Commonsense Entailment: A Modal Theory of Nonmonotonic Reasoning

Nicholas Asher
Center for Cognitive Science, GRG 220
University of Texas at Austin
Austin TX 78712, USA
asher@sygmund.cgs.utexas.edu

Michael Morreau
IMS, University of Stuttgart
Keplerstrasse 17
7000 Stuttgart 1, Deutschland
mimo@adler.philosophie.uni-stungart.de

Abstract

In this paper, we construct a truth conditional semantics for generic sentences, which treats arbitrarily deep nestings of generic sentences. The resulting notion of logical entailment captures intuitively valid argument forms involving generics. A dynamic semantics is built on top of the truth conditional one, and the resulting inference notion captures nonmonotonic argument patterns familiar from the artificial intelligence literature by exploiting constraints on modal frames alone, without the use of ordering principles on rules, abnormality or "relevance" predicates.

1. Introduction

Potatoes contain vitamin C, *amino acid, protein and thiamin* expresses a true generalization about potatoes. *John smokes a cigar after dinner,* understood in its generic sense as expressing a regularity in John's behaviour after dinner, can be true, and it can be false. This realist conviction inspires the theory of generic propositions which is the subject of this paper. The difficulty with generic propositions is that their truth is loosly but clearly connected with particular facts. For instance, potatoes contain vitamin C, even though large numbers of them are boiled for so long that it is lost Potatoes would contain vitamin C even if *all* of them were to be boiled for so long that it is lost Nevertheless, the generic fact that potatoes contain vitamin C furnishes all the evidence we need to be justified in concluding that this potato, in the abscence of any other conflicting information, contains vitamin C. The curious relation between generic and particular facts has for a decade or more frustrated efforts in artificial intelligence, linguistics, and philosophy to provide generic sentences with a rigorous semantics. We offer such a semantics here for one kind of genericity.

Researchers in AI have produced many theories of non-monotonic reasoning that be seen also as attempting to give a semantics for genericity. As motivation for our *theory,* we argue for three desiderata of a semantics for genericity and defaults which other theories of genericity do not simultaneously satisfy. A first requirement is that any theory of genericity should explain the ways in which we reason with generic sentences. Logical entailment is one form of reasoning. We think that the truth conditions of generic sentences can be captured by using ordinary quantification and a non-monotonic conditional operator; so in what follows we shall write O*'s normally* ψ as $\forall x(\varphi > \psi)$.[1] There are a few forms involving generic sentences that seem clearly to be cases of logical entailment ($\vdash$). One is WEAKENING OF THE CONSEQUENT. Suppose ζ is a logical consequence of ψ; then:

$$\forall x(\varphi > \psi) \vdash \forall x(\varphi > \zeta).$$

Among the intuitively valid generic sentences, those which are entailed by everything, we count *Lions are lions,* and the nested generic sentence *People who don't like to eat out don't like to eat out* We take this last sentence to be nested because it says that people possessing a characteristic property - namely the property of typically not liking to eat out- typically have this property.

While the logic of generic sentences seems to support few valid argument forms, it does seem to support many "reasonable inference patterns." Among the things not entailed by the generic statement that potatoes contain vitamin C is the particular conclusion that *this* potato contains vitamin C. Nevertheless, the generic fact makes it somehow reasonable to expect this potato to contain vitamin C, without at the same time making it reasonable to expect any number of other things which are not entailed, like say that the moon is made of green cheese. Researchers in the field of nonmonotonic reasoning have discovered a wealth of such patterns of invalid but reasonable generic inference, of which some examples are given below. We will symbolize reasonable inference by $\mid\!\sim$.

The main patterns come in three distinguishable groups. In this paper we will present in some detail an interpretation of generic sentences in which the following generally acknowledge patterns of nonmonotonic reasoning hold.

DEFEASIBLE MODUS PONENS $\forall x(\varphi > \psi), \varphi(\delta) \mid\!\sim \psi(\delta)$,
but not $\forall x(\varphi > \psi), \varphi(\delta), \neg \psi(\delta) \mid\!\sim \psi(\delta)$

NIXON DIAMOND not{ $\forall x(\varphi > \psi), \forall x(\zeta > \neg\psi), \varphi(d), \zeta(d) \mid\!\sim$ $\psi(d)$ (or $\neg \psi(d)$)}

POINTWISE DEFEASIBLE TRANSITIVITY $\forall x(\varphi > \psi), \forall x(\psi > \zeta)$, $\varphi(\delta) \mid\!\sim \zeta(\delta)$
but not { $\forall x(\varphi > \psi), \forall x(\psi > \zeta), \forall x(\varphi > \neg\zeta), \varphi(\delta) \mid\!\sim$ $\varphi(\delta)$}

[1] Assuming a generic quantifier makes more linguistic sense than what we have done above; however, such a quantifier is definable in terms of V and >. given our semantics.

POINTWISE DEFEASIBLE STRENGTHENING OF THE ANTECEDENT: $\forall x(\varphi > \psi), \varphi(\delta) \& \zeta(\delta) \mid\sim \psi(\delta)$, but not $\{\forall x(\varphi > \psi), \varphi(\delta) \& \zeta(\delta), \forall x((\varphi \& \zeta) > \neg\psi) \mid\sim \psi(\delta)\}$.

PENGUIN PRINCIPLE: $\forall x(\varphi > \psi), \forall x(\psi > \zeta), \forall x(\varphi > \neg\zeta), \varphi(\delta) \mid\sim \neg\psi(\delta)$.

EMBEDDED DEFEASIBLE MODUS PONENS: $\forall x(\zeta > \forall y(\varphi > \psi)), \gamma(\delta), \forall u(\gamma > \exists y\varphi) \mid\sim \exists y\psi(\delta)$.

In all of these forms of reasoning, we observe that $\mid\sim$ is very sensitive to "relevant premises" that may affect conclusions that may be reasonably drawn from a subset of those premises. Capturing such inferences has been a major challenge for AI. But further these forms of reasoning may interact, but they should not interact in "bizarre" ways. Here is one bizarre interaction that we want to avoid. For while we have

$\forall x(\varphi > \psi) \models \forall x(\varphi > (\psi \vee \zeta))$

and we have defeasible modus ponens, we do not get the undesirable inference,

DEFEASIBLE IRRELEVANCE: $\forall x(\varphi > \psi), \varphi(\delta), \neg\psi(\delta) \mid\sim \zeta(\delta)$

for then we would defeasibly conclude that from *birds fly* and *Tweety is a bird that does not fly*, that Tweety is a locomotive!

A second class of inferences that we believe to be plausible are the following "rule versions" of the pointwise, defeasible transitivity and strengthening of the antecedent rules.

DEFEASIBLE TRANSITIVITY $\forall x(\varphi > \psi), \forall x(\psi > \zeta) \mid\sim \forall x(\varphi > \zeta)$, but not $[\forall x(\varphi > \psi), \forall x(\psi > \zeta), \forall x(\varphi > \neg\zeta) \mid\sim \forall x(\varphi > \zeta)]$.

DEFEASIBLE STRENGTHENING OF THE ANTECEDENT: $\forall x(\varphi > \psi) \mid\sim \forall x((\varphi \& \zeta) > \psi)$ but not $\{\forall x(\varphi > \psi), \forall x((\varphi \& \zeta) > \neg\psi) \mid\sim \forall x((\varphi \& \zeta) > \psi)$.

Finally there is a third group of reasonable inferences that depend on a view of normality on which normality comes in degrees.

GRADED NORMALITY AND DEFASIBLE MODUS PONENS: $\forall x(\varphi > \psi_1), \ldots, \forall x(\varphi > \psi_n), \varphi(\delta), \neg\psi_1(\delta) \mid\sim \psi_2(\delta) \& \ldots \& \psi_n(\delta)$ but not $\{\forall x(\varphi > \psi_1), \ldots, \forall x(\varphi > \psi_n), \varphi(\delta), \neg\psi_1(\delta) \mid\sim \psi_1(\delta) \& \ldots \& \psi_n(\delta)\}$.

GRADED NORMALITY AND DEFASIBLE TRANSITIVITY: $\forall x(\varphi > \psi), \forall x(\psi > \zeta), \forall x(\psi > \psi_1), \forall x(\varphi > \neg\psi_1) \mid\sim \forall x(\varphi > \zeta)$, but not $\{\forall x(\varphi > \psi), \forall x(\psi > \zeta), \forall x(\psi > \psi_1), \forall x(\varphi > \neg\psi_1) \mid\sim \forall x(\varphi > \psi_1)\}$.

Although we have worked out a semantics on which these last two groups of inferences hold, we cannot give a detailed account here. We give a brief description in the penultimate section of the paper; we will concentrate on presenting a minimal system of commonsense entailment

In any case, an acceptable theory of genericity will respect the distinction between logical entailment and reasonable inference, and it must realistically model the reasoning belonging to each area. In particular, the semantics must capture the feature that these defeasible patterns of inference introduce a dependence on epistemic contexts which is not at all present in the case of the valid patterns. Their conclusions are defeated *as one obtains information* which brings them into doubt, not as the world changes in whatever way. Realistically modelling such defeasible reasoning will lead us to model these epistemic contexts explicitly in the semantics of generic sentences.

A second requirement for theories of genericity is that they be sufficiently general. The correct theory must provide interpretations not only for simple generic sentences like most of those we have seen up until now, but also for composite sentences in which genericity mixes with counterfactuality, belief, knowledge, and even with more genericity, as in the case of nested generic sentences like *healthy cats jump at small moving objects,* or peopie who work late *at nights do not wake up early.* Here is an example of generics interacting with counterfactuality and propositional attitudes: *John knows that Mary loves kissing him, and he would be unhappy if she were were to like it less.* A theory of what generics mean ought at least to extend to a theory of what they mean in such contexts.

The third desideratum for theories of genericity is a methodological one. In one respect most of the formalisms (including the one to be presented below) for representing and reasoning with generic information are alike: they are empirically inadequate. In another respect, however, they differ greatly. Some theories cover up the deficiencies of the underlying mechanisms that purport to explain nonmonotonic reasoning by introducing devices foreign to those mechanisms. One case in point is the way in which theories treat the penguin principle, the pattern of reasoning where specific information takes precedence. This is the familiar problem of multiple extensions. Formalizing the premises of the penguin principle in the way done in circumscription by means of a multitude of "abnormality predicates," for example. we find that minimization of abnormality results in two kinds of minimal models: there are models where Tweety is an abnormal bird but a normal penguin, and so does not fly. But in addition there are others where he is a normal bird but an abnormal penguin, and does fly. Because of these latter, undesirable models it then does not follow that Tweety does not fly, and we see that circumscription does not handle the penguin principle adequately. Similar problems confront default logic and autoepistemic logic. The solution which proponents of these theories have suggested is as familiar as the problem: the order in which default rules fire needs to be constrained; the predicates to be minimized in the case of circumscription need to be prioritized. They thus commit themselves to the

HYPOTHESIS OF THE GHOST IN THE MACHINE

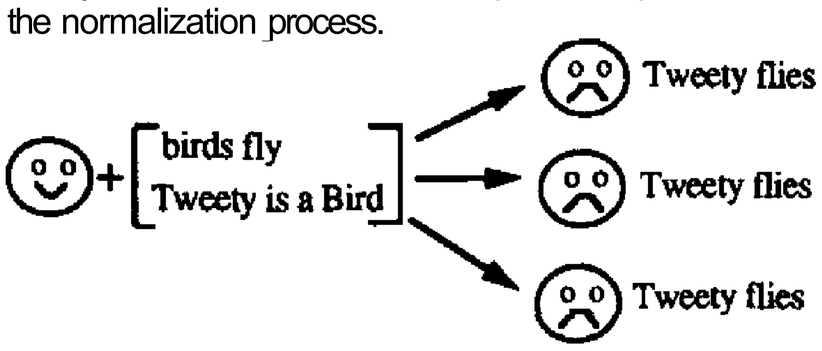
> That specific information takes precedence over general information is not to be accounted for by the semantics of generic statements itself. Rather, it is due to the intervention of a power which is extraneous to the semantic machinery, but which guides this machinery to have this effect (by ordering the defaults* deciding the priorities of predicates to be minimized, or whatever).

Whatever kind of reasoning generic reasoning is, more specific information takes precedence is intrinsic to it. The penguin principle should emerge naturally from the semantics of generic sentences without the intervention of a user who

decides how the reasoning is to be applied. We want to exorcise the ghost from the machine.

Emulating the possible worlds analysis of conditional sentences, we construct first a semantics assigning truth values to (nested) generic sentences relative to possible worlds. The truth conditional part of our semantics, with its standard notion of entailment, accounts for the valid argument forms we mentioned. Because it is a conventional possible worlds theory, we can insert the semantics of generics within general, possible worlds frameworks that yield semantics for counterfactuals, prepositional attitudes and so on. It is thus in principle very clear how to interpret complex sentences such as those discussed above.

On top of the truth conditional semantics, we build a second, dynamic, partial theory, which accounts for defeasible patterns of reasonable inference. Our intuitive picture of what goes on when one reasons by defeasible modus ponens is this: first one assumes the premises *birds fly* and *Tweety is a bird,* and no more than this. Second, one assumes that Tweety is as normal a bird as is consistent with these premises. Finally, one then looks to see whether one believes that Tweety flies or not, and finds that he does. In our view *all* of the patterns of defeasible reasoning outlined in the introduction arise in this way, from assuming just their premises, then assuming everyone and everything as normal as is epistemically possible, and finally seeing whether one believes their conclusions. The dynamic semantic models such epistemic reasoning by means of information states, which are sets of possible worlds taken from the truth conditional model theory. We define two functions on these information states: updating and normalization. The first of these is eliminative, simply removing from information states all those possible worlds where the sentences with which one is updating are not true. Assuming just the premises of an argument can then be modelled as updating a distinguished informauonally minimal state ☺ with those premises. The second of these functions, normalization, encodes in the semantics the notion of assuming everyone and everything as normal as possible. Normalizing the result of updating ☺ with a set of premises *T* yields a set of information states which are fixpoints of the normalization process. The conclusions of reasonable inferences from premises T are all those sentences that are true at all the worlds in all of these fixpoints. The figure below graphically depicts our dynamic theory of reasonable inference;+ represents update, the arrows the normalization process.



2, The Truth Conditional Semantics of Generics

2.1 The Language

In the following we are working with a first-order language L augmented with a binary conditional operator >.

The formulas of the language $L_>$ are defined by the usual clauses, tögether with the following one:

If φ and ψ are formulas, then φ > ψ is a formula.

This definition allows for arbitrarily deep nesting of the conditional > in formulas.

2.2 The Truth Conditional Semantics

The underlying semantic idea in interpreting the language $L_>$ is that a generic sentence $\forall x(\varphi > \psi)$ is true at a possible world just in case, at that world, being a normal φ involves being ψ. The modal frames encode what being a normal O invariably involves by means of a worlds accessibility function *, which assigns to each possible world w and proposition p a set of worlds. This gives a basic semantics for $L_>$ and a simple system of commonsense entailment

DEFINITION: **A $L_>$ frame is a triple $F = \langle W, D, * \rangle$ where**

i) W is a non-empty set of worlds,

ii) D is a non-empty set of individuals, and

iii) $*: W \times \wp(W) \rightarrow \wp(W)$

*(w, p) contains only worlds in which, intuitively speaking, the proposition p holds together with everything else which, in world w, is normally the case when p holds. For example, let p be the proposition that Big Bird of *Sesame Street* fame is a bird. Let w be the actual world, where it is true that birds fly, that birds have feathers, and that birds lay eggs. Then *(w, p) contains only worlds where Big Bird flies, has feathers and lays eggs. World w may in fact not be in *(w, p), given that the television character is, as birds go, not at all typical.

Note that what is normally the case when a proposition p holds is allowed to vary from possible world to possible world. We want there to be, for example, possible worlds where Tweety is a perfectly normal bird and doesn't fly, these quite simply being those possible worlds where it is not true that birds fly. Note also that it is important that we do not suppose any absolute normality order on possible worlds. In particular, we explicitly reject the idea that *(w, p) is to be identified with those most normal of all possible worlds where p holds.

Up until now * has been left virtually unconstrained, and indeed in view of the weak logic of generics only a few constraints are needed.[2] The first of these is

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

Worlds where p holds together with other propositions which are normally associated with p are, no matter how few of these other propositions hold, in any case worlds where p holds. One of the most important patterns of defeasible reasoning which we want to capture is the penguin principle. We capture the penguin principle by introducing a constraint on *, When we build the dynamic semantics, this constraint

[2]One constraint on * which is familiar from the literature on conditional logic but which we certainly do not want to impose here is

CENTERING **If $w \in p$, then $w \in *(w, p)$**

That w is a world where p holds is no guarantee that w is a world where everything holds which is normally associated with p.

will interact with the normalization function in such a way as to give us the penguin principle.

SPECIFICITY: If $*(w, p) \subseteq q$, $*(w, p) \cap *(w, q) = \emptyset$, and $*(w, p) \neq \emptyset$, then $*(w, q) \cap p = \emptyset$.

Suppose that p is the proposition that Tweety is a penguin, q the proposition that he is a bird; suppose further that if Tweety is a normal bird he flies but that if he is a normal penguin he doesn't fly. Then specificity says that for any world in which Tweety is a normal bird, he's not a penguin. We don't think specificity is all that intuitive by itself, but it yields plausible results when combined with our operation of normalization. In all of the following we will restrict $L_>$ to those which satisfy facticity and specificity.

We now define $L_>$ interpretations in $L_>$ frames in the standard way.

DEFINITION: A *base model* M is a pair $M = \langle F, [\![\,]\!] \rangle$, where
i) *F* is an $L_>$ frame, and
ii) $[\![\,]\!]$ is a function from nonlogical constants of L to appropriate intensions (functions from worlds to appropriate extensions)

The satisfaction definition for $L_>$ sentences islargely familiar and uses assignments α_i: Var → D_M. Truth is deiined as satisfaction with respect to all assignments.

DEFINITION: For any possible world $w \in W_M$ and model M:
$M, w, \alpha_i \models \varphi$, as usual, for φ an atomic formula of FOL.
$M, w, \alpha_i \models \varphi > \psi$ iff $*(w, |\varphi|_{M, \alpha_i}) \subseteq |\psi|_{M, \alpha_i}$
The usual clauses for complex formulas involving $\forall, \exists, \vee, \&, \rightarrow, \neg$.
$|\varphi|_{M, \alpha_i} = \{w \in W_M: M, w, \alpha_i \models \varphi\}$.

This truth definition implies that a generic sentence is true in a model M at a world w iff for all $\delta \in D_M$, $*(w, |\varphi\underline{\delta}|_M) \subseteq |\psi\underline{\delta}|_M$. One important novelty in this definition is that unlike the usual definitions in modal or conditional logic, the semantics for generics exploits the instances of the formulas on the left and the right. As a result, the truth of *birds fly* does not imply the existence of worlds where *all* birds are normal fliers. This is an unnatural assumption in many cases.

$L_>$ logical consequence, $\models$, is defined in a completely standard way: $\Gamma \models \varphi$ iff in all $L_>$ models M extending $L_>$ frames satisfying facticity and specificity, if $\forall \gamma \in \Gamma$ $M \models \gamma$, then $M \models \varphi$. A corresponding derivability notion $\vdash$ is given by the following rules and axioms; the resulting derivability notion is denoted $\vdash$.

The Logic T_1
(A1) Truth functional $L_>$ tautologies
(A2) $\forall x \varphi \rightarrow \varphi[t/x]$ for any term t
(A3) $\forall x\, \varphi \leftrightarrow \neg \exists x \neg \varphi x$
(A4) $\forall x(\varphi \rightarrow \psi) \rightarrow (\exists x \varphi \rightarrow \psi)$
(A5) $\varphi > \varphi$.
(A6) $(\varphi > \psi \,\&\, \psi > \zeta \,\&\, \varphi > \neg\zeta) \rightarrow \psi > \neg\varphi$.
(A7) $\forall x(\varphi > \psi) \rightarrow (\varphi > \forall x \psi)$, for x not free in φ.
(R1) $\vdash \varphi$ and $\vdash \varphi \rightarrow \psi \Rightarrow \vdash \psi$
(R2) $\vdash (\psi_1 \,\&\, .. \,\&\, \psi_n) \rightarrow \psi \Rightarrow \vdash (\varphi > \psi_1 \,\&\, .. \,\&\, \varphi > \psi_n) \rightarrow \varphi > \psi$
(R3) $\vdash \varphi \rightarrow \psi[t/x]$, where t is a constant not in φ or $\psi \Rightarrow \vdash \varphi \rightarrow \forall x \psi$
(R4) $\vdash \varphi \Rightarrow \vdash \varphi[t/x]$ where t is a term not in φ.
(R5) $\vdash \varphi \leftrightarrow \psi$ and φ a subformula of $\zeta \Rightarrow \vdash \zeta \leftrightarrow \zeta[\psi/\varphi]$

Using a slight extension of the present language so as to express the Penguin principle and Henkin's technique to construct a canonical model yields the following completeness theorem for T_1.

THEOREM 1: $\Gamma \vdash \varphi$ iff $\Gamma \models \varphi$

3. The Dynamic Semantics of Generic Reasoning

We now show how to model the patterns of invalid but reasonable inference by building on top of the truth conditional semantics a dynamic semantics. We will use four concepts: information states, updating, the state of ignorance, and normalization, which we now spell out

3.1. Information States, Updates and Ignorance

We take *information states* to be sets of possible worlds taken from the base models already defined. Accordingly, our approach to updating information states with new information is very simple and follows Stalnaker's definition. On updating with *Sam is a dodo,* the set of one's informational possibilities is reduced to those possible worlds where Sam is a dodo. We will define update functions + of this kind. We also will define a support relation h between belief states and sentences of L>.

In order to capture the notion of believing no more than one has been told, we must define a very particular information state. This is the *informationaUy minimal* state of the introduction, ☺. This informational!y minimal state must support only logical truths. This state must have some particular properties: it must contain enough worlds to verify every possible consistent combination of L sentences, for instance. Furthermore, it must contain worlds w to which are assigned sets of p normal worlds for every proposition p so that the T_1 axioms hold at every w. The Henkin construction procedure for the canonical T_1 model yields the appropriate model M_o in which to define ignorance. Wo, the set of worlds in Mo, is just ☺. "Knowing no more than *V* comes to being in the information state ☺ + T

3.2 Normalization

Normalization is the most complex part of our dynamic semantics. In normalization we assume that various individuals in a certain *situation* are normal. Situations contain objects with properties and standing in relations to each other. In normalizing with respect to a possible situation, we will be assuming the individuals in that situation to be normal with respect to the properties they have in that situation. Since such situations need not be actual, the individuals may not actually possess those properties. We limit ourselves here to *simple situations,* A simple situation is one in which a single individual has a simple atomic property; in virtue of the canonical model in which we are working, we may identify a simple situation with an atomic formula or a negation of an atomic formula paired with an individual. To

model the notion of assuming everything to be as normal as epistemically possible, we iterate normalization with respect to a set of simple situations.

To get an intuitive feel for normalization with respect to a single, simple situation consider for example a state which contains only the information that birds fly and that Tweety is a bird. Strengthening such a state with the assumption that Tweety is as normal a bird as is consistent with that state will return the information that Tweety flies. An initial state which contains the additional information that Tweety does not fly will, when thus strengthened, not return the information that Tweety flies, since in this case the new assumption is not consistent

The normalization function makes use of the information about normality contained within whole information states. So we define the notion of a set of normal worlds for information states as follows:

DEFINITION: $*(s, p) := \cup_{w \in s} *(w, p)$.

Consider a simple situation ψ $\psi = <\varphi(x), \delta>$ and $\delta \in D_{M_0}$. We will write $|\psi|$ to denote the set of worlds in which δ has the property of being a φ. $\underline{N}(s, \psi)$, defined below, stands for the result of strengthening s by assuming δ to be a normal ψ, if this is consistent with s. The definition of normalization uses information about normality to characterize worlds that the state does *not* take to be normal. For ψ as defined above, $|\psi| \setminus *(w, |\psi|)$ stands for the set of possible worlds in which, according to an isolated w, an individual δ is, though a φ, not a normal φ. So $|\psi| \setminus *(s, |\psi|))$ is the set of worlds where, according to *all* of the worlds in s, δ is not a normal φ. They all agree that in each world in $|\psi| \setminus *(s, |\psi|))$, δ lacks at least one of the properties had by a normal φ.

DEFINITION: $\underline{N}(s, \psi) := \{w \in s : w \notin (|\psi| \setminus *(s, |\psi|))\}$, if $s \cap *(s, |\psi|) \neq \emptyset$;
$:= s$, otherwise.

$\underline{N}$ isolates those worlds in s where δ is a normal φ, if this is consistent with s: a world v is excluded from $\underline{N}(s, \psi)$ just in case δ is, *at v itself*, not a normal φ, if this is consistent with s. To see what this means concretely, let us for the moment suppose that it is, intuitively speaking, consistent with s that δ is what (in s) is taken to be a completely normal φ. For example, where φ is *bird*(x) and s contains the information *bird*(δ) and *bird*(x) $>_X$ *fly*(x), we are supposing the sentence *fly*(δ) to be consistent with s. This state of affairs amounts formally to the case where $s \cap *(s, |\psi|) \neq \emptyset$, so then, given the above explanation of what $|\psi| \setminus *(s, |\psi|)$ stands for, $\underline{N}(s, \psi)$ clearly contains only those worlds in s where δ is, if at all a φ, then as normal a φ as is consistent with s (namely a completely normal one). It is not always consistent with s that δ is a normal φ. *This is the case in which* $s \cap *(s, |\psi|) = \emptyset$. Assuming δ to be a normal φ is in this case hopeless, so the normalization function "gives up" and simply returns the original state s.

Let us now turn to iterated normalization. In normalizing with respect to many individuals and many properties, we iterate the normalization function on information states. One factor affecting such iterated normalization is the order in which the normalizations of the sort discussed in the previous section are performed. Consider for instance a Nixon diamond situation, where the outcome depends on the order in which the different respects in which individuals are assumed normal show up in the iteration. If one first assumes that Dick is a normal republican and only then assumes that he is as normal a quaker as possible, then one will end up with the information that he is a non-pacifist If one first assumes him to be a normal quaker and then assumes that he is as normal a republican as possible, then one will end up with the information that he is a pacifist So it is desirable that the order sensitivity of the iterated normalization be cancelled out by taking all different orderings of the iterations into account

A second point about iterated normalization concerns the situations relevant to normalization. The relevant situations are defined with repsect to the premises T in a nonmonotonic inference that one wishes to verify. The relevant situations relative to a set of premises *T* are those situations «p(x), 8> where 6 E DM_0 and O(x) is either the antecedent of a "positive" occurrence of a universally quantified > conditional in T* or identical with 5(x/t)(where 6 is the antecedent of a non quantified > conditional in T*, where T* contains the conjunctive normal form of each sentence in T. We call the set of such antecedents Subst(Ant(T))-

DEFINITION: For $\Gamma \subseteq L_>$, the *Γ-normalization chain* with respect to an enumeration ν of Subst(Ant(Γ)) is defined to be the following sequence:
$\mathbf{N}^0_\nu = s$
$\mathbf{N}^{\alpha+1}_\nu = \underline{N}(\mathbf{N}^\alpha_\nu, \varphi_i)$, where $\alpha = \lambda + n + 1$, and $\nu(\varphi_i) = n + 1$.
$\mathbf{N}^\lambda_\nu = \cap_{\mu \in \lambda} \mathbf{N}^\mu_\nu$

There is one such Γ-normalization chain for each state and each enumeration of Subst(Ant(Γ)). Every Γ-normalization chain beginning in a state s reaches a fixed point. That is: for all s and ν, $\exists\alpha\forall\beta \geq \alpha\ \mathbf{N}^\beta_\nu = \mathbf{N}^\alpha_\nu$.

3.3 Reasonable Inference

Armed with the notions of information states, ignorance, updating and maximal normality, we can now put together our model theory of non-monotonic reasoning. We define the relevant dynamic information model, as well as the support relation for information states, and then finally our notion of commonsense entailment.

DEFINITION: The *dynamic information model* $\mathcal{A}_0$ is a triple $<\wp(W_{M_0}), +, \underline{N}>$, and:
(i) $+: \wp(W_{M_0}) \times \wp(L_>) \to \wp(W_{M_0})$ is defined such that $\forall s \in \wp(W_{M_0}), \forall \Gamma \subseteq L_>$ $s+\Gamma = s \cap (\cap_{\gamma \in \Gamma} |\gamma|_{M_0})$.
(ii) $\underline{N}$ is the normalization function on information states.

Note that $\textcircled{\smile} \in \wp(W_{M_0})$.

DEFINITION: For $s \in \wp(W_{M_0})$, $s \models \varphi$ iff for all $w \in s$, $M_0, w \models \varphi$.

DEFINITION: $\Gamma \mid\!\sim \varphi$ iff for any Γ-normalization chain C beginning from $\textcircled{\smile} + \Gamma$, $C^* \models \varphi$, where C^* is the fixpoint of C.

Since $\Gamma \vdash \varphi \Rightarrow \Gamma \mid\sim \varphi$, our non-monotonic consequence relation is supraclassical. Further all the T_1 theorems are also $\mid\sim$ valid. Theorem 2 shows that $\mid\sim$. captures the first group of reasonable inference patterns from the introduction.

DEFINITION: φ, ψ and ζ are *independent formulas* just in case for all $\underline{\delta}$, each boolean combination containing at most one instance of each of $\varphi\underline{\delta}$, $\psi\underline{\delta}$, $\zeta\underline{\delta}$ is satisfiable.

THEOREM 2: When φ, ψ, and ζ are restricted to independent formulas, Defeasible Modus Ponens, Pointwise Defeasible Transitivity, Pointwise Defeasible Strengthening of the Antecedent, Nixon Diamond, Penguin Principle, the Embedded Normality patterns as they are stated in the introduction all hold given our interpretation of $\mid\sim$.

Another welcome fact about our interpretation of I= is that Defeasible Irrelevance does not hold. While this fact follows straightforwardly from the definition of I=, the proof of theorem 3 exploits a construction of particular worlds for each inference pattern that (a) verify the premises of the inference pattern and exist in the state of ignorance, (b) survive the process of normalization, and (c) have the requisite properties to force the normalized state to verify the desired conclusion.

4. Subtheories and Extensions

There are several extensions and subtheories of the basic system, of commonsense entailment we have just described thai we will describe at length in a longer paper. For instance, there is a quantifier free version of our system in which the set of nonmonotonic consequences of a finite set of premises r is decidable.

We can also extend the system of commonsense entailment in $L_>$ with stronger monotonic inferences. We think, for example, that this inference pattern is valid.

$$\forall x(\varphi > \psi) \;\&\; \forall x(\zeta > \psi) \vdash \forall x((\varphi \vee \zeta) > \psi)$$

The constraint on * required to validate the Dudley Doorite argument scheme in an $L_>$ complicates, however, the construction of a canonical model.

We may also extend the system by complicaing the definition of normalization.

DEFINITION: $\underline{N}^*(s, \psi) = \underline{N}(s, \psi) \cap X$, where

$X :=$ { $w \in s$: for all situations ζ $*(w, |\zeta|) = *(w, |\zeta|) \setminus *(s, |\psi|)$}, if $*(w, |\zeta|) \setminus *(s, |\psi|) \neq \emptyset$;

$:= s$ otherwise.

Where $\psi = \langle\varphi(x), \delta\rangle$ and $\delta \in D_{M_0}$, a world v is excluded from $\underline{N}(s, \psi)$ just in case δ is, *at v itself*, not a normal φ and this is consistent with s. A world v is excluded from set X just in case *the set of normal ζ worlds* of v contains worlds where δ is not a normal φ if this is consistent with other normality information in s. Each set of ζ normal worlds for worlds of s should encode as much normality information of s as is consistent with it. A nonmonotonic consequence relation $\mid\sim^*$ based on $\underline{N}^*$ extends $\mid\sim$ and verifies the rule forms of defeasible transitivity and strengthening of the antecedent.

Yet another extension refines the basic semantics by replacing the sets of normal worlds * (w, p) in L> models with concentric spheres of worlds $(w, p). to capture a limited version of the graded normality inferences.[3]

5. Comparisons with Other Work

The theory presented above derives heavily from two traditions in logic which many have thought closely related: modal logic and the theories of nonmonotonic and default reasoning developed during the last decade. Our theory differs considerably from Reiter's default logic. Reiter's default logic augments classical logic with default rules of the form: if O and it is not inconsistent to assume y, then L. These rules offer a representation of generic facts, but significantly the theory has no representations for generic sentences in the object language. Thus, there seems to be no way to write down sensible representations for nested defaults and nested generic statements or to reason about them within Reiter's formalism. McCarthy's original proposal concerning circumscription— or the model theoretic minimization of certain predicates— has in principle the resources to represent nested generic sentences, on the other hand; but the formalism does not capture most of the desired inferences. Further, the original system as well as subsequent refinements are very unstable. Defeasible modus ponens fails, for example, as soon as we add the new premise that there is a bird that doesn't fly. Further, the refinements get the desired inferences only to the extent that ghosts are imported into the logic. We get the inferences without ghosts.

With Delgrande, we exploit an analogy with conditional logic. Again, however, Delgrande does not give a theory that permits us to reason about or even assign meanings to nested generic statements. Delgrande's theory does not, insofar as we understand it, handle adequately inferences like defeasible strengthening of the antecedent The reason we get defeasible strengthening in the same breath (though it's a deep one) as default transitivity is because of the very special use of the informationally minimal state and updating. Whereas Delgrande has to appeal to complicated mechanisms to test far "irrelevance," the update of the informationally minimal state with *birds fly* will yield upon normalization that *white birds*

References

[Delgrande, 1987] Jim Delgrande. A Semantics for Defaults Using Conditional Logic. *Artificial Intelligence,* 1987.

[McCarthy, 1980] John McCarthy. Circumscription- A Form of Non-Monotonic Reasoning. *Artificial Intelligence* 13,pp.27-39,I980.

[Reiter, 1980] Ray Reiter. A Logic for Default Reasoning. *Artificial Intelligence* 13, pp.81-132,1980.

[Stalnaker, 1968J Robert Stalnaker. A Theory of Conditionals. Nicholas Resetter, editor. *Studies in Logical Theory,* Oxford: Basil Blackwell.

[3]We have shown that these inferences may hold if the consequents of conditionals in the premises are restricted to atomic formulas or negations of atomic formulas. These limits are significant; we know that we cannot give a fully general version of the graded normality inferences without trivializing the theory.