Towards a logical framework for reasoning with stereotypes

Florence Dupin de Saint-Cyr, Francis Faux, Sabine Frittella

Towards a logical framework for reasoning with stereotypes*

Florence Dupin de Saint-Cyr[1] Francis Faux[1] Sabine Frittella[2]
[1] IRIT, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France
[2] INSA Centre Val de Loire, Univ. Orléans, LIFO EA 4022, France

October 10, 2024

Abstract

Stereotypes are necessary for human cognition. Indeed our limited computational capabilities and our need for quick decision making require using shortcuts for reasoning. In this work, we discuss how to formalize reasoning with stereotypes using uncertain default rules with an anchorage degree.

Keywords: Uncertainty, Stereotypes, Belief revision.

1 Introduction

Stereotypes are a widespread form of prejudice. A prejudice can be defined as "an a priori favourable or unfavourable opinion, adopted without examination, imposed by an environment, an education" (Montaigne, Essais, II, 12, ed. P. Villey and V.-L. Saulnier, p.506). Prejudices have the characteristic of being more or less entrenched: the more entrenched they are, the more difficult it is to refute them and the more they influence reasoning.[1] We can consider *dogmatism*, as the "extreme" prejudice, because it refers to closed mind, rigidity, inflexible system of thought not open to new information, new experiences or new environments [11].

Stereotypes are defined as "characterisations of social categories by which membership of a group is associated with the possession of certain attributes" (for example, scientists are intellectuals, Scots are greedy, men like the colour blue). Consequently, when faced with new information, prejudices can be seen as filters influenced by their polarities. In this paper, we consider that stereotypes are anchored beliefs of the form "if A, then generally B", some of them are objectively mostly true like "birds fly" and

*This is a draft version, this short paper is published in the proceedings of the 16th International Conference on Scalable Uncertainty Management (SUM 2024). This research is part of the MOSAIC project financed by the European Union's Marie Skłodowska-Curie grant No. 101007627.

[1] In his book *The Nature of Prejudice*, the American psychologist Gordon [1] argues that "prejudice is essentially derived from the necessary mental shortcuts that the human brain uses to process the large amount of information it receives".

some are not, like "French are rude". We use default rules enriched with a degree of certainty and anchorage to encode prejudices.

Within the framework of evidence theory, which is well suited to modeling uncertain or incomplete testimonies from different sources, the notion of prejudice was introduced and formalized by a state of belief [8, 5]: prejudices can evolve, and their effect consists in weakening or fragmenting some focal sets of incoming information represented by a belief function. In this article, we choose to study stereotypes within the logical framework of uncertain default rules [9] based on possibility theory. It provides a natural way to encode a potentially unreliable, but still used, deduction rule without mentioning all its exceptions. Default rules are compact ways to express generic laws that are easy to apply even in the case of incomplete information, which is similar to what human beings expect from stereotypes: compactness means easy to memorize and easy to transmit to other people, genericity means easy to apply. In our approach, a stereotype is considered to be an information that is more or less reliable and more or less anchored, i.e., the prejudice is more or less likely to disappear or diminish. This is why we propose to represent stereotypes using two parameters, one for reliability and one for anchorage. The objective is to be able to simulate the behaviour of an agent with prejudices on the one hand and more or less certain information on the other.

In section 2, we provide notations and present the notion of *belief base* used to encode the knowledge and stereotypes of an agent. We end this part by recalling a method proposed in [9] for transforming uncertain default rules into uncertain strict rules. In section 3, we discuss the evolution of an agent's belief base given a new piece of information. The paper concludes with a look ahead to future work.

2 Definitions

2.1 Preliminaries

Let $\mathcal{X}, \mathcal{C}, \mathcal{P}$ be respectively sets of variables, constants, and predicates. Let $\mathcal{L}$ be the first-order language s.t.:
$\mathcal{T} \ni t := x \mid a$,
$\mathcal{L} \ni \phi := P(t_1, \ldots, t_n) \mid \neg\phi \mid \phi \wedge \phi \mid \forall x, \phi$
with $x \in \mathcal{X}$, $a \in \mathcal{C}$, $P \in \mathcal{P}$ a predicate of arity n and $t_1, \ldots, t_n \in \mathcal{T}$. The other standard connectives ($\bot, \top, \vee, \rightarrow, \leftrightarrow, \exists$) are defined as usual.

A *structure* for the language $\mathcal{L}$ is a tuple $\mathbb{S} = (\mathcal{D}, \mathcal{I}_\mathcal{C}, \mathcal{I}_\mathcal{P})$ such that $\mathcal{D}$ is a non-empty *domain*, $\mathcal{I}_\mathcal{C}$ is the *interpretation of constant symbols*, i.e., for every $a \in \mathcal{C}$, $\mathcal{I}_\mathcal{C}(c) \in \mathcal{D}$, $\mathcal{I}_\mathcal{P}$ is the *interpretation of predicate symbols*, i.e., for every $P \in \mathcal{P}$ of arity n, $\mathcal{I}_\mathcal{C}(P) \subseteq \mathcal{D}^n$. A *model* for the language $\mathcal{L}$ is a tuple $\mathbb{M} = (\mathbb{S}, \mathcal{I}_\mathcal{X})$ such that $\mathbb{S}$ is a structure for the language $\mathcal{L}$ and $\mathcal{I}_\mathcal{X} : \mathcal{X} \rightarrow \mathcal{D}$ is the *interpretation of variables*. The interpretation of formulas over a model $\mathbb{M}$ is defined in the usual manner. A *structure* $\mathbb{S}$ *satisfies a formula*, noted $\mathbb{S} \models \phi$, if $(\mathbb{S}, \mathcal{I}_\mathcal{X}) \models \phi$ for every interpretation $\mathcal{I}_\mathcal{X}$. In this article, we assume that the domain $\mathcal{D}$ is fixed and finite.

Let $\mathcal{L}_x$ be the restriction of $\mathcal{L}$ without the use of quantifiers and with one single variable x. Let $\mathcal{L}_\mathrm{C}$ be the set of closed formulas in $\mathcal{L}$. We will only consider models with finite domains, therefore every formula can be translated into propositional logic.

We use an additional symbol $\rightsquigarrow$ that connects two formulas ϕ and ψ from $\mathcal{L}_x$. Formulas of the form $\phi \rightsquigarrow \psi$ will be used to encode stereotypes via so-called *belief rules*. We will reason with those rules following the non-monotonic reasoning approaches of default rules [2].

Possibility theory enables us to associate two measures Π and N with a formula φ [7]. $\Pi(\varphi)$, called the *possibility* of φ, quantifies *how unsurprising φ is* ($\Pi(\varphi) = 0$ means that φ is necessarily false), and N, called the *necessity*, is the dual of Π defined by $N(\varphi) = 1 - \Pi(\neg\varphi)$ ($N(\varphi) = 1$ means that φ is necessarily true). In possibilistic logic, the resolution rule [6] is written $(a \vee b, \rho_1); (\neg a \vee c, \rho_2) \vdash_\pi (b \vee c, \min(\rho_1, \rho_2))$. A *possibilistic propositional belief base* is a set $B_\pi = \{(\varphi_i, \rho_i)\}_{i\in[1,m]}$ where each propositional formula φ_i is associated with a weight $\rho_i \in]0, 1]$ representing its certainty level and such that $N(\varphi) \geq \rho_i$. To compute the maximal certainty level to attach to a formula φ w.r.t constraints expressed in B_π, the user can add to B_π the clauses corresponding to the refutation of φ with a necessity level of 1 and then deduce $(\bot, \rho)$ with the possibilistic resolution rule $\vdash_\pi$. The weight ρ is a lower bound of the necessity of φ. In the following, we use $sk(B_\pi)$ to denote the skeleton of B_π, it is the set of formulas of B_π without their weights. Let $B_{\pi\rho} = \{\varphi_i \in \mathcal{L} \mid (\varphi_i, \rho_i) \in B_\pi$ and $\rho_i \geq \rho\}$ denote the "ρ-cut of the possibilistic base B_π", i.e., formulas of B_π (without their weights) that have a certainty level higher or equal to ρ. Now, $\text{Inc}(B_\pi) = \max\{\rho \in [0, 1] \mid B_{\pi\rho} \vdash \bot\}$ is called the inconsistency level of the possibilistic base B_π. When $sk(B_\pi)$ is inconsistent (then $\text{Inc}(B_\pi) > 0$), a refutation using only formulas strictly above the inconsistency level yields non-trivial conclusions (i.e., with a certainty strictly above the inconsistency level). We use $B_\pi \vdash_{>\text{Inc}} \varphi$ to denote a non-trivial inference, i.e., the existence of a value $\rho' \in [0, 1]$ and a refutation from $B_{\pi>\text{Inc}} = \{(\varphi_i, \rho_i) \in B_\pi \mid \rho_i > \text{Inc}(B_\pi)\}$ with $(\neg\varphi, 1)$ that leads to $(\bot, \rho')$. In the following, $\vdash_\pi$ is used to represent syntactic inference between possibilistic propositional formulas.

2.2 Stereotypes

The following definition proposes a formal structure, called a *belief base*, to encode both the *beliefs* and the *stereotypes* an agent uses to reason. Here the term *beliefs* refers to statements the agent believes to some degree of certainty. We choose to encode stereotypes by uncertain default rules [9] extended with an anchorage factor, below named *belief rules*.

Definition 1 (Belief Base) *A* belief base *is a tuple $(\mathcal{B}, \mathcal{R})$ such that*

- *$\mathcal{B}$ is a finite* belief set*: $\{(\chi_j, \gamma_j) \mid \chi_j \in \mathcal{L}_\text{C}$ and $\gamma_j \in]0, 1]\}_{j\in[\![1,k]\!]}$*
- *$\mathcal{R}$ is a set of* belief rules *$\{(\varphi_i \rightsquigarrow \psi_i, \rho_i, \alpha_i)\}_{i\in[\![1,n]\!]}$, where*
  - *$\varphi_i, \psi_i \in \mathcal{L}_x$, they share the same free variable x, $\varphi_i \rightsquigarrow \psi_i$ is called a* default rule*,*
  - *$\rho_i \in]0, 1]$ encodes the* reliability *of the rule, i.e. the certainty of the conclusion when the rule is applied (in a non-exceptional context),*

– $\alpha_i \in [0,1]$ *encodes the* anchorage *of the rule, i.e. how willing the agent is to update the reliability of the rule.*[2]

Example 1 (All birds fly) *Consider a child who has only seen birds that fly and therefore is convinced that all birds fly, but who is open to learning that ostriches and penguins are birds that do not fly. Then the stereotype (before the arrival of the evidence about ostriches and penguins) could be encoded as follows* $(\texttt{Bird}(x) \rightsquigarrow \texttt{Fly}(x), 1, 0)$.

Example 2 (French people are rude) *Consider an American moving to France who has heard that French people are rude and holds a strong belief in that statement while being open to being convinced otherwise if given enough evidence. Then the stereotype could be encoded as follows* $(\texttt{French}(x) \rightsquigarrow \texttt{Rude}(x), 0.7, 0.4)$.

Example 3 *An example of dogmatism is illustrated on "Flat Earth Society" group which currently has 200,000 members. This community believes that the Earth is a flat disk, surrounded by an enormous wall of ice (Antarctica) which would prevent us from falling when we reach the end. The flatists are convinced that we are trying to impose fake belief on them. Then the stereotype could be encoded as follows* $(\neg\texttt{FlatEarther}(x) \rightsquigarrow \texttt{Wrong}(x), 1, 1)$. *This belief rule is characterized by maximum reliability and anchorage what is specific to dogmatism.*

2.3 Semantics of belief bases

Translation of belief bases into classical propositional logic. Let $(\mathcal{B}, \mathcal{R})$ be a belief base. Assume that $\mathcal{L}$ contains a constant symbol a_d for each $d \in \mathcal{D}$. Therefore, given a structure $\mathbb{S} = (\mathcal{D}, \mathcal{I}_\mathcal{C}, \mathcal{I}_\mathcal{P})$ over $\mathcal{L}$ such that $\mathcal{I}_\mathcal{C}(a_d) = d$, every closed formula of the knowledge base can be equivalently encoded into a formula of classical propositional logic as follows. Let the set $\texttt{Prop} := \{z_{P(a_1,\ldots,a_n)} \mid P \in \mathcal{P}$ of arity n and $a_1, \ldots, a_n \in \mathcal{C}\}$ be a finite set of propositional variables on which is defined a propositional language denoted $\mathcal{L}_{\mathrm{CL}}$. Let us define the translation $\tau : \mathcal{L}_{\mathrm{C}} \to \mathcal{L}_{\mathrm{CL}}$ as follows: $\tau(\forall x.\phi) = \bigwedge_{d\in\mathcal{D}} \tau(\phi[a_d/x])$, $\tau(\neg\phi) = \neg\tau(\phi)$, $\tau(\phi \wedge \psi) = \tau(\phi) \wedge \tau(\psi)$ and $\tau(P(a_1, \ldots, a_n)) = z_{P(a_1,\ldots,a_n)}$, where $\phi[a_d/x]$ is the formula ϕ in which every occurrence of x has been replaced by a_d. Since we only consider closed formulas, $\tau(P(a_1, \ldots, a_n))$ is only applied in situations where $a_i \in \mathcal{C}$. The valuation $v_\mathbb{S} : \texttt{Prop} \to \{0, 1\}$ is defined as follows: $v_\mathbb{S}(z_{P(a_1,\ldots,a_n)}) = 1$ iff $\mathbb{S} \models P(a_1, \ldots, a_n)$. $v_\mathbb{S}$ is extended to every propositional formula in the standard way. We have $\mathbb{S} \models \phi$ iff $v_\mathbb{S}(\tau(\phi)) = 1$ for every $\phi \in \mathcal{L}_{\mathrm{C}}$. Belief rules can be translated in a similar way in classical propositional logic rules, by considering that the rule $\varphi_i \rightsquigarrow \psi_i$ encodes the statement $\forall x, \varphi_i \rightsquigarrow \psi_i$.

Semantics of belief bases. Let Ω denote the set of valuations associated with the propositional language $\mathcal{L}_{\mathrm{CL}}$. Let $\omega, \omega_i, \ldots$ denote the elements of Ω, and $[\omega]$ be the set of formulas of $\mathcal{L}_{\mathrm{CL}}$ satisfied by ω.

[2] If $\alpha = 0$, then the agent reasons very scientifically about the rule and will agree to update the reliability ρ of the rule when given new trusted evidence. If $\alpha = 1$, then the agent's opinion about the reliability of the rule cannot be changed.

We interpret a set of default rules Δ in terms of the *Lexicographic-ordering* $\succ_\Delta$ (read "is more plausible than") on valuations as defined in [3]. Given a set of default rules of the form $\alpha \rightsquigarrow \beta$, it is possible to compute a stratification of this set according to their specificity: here a rule has more (or less) specificity when it describes a more (or less) exceptional case, the rules that have no exception being in the most specific stratum. Note that a set of default rules may not admit a stratification. We assume that this is not the case here, i.e., that Δ can always be stratified which is called *consistent* by [10]. Pearl in [10] provides an algorithm named System Z[3] to compute automatically the strata, i.e., the subsets of formulas with the same rank. As shown in [4], the same ordering on valuations can be obtained by interpreting each default rule $\varphi_i \rightsquigarrow \psi_i$ by a constraint $\Pi(\varphi_i \wedge \psi_i) > \Pi(\varphi_i \wedge \neg\psi_i)$ on a possibility measure Π.

Definition 2 (Lexicographic ordering $\succ_\Delta$ on Ω) *Let $\Delta = \Delta_1 \cup \cdots \Delta_n$ be a stratified default base with n strata ordered from the most specific stratum Δ_1 to the least one Δ_n, let α, β be two formulas of $\mathcal{L}_{\mathrm{CL}}$, let $\omega, \omega' \in \Omega$,*

- *Notations: str (for "strict") is a function that translates a set of default rules into a set of formulas of $\mathcal{L}_{CL}$ as follows $str(E) = \bigcup_{\varphi_i \rightsquigarrow \psi_i \in E} \{\neg\varphi_i \vee \psi_i\}$.*
- *$\omega \succ_\Delta \omega'$ iff there exists $k \in [1, n]$ s.t.* $\begin{cases} |str(\Delta_k) \cap [\omega]| > |str(\Delta_k) \cap [\omega']| \text{ and} \\ \forall i < k, |str(\Delta_i) \cap [\omega]| = |str(\Delta_i) \cap [\omega']| \end{cases}$

The last item of Definition 2 explains the conditions for $\omega \succ_\Delta \omega'$: we compare by lexicographic order the tuples obtained by computing the number of formulas satisfied by each valuation (ω and ω') in each stratum.

Definition 3 (Models associated to a consistent belief base) *When $sk(\mathcal{B})$ is* consistent*, the set of models $\mathbb{M}$ associated with the belief base $(\mathcal{B}, \mathcal{R})$ (in which Δ is the set of rules of $\mathcal{R}$ without considering their reliability and anchorage) is $\mathbb{M} = \{v \mid v$ is a valuation such that $v(\phi) = 1$ for all $\phi \in sk(\mathcal{B})\}$. Moreover, the rules enable us to compare any valuation thanks to $\succ_\Delta$ defined in Definition 2.*

Example 4 *We consider the first-order language $\mathcal{L}$ such that the predicates (indexed with their arity) are $\mathcal{P} = \{\mathtt{Bird}_1, \mathtt{Penguin}_1, \mathtt{Fly}_1, \mathtt{French}_1, \mathtt{Rude}_1\}$ and with no constants and the following belief base $(\mathcal{B}, \mathcal{R})$:*
$\mathcal{B} = \{(\forall x.(\mathtt{Penguin}(x) \rightarrow \mathtt{Bird}(x)), 0.8); (\forall x.(\mathtt{Bird}(x) \rightarrow \neg\mathtt{French}(x)), 1)\}$,
$\mathcal{R} = \{(\mathtt{Bird}(x) \rightsquigarrow \mathtt{Fly}(x), 0.9, 0), (\mathtt{Penguin}(x) \rightsquigarrow \neg\mathtt{Fly}(x), 0.99, 0), (\mathtt{French}(x) \rightsquigarrow \mathtt{Rude}(x), 0.7, 0.5)\}$.

First, we fix a finite domain of interpretation $\mathcal{D}$, then we translate $\mathcal{B}$ into $\mathcal{L}_{\mathrm{CL}}$ over the variables $\mathtt{Prop} = \{\mathtt{Bird}(d), \mathtt{Penguin}(d), \mathtt{Fly}(d), \mathtt{French}(d), \mathtt{Rude}(d) \mid d \in \mathcal{D}\}$. From this, we build the set $\mathbb{M}$ of valuations satisfying the formulas of the belief base $\mathcal{B}$. Now we translate $\mathcal{R}$ (we drop the anchoring for the moment) into $\mathcal{L}_{\mathrm{CL}}$. First, notice that $\mathtt{Bird}(x) \rightsquigarrow \mathtt{Fly}(x)$ is less specific than $\mathtt{Penguin}(x) \rightsquigarrow \neg\mathtt{Fly}(x)$ since $\forall x, (\mathtt{Penguin}(x) \rightarrow \mathtt{Bird}(x))$. We get $\mathcal{R}_{\mathrm{CL}} = \Delta_1 \cup \Delta_2$ with
$\Delta_1 = \{\mathtt{Penguin}(d) \rightsquigarrow \neg\mathtt{Fly}(d) \mid d \in D\}$,
$\Delta_2 = \{\mathtt{Bird}(d) \rightsquigarrow \mathtt{Fly}(d) \mid d \in D\} \cup \{\mathtt{French}(d) \rightsquigarrow \mathtt{Rude}(d) \mid d \in D\}$.

[3] In this paper, the indices of the strata are considered in reversed order compared to system Z.

Consider $\mathcal{D} = \{a, b, c\}$ *and* $v_1, v_2 \in \mathbb{M}$ *such that* $v_1(\texttt{Penguin}(a)) = v_1(\texttt{Bird}(b)) = v_1(\texttt{French}(c)) = v_1(\texttt{Fly}(b)) = v_1(\texttt{Rude}(c)) = 1$, $v_1(\texttt{Fly}(a)) = 0$, *and* $v_2(\texttt{Penguin}(a)) = v_2(\texttt{Bird}(b)) = v_2(\texttt{French}(c)) = v_2(\texttt{Fly}(b)) = 1$, $v_2(\texttt{Fly}(a)) = v_2(\texttt{Rude}(c)) = 0$. *Then, we get* $v_1 \succ_\Delta v_2$, *since*

$$
\begin{array}{ll}
|str(\Delta_1) \cap [v_1]| & = |\{\neg\texttt{Penguin}(x) \vee \neg\texttt{Fly}(x)\}_{x\in\{a,b,c\}}| = 3 \\
|str(\Delta_2) \cap [v_1]| & = |\{\neg\texttt{Bird}(x) \vee \texttt{Fly}(x)\}_{x\in\{b,c\}} \cup \{\neg\texttt{French}(x) \vee \texttt{Rude}(x)\}_{x\in\{a,b,c\}}\}| = 5 \\
|str(\Delta_1) \cap [v_2]| & = |\{\neg\texttt{Penguin}(x) \vee \neg\texttt{Fly}(x)\}_{x\in a,b,c}\}| = 3 \\
|str(\Delta_2) \cap [v_2]| & = |\{\neg\texttt{Bird}(x) \vee \texttt{Fly}(x)\}_{x\in\{b,c\}} \cup \{\neg\texttt{French}(x) \vee \texttt{Rude}(x)\}_{x\in\{a,b\}}\}| = 4.
\end{array}
$$

In [9], an algorithm is proposed to translate a set of uncertain default rules UΔ of the form $(\varphi \rightsquigarrow \psi, \rho)$ into a set of classic formulas associated with weights representing a lower bound of the necessity of the rule. Since default rules enable us to reason by assuming a non-exceptional situation, first each rule is rewritten into a formula $(\varphi \wedge \bigwedge_{i\in[1,k]} \neg e_i \rightarrow \psi, \rho)$ where $e_1, \ldots, e_k$ are exceptions to this rule. For computing exceptions, the algorithm starts from the second most specific stratum UΔ_2 (since the rules of UΔ_1 have no exception). At stratum UΔ_s each rule can only admit exceptions in strata lower than s. Adapted to our context where we consider also a belief base $\mathcal{B}$, exceptions to a rule $(\varphi \rightsquigarrow \psi, \rho)$ are rules $(e_i \rightsquigarrow \psi_i, \rho_i) \in$ UΔ with compatible premises but incompatible conclusions (i.e., $\mathcal{B} \cup T_{s-1} \cup \{(\varphi \wedge e_i, \min(\rho, \rho_i))\} \nvdash_{>\text{Inc}} \bot$ and $\mathcal{B} \cup T_{s-1} \cup \{(\psi \wedge \psi_i, \min(\rho, \rho_i)\} \vdash_{>\text{Inc}} \bot$[4]), where T_{s-1} is the set of strict rules with explicit exceptions (and with certainty levels) coming from the translation of the rules of all the previous strata U$\Delta_1 \cup \ldots \cup$ UΔ_{s-1}. In addition to the translation of a rule, $(\varphi \rightsquigarrow \psi, \rho)$, k formulas of the form $\varphi \rightarrow \neg e_i$ are added in order to impose that when nothing says the contrary, the situation is not exceptional. [9] showed that each new formula $\varphi \rightarrow \neg e_i$ can be attributed a degree equal to $\min(\rho, \rho_i)$. Let $tr(\mathcal{B}, \text{U}\Delta)$ be the translation of a set of uncertain rules UΔ given a possibilistic base $\mathcal{B}$. In our setting, rules have also an anchorage, we transfer the anchorage to the new formulas in the same way that the degree is transferred, more precisions are given in the next section.

Example 5 (Exemple 4 continued) *For the moment we forget the anchoring degrees*

$$
\mathcal{B} = \left\{ \begin{array}{l} (\texttt{Penguin}(x) \rightarrow \texttt{Bird}(x), 0.8) \\ (\texttt{Bird}(x) \rightarrow \neg\texttt{French}(x), 1) \end{array} \right\}, \mathcal{R} = \left\{ \begin{array}{ll} (\texttt{Penguin}(x) \rightsquigarrow \neg\texttt{Fly}(x), 0.99) & U\Delta_1 \\ (\texttt{Bird}(x) \rightsquigarrow \texttt{Fly}(x), 0.9) & \\ (\texttt{French}(x) \rightsquigarrow \texttt{Rude}(x), 0.7) & U\Delta_2 \end{array} \right\}
$$

In order to rewrite $\mathcal{R}$ *it is enough to consider the two rules of U*Δ_2 *and find their exceptions, here only* $Bird(x) \rightsquigarrow Fly(x)$ *admits exceptions (in stratum U*Δ_1*) which gives:*

$$
tr(\mathcal{B}, \mathcal{R}) = \left\{ \begin{array}{ll} (\texttt{Penguin}(x) \rightarrow \neg\texttt{Fly}(x), 0.99, 0) & (\texttt{Bird}(x) \wedge \neg\texttt{Penguin}(x) \rightarrow \texttt{Fly}(x), 0.9, 0) \\ (\texttt{French}(x) \rightarrow \texttt{Rude}(x), 0.7, 0.5) & (\texttt{Bird}(x) \rightarrow \neg\texttt{Penguin}(x), 0.9, 0) \end{array} \right.
$$

Note that here, forgetting anchoring, $\mathcal{B} \cup tr(\mathcal{B}, \mathcal{R})$ *is consistent: Inc*$(B \cup tr(\mathcal{B}, \mathcal{R})) = 0$.

3 Arrival of a new piece of information

In this section, we propose a strategy for considering the arrival of a new piece of information $\phi \in \mathcal{L}_\text{C}$ with a certainty degree γ. Since the anchorage reinforces the

[4]Or equivalently Inc$(\{(\varphi \wedge e_i, \min(\rho, \rho_i))\} \cup \mathcal{B} \cup T_{s-1}) =$ Inc$(B \cup T_{s-1})$ and Inc$(\{(\psi \wedge \psi_i, \min(\rho, \rho_i))\} \cup \mathcal{B} \cup T_{s-1}) >$ Inc$(B \cup T_{s-1})$

certainty of a rule, when ϕ is incompatible with a rule, we compare its certainty level γ with the aggregation of the level of certainty and the anchorage of this rule and react accordingly either to reject ϕ or to modify the rule. This aggregation, denoted $\oplus$, can be a simple addition or a more sophisticated operation. We leave its study for further research. Let us introduce the transformation tra of the belief base $(\mathcal{B}, \mathcal{R})$ into a possibilistic base $\{(\phi_i, \gamma_i)\}_i \in [1, m]$ by aggregating certainty and anchorage levels:

$$tra(\mathcal{B}, \mathcal{R}) = \left\{ (\phi_i, \gamma_i) \in \mathcal{L} \times [0,1] \;\middle|\; \begin{array}{l} (\phi_i, \gamma_i) \in \mathcal{B} \text{ or} \\ (\phi_i, \rho_i, \alpha_i) \in tr(\mathcal{B}, \mathcal{R}) \text{ and } \gamma_i = \rho_i \oplus \alpha_i \end{array} \right\}$$

Given a belief base $(\mathcal{B}, \mathcal{R})$, compute $\mathcal{B}' = tra(\mathcal{B}, \mathcal{R})$

1. If ϕ is *inconsistent* with the 1-cut of $\mathcal{B}'$, $\mathcal{B}'_1$, i.e., $sk(\mathcal{B}'_1) \cup \{\phi\} \vdash \bot$, then ϕ is not integrated
2. If (ϕ, γ) is *compatible* with $(\mathcal{B}, \Delta)$, in other words, if the level of inconsistency do not increase when adding (ϕ, γ), i.e., $\text{Inc}(\mathcal{B}' \cup \{(\phi, \gamma)\}) = \text{Inc}(B')$ then (ϕ, γ) is added to $\mathcal{B}'$ giving $(\mathcal{B}' \cup (\phi, \gamma), \mathcal{R})$
3. Otherwise (when the new information is more certain than some contradicting formula of $(\mathcal{B}, \Delta)$), let us consider the rule (i.e., the formula of $\mathcal{B}'$ that comes from a rule $(\varphi_i \rightsquigarrow \psi_i, \rho_i, \alpha_i) \in \mathcal{R}$) with a maximum certainty in $\mathcal{B}'$, violated by ϕ i.e, $\mathcal{B}' \vdash_{>Inc} \phi \rightarrow (\varphi_i \wedge \neg\psi_i)$ [5] and $\rho_i \oplus \alpha_i$ is maximum among such rules
   (a) if $\gamma < \rho_i \oplus \alpha_i$ the piece of information ϕ is rejected (but the anchoring of the rule can be decreased $\varphi_i \rightsquigarrow \psi_i$, because even rejected information may influence the reasoner)
   (b) otherwise a rule $(\varphi_i \wedge \phi \rightsquigarrow \neg\psi_i, \gamma, 0)$ is added encoding this exception. The certainty and anchorage of $\varphi_i \rightsquigarrow \psi_i$ are lowered yielding $(\varphi_i \rightsquigarrow \psi_i, \rho'_i, \alpha'_i)$ such that $(\rho'_i \oplus \alpha'_i < \rho_i \oplus \alpha_i)$. The same exception adding and degree lowering is done for all the remaining rules violated by ϕ.

This strategy proposes a way to take anchorage into account when revising a belief base. In our opinion, anchorage must reinforce the barrier to the addition of an exception. Hence, to determine whether to add an exception or not, we must compare the certainty level (γ) of the new information with the certainty of the intern rule combined with its anchorage ($\rho \oplus \alpha$). In the case where the information is stronger, the anchorage must decrease and the exception be added. The operator $\oplus$ should translate an effect of reinforcing certainty (thus preventing a rational reasoning which would possibly allow for a revision). As soon as a rule dismantles the information then it is discarded, hence it is not necessary to look at less important rules violated by this information.

Example 6 (Example 4 continued) *Assume that we take $\oplus = \max$, then the rewriting of the belief base gives a new possibilistic base $\mathcal{B}' = tra(\mathcal{B}, \mathcal{R})$ which contains:*

$(\texttt{Bird}(x) \rightarrow \neg\texttt{French}(x), 1)$
$(\texttt{Bird}(x) \rightarrow \neg\texttt{Penguin}(x), 0.9)$
$(\texttt{Penguin}(x) \rightarrow \neg\texttt{Fly}(x), 0.99)$
$(\texttt{Penguin}(x) \rightarrow \texttt{Bird}(x), 0.8)$
$(\texttt{Bird}(x) \wedge \neg\texttt{Penguin}(x) \rightarrow \texttt{Fly}(x), 0.9)$
$(\texttt{French}(x) \rightarrow \texttt{Rude}(x), 0.7)$

[5] When $sk(\mathcal{B}')$ is consistent, it amounts to check whether $\mathbb{M} \models \phi \rightarrow (\varphi_i \wedge \neg\psi_i)$.

We first consider the piece of information $\texttt{French}(Jeanne) \wedge \texttt{Bird}(Jeanne)$ *meaning that we have an individual Jeanne that is both a human (who is French) and a bird. Then this information is not integrated (case 1) because it is contradictory to the only formula of certainty 1. Now, assume that* $\phi_0 = (\texttt{Rude}(x) \rightarrow \texttt{RudeUS}(x) \wedge \texttt{RudeF}(x)$ *with certainty 1 arrives, meaning that any rude person is rude for everyone, i.e, she is rude for French people and for US citizens. This formula being consistent with all the formulas of* $\mathcal{B}'$*, thus* $Inc(\mathcal{B}' \cup (\phi_0, 1)) = Inc(\mathcal{B}') = 0$*. Assume now, that the user learns that Jeanne is French and not considered rude by French people:* $\phi_1 = French(Jeanne) \wedge \neg RudeF(Jeanne)$ *with a certainty degree of 0.8. Then we apply the case 3b,* ϕ_1 *violates the rule* $(French(x) \rightsquigarrow Rude(x), 0.7, 0.4)$ *and* $0.8 > 0.7 = 0.7 \oplus 0.4$ *this leads us to add the rule* $(French(x) \wedge French(Jeanne) \wedge \neg RudeF(Jeanne) \rightsquigarrow \neg Rude(x), 0.8, 0)$*, i.e.,* $(French(Jeanne) \wedge \neg RudeF(Jeanne) \rightsquigarrow \neg Rude(Jeanne), 0.8, 0)$*, namely, Jeanne who is not considered rude in France is not rude at all (here, we propose to consider the anchorage as null due to the novelty of this information). Additionally we have to reduce the certainty and anchorage of the initial rule. We could get for instance* $(French(x) \rightsquigarrow Rude(x), 0.65, 0.3)$ *in such a way that* $0.65 \oplus 0.4 < 0.7 \oplus 0.4$ *(witch holds when* $\oplus = \max$*).*

Note that this is a preliminary example, to explain what we have planned, but for the moment we have only considered the decrease in anchorage and not its increase, since this amounts to studying the initial creation of prejudice, which is a complex, social phenomenon that we are leaving aside for the moment.

4 Concluding remark

This article explores the way to encompass the handling of stereotypes in a logical framework. We showed that disposing of a framework where we can express both defeasibility, certainty and anchorage strength would be suitable for this purpose. The possibilistic setting seems suitable for dealing with these three notions. A lot of work remains to be done in order to consolidate the foundations of this new line of research. Two crucial points are the study of the different ways of defining the aggregation of certainty and anchorage, and the management of the reduction of these degrees.

References

[1] Gordon Willard Allport, Kenneth Clark, and Thomas Pettigrew. *The nature of prejudice*. Addison-Wesley Reading, MA, 1954.

[2] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artif. Intell.*, 92(1-2):259–276, 1997.

[3] Salem Benferhat, Claudette Cayrol, Didier Dubois, Jérôme Lang, and Henri Prade. Inconsistency management and prioritized syntax-based entailment. In *Proc. of Int. Joint. Conf. on Artificial Intelligence (IJCAI'93)*, volume 93, pages 640–645, 1993.

[4] Salem Benferhat, Didier Dubois, and Henri Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1-2):259–276, 1997.

[5] Didier Dubois, Francis Faux, and Henri Prade. Prejudice in uncertain information merging: Pushing the fusion paradigm of evidence theory further. *International Journal of Approximate Reasoning*, 121:1–22, 2020.

[6] Didier Dubois, Jérôme Lang, and Henri Prade. Possibilistic logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 439–513. Oxford University Press, 1994.

[7] Didier Dubois and Henri Prade. *Possibility theory – An Approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.

[8] Florence Dupin de Saint-Cyr and Francis Faux. Integrating evolutionary prejudices in belief function theory. In Zied Bouraoui, Said Jabbour, and Srdjan Vesic, editors, *ECSQARU 2023*, Lecture Notes in Artificial Intelligence (LNCS/LNAI), pages 400–414, Arras, France, September 2023. CRIL, Springer.

[9] Florence Dupin De Saint Cyr and Henri Prade. Handling uncertainty and defeasibility in a possibilistic logic setting. *International Journal of Approximate Reasoning, Special Section on Logical Approaches to Imprecise Probabilities*, 49(1):67–82, septembre 2008.

[10] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proc. 3rd Conf. on Theoretical Aspects of Reasoning about Knowledge*, pages 121–135, 1990.

[11] Milton Rokeach. The nature and meaning of dogmatism. 1954.