Making the right exceptions

Frank Veltman & Harald Bastiaanse
Institute for Logic, Language and Computation.
University of Amsterdam

LIRa Workshop, March 28, 2011
Abstract

In non-monotonic reasoning conflicts between default rules abound. I will present a principled account to deal with them. I will do so in two ways:

- *semantically*, within a circumscripive theory

- *syntactically*, by supplying an algorithm for inheritance networks

The latter is sound and complete with respect to the first.
This talk is about sentences of the form

\[ P\text{'s are normally } Q \]

Such sentences express default rules. Roughly, what they mean is this. Whenever you are confronted with an object with the property \( P \), you may assume it has the property \( Q \) as well, provided you have no evidence to the contrary.
Default Reasoning 2

| premise 1  | P's are normally R |
| premise 2  | x is P              |
| by default | x is R              |
Default Reasoning 3

premise 1  Master students normally are full time students
premise 2  John is a master student
by default  John is a full time student
Default Reasoning 4

premise 1  P's are normally $R$
premise 2  $x$ is $P$ and $x$ is $Q$
by default  $x$ is $R$
Default Reasoning 5

premise 1  \( Q \)'s are normally not \( R \)
premise 2  \( P \)'s are normally \( R \)
premise 3  \( x \) is \( P \) and \( x \) is \( Q \)

by default  ???
premise 1  Republicans are normally not pacifists
premise 2  Quakers are normally pacifists
premise 3  Nixon is a republican and a quaker

by default  ???
Default Reasoning 6

premise 1  Q's are normally $P$
premise 2  Q's are normally not $R$
premise 3  P's are normally $R$
premise 4  $x$ is $P$ and $x$ is $Q$

by default  $x$ is not $R$
Default Reasoning 7

premise 1  Q's are normally $P$
premise 2  Q's are normally not $R$
premise 3  $P$'s are normally $R$
premise 4  $x$ is $P$

by default  $x$ is $Q$, but $x$ is not $R$
Weak Tweety Triangle

\begin{align*}
\text{premise 1} & \quad \text{Master students are normally adults} \\
\text{premise 2} & \quad \text{Master students are normally not employed} \\
\text{premise 3} & \quad \text{Adults are normally employed} \\
\text{premise 4} & \quad \text{John is a master student} \\
\text{by default} & \quad \text{John is an adult, but not employed}
\end{align*}
Strong Tweety Triangle

premise 1    Penguins are birds
premise 2    Penguins cannot fly
premise 3    Birds normally fly
premise 4    Tweety is a penguin

by default   Tweety is a bird, but Tweety cannot fly
Circumscription 1

A sentence of the form

\[ P's \text{ are normally } Q \]

will be represented by a formula of the form

\[ \forall x((Px \land \lnot Ab_{Px,Qx} x) \rightarrow Qx) \]

If an object satisfies the formula \( Ab_{Px,Qx} x \) this means that it behaves \textit{abnormally} with respect to this rule.
More precisely

Let $\mathcal{L}_0$ be a language of monadic first order logic with finitely many one-place predicates.

We extend the language $\mathcal{L}_0$ with exception predicates $A b \varphi(x), \psi(x)$. Here $\varphi(x)$ and $\psi(x)$ are both formulas of $\mathcal{L}_0$ with one and the same free variable $x$.

(I omit some technical proviso’s here)
**Default Rules**

A *default rule* is a formula of the form

\[ \forall x((\varphi(x) \land \neg Ab_{\varphi(x),\psi(x)}x) \rightarrow \psi(x)) \]

- \( \varphi(x) \) and \( \psi(x) \) are formulas of \( \mathcal{L}_0 \) in which \( x \) is the only free variable.

- \( \varphi(x) \) is the antecedent and \( \psi(x) \) is the consequent of the rule.

- \( Ab_{\varphi(x),\psi(x)}x \) is the abnormality clause of the rule.
**Circumscription 2**

Let the models $\mathfrak{A} = \langle D, I \rangle$ and $\mathfrak{A}' = \langle D, I' \rangle$ be based on the same domain $D$. Then $\mathfrak{A}$ is at least as normal as $\mathfrak{A}'$ iff for all predicates $Ab_\varphi(x), \psi(x)$, $I(Ab_\varphi(x), \psi(x)) \subseteq I'(Ab_\varphi(x), \psi(x))$.

Let $S$ be a set of models. Then $\mathfrak{A}$ is optimal in $S$ iff there is no $\mathfrak{A}' \in S$ such that $\mathfrak{A}'$ is more normal than $\mathfrak{A}$.
Naive Circumscription

\[ \Delta \models_{d} \varphi \text{ iff for all nonempty domains } \mathcal{D}, \text{ and all models } \mathcal{A} \text{ based on } \mathcal{D} \text{ it holds that if } \mathcal{A} \text{ is an optimal model of } \Delta, \text{ then } \mathcal{A} \text{ is a model of } \varphi. \]
Normal in some respects, but not in other

premise 1 Adults normally have a bank account
premise 2 Adults normally have a driver’s licence
premise 3 John is an adult without a driver’s licence
by default John is an adult with a bank account
Normal in some respects, but not in other

\[\forall x((Ax \land \neg Ab_{Ax,Bx}x) \rightarrow Bx)\]
\[\forall x((Ax \land \neg Ab_{Ax,Dx}x) \rightarrow Dx)\]
\[Aj \land \neg Dj\]

by default Bj

The example illustrates why the abnormality predicates have two indices, and not just one.
Naive Approach (continued)

This way

\[ \forall x ((S_x \land \neg Ab_{S_x,A_x} x) \rightarrow A_x) \]
\[ \forall x ((Ax \land \neg Ab_{A_x,Ex} x) \rightarrow E_x) \]
\[ \forall x ((S_x \land \neg Ab_{S_x,\neg Ex} x)) \rightarrow \neg E_x \]

\[ \text{by default} \quad \neg E_a \]

is not valid.
What we would like

\[
\forall x((Sx \land \neg Ab_{Sx,Ax}x) \rightarrow Ax)
\]
\[
\forall x((Ax \land \neg Ab_{Ax,Ex}x) \rightarrow Ex)
\]
\[
\forall x((Sx \land \neg Ab_{Sx,\neg Ex}x)) \rightarrow \neg Ex)
\]
\[
\therefore \forall x(Sx \rightarrow Ab_{Ax,Ex}x)
\]
Exemption

We will only admit models in which the formula $\forall x(Sx \rightarrow Ab_{Ax,Ex} x)$ is true. This way we enforce the idea that objects with property $S$, are \textit{exempted} from the default rule that $A$’s are normally $E$.

(Think of default rules as normative rules. Students \textit{have to be} adults, adults \textit{have to be} employed, but here an exception is made for students, they \textit{don’t have to} be employed, they are not subjected to this rule.)
Strict Rules

Henceforth, I will often write $\forall x (\varphi(x) \sim \psi(x))$ to abbreviate $\forall x ((\varphi(x) \land \neg Ab_{\varphi(x),\psi(x)}x) \rightarrow \psi(x))$. (Since the abnormality clause is determined by the antecedent and the consequent, we can do so)

Some sentences of the form $\forall x (\varphi(x) \rightarrow \psi(x))$ will get a special status as *strict rules*, rules that don’t allow for exceptions.

They are to be distinguished from universal sentences that are accidentally true, and will be treated different from these.
Set up

Let $\Sigma$ be a set of rules, and $\Delta$ be a set of sentences. Think of $I = \langle \Sigma, \Delta \rangle$ as the information of some agent at some time, where $\Sigma$ is the set of rules the agent is acquainted with, and $\Delta$ his/her factual information.

We will correlate with $I$ a pair $\langle U_I, F_I \rangle$, and call this the (information) state generated by $I$.

$U_I$ is called the universe of the state. The elements of $U_I$ are models of $\Sigma$, but not all models of $\Sigma$ are allowed. $U_I$ has to satisfy some additional constraints.

$F_I$ consists of all models in $U_I$ that are models of $\Delta$. 

17
Set up (continued)

Given this set up we can define validity as follows:

\[ \Sigma, \Delta \models d \varphi \iff \text{for all optimal models } \mathcal{A} \in \mathcal{F}_I, \mathcal{A} \models \varphi. \]
Some (technical) notions

• Suppose $\mathfrak{A} \models \forall x (\varphi(x) \leadsto \psi(x))$, and let $d$ be an element of the domain of $\mathfrak{A}$. Then $d$ complies with $\forall x (\varphi(x) \leadsto \psi(x))$ (in $\mathfrak{A}$) iff $d$ does not satisfy $A\varphi(x),\psi(x)x$.

Let $\Delta$ be a set of default rules, and $d$ an element of the domain of some model $\mathfrak{A}$ for $\Delta$. Then $d$ complies with $\Delta$ (in $\mathfrak{A}$) iff $d$ complies with all $\delta \in \Delta$. 
Compliance

Notice that the definition allows for the following situations

- The object $d$ complies with $\forall x (\varphi(x) \implies \psi(x))$, but $d$ does not satisfy $\varphi(x)$.

- The object $d$ satisfies $\varphi(x)$ and $\psi(x)$, but $d$ does not comply with $\forall x (\varphi(x) \implies \psi(x))$.

We will see examples later on. For now ‘just’ notice that this can happen.
Some (technical) notions 2

- Let $\Sigma$ be a set of rules and $\varphi(x)$ be some formula with one free variable $x$. $\Sigma \varphi(x)$ is the set of all defaults $\delta \in \Sigma$ with antecedent $\varphi(x)$.

$\Sigma \varphi(x)$ is called the default theory of $\varphi(x)$ in $\Sigma$. 
What we want

Minimal Requirement

*Suppose* it is logically possible for there to exist objects with property $P$ that comply with all rules for objects with property $P$.

*Then* if the only factual information about some object is that it has property $P$, it must at least be valid to infer (by default) that it does comply with all rules for objects with property $P$. 
**Exemption Constraint 1**

One of the constraints that we have to impose for the Minimal Requirement to be satisfied is this.

Let $\varphi(x)$ a formula with one free variable $x$ and let $\Sigma' \subseteq \Sigma$.

*Suppose* for all $\mathcal{A} \in \mathcal{U}_I$ it holds that no object in the domain of $\mathcal{A}$ satisfies $\varphi(x)$ and complies with $\Sigma' \cup \Sigma\varphi(x)$.

*Then* for all $\mathcal{A} \in \mathcal{U}_I$ it holds that no object in the domain of $\mathcal{A}$ satisfies $\varphi(x)$ and complies with $\Sigma'$. 


Exemption Constraint 2

Example

Consider $\Sigma = \{\forall x (Sx \Rightarrow Ax), \forall x (Sx \Rightarrow \neg Ex), \forall x (Ax \Rightarrow Ex)\}$

Then $\Sigma^{Sx} = \{\forall x (Sx \Rightarrow Ax), \forall x (Sx \Rightarrow \neg Ex)\}$

Let $\Sigma' = \{\forall x (Ax \Rightarrow Ex)\}$

Clearly, there is no $A$ such that some object in the domain of $A$ satisfies $Sx$ and complies with $\Sigma' \cup \Sigma^{Sx}$.

This means that all $A \in \mathcal{U}_I$ have the property that all objects in the domain of $A$ that satisfy $Sx$, satisfy $Ab_{Ax,Ex}x$. 
Exemption Constraint 3

Consider $I = \langle \Sigma, \Delta \rangle$ and let $\Sigma' \subseteq \Sigma$.

Suppose

$\mathcal{U}_I \models \forall x (\varphi(x) \rightarrow \bigvee_{\delta \in \Sigma' \cup \Sigma} Ab_{\delta} x)$,

then

$\mathcal{U}_I \models \forall x (\varphi(x) \rightarrow \bigvee_{\delta \in \Sigma'} Ab_{\delta} x)$
**Inheritance constraint (simple form)**

The next constraint goes beyond the Minimal Requirement.

Suppose
\[ U_I \models \forall x(\varphi(x) \leadsto \psi(x)) \text{ and } U_I \models \forall x(\psi(x) \rightarrow Ab_{\chi(x),\theta(x)}x), \]

then
\[ U_I \models \forall x(\varphi(x) \rightarrow Ab_{\chi(x),\theta(x)}x) \]

So, if the \( \varphi \)'s are normally \( \psi \) then the \( \varphi \)'s are exempted from all the rules the \( \psi \)'s are exempted from.
Inheritance constraint (example)

Let $\Sigma$ be the theory consisting of the following five default rules

\[
\begin{align*}
\forall x ((Ax \Rightarrow Bx) \\
\forall x ((Bx \Rightarrow Cx) \\
\forall x ((Cx \Rightarrow Dx) \\
\forall x ((Bx \Rightarrow \neg Dx) \\
\forall x ((Ax \Rightarrow Dx)
\end{align*}
\]

The exemption constraint enforces $\forall x (Bx \Rightarrow Ab_{Cx,Dx} x)$. 
By the exemption constraint we also have $\forall x (Ax \Rightarrow Ab_{Bx,\neg Dx} x)$. 
But, exceptions to exceptions do not count as normal: Applying 
the inheritance constraint we get $\forall x (Ax \Rightarrow Ab_{Cx,Dx} x)$. 

27
Inheritance constraint (example continued)

In this case we will find that \( \Sigma, A_c \models_d C_c \land D_c \land A_b^{C_x,D_x} \).

(The object named \( c \) satisfies \( C_x \) and \( D_x \), but does not comply with the rule \( \forall x (C_x \rightarrow D_x) \).)
Inheritance constraint 3

Consider $I = \langle \Sigma, \Delta \rangle$ and let $\Sigma' \subseteq \Sigma$.

Suppose
\[ \mathcal{U}_I \models \forall x (\varphi(x) \leadsto \psi(x)) \quad \text{and} \quad \mathcal{U}_I \models \forall x (\psi(x) \rightarrow \bigvee_{\delta \in \Sigma} Ab_{\delta} x), \]

then
\[ \mathcal{U}_I \models \forall x (\varphi(x) \rightarrow \bigvee_{\delta \in \Sigma'} Ab_{\delta} x) \]
States

Let $\Sigma$ be a set of rules, and $\Delta$ be a set of sentences. The state generated by $I = \langle \Sigma, \Delta \rangle$ is the pair $\langle U_I, F_I \rangle$ where

- $U_I$ is the largest class of models of $\Sigma$ satisfying the three constraints (Exemption, Inheritance) discussed.

- $F_I$ is the class of all models in $U_I$ that are models of $\Delta$. 
Some examples

Both *Defeasible Modus Ponens* and *Defeasible Modus Tollens* are valid.

\[ \forall x((Px \leadsto Qx) \to Pa) \quad \therefore Qa \]

\[ \forall x((Qx \leadsto \neg Px) \to Pa) \quad \therefore \neg Qa \]
Some Examples 2

\[
\begin{align*}
\forall x((P_x \rightsquigarrow Q_x) \\
\forall x((Q_x \rightsquigarrow \neg P_x))
\end{align*}
\]

\[
\begin{array}{c}
Pa \\
\hline
\therefore Qa
\end{array}
\]

Defeasible Modus Ponens beats Defeasible Modus Tollens! It does not follow from the premises that \( \neg Pa \). The exemption constraint enforces that \( \mathcal{U}_I \models \forall x(P_x \rightarrow Ab_{Q_x,\neg P_x} x) \).
Some examples 3

This example illustrates the Inheritance Principle

\[ \forall x (Rx \implies \neg Px) \]
\[ \forall x (Qx \implies Px) \]
\[ \forall x (Sx \implies Rx) \]
\[ \forall x (Sx \implies Qx) \]
\[ \forall x (Tx \implies Sx) \]
\[ \forall x (Ux \implies Tx) \]
\[ Ua \]
\[ \therefore Ra \land Qa \]

Exemption enforces \[ \forall x (Sx \implies (\text{Ab}_{Rx,\neg Px} x \lor \text{Ab}_{Qx,Px} x)) \].

2 \times Inheritance gives \[ \forall x (Ux \implies (\text{Ab}_{Rx,\neg Px} x \lor \text{Ab}_{Qx,Px} x)) \].
Quakers are normally doves
Republicans are normally hawks
Nobody can be both a hawk and a dove
Hawks are normally politically motivated
Doves are normally politically motivated
Nixon is a republican quaker

Is Nixon politically motivated?
The exemption constraint enforces that in all models Nixon has either the property $Ab_{Rx,Hx}$ or the property $Ab_{Qx,Dx}$.

In the optimal models he will be abnormal in only one of these respects and perfectly normal in the other respect.

So, yes, presumably Nixon is politically motivated.
An inheritance network is a directed graph where the arrows represent default rules. Nodes may represent individuals or properties. Specifically marked arrows are used for negative rules and for strict rules.
Paths bring you from a given premise to a ‘prima facie’ conclusion. There are positive paths and negative paths. Where these contradict, some arrows must be eliminated.
For any node $x$, $Min(x)$ consists of the strict rules of the network and the arrows starting at $x$. Where a set of rules allows for contradicting conclusions when starting from $x$, it is concluded that $x$ is an exception to one of the other rules in that set but not in $Min(x)$.

Exceptions are inherited: if $Q$’s are an exception to a given rule (or to at least one rule in a given set) and $P$’s are normally $Q$’s, then $P$’s are an exception to that rule (to one of those rules).

The inheritance principle makes a Backward Induction approach ideal.
Networks - algorithm

Rather than spelling the algorithm out, I will show you how it works on the blackboard.
An example with a ‘zombie path’

Quakers are normally pacifists
Republicans are normally not pacifists
Republicans are normally football fans
Pacifists are normally anti-military
Football fans are normally not anti-military
Nixon is a republican quaker
Appendix
Recall the Minimal Requirement

*Suppose* it is possible for there to exist objects with property $P$ that comply with all rules for objects with property $P$.

*Then* if the only factual information about some object is that it has property $P$, it must at least be valid to infer (by default) that it does comply with all rules for objects with property $P$. 
Recall the Minimal Requirement

*Suppose* it is possible for there to exist objects with property $P$ that comply with all rules for objects with property $P$.

*Then* if the only factual information about some object is that it has property $P$, it must at least be valid to infer (by default) that it does comply with all rules for objects with property $P$. 


Consider the following example

We would want to conclude $Qa$ and $Ra$, but we cannot. By the exemption constraint we get $\forall x(Qx \rightarrow Ab_{Px,Rx}x)$. As a consequence there are no models in which the object $a$ complies with both the rule $\forall x(Px \leadsto Qx)$ and the rule $\forall x(Px \leadsto Rx)$.
Equivalence Constraint (simple form)

We can avoid that such situations can consistently arise by adopting the following constraint.

Suppose both $\forall x (\varphi(x) \leadsto \psi(x))$ and $\forall x (\psi(x) \leadsto \varphi(x))$ hold in $\mathcal{U}_I$.

Then if $\forall x (\varphi(x) \leadsto \chi(x))$ holds in $\mathcal{U}$, also $\forall x (\psi(x) \leadsto \chi(x))$ holds in $\mathcal{U}_I$. 
Equivalence Constraint (general form)

In fact we will adopt something more general.

Let \( n > 1 \)

\[ \text{Suppose} \text{ for all } 1 \leq i < n \]
\[ \mathcal{U}_I \models \forall x (\varphi_i(x) \sim \varphi_{i+1}(x)), \text{and } \mathcal{U}_I \models \forall x (\varphi_n(x) \sim \varphi_1(x)), \]

\[ \text{then} \text{ for all } 1 \leq i, j \leq n \]
\[ \text{if } \mathcal{U}_I \models \forall x (\varphi_i(x) \sim \psi(x)), \mathcal{U}_I \models \forall x (\varphi_j(x) \sim \psi(x)) \]
Minimal Requirement (Strengthened form)

**Theorem**
Consider an inheritance net representing a set of rules $\Delta$ satisfying the Equivalence Constraint.

*Suppose* that there are no nodes $N$ and $M$ so that there is both a positive link $N \rightarrow M$ and a negative link $N \not\rightarrow M$ between $N$ and $M$.

*Then* if the net contains the link $P \rightarrow Q$, the net supports the argument $\Delta, Pa/ \therefore Qa$. 

47