A Note on Cumulative Stereotypical Reasoning

Giovanni Casini[1] and Hykel Hosni[2]

[1] University of Pisa
giovanni.casini@gmail.com
[2] Scuola Normale Superiore, Pisa
hykel.hosni@sns.it

Abstract. We address the problem of providing a logical characterization of reasoning based on stereotypes. Following [7] we take a semantic perspective and we base our model on a notion of semantic distance. While still leading to cumulative reasoning, our notion of distance does, unlike Lehmann's, allow reasoning under inconsistent information.

Keywords: Stereotypes, prototypes, cumulative reasoning, nonmonotonic logic, default-assumption logic.

1 Introduction

One important feature of intelligent reasoning consists in the capability of associating specific situations to general patterns and by doing so, extending one's initial knowledge. Reasoning based on stereotypes is a case in point. Loosely speaking, a stereotype can be thought of as an individual whose characteristics are such that it represents a typical (i.e. generic) individual of the class it belongs to. For this reason a stereotypical individual can be expected to satisfy the key properties which are typically true of the class to which the individual belongs (see Section 2 below for an example). Of course exceptions might be waiting just around the corner and an intelligent agent must be ready to face a situation in which the properties projected on a specific individual by using stereotypical information turn out not to apply. Stereotypical reasoning is therefore *defeasible.*

The purpose of this paper is to provide a logical insight on the problem of modelling rational stereotypical reasoning. Our central idea consists in representing the latter as a two-stage inference process along the following lines. Given a piece of specific information, an agent *selects* among some background information available to it, those stereotypes which better fit the factual information at hand. We expect this to normally expand the initial information available to the agent. The second step is properly inferential: using the new (possibly expanded) information set the agent draws defeasible conclusions about the situation at hand. The key ingredient in the formalization of the first stage is a function which ranks the fitness of a set of stereotypes with respect to some factual information. Following [7] we interpret fitness in terms of a semantic distance function. Due to the defeasible nature of reasoning based on stereotypes, the inferential stage will have to be formalized by a non monotonic consequence

C. Sossai and G. Chemello (Eds.): ECSQARU 2009, LNAI 5590, pp. 590–601, 2009.

relation. Since we are interested in representing *rational* reasoning, we shall be asking for this consequence relation to be particularly well-behaved. In our model this amount to requiring that stereotypical reasoning should be *cumulative*.

The paper is organized as follows. Section 2 sets the stage for our discussion on stereotypes and provides a general characterization of semantic distance. Section 3 is devoted to recall some basic facts about non monotonic reasoning in general and default-assumption consequence in particular. We then review in Section 4 Lehmann's original proposal for distance-based stereotypical reasoning and we highlight a basic shortcoming of such model. We attempt to fix this in Section 5 where we propose a semantic distance for information which is potentially inconsistent with an agent's defaults. While overcoming the limitation of Lehmann's model, the distance function introduced there fails to lead to full cumulative reasoning. We then combine the intuition behind both distances in Section 6 where we define a lexicographic distance function which at the same time admits inconsistency and leads to cumulative consequence relations.

Before getting into the main topic of this paper, let us fix some notation. We denote by ℓ the set of sentences built-up from the finite set propositional letters $P = \{p_1, \ldots, p_n\}$ using the classical propositional connectives $\{\neg, \wedge, \vee, \rightarrow\}$ in the usual, recursive way. We denote by lowercase Greek letters $\alpha, \beta, \gamma, \ldots$, the sentences of ℓ while sets of such sentences will be denoted by capital Roman letters $A, B, C, \ldots$. As usual we denote consequence relations by $\vDash$ and $\mid\!\sim$. In particular, $\vDash$ denotes the classical (Tarskian) consequence relation while we use $\mid\!\sim$ (with various decorations) for non-monotonic consequence relations. Since it is sometimes handier to work with inference operations rather than consequence operations, we shall use Cl for the classical inference operation, that is $Cl(A) = \{\beta | A \vDash \beta\}$ and C (with decorations) for the non monotonic ones, that is $C(A) = \{\beta | A \mid\!\sim \beta\}$.

Semantically, we take sets of classical (binary) propositional valuations on the language $W = \{w, v, \ldots\}$ interpreted as possible *states* of the world. Then we also use $\vDash$ for the satisfaction relation between valuations and formulae where $w \vDash \alpha$ reads as 'The valuation w satisfies the formula α'. Given $A \subseteq \ell$ and a set W, we shall write $[A]_W$ to indicate the set of the valuations in W which satisfy all the sentences in A ($[A]_W = \{w \in W \mid w \vDash \phi \quad \text{for every} \quad \phi \in A\}$). We shall drop the subscript and write simply $[A]$ whenever the reference to the particular set of valuations is irrelevant.

2 Stereotypes

Stereotypes have been vastly investigated in a number of areas, from the philosophy of mind to the cognitive sciences, for their key role in the development of theories of concept-formation and commonsense reasoning (for an overview, see e.g. [6]). Stereotypes feature prominently in Putnam's social characterization of meaning (see, e.g. [11]) as well as in most of the current approaches to conceptualization while Lackoff [5] points out their fundamental importance in commonsense and uncertain reasoning.

To fix a little our ideas on stereotypes, let us take a class of individuals, 'birds', for example; a *stereotype bird* can be thought as a set of properties defining an individual bird that we consider to be particularly representative of the very concept of a bird. In this case, then, those properties could be identifying a robin or some other little tree-bird. Hence, if we take a logical perspective on the problem, we can think of stereotypes as a set of states that typically, but not necessarily, are true of some particularly representative members of a class (a stereotypical bird will be a flying winged animal, of little dimensions, covered with feathers, with a beak, laying eggs, singing, nesting on trees, etc.). This idea suggests identifying a stereotype with a finite set of sentences $\Delta = \{\alpha_1, \ldots, \alpha_m\}$ which are true exaclty at those states which characterize the stereotype. We denote by $\mathfrak{S}, \mathfrak{T}, \ldots$ finite sets of stereotypes ($\mathfrak{S} = \{\Delta_1, \ldots, \Delta_n\}$).

In our interpretation, a set of stereotypes represents the *stereotypical* or *default* information available to an agent. This interpretation is justified by recalling that in defeasible reasoning, *defaults* refer to those pieces of information that an agent considers to be typically, normally, usually, etc. true. So, by taking stereotypical properties as defaults, we capture the idea that an agent considers stereotypical information as defeasible, and hence possibly revisable in the event of evidence to the contrary.

The close connection between defeasible and stereotypical reasoning has been brought to the logician's attention by D. Lehmann ([7]) who proposes a model for stereotypical reasoning along the following lines. An agent starts with a set of n stereotypes, $\mathfrak{S} = \{\Delta_1, \ldots, \Delta_n\}$, and is then given information about some particular individual, represented by a *factual* formula α, that we assume is consistent. This fixes what the agent considers true of the state of the world at hand. The idea then is that an agent's reasoning depends on "how good" α is as a stereotype in $\mathfrak{S}$. In order to capture this formally, Lehmann introduces a notion of semantic distance $d(\alpha, \Delta)$ between the factual and the stereotypical information available to the agent. The smaller the distance d, the better factual information "fits" the stereotype Δ. To take good advantage of stereotypical reasoning, then, the agent should associate to the factual information at hand the *nearest* stereotype. More precisely, given α and every stereotype Δ_i in $\mathfrak{S}$, the agent *selects* a subset of $\mathfrak{S}$, $\mathfrak{S}^\alpha_d$, of maximally close (i.e. nearest) elements of $\mathfrak{S}$ to α with respect to d. This is interpreted as the set of stereotypes which is natural for the agent to associate to α. Formally:

$$\mathfrak{S}^\alpha_d = \{\Delta_i \in \mathfrak{S} |\ d(\alpha, \Delta_i) \leq d(\alpha, \Delta_j) \text{ for every } \Delta_j \in \mathfrak{S})\} \qquad (\sharp)$$

The selection of the nearest stereotypes to a formula α leads naturally to defeasible reasoning which is captured by the consequence relation $\mid\!\sim_{\mathfrak{S},d}$. For obvious reasons we refer to this latter as the *consequence relation generated by* $\mathfrak{S}$ *and* d. To recap, the model goes as follows. An agent is equipped with a finite set of stereotypes $\mathfrak{S}$ and a semantic distance function d. Given a factual formula α, the agent selects the set $\mathfrak{S}^\alpha_d$ of stereotypes which d ranks nearest to α. Now this set $\mathfrak{S}^\alpha_d$ is used to *generate* a consequence relation $\mid\!\sim_{\mathfrak{S},d}$, which, as we shall shortly see, provides an adequate tool to produce defeasible conclusions from α and the default information contained in $\mathfrak{S}^\alpha_d$.

Thus, before recalling the constraints imposed by Lehmann on the distance function d and the properties of the generated consequence relation, we need to recall some basic facts about non monotonic consequence relations.

3 Cumulative and Default-Assumption Consequence

Among the many proposals to characterize defeasible reasoning (see [2] for an overview) some core structural properties emerge as particularly compelling (see [4], [9], and [10]). In particular, the class of *cumulative consequence relations* has gained quite a consensus in the community as the industry standard.

Definition 1 (Cumulative Consequence Relations). *A consequence relation* $\mid\!\sim$ *is* cumulative *if and only if it satisfies the following properties:*

REF	$\alpha \mid\!\sim \alpha$	*Reflexivity*
LLE	$\dfrac{\alpha \mid\!\sim \gamma \quad \models \alpha \leftrightarrow \beta}{\beta \mid\!\sim \gamma}$	*Left Logical Equivalence*
RW	$\dfrac{\alpha \mid\!\sim \beta \quad \beta \models \gamma}{\alpha \mid\!\sim \gamma}$	*Right Weakening*
CT	$\dfrac{\alpha \mid\!\sim \beta \quad \alpha \wedge \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma}$	*Cut (Cumulative Transitivity)*
CM	$\dfrac{\alpha \mid\!\sim \beta \quad \alpha \mid\!\sim \gamma}{\alpha \wedge \beta \mid\!\sim \gamma}$	*Cautious Monotony*

where $\models$ *denotes as usual the tarskian consequence relation of classical logic.*

Combining the flexibility of nonmonotonic (i.e. default, revisable, defeasible, etc.) reasoning with many desirable metalogical properties, such as *idempotence*, *supraclassicality*, and *full-absorption* (see [9]), cumulative consequence relations constitute a tool of choice in the formalization of commonsense inference.

Among the class of cumulative consequence relations are the so-called *default-assumption consequence relations*, which will play a key role in our model and which we therefore turn to recall. The idea behind default-assumption reasoning is that an agent's information can be viewed as being two-fold. On the one hand agents have *defeasible information*, a set Δ of defaults that an agent *presumes* to be typically true. On the other hand agents might acquire *factual information*, that is, information that the agent *takes as true* of the particular situation at hand, and which is represented, in our setting, by a single formula α. Intuitively, then, default-assumption reasoning takes place when an agent extends its factual information α with those defaults which are compatible with α and takes the result as premises for its inferences.

In order to formalize this we need to define the set of *maximally α-consistent subsets* of Δ, or, equivalently, the notion of *remainder set* (see [3], p.12).

Definition 2 (Remainder Sets). *For B a set of formulae and α a formula, the* remainder set $B \bot \alpha$ *('B less α') is the set of sets of formulae such that* $A \in B \bot \alpha$ *if and only if:*

1. $A \subseteq B$
2. $\alpha \notin Cl(A)$
3. There is no set A' *such that* $A \subsetneq A' \subseteq B$*, and* $\alpha \notin Cl(A')$

Thus, for a set of defaults Δ, $\Delta \bot \neg\alpha$ is the set of every maximal subsets of Δ consistent with α. Default-assumption consequence relation can then be defined as follows:

Definition 3 (Default-assumption consequence relation). *β is a default-assumption consequence of α given a set of default-assumptions Δ, (written $\alpha \mid\!\sim_\Delta \beta$) if and only if β is a classical consequence of the union of α and every set in $\Delta \bot \neg\alpha$:*

$$\alpha \mid\!\sim_\Delta \beta \text{ iff } \alpha \cup \Delta' \vDash \beta \text{ for every } \Delta' \in \Delta \bot \neg\alpha$$

It is well-known that default-assumption consequence relations are cumulative (see e.g. [1] and [10]).

4 Lehmann's Model

In [7] Lehmann proposes a set of intuitive constraints that the semantic distance d should satisfy in order to generate a well-behaved consequence relation $\mid\!\sim_{\mathfrak{S},d}$. We denote by δ Lehmann's distance function and by $\mathfrak{S}^\alpha_\delta$ the set of stereotypes in $\mathfrak{S}$ selected by δ with respect to a factual formula α as in ($\sharp$) above. Finally we denote by $\mid\!\sim_{\mathfrak{S},\delta}$ the consequence relation generated by $\mathfrak{S}$ and δ.

Recall that the stereotypes in $\mathfrak{S}^\alpha_\delta$, are meant be those which fit better the factual information represented by α i.e. those with minimal semantic distance from α. Thus, it is natural to capture this by looking at the overlap between the states of the world which make α and a set of stereotypes Δ true. But such states are precisely the models of α and the models of Δ ($[\alpha]$ and $[\Delta]$, respectively). The idea is obviously that greatest overlap means maximal closeness. So, given a set of stereotypes $\mathfrak{S}$ and a factual formula α, Lehmann requires that:

- For every $\Delta \in \mathfrak{S}$, $\delta(\alpha, \Delta)$ should be anti-monotonic with respect to $[\Delta] \cap [\alpha]$ (the larger the overlap, the smaller the distance).
- For every $\Delta \in \mathfrak{S}$, $\delta(\alpha, \Delta)$ should be monotonic with respect to $[\Delta] - [\alpha]$ (the larger the set of states which satisfy the defaults but not the factual information, the larger the distance).

The following simplifying assumption is also made:

- $|\mathfrak{S}^\alpha_\delta| = 1$ (i.e. for every α and $\mathfrak{S}$, the agent selects exactly one element in $\mathfrak{S}$).

The above constraints are formalized by:

$$[\Delta'] \cap [\alpha'] \subseteq [\Delta] \cap [\alpha] \text{ and } [\Delta] - [\alpha] \subseteq [\Delta'] - [\alpha'] \;\Rightarrow\; \delta(\alpha, \Delta) \leq \delta(\alpha', \Delta') \quad \text{(L1)}$$

The generated consequence relation $\mid\!\sim_{\mathfrak{S},\delta}$ is then defined by adding the information of the only default set Δ^α_δ in $\mathfrak{S}^\alpha_\delta$ to the premise set α:

$$\alpha \mid\!\sim_{\mathfrak{S},\delta} \beta \text{ iff } \{\alpha\} \cup \Delta^\alpha_\delta \vDash \beta. \quad (\text{L}\mid\!\sim)$$

For any distance function δ satisfying (L1), Lehmann proves the following result:

Theorem 1 ([7], Theorem 5.5). *If* $([\alpha] \cap [\Delta_\delta^\alpha]) \subseteq [\alpha'] \subseteq [\alpha]$, *then* $\mathfrak{S}_\delta^{\alpha'} = \mathfrak{S}_\delta^\alpha$.

That is, if the agent becomes aware of new factual information α' that is not inconsistent with the stereotype previously selected $(([\alpha] \cap [\Delta_\delta^\alpha]) \subseteq [\alpha'])$, then the agent should not abandon the selected stereotype (Δ_δ^α) to extend its factual information (that is, $\mathfrak{S}_\delta^{\alpha'} = \mathfrak{S}_\delta^\alpha$). From this theorem Lehmann proves his main result: given a set of stereotypes $\mathfrak{S}$ and a distance function δ, if the distance function δ satisfies the constraint (L1), then the generated consequence relation $\mid\!\sim_{\mathfrak{S},\delta}$ is cumulative ([7], Corollary 5.6).

To see the importance of the result, let us observe one of its consequences through namely the fact that $\mid\!\sim_{\mathfrak{S},\delta}$ satisfies Cautious Monotonicity. Suppose that α stands for the fact that Sherkan is a big feline with a black-striped, tawny coat. Then it is natural to associate Sherkan to the stereotype of the tiger and then using this information to conclude that Sherkan has also long teeth $(\alpha \mid\!\sim_{\mathfrak{S},\delta} \beta)$ and is a predator $(\alpha \mid\!\sim_{\mathfrak{S},\delta} \gamma)$. Reasonably then, if we add to our premises the information that Sherkan has long teeth, we should continue to consider it to be a tiger and, consequently, a predator $(\alpha \wedge \beta \mid\!\sim_{\mathfrak{S},\delta} \gamma)$.
Intuitive as (L1) may be, Lehmann's model has a significant shortcoming. The problem lies in the requirement that in order for stereotypical reasoning to take place, there should be a nonempty intersection between the factual information at hand and the agent's set of stereotypes. In other words, Lehmann does not take into account the possibility that every stereotype in $\mathfrak{S}$ is inconsistent with the premise α. In such a case, then by the definition of $\mid\!\sim_{\mathfrak{S},\delta}$, we could set the distance between the premise and a stereotype to ∞ (i.e. the largest the distance according to δ):

$$\delta(\alpha, \Delta) = \infty \text{ iff } [\alpha] \cap [\Delta] = \emptyset.$$

So, any choice of stereotypes here is admissible, making stereotypical reasoning basically vacuous (that is, $\mathfrak{S}_\delta^\alpha = \mathfrak{S}$). This shortcoming reduces significantly the scope of Lehmann's model for one key feature of stereotypical reasoning is precisely the fact that an individual can be related to a stereotype even if its properties do not match all the properties of the stereotype so that we can derive defeasible conclusions on the basis of the pieces of stereotypical information compatible with the premises. For example, knowing that Tweety is a penguin, we can reason about it using the information contained in the stereotype of a bird excluding the information that is known to be inconsistent with being a penguin (flying, nesting in trees, etc.).

5 A Semantic Distance for Inconsistent Information.

So we now focus on the situation in which every stereotype available to an agent turns out to be inconsistent with its factual information. More precisely we define a notion of semantic distance, ε, with the idea of capturing the distance between a formula α and a set of α-inconsistent default sets. If there are no α-consistent stereotypes, we allow the choice of the 'nearest' α-inconsistent default sets. This new notion of distance has clearly an effect on the associated consequence relation $\mid\!\sim_{\mathfrak{S},\varepsilon}$: as $\alpha \cup \Delta$ might be inconsistent, we need to move from the classical relation $\models$, used by Lehmann in (L$\mid\!\sim$), to a default-assumption consequence relation $\mid\!\sim_\Delta$. Note that by definition 3 above, if the set $\{\alpha\} \cup \Delta$ is consistent, we have $\alpha \mid\!\sim_\Delta \beta$ if and only if $\{\alpha\} \cup \Delta \models \beta$, as in Lehmann's definition.

We begin by recalling the notion of semantic distance proposed by Lehmann, Magidor and Schlechta [8] in the context of belief revision. We claim that this is appropriate as a measure of 'consistency distance' between formulae. For an arbitrary set Z we say that ε is *a semantic pseudo-distance function*

$$\varepsilon : W \times W \to Z$$

if it satisfies the following:

(ε1) The set Z is totally ordered by a strict order $<$
(ε2) Z has a $<$-smallest element 0, and $\varepsilon(w, v) = 0$ if and only if $w = v$

Note that ε is not required to be symmetric (i.e. $\varepsilon(w, v) = \varepsilon(v, w)$ for every $w, v \in U$). This is matches our intuitive interpretation of distance. Indeed, as we shall shortly see, an agent should have different attitudes towards the information represented by the fist argument of the distance function, that refers to what the agent takes to be certainly true, and the second argument, which concerns default information.

Again, the distance between two given sets of formulae A and B is semantic as it is defined with respect to their models $[A]$ and $[B]$, and the distance between two sets of valuations U and U' ($U, U' \subseteq W$) is set to be the minimal distance between the valuations in U and U':

$$\varepsilon(U, U') = min\{\varepsilon(w, v) | w \in U, v \in U'\}.$$

In analogy with equation ($\sharp$) above, given a finite set $\mathfrak{S}$ of default sets $\{\Delta_1, \dots, \Delta_n\}$ and a formula α, $\mathfrak{S}^\alpha_\varepsilon$ is identified with the set of ε-'nearest' default sets in $\mathfrak{S}$ to α.

From now on we relax Lehmann's assumption that $\mathfrak{S}^\alpha$ must contain a single default set, thus allowing the possibility that, under the uncertainty connected to the presence of inconsistencies, a set of premises is taken to be equally distant from distinct default sets.

A few observations are in order. Note that since ε is a total function, $\mathfrak{S} \neq \emptyset$ implies $\mathfrak{S}^A_\varepsilon \neq \emptyset$. Note also that it makes no difference if we use as arguments ε sets of formulae A or sets of valuations $[A]$. That is, we $\varepsilon(\alpha, \Delta) = \varepsilon([\alpha], [\Delta])$ for every α,Δ. Finally, since ε satisfies (ε2), if a factual formula and a default set are mutually consistent, then the distance between them is 0, as they share at least a valuation. Hence, the default sets which turn out to be consistent with our set of premises have, intuitively, priority over those which are inconsistent.

We can now define $\mid\!\sim_{\mathfrak{S},\varepsilon}$ using default-assumption consequence relations:

$$\alpha \mid\!\sim_{\mathfrak{S},\varepsilon} \beta \text{ iff } \alpha \mid\!\sim_\Delta \beta \text{ for every } \Delta \in \mathfrak{S}^\alpha_\varepsilon. \qquad (*)$$

Note that the corresponding inference operation is

$$C_{\mathfrak{S},\varepsilon}(\alpha) = \bigcap\{C_\Delta(\alpha) | \Delta \in \mathfrak{S}^\alpha_\varepsilon\}$$

where C_Δ is the inference operation corresponding to the default-assumption consequence relation $\mid\!\sim_\Delta$. $\mid\!\sim_{\mathfrak{S},\varepsilon}$ so defined satisfies some properties of cumulative consequence relations.

Lemma 1. *Assume a pseudo-distance ε and a set of stereotypes $\mathfrak{S}$. The consequence relation $\mid\!\sim_{\mathfrak{S},\varepsilon}$ satisfies REF, LLE, RW.*

Proof. Assume a formula α and a set of stereotypes $\mathfrak{S}$. By means of our distance function ε, we can identify the set $\mathfrak{S}^\alpha_\varepsilon$. Since default-assumption consequence relations are cumulative (see section 3), we have that, for every default set Δ, the default-assumption relation $\mid\!\sim_\Delta$ satisfies REF, LLE, and RW. The inference operation $C_{\mathfrak{S},\varepsilon}(\alpha)$ is defined as the intersection of every default-assumption inference operation $C_\Delta(\alpha)$, s.t. $\Delta \in \mathfrak{S}^\alpha_\varepsilon$ ($C_{\mathfrak{S},\varepsilon}(\alpha) = \bigcap\{C_\Delta(\alpha)|\Delta \in \mathfrak{S}^\alpha_\varepsilon\}$) and it is straightforward to prove that REF, LLE, and RW are preserved under intersection.

To see that $\mid\!\sim_{\mathfrak{S},\varepsilon}$ is not cumulative, suppose that we have a set of stereotypes $\mathfrak{S} = \{\Delta, \Delta'\}$, where $\Delta = \{\neg p, p \to r, p \to t\}$ and $\Delta' = \{\neg p, p \wedge r \to \neg t\}$. Since we have $[\Delta] \cap [p] = [\Delta'] \cap [p] = \emptyset$, we have that $\varepsilon(p, \Delta) \neq 0$ and $\varepsilon(p, \Delta') \neq 0$. Without loss of generality let $\varepsilon(p, \Delta) < \varepsilon(p, \Delta')$ and $\varepsilon(p \wedge r, \Delta') < \varepsilon(p \wedge r, \Delta)$. Note that this satisfies both (d1) and (d2).

Now, from these assumptions we get $p \mid\!\sim_{\mathfrak{S},\varepsilon} r$, $p \mid\!\sim_{\mathfrak{S},\varepsilon} t$, since $C_{\mathfrak{S},\varepsilon}(p) = C_\Delta(p)$, but, since $C_{\mathfrak{S},\varepsilon}(p \wedge r) = C_{\Delta'}(p \wedge r)$, we also get $p \wedge r \mid\!\sim_{\mathfrak{S},\varepsilon} \neg t$, violating cautious monotony.

To get cumulativity, we need ε to satisfy a further constrain which intuitively ensures that given a premise α and a default set Δ, there is some valuation satisfying both α and a maximal α-consistent subset of Δ that is at least as near to Δ as any other valuation in $[\alpha]$. To formalize this we first define $[\Delta\bot\alpha]$ as the set of valuations satisfying at least one element of the remainder set $\Delta\bot\alpha$ (see definition 2):

$$[\Delta\bot\alpha] = \bigcup\{[B]|B \in \Delta\bot\alpha\}$$

We can now define the required new constraint

(ε3) For every α and Δ, there is a $w \in [\alpha] \cap [\Delta\bot\neg\alpha]$ s.t. $\varepsilon(w, [\Delta]) \leq \varepsilon(v, [\Delta])$ for every $v \in [\alpha]$.

In order to guarantee that $[\Delta\bot\neg\alpha] \neq \emptyset$ (and hence that $[\alpha] \cap [\Delta\bot\neg\alpha] \neq \emptyset$), we can simply assume that every default set Δ contains a tautology ($\top \in \Delta$, for every Δ). We now prove a series of lemmas leading to the result that if ε satisfies (ε1) – (ε3), then the generated consequence relation $\mid\!\sim_{\mathfrak{S},\varepsilon}$ is cumulative.

Lemma 2. *If $[\alpha] \subseteq [\alpha']$, then $\varepsilon(\alpha', \Delta) \leq \varepsilon(\alpha, \Delta)$ for every Δ.*

Proof. $\varepsilon(\alpha, \Delta) = min\{\varepsilon(w, v)|w \in [\alpha], v \in [\Delta]\}$. Since $w \in [\alpha]$ implies $w \in [\alpha']$, we have that $min\{\varepsilon(w, v)|w \in [\alpha'], v \in [\Delta]\} \leq min\{\varepsilon(w, v)|w \in [\alpha], v \in [\Delta]\}$, i.e. $\varepsilon(\alpha', \Delta) \leq d(\alpha, \Delta)$.

We now want to prove that if we add to the factual information information which is itself derivable by means of $\mid\!\sim_{\mathfrak{S},\varepsilon}$, then we continue to associate the same stereotypes to our premise (see the example about the tiger Sherkan in section 4).

Lemma 3. *If* $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$ *and* $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$*, then* $\varepsilon(\alpha \wedge \beta, \Delta) = \varepsilon(\alpha, \Delta)$*.*

Proof. Recall that $\mathfrak{S}^{\alpha}_{\varepsilon} = \{\Delta \in \mathfrak{S} | \ \varepsilon(\alpha, \Delta) \leq \varepsilon(\alpha, \Delta') \text{ for every } \Delta' \in \mathfrak{S}\}$.
By ($\varepsilon 3$), we have that $\varepsilon([\alpha], [\Delta]) = \varepsilon(w, [\Delta])$ for some $w \in [\alpha] \cap [\Delta \bot \neg\alpha]$.
$\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$ implies that $\alpha \mid\sim_{\Delta} \beta$ for every $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$, which implies that if $w \in [\alpha] \cap [\Delta \bot \neg\alpha]$, then $w \in [\alpha \wedge \beta]$.

Given that $\varepsilon([\alpha], [\Delta]) = \varepsilon(w, [\Delta])$, we have that $\varepsilon(w, [\Delta]) \leq \varepsilon(v, [\Delta])$ for every $v \in [\alpha]$. Since $[\alpha \wedge \beta] \subseteq [\alpha]$, we have that $\varepsilon(w, [\Delta]) \leq \varepsilon(v, [\Delta])$ for every $v \in [\alpha \wedge \beta]$, that is, $\varepsilon(\alpha \wedge \beta, \Delta) = \varepsilon(w, [\Delta]) = \varepsilon(\alpha, \Delta)$.

Lemma 4. *If* $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$*, then* $\mathfrak{S}^{\alpha}_{\varepsilon} = \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$*.*

Proof. Assume $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$. We show that $\mathfrak{S}^{\alpha}_{\varepsilon} \subseteq \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$. If $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$, then $\varepsilon(\alpha, \Delta) \leq \varepsilon(\alpha, \Delta')$ for every $\Delta' \in \mathfrak{S}$. By Lemma 2, since $[\alpha \wedge \beta] \subseteq [\alpha]$, we have that $\varepsilon(\alpha, \Delta) \leq \varepsilon(\alpha \wedge \beta, \Delta')$ for every $\Delta' \in \mathfrak{S}$. Since $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$, by Lemma 3, we obtain $\varepsilon(\alpha \wedge \beta, \Delta) \leq \varepsilon(\alpha \wedge \beta, \Delta')$ for every $\Delta' \in \mathfrak{S}$, i.e. $\Delta \in \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$.

For $\mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon} \subseteq \mathfrak{S}^{\alpha}_{\varepsilon}$, if $\Delta \notin \mathfrak{S}^{\alpha}_{\varepsilon}$, then $\varepsilon(\alpha, \Delta') < \varepsilon(\alpha, \Delta)$ for some $\Delta' \in \mathfrak{S}^{\alpha}_{\varepsilon}$. By Lemma 2, we have that $\varepsilon(\alpha, \Delta') < \varepsilon(\alpha \wedge \beta, \Delta)$. Since $\Delta' \in \mathfrak{S}^{\alpha}_{\varepsilon}$, by Lemma 3, we obtain $\varepsilon(\alpha \wedge \beta, \Delta') < \varepsilon(\alpha \wedge \beta, \Delta)$, i.e. $\Delta \notin \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$.

We are now ready to prove the key result about our notion of distance ε.

Theorem 2. *Given a set of stereotypes* $\mathfrak{S}$ *and a distance function* ε *satisfying (*$\varepsilon 1$*)-(*$\varepsilon 3$*), the generated consequence relation* $\mid\sim_{\mathfrak{S},\varepsilon}$ *is cumulative.*

Proof. We have to show that $\mid\sim_{\mathfrak{S},\varepsilon}$ satisfies CM and CT.
CM: assume $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$ and $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \gamma$, which correspond to saying that $\alpha \mid\sim_{\Delta} \beta$ and $\alpha \mid\sim_{\Delta} \gamma$ for every $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$. Since every default-assumption consequence relation $\mid\sim_{\Delta}$, being cumulative (see section 3), satisfies CM, we have $\alpha \wedge \beta \mid\sim_{\Delta} \gamma$ for every $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$. Given $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$, we have, by Lemma 4, that $\mathfrak{S}^{\alpha}_{\varepsilon} = \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$, which implies that $\alpha \wedge \beta \mid\sim_{\Delta} \gamma$ for every $\Delta \in \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$, i.e. $\alpha \wedge \beta \mid\sim_{\mathfrak{S},\varepsilon} \gamma$.

CT: assume $\alpha \wedge \beta \mid\sim_{\mathfrak{S},\varepsilon} \gamma$ and $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$. Note again that $\alpha \wedge \beta \mid\sim_{\mathfrak{S},\varepsilon} \gamma$ means that $\alpha \wedge \beta \mid\sim_{\Delta} \gamma$ for every $\Delta \in \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$. $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \beta$ implies, again by Lemma 4, that $\mathfrak{S}^{\alpha}_{\varepsilon} = \mathfrak{S}^{\alpha \wedge \beta}_{\varepsilon}$. Hence, we have that $\alpha \wedge \beta \mid\sim_{\Delta} \gamma$ and $\alpha \mid\sim_{\Delta} \beta$ for every $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$. Since every such $\mid\sim_{\Delta}$, being cumulative, satisfies CT, we have $\alpha \mid\sim_{\Delta} \gamma$ for every $\Delta \in \mathfrak{S}^{\alpha}_{\varepsilon}$, i.e. $\alpha \mid\sim_{\mathfrak{S},\varepsilon} \gamma$.

Thus our notion of distance captures the stereotypical reasoning underlying Lehmann's approach while preserving the cumulativity of the generated consequence relation in the general case in which an agent's factual information comes out to be inconsistent with its stereotypical information. However, this revised distance function looses its appeal if more than one stereotype is consistent with an agent's factual information. In such a case, by ($\varepsilon 2$), an agent cannot distinguish between the stereotypes in $\mathfrak{S}$ that are consistent with α, since their mutual distance is always 0.

6 A Lexicographic Combination of δ and ε

Summing up what has been done so far, we started by reviewing Lehmann's notion of distance δ and noted that while logically well-behaved (generates cumulative consequence relations) it suffers from the drawback of not handling inconsistency between factual and default information. In order to overcome this limitation we considered a semantic pseudo-distance ε, which again leads to cumulative reasoning, but which, at the same time, allows an agent to face the situation in which its factual information turns out to be inconsistent with its default information. The purpose of the remainder of this paper is to study a combination of the two approaches which enables us to refine pseudo-distance ε in order to let this latter distinguish between the stereotypes consistent with α. To do this we define a lexicographic ordering of a distance d_{lex}, with the idea that the precedence should be given whenever possible to ε over δ. More precisely:

$$d_{lex}(\alpha, \Delta) \leq d_{lex}(\alpha', \Delta') \Leftrightarrow \begin{cases} \varepsilon(\alpha, \Delta) < \varepsilon(\alpha', \Delta') \\ or \\ \varepsilon(\alpha, \Delta) = \varepsilon(\alpha', \Delta') \text{ and } \delta(\alpha, \Delta) \leq \delta(\alpha', \Delta') \end{cases}$$

Given a set of stereotypes $\mathfrak{S}$, a semantic distance d_{lex} and a formula α, we define, again in analogy with equation ($\sharp$), the set $\mathfrak{S}^\alpha_\varepsilon$ of the stereotypes in $\mathfrak{S}$ which are nearest to α. $\mid\!\sim_{\mathfrak{S},d_{lex}}$ is defined analogously to $\mid\!\sim_{\mathfrak{S},\varepsilon}$, using default-assumption consequence relations as in equation (*). We devote the rest of this paper to show that d_{lex} does indeed combine the best of ε and δ since it eventually leads to cumulative reasoning.

Recall that, given a set of stereotypes $\mathfrak{S}$, a semantic distance d_{lex}, defined lexicographically over two distances ε and δ, and a formula α, we indicate by $\mathfrak{S}^\alpha_{d_{lex}}$ the set of the nearest stereotypes to α with respect to d_{lex}, by $\mathfrak{S}^\alpha_\varepsilon$ the nearest stereotypes with respect to ε, and by $\mathfrak{S}^\alpha_\delta$ the nearest stereotypes with respect to δ. As we have seen, they define, respectively, three consequence relations: $\mid\!\sim_{\mathfrak{S},d_{lex}}$, $\mid\!\sim_{\mathfrak{S},\varepsilon}$, and $\mid\!\sim_{\mathfrak{S},\delta}$ (and the correspondent inference operations $C_{\mathfrak{S},d_{lex}}$, $C_{\mathfrak{S},\varepsilon}$, and $C_{\mathfrak{S},\delta}$). Note that by the lexicographic definition of the d_{lex}-ordering, we have:

$$\mathfrak{S}^\alpha_{d_{lex}} = \begin{cases} \mathfrak{S}^\alpha_\varepsilon \text{ if } |\mathfrak{S}^\alpha_\varepsilon| = 1 \\ \\ (\mathfrak{S}^\alpha_\varepsilon)^\alpha_\delta \text{ if } |\mathfrak{S}^\alpha_\varepsilon| > 1 \end{cases}$$

where $(\mathfrak{S}^\alpha_\varepsilon)^\alpha_\delta$ is the composition of the selection functions of the stereotypes in $\mathfrak{S}$ that ε and δ associate to α: we first select the subset of $\mathfrak{S}$ nearest to α with respect to ε (that is, $\mathfrak{S}^\alpha_\varepsilon$), and, if using ε we have not been able to distinguish between distinct stereotypes ($|\mathfrak{S}^\alpha_\varepsilon| > 1$), we refine our procedure by selecting the δ-nearest stereotypes to α between those in $\mathfrak{S}^\alpha_\varepsilon$ (that is, $(\mathfrak{S}^\alpha_\varepsilon)^\alpha_\delta$).

Lemma 5. *If* $\alpha \mid\!\sim_{\mathfrak{S},d_{lex}} \beta$, *then* $\mathfrak{S}^\alpha_{d_{lex}} = \mathfrak{S}^{\alpha\wedge\beta}_{d_{lex}}$

Proof. We have three possibilities.

(1) $|\mathfrak{S}_{\varepsilon}^{\alpha}| = 1$.
(2) $|\mathfrak{S}_{\varepsilon}^{\alpha}| > 1$ and $\varepsilon(\alpha, \Delta) > 0$ for every $\Delta \in \mathfrak{S}_{\varepsilon}^{\alpha}$.
(3) $|\mathfrak{S}_{\varepsilon}^{\alpha}| > 1$ and $\varepsilon(\alpha, \Delta) = 0$ for every $\Delta \in \mathfrak{S}_{\varepsilon}^{\alpha}$.

(1): $|\mathfrak{S}_{\varepsilon}^{\alpha}| = 1$ implies that $\mathfrak{S}_{d_{lex}}^{\alpha} = \mathfrak{S}_{\varepsilon}^{\alpha}$ and $C_{\mathfrak{S},d_{lex}}(\alpha) = C_{\mathfrak{S},\varepsilon}(\alpha)$,that is, $\alpha \mathrel{|\!\sim}_{\mathfrak{S},d_{lex}} \beta$ iff $\alpha \mathrel{|\!\sim}_{\mathfrak{S},\varepsilon} \beta$. By Lemma 4, from $\alpha \mathrel{|\!\sim}_{\mathfrak{S},\varepsilon} \beta$ we obtain $\mathfrak{S}_{\varepsilon}^{\alpha} = \mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta}$, that implies $|\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta}| = 1$ and $\mathfrak{S}_{d_{lex}}^{\alpha\wedge\beta} = \mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta}$, that is, $\mathfrak{S}_{d_{lex}}^{\alpha\wedge\beta} = \mathfrak{S}_{d_{lex}}^{\alpha}$.

(2): $\varepsilon(\alpha, \Delta) > 0$ for every $\Delta \in \mathfrak{S}_{\varepsilon}^{\alpha}$ implies that the default sets in $\mathfrak{S}_{\varepsilon}^{\alpha}$ are not consistent with the premise α. Therefore, δ cannot distinguish them out and we have $(\mathfrak{S}_{\varepsilon}^{\alpha})_{\delta}^{\alpha} = \mathfrak{S}_{\varepsilon}^{\alpha}$. Again we have that $\mathfrak{S}_{d_{lex}}^{\alpha} = \mathfrak{S}_{\varepsilon}^{\alpha}$ and $C_{\mathfrak{S},d_{lex}}(\alpha) = C_{\mathfrak{S},\varepsilon}(\alpha)$, and the case is already covered by (1).

(3): $\varepsilon(\alpha, \Delta) = 0$ for every $\Delta \in \mathfrak{S}_{\varepsilon}^{\alpha}$ implies that the stereotypes in $\mathfrak{S}_{\varepsilon}^{\alpha}$ are consistent with α, and we can refine the choice by means of δ.
Since $\alpha \mathrel{|\!\sim}_{\mathfrak{S},d_{lex}} \beta$, we have that some default sets in $\mathfrak{S}_{\varepsilon}^{\alpha}$ are consistent with $\alpha\wedge\beta$ (surely the one in $(\mathfrak{S}_{\varepsilon}^{\alpha})_{\delta}^{\alpha}$). Since $\mathfrak{S}_{\varepsilon}^{\alpha}$ is composed by every set in $\mathfrak{S}$ consistent with α, every default set consistent with $\alpha\wedge\beta$ is in $\mathfrak{S}_{\varepsilon}^{\alpha}$. Hence, we have $(\mathfrak{S}_{\varepsilon}^{\alpha})_{\delta}^{\alpha} \subseteq \mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta} \subseteq \mathfrak{S}_{\varepsilon}^{\alpha}$, that is,

$$\mathfrak{S}_{d_{lex}}^{\alpha} \subseteq \mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta} \subseteq \mathfrak{S}_{\varepsilon}^{\alpha}$$

Since every element in $(\mathfrak{S}_{\varepsilon}^{\alpha})_{\delta}^{\alpha}$ is in $\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta}$, we have that $(\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta})_{\delta}^{\alpha} = (\mathfrak{S}_{\varepsilon}^{\alpha})_{\delta}^{\alpha} = \mathfrak{S}_{d_{lex}}^{\alpha}$.

Now take theorem 1, that is,

if $([\alpha] \cap [\Delta]) \subseteq [\alpha'] \subseteq [\alpha]$, then $\mathfrak{S}_{\delta}^{\alpha'} = \mathfrak{S}_{\delta}^{\alpha}$, where $\mathfrak{S}_{\delta}^{\alpha} = \{\Delta\}$.

Let α' be $\alpha\wedge\beta$ and $\mathfrak{S}$ be $\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta}$, and, consequently, let $(\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta})_{\delta}^{\alpha} = \{\Delta\}$. Given that $\alpha \mathrel{|\!\sim}_{\mathfrak{S},d_{lex}} \beta$ and $\mathfrak{S}_{d_{lex}}^{\alpha} = (\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta})_{\delta}^{\alpha}$, we have that $([\alpha] \cap [\Delta]) \subseteq [\alpha \wedge \beta] \subseteq [\alpha]$, and this, by theorem 1, implies $(\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta})_{\delta}^{\alpha\wedge\beta} = (\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta})_{\delta}^{\alpha}$.
Combining the equations, we have $(\mathfrak{S}_{\varepsilon}^{\alpha})_{\delta}^{\alpha} = (\mathfrak{S}_{\varepsilon}^{\alpha\wedge\beta})_{\delta}^{\alpha\wedge\beta}$, that is

$$\mathfrak{S}_{d_{lex}}^{\alpha} = \mathfrak{S}_{d_{lex}}^{\alpha\wedge\beta}$$

as desired.

We now have all the ingredients to prove our central result.

Theorem 3. *Given a set of stereotypes $\mathfrak{S}$ and a distance function d_{lex}, the consequence relation $\mathrel{|\!\sim}_{\mathfrak{S},d_{lex}}$ is cumulative.*

Proof. Since $C_{\mathfrak{S},d_{lex}}(\alpha)$ is obtained by the intersection of default-assumption inference operations, it satisfies REF, LLE, RW (see Lemma 1).
Cumulativity then follows by Lemma 5 with exactly the same procedure argument used in the proof of Theorem 2.

7 Conclusions

We have addressed the problem of providing a logical characterization of reasoning based on stereotypes and we presented a model which combines two basic intuitions. On the one hand, stereotypical reasoning requires an agent to choose, given a piece of factual information, how this can be extended by relying on some background information about its class. This puts the agent in a new epistemic state (usually richer than the original one) which can be used to reason non-monotonically. Our central result shows that if we put appropriate constraints on the selection of stereotypes – in our case by using appropriate distance functions – we can generate a cumulative non monotonic consequence relation which is widely regarded in the field as capturing some fundamental aspects of commonsensical reasoning.

References

1. Freund, M.: Preferential Reasoning in the Perspective of Poole Default Logic. Artificial Intelligence 98(1-2), 209–235 (1998)
2. Gabbay, D., Woods, J. (eds.): Handbook of the History of Logic, vol. 8. North Holland, Amsterdam (2007)
3. Hansson, S.O.: A Textbook of Belief Dynamics. Kluwer, Dordrecht (1999)
4. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logic. Artificial Intelligence 44, 167–207 (1990)
5. Lakoff, G.: Cognitive Models and Prototype Theory. In: Laurence, S., Margolis, E. (eds.) Concepts, pp. 391–431. MIT Press, Cambridge (1999)
6. Laurence, S., Margolis, E.: Concepts and Cognitive Science. In: Laurence, S., Margolis, E. (eds.) Concepts, pp. 3–81. MIT Press, Cambridge (1999)
7. Lehmann, D.: Stereotypical Reasoning: Logical Properties. L. J. of the IGPL 6(1), 49–58 (1998)
8. Lehmann, D., Magidor, M., Schlechta, K.: Distance Semantics for Belief Revision. The Journal of Symbolic Logic 66(1), 295–317 (2001)
9. Makinson, D.: General Patterns in Nonmonotonic Reasoning. In: Gabbay, D., Hogger, C., Robinson, J. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3, pp. 35–110. Clarendon Press, Oxford (1994)
10. Makinson, D.: Bridges from Classical to Nonmonotonic Logic. King's College Publications, London (2005)
11. Putnam, H.: Representation and Reality. MIT Press, Cambridge (1988)