Making the right exceptions

Frank Veltman & Harald Bastiaanse

Institute for Logic, Language and Computation.

University of Amsterdam

Göttingen, June 5, 2010

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

The latter is sound and complete with respect to the first.

Default Reasoning 1

This talk is about sentences of the form

P'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 P's are normally R
premise 2 x is P and x is not R

---

by default

Default Reasoning 6

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 ???

Nixon Diamond

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 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 and x is Q
by default	x is not R

Default Reasoning 8

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

premise 1	Master students are normally adults
premise 2	Master students are normally not employed
premise 3	Adults are normally employed
premise 4	John is a master student
by default	John is an adult, but not employed

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 are normally Q

will be represented by a formula of the form

$$\forall x((Px \wedge \neg Ab_{Px,Qx}\, x) \rightarrow Qx)$$

If an object satisfies the formula $Ab_{Px,Qx}x$ this means that it behaves *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 $Ab_{\varphi(x),\psi(x)}$. 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

$$\forall x((\varphi(x) \wedge \neg Ab_{\varphi(x),\psi(x)}x) \rightarrow \psi(x))$$

- $\varphi(x)$ 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 of the rule.

Circumscription 2

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

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

Naive Circumscription

$\Delta \models_d \varphi$ iff for all nonempty domains $\mathcal{D}$, and all models $\mathfrak{A}$ based on $\mathcal{D}$ it holds that if $\mathfrak{A}$ is an *optimal* model of Δ, then $\mathfrak{A}$ is a model of φ.

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

premise 1 $\forall x((Ax \wedge \neg Ab_{Ax,Bx}\, x) \rightarrow Bx)$

premise 2 $\forall x((Ax \wedge \neg Ab_{Ax,Dx}\, x) \rightarrow Dx)$

premise 3 $Aj \wedge \neg Dj$

---

by default Bj

The example illustrates why the abnormality predicates have two indices, and not just one.

Naive Approach (continued)

This way

$$
\begin{array}{rl}
 & \forall x((Sx \wedge \neg Ab_{Sx,Ax}\, x) \rightarrow Ax) \\
 & \forall x((Ax \wedge \neg Ab_{Ax,Ex}\, x) \rightarrow Ex) \\
 & \forall x((Sx \wedge \neg Ab_{Sx,\neg Ex}\, x)) \rightarrow \neg Ex) \\
 & Sa \\
\hline
\textit{by default} & \neg Ea
\end{array}
$$

is invalid.

What we would like

$$
\begin{array}{ll}
 & \forall x((Sx \land \neg Ab_{Sx,Ax}\, x) \rightarrow Ax) \\
 & \forall x((Ax \land \neg Ab_{Ax,Ex}\, x) \rightarrow Ex) \\
 & \forall x((Sx \land \neg Ab_{Sx,\neg Ex}\, x)) \rightarrow \neg Ex) \\
\hline
\therefore & \forall x(Sx \rightarrow Ab_{Ax,Ex}\, x)
\end{array}
$$

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 *exempted* from the default rule that A's are normally E.

(Think of default rules as normative rules. Students *have to be* adults, adults *have to* be employed, but here an exception is made for students, they *don't have to* be employed, they are not subjected to this rule.)

Strict Rules

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

Some sentences of the form $\forall x(\varphi(x) \to \psi(x))$ 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 Σ be a set of rules, and Δ be a set of sentences. Think of $I = \langle \Sigma, \Delta \rangle$ as the *information* of some agent at some time, where Σ is the set of rules the agent is acquainted with, and Δ his/her factual information.

We will correlate with I a pair $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$, and call this the (information) *state* generated by I.

$\mathcal{U}_I$ is called the universe of the state. The elements of $\mathcal{U}_I$ are models of Σ, but not all models of Σ are allowed. $\mathcal{U}_I$ has to satisfy some additional constraints.

$\mathcal{F}_I$ consists of all models in $\mathcal{U}_I$ that are models of Δ.

Set up (continued)

Given this set up we can define validity as follows :

$\Sigma, \Delta \models_d \varphi$ iff for all *optimal* models $\mathfrak{A} \in \mathcal{F}_I$, $\mathfrak{A} \models \varphi$.

Some (technical) notions

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

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

Some (technical) notions 2

- Let Σ 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 Σ*.

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

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

Exemption Constraint 2

Example

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

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

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

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

This means that all $\mathfrak{A} \in \mathcal{U}_{\mathcal{I}}$ have the property that all objects in the domain of $\mathfrak{A}$ that satisfy Sx, satisfy $Ab_{Ax,Ex}x$.

Exemption Constraint 3

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

Suppose

$$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi(x) \rightarrow \bigvee\nolimits_{\delta \in \Sigma' \cup \Sigma^{\varphi(x)}} Ab_\delta \, x),$$

then

$$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi(x) \rightarrow \bigvee\nolimits_{\delta \in \Sigma'} Ab_\delta \, x)$$

Inheritance constraint (simple form)

The next constraint goes beyond the Minimal Requirement.

Suppose

$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi \rightsquigarrow \psi)$ and $\mathcal{U}_{\mathcal{I}} \models \forall x(\psi(x) \rightarrow Ab_{\chi(x),\theta(x)}\, x)$,

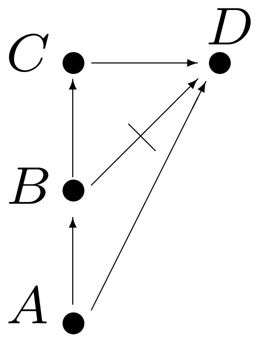
then

$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi(x) \rightarrow Ab_{\chi(x),\theta(x)}\, x)$

So, if the φ’s are normally ψ then the φ’s are exempted from all the rules the ψ’s are exempted from.

Inheritance constraint (example)

Suppose we have the following five default rules



$$\forall x((Ax \rightsquigarrow Bx)$$
$$\forall x((Bx \rightsquigarrow Cx)$$

$$\forall x((Cx \rightsquigarrow Dx)$$
$$\forall x((Bx \rightsquigarrow \neg Dx)$$
$$\forall x((Ax \rightsquigarrow Dx)$$

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

Inheritance constraint 3

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

Suppose

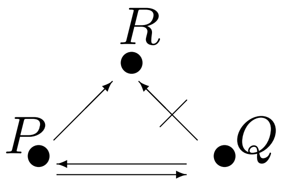
$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi \rightsquigarrow \psi)$ and $\mathcal{U}_{\mathcal{I}} \models \forall x(\psi(x) \to \bigvee_{\delta \in \Sigma'} Ab_{\delta}\, x)$,

then

$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Sigma'} Ab_{\delta}\, x)$

Equivalence constraint 1

As things stand now, the Minimal Requirement is not satisfied. Consider the following example



$$\forall x(Px \rightsquigarrow Qx)$$
$$\forall x(Qx \rightsquigarrow Px)$$
$$\forall x(Px \rightsquigarrow Rx)$$
$$\forall x(Qx \rightsquigarrow \neg Rx)$$
$$Pa$$

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 $\mathcal{F}_I$ in which the object a complies with both the rule $\forall x(Px \rightsquigarrow Qx)$ and the rule $\forall x(Px \rightsquigarrow Rx)$.

Equivalence Constraint (simple form)

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

Suppose both $\forall x(\varphi(x) \rightsquigarrow \psi(x))$ and $\forall x(\psi(x) \rightsquigarrow \varphi(x))$ hold in $\mathcal{U}_{\mathcal{I}}$.

Then if $\forall x(\varphi(x) \rightsquigarrow \chi(x))$ holds in $\mathcal{U}$, also $\forall x(\psi(x) \rightsquigarrow \chi(x))$ holds in $\mathcal{U}_{\mathcal{I}}$.

Equivalence Constraint (general form)

In fact we will adopt something more general.

Let $n > 1$

Suppose for all $1 \leq i < n$
$\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi_i(x) \rightsquigarrow \varphi_{i+1}(x))$,and $\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi_n(x) \rightsquigarrow \varphi_1(x))$,

then for all $1 \leq i, j \leq n$
if $\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi_i(x) \rightsquigarrow \psi(x))$, $\mathcal{U}_{\mathcal{I}} \models \forall x(\varphi_j(x) \rightsquigarrow \psi(x))$

States

Let Σ be a set of rules, and Δ be a set of sentences. The *state* generated by $I = \langle \Sigma, \Delta \rangle$ is the pair $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ where

- $\mathcal{U}_I$ is the largest class of models of Σ satisfying the three constraints (Exemption, Inheritance, Equivalence) discussed.

- $\mathcal{F}_I$ is the class of all models in $\mathcal{U}_I$ that are models of Δ.

Some examples

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

$$\begin{array}{ll} & \forall x((Px \leadsto Qx) \\ & Pa \\ \hline \therefore & Qa \end{array}$$

$$\begin{array}{ll} & \forall x((Qx \leadsto \neg Px) \\ & Pa \\ \hline \therefore & \neg Qa \end{array}$$

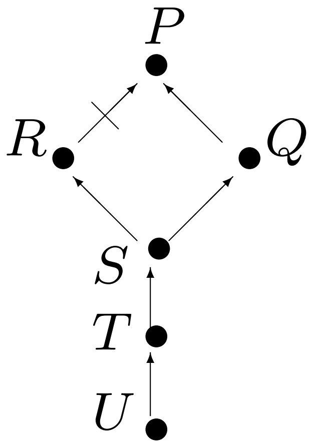
Some Examples 2

$$\begin{array}{ll} & \forall x((Px \leadsto Qx) \\ & \forall x((Qx \leadsto \neg Px) \\ & Pa \\ \hline \therefore & Qa \end{array}$$

Defeasible Modus Ponens beats *Defeasible Modus Tollens*! It does not follow from the premises that $\neg Pa$. The exemption constraint enforces that $\mathcal{U}_{\mathcal{I}} \models \forall x(Px \rightarrow Ab_{Qx,\neg Px}x)$.

Some examples 3

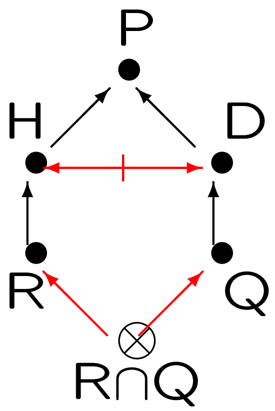
This example illustrates the Inheritance Principle



$$\begin{array}{rl}
 & \forall x(Rx \rightsquigarrow \neg Px) \\
 & \forall x(Qx \rightsquigarrow Px) \\
 & \forall x(Sx \rightsquigarrow Rx) \\
 & \forall x(Sx \rightsquigarrow Qx) \\
 & \forall x(Tx \rightsquigarrow Sx) \\
 & \forall x(Ux \rightsquigarrow Tx) \\
 & Ua \\
\hline
\therefore & Ra \wedge Qa
\end{array}$$

Exemption enforces $\forall x(Sx \rightarrow (Ab_{Rx,\neg Px}x \ \vee \ Ab_{Qx,Px}x))$.
2 x Inheritance gives $\forall x(Ux \rightarrow (Ab_{Rx,\neg Px}x \ \vee \ Ab_{Qx,Px}x))$.

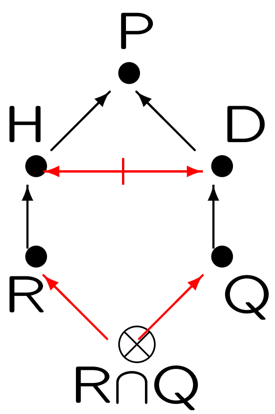
A floating conclusion



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 polically motivated?

A floating conclusion (continued)



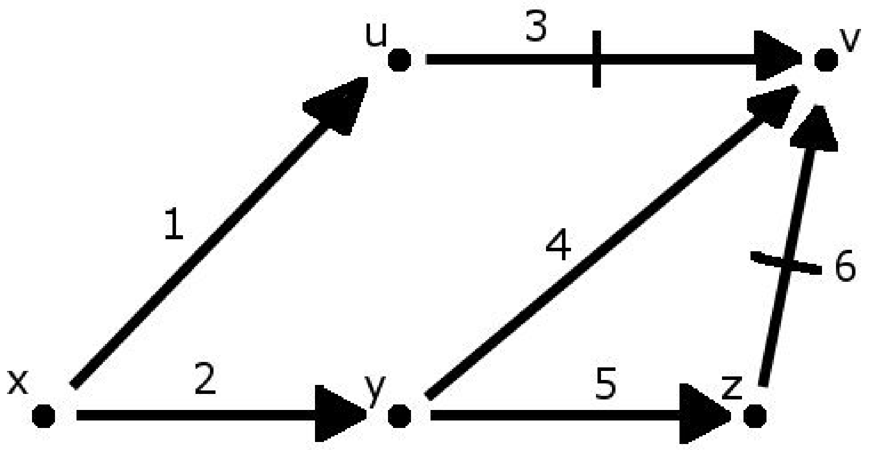
The exemption constraint enforces that in all models Nixon has either the property $Ab_{Rx,Hx}$ or the property $Ab_{Qx,Dx}$.

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 policallly motivated.

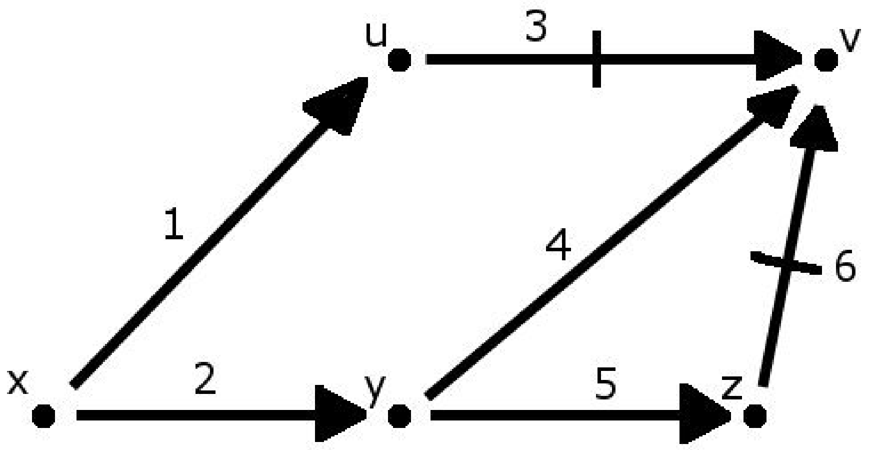
Networks - basics

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.



Networks - basics

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.



Networks - algorithm

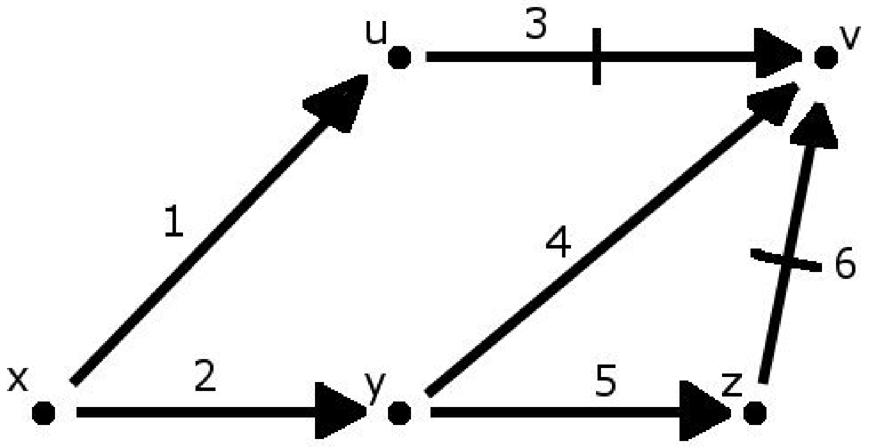
For any node x, $Min(x)$ consists of the strict rules of the network and the arrows starting at x. Where a set of rules allows for contradicting conclusions when starting from x, it is concluded that x is an exception to one of the other rules in that set but not in $Min(x)$.

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).

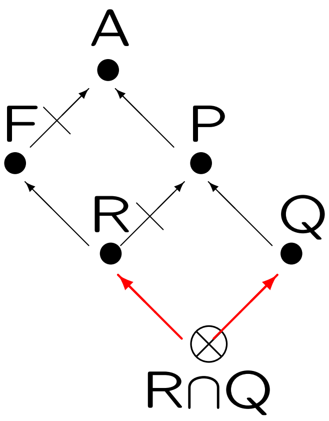
The inheritance principle makes a Backward Induction approach ideal.

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