B \cup C) \cap (A \cup C) \neq \emptyset$, and therefore $N(A \cup B \cup C) \cap (A \cup C) = N(A \cup C)$. With Lemma 1, we know that $N(A \cup C) \subseteq N(A \cup B \cup C)$. Thus, $N(A) = N(A \cup C) = N(A \cup B \cup C)$ (using (iv)).

(vii) If $N(A \cup C) \cap C$ was $\neq \emptyset$; we'd get that $N(A \cup B \cup C) \cap C \neq \emptyset$, that is $C \cap N((A \cup B) \cup C) \neq \emptyset$ and thus $N(C) \subseteq N(A \cup B \cup C)$.

(viii) However, with Lemma 8: $N(C) \cap N(A \cup B \cup C) = \emptyset$, in contradiction to (vii). Thus, $N(A \cup C) \cap C = \emptyset$, that is $N(A) < N(C)$, as required.

Definition 2: *For all* $A,B \subseteq M$ *let* $A \leq B$ *iff* $N(A \cup B) \subseteq A$.

We say that $A \sim B$ *iff* $A \leq B$ *and* $B \leq A$.

Lemma 9: *The relation* $\leq$ *is transitive and reflexive.*

Proof: We assumed that $N(A) \subseteq A$ for all A. This shows reflexivity.

Assume that $A \leq B$ and $B \leq C$. We have to show that $A \leq C$. Let us spell out what this means. If $A \leq B$, this means that $N(A \cup B) \subseteq A$ and either $N(B) \cap N(A \cup B) = \emptyset$ (that is, $N(A) < N(B)$) or else $N(B) \cap N(A \cup B) \neq \emptyset$, which means that $N(B) \subseteq N(A)$. The same holds true for $B \leq C$. Thus, we have four subcases to consider:

Case 1: $N(A) < N(B)$ and $N(B) < N(C)$. This implies $N(A) < N(C)$ which means that $N(A \cup B) = N(A)$ and $C \cap N(A \cup C) = \emptyset$. Especially, $N(A \cup C) \subseteq A$ and therefore $A \leq C$.

Case 2: $N(A) < N(B)$ and $N(C) \subseteq N(B)$. We conclude that $N(A) < N(C)$ by Lemma 5. As in case 1, we infer from $N(A) < N(C)$ that $A \leq C$.

Case 3: $N(B) \subseteq N(A)$ and $N(B) < N(C)$. We conclude with Lemma 6 that $N(A) < N(C)$ and thus $A \leq C$.

Case 4: $N(B) \subseteq N(A)$ and $N(C) \subseteq N(B)$. This implies $N(C) \subseteq N(A)$ and so $A \leq C$.

Lemma 9 shows that even in the weaker first-order variant, axiom (N3) is strong enough to considerably limit the shape of the normality function. It ensures that the normality function implicitly introduces an ordering at least on all L-definable sets in the model. This suffices to show that the normality function corresponds to a suitable ranking on the definable subsets of the model—which are all sets we can talk about, anyway. We can now proceed to prove Theorem 1', which I repeat here for the sake of convenience:

Theorem 11: *Let* L *be a logic language and* L_{norm} *its extension to a language with normality functors. Let* L_{rank} *its extension to a language with default implication based on ranking. We will use*

ϕ^* to denote the inverse mapping to $*$ which we defined above. Evidently, # does not map the entire of L_{norm} into L_{rank} but only the range of $*$. This shows that L_{norm} is richer than ranked languages because explicit reference to normality can be made. We can now prove the following: for each L_{norm}—model M which satisfies $N(1)$ to $N(3)$, we can define an L_{rank}-model (M, R) such that

$$M \models \phi \text{ iff } (M, R) \models \phi^{\#}$$

Proof: Let M be an L_{norm}—model which satisfies (N1) to (N3). Let me use $P_d(D_e^n)$ to denote the definable part of the power set over D_e^n. We adopt definition 2 to get a preorder on $P_d(D_e^n)$: For all A, B $P_d(D_e^n)$ let

$$A \leq B \text{ iff } N(A \cup B) \subseteq A$$
$$A \sim B \text{ iff } A \leq B \text{ and } B \leq A$$

Now we take the set of equivalence classes of sets: For all $A \in P_d(D_e^n)$ define

$$[A] := \{X \mid X \in P_d(D_e^n) \ \& \ X \sim A\}$$
$$\Theta := \langle \{[A] \mid A \in P_d(D_e^n)\}, \leq \rangle$$

where the relation $\leq$ is lifted to the set of equivalence classes in the evident way. The resulting relation is a linear ordering relation (reflexive, transitive, antisymmetric, and any two elements are mutually comparable) on Θ. We are now in the position to be able to map all definable sets into the ordered set Θ in a way which respects the normality relations: If A is a definable subset of D_e^n, let $R_n(A) := [A]$. However, this is not enough: n order to come to a ranking, we have to map all *elements* in D_e^n into an ordered set. As the singleton sets in D_e^n need not be definable, this mapping is not yet immediate. Let $\langle \Omega; \leq \rangle$ denote the Dedekind completion of the linear ordering $\Theta = \langle \{[A] \mid A \in P_d(D_e^n)\}; \leq \rangle$. Thus for each $m \in D_e^n$, the supremum $\sup\{[A] \mid A \in P_d(D_e^n) m \in A\}$ exists in Θ. Define $R_n(m) := \sup\{[A] \mid A \in P_d(D_e^n) \wedge m \in A\}$. The rationale behind this defniition is quite simple: if m is in A, it can't have lower rank than A as a whole, because the normality of A is measured by the normality of its most normal elements.

Clearly, $R = (R_n)_{n \in \omega}$ is a ranking function. Thus, (M, R) is a model of L_{rank}. We will show the equivalence of the two models by induction on the complexity of formulae. As the models are equal in their atomic part, the interesting step is the inductive step from formulae ϕ^*, ϕ and ψ^*, ψ to formulae of the shape $\forall x (N(\phi)(x) \rightarrow \psi(x))$ and $\phi^{\#} \rightarrow_x \psi^{\#}$ respectively.

Let A be a definable subset of D_e^n. We know that for any B definable in (D_e^n) the following holds true: If $[A] < [B]$, then $N(A) \subseteq A \cap \neg B$. Therefore

$$N(A) \subseteq A \cap \bigcap_{B \text{ definable, } [A] < [B]} \neg B$$

Assume that $a \in A$ and $a \in N(A)$. It follows that a is in no set B such that $[A] < [B]$.

Thus, $R_n(a) = [A]$, according to definition. We therefore know that

$$N(A) \subseteq \{ a \mid \forall b \in A : R(b) \geq R(a)\}$$

For this reason

$$(M, R) \models \phi^{\#} \rightarrow_x \psi^{\#} \text{ implies } M \models \forall x (N(\phi)(x) \rightarrow \psi(x))$$

Generally, the set $A \cap \bigcap_{B \text{ definable, } [A] < [B]} \neg B$ might contain more elements than $N(A)$.

However, we can show that this does not make any difference with respect to what can be expressed in L_{rank}. We will show this by contraposition. Assume that (**) held true:

(**) $M \models \forall x \, (N(\phi)(x) \rightarrow \psi)$, but $(M, R) \models \neg (\phi^{\#} \rightarrow_x \psi^{\#})$

This would mean that there is an a in ϕ such that $a \notin N(\phi)$, $a \notin \psi$, and $R(a) = [\phi]$, which means that $a \in \phi \cap \bigcap_{B \, definable, \, [\phi] < [B]} \neg B$. But $N(\phi) \subseteq \psi$ and therefore $N(\phi) = N((\phi \cap \neg \psi) \cup (\phi \cap \psi)) = N(\phi \cap \psi)$. That is, $N(\phi \cap \psi) < N(\phi \cap \neg \psi)$. As $a \in (\phi \cap \neg \psi)$ and $R(a) = sup\{[X] \, | \, X \, definable, \, a \in X\}$, we can conclude that $R(a)$ is not minimal in ϕ, in contradiction to the assumption that $R(a) = [\phi]$. Therefore, there can not be an element a of minimal rank in ϕ which falsifies $\phi^{\#} \rightarrow_x \psi^{\#}$ in (M, R) while $\forall x (N(\phi)(x) \rightarrow \psi(x))$ holds true in M. This shows the converse direction of the main claim:

$$M \models \forall x(N(\phi)(x) \rightarrow \psi(x)) \text{ implies that } (M, R) \models \phi^{\#} \rightarrow_x \psi^{\#}$$

Which finishes the proof.

REFERENCES

Asher, N. & Morreau, M. (1995), 'What some generic sentences mean', in G. Carlson & J. Pelletier (1995), 300—38.

Brafman, R. I. (1996), '"Statistical" first order conditionals', in *Proceedings of the Fifth Conference on the Principles of Knowledge Representation and Reasoning KR 96*, Morgan Kaufmann, Los Altos, CA.

Brehm, A. E. (1876), *Brehms Thierleben*, Großausgabe, 2nd edn., Vol. 1, *Säugethiere*, Verlag des Bibliographischen Instituts, Leipzig.

Carlson, G. (1995), 'Truth conditions of generic sentences: two contrasting views', in G. Carlson & J. Pelletier (1995), 224—38.

Carlson, G. & Pelletier, J. (1995), *The Generic Book*, Chicago University Press, Chicago.

Chellas, B. (1974), 'Basic conditional logic', *Journal of Philosophical Logic*, 4, 133—53.

Chierchia, G.(1995), *Dynamics of Meaning*, Chicago University Press, Chicago.

Cohen, A. (1999), 'Generics, frequency adverbs, and probability', *Linguistics and Philosophy*, 22, 221—54.

Delgrande, J. (1987), 'A semantics for defaults using conditional logic', *Artificial Intelligence*, 33, 105—30.

Delgrande, J. (1988), 'An approach to default reasoning based on a first-order conditional logic: a revised report', *Artificial Intelligence*, 36, 63—90.

Diesing, M. (1992), *Indefinites*, MIT Press, Cambridge, MA.

von Fintel, K. (1998), 'Bare plurals, bare conditionals, and only', *Journal of Semantics*, 14, 1—56.

Gabbay, D., Hogger, C. & Robinson, J. (1994), *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. III*, Clarendon Press, Oxford.

Gallin, D. (1975), *Intensional and Higher-order Logic*, North-Holland, Amsterdam.

Ginsberg, M. (1994), 'AI and nonmonotonic reasoning', in Gabbay et al. (1994), Ch. 1.

Grizmek, B. (1987): *Grizmeks Enzyklopädie der Säugetiere, Vol. 4*, Kindler Verlag, Munich.

Heim, I. (1982), 'The semantics of definite and indefinite noun phrases', Ph.D. dissertation, UMass, Amherst, also SFB-Papier 73, University of Konstanz.

Kratzer, A. (1978), *Semantik der Rede, Kontexttheorie, Modalwörter, Konditionalsätze*, Scriptor, Königstein.

Kraus, S., Lehmann, D. & Magidor, M. (1990), 'Nonmonotonic reasoning, preferential models and cumulative Logics', *Artificial Intelligence*, 44, 167–207.

Krifka, M., Pelletier, F. J., Carlson, G. N. A., ter Meulen, Chierchia, G. & Link, G. (1995), 'Genericity: an introduction', in G. Carlson & J. Pelletier (1995), 1–124.

Lehmann, D. (1989), 'What does a conditional knowledge base entail?', in *Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference*, R. Brachmann & H. Levesque (eds.), Morgan Kaufmann, Los Altos, CA, 212–22.

Lewis, D. (1973a), *Counterfactuals*, Blackwell, Oxford.

Lewis, D. (1973b), 'Counterfactuals and comparative possibility', *Journal of Philosophical Logic*, 2, 418–46.

Morreau, M. (1996), 'Allowed arguments', in *Proceedings of IJCAI '96*, Morgan Kaufmann, Los Altos, CA.

Pelletier, F. J. & Asher, N. (1997), 'Generics and defaults', in J. van Benthem & A. ter Meulen (eds.), *Handbook of Logic and Language*, North Holland, Amsterdam, and MIT Press, Cambridge, MA.

Rooth, M. (1995), 'Indefinites, adverbs of quantification, and focus semantics', in G. Carlson & J. Pelletier (1995).

Rott, H. (1996), 'Making up one's mind', Habilitationsschrift, University of Konstanz (revised version to appear as *Change, Choice and Inference*, Oxford University Press).

Spohn, W. (1988), 'Ordinal conditional functions: a dynamic theory of epistemic states', in W. Harper & B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics, Vol II*, Kluwer, Dordrecht.

Stalnaker, R. (1984), *Inquiry*, MIT Press, Cambridge, MA.

Weydert, E. (1997), 'Rational default quantifier logic', extended abstract, MPI for Computer Science, Saarbrücken.

Zimmermann, E. (1989), 'Intensional logic and two-sorted type theory', *Journal of Symbolic Logic*, 54.