Making the right exceptions

Harald Bastiaanse & Frank Veltman
ILLC University of Amsterdam

Abstract

This paper provides a principled answer to the question of how to deal with conflicting default rules. It does so in two ways: semantically within a circumscriptive theory, and syntactically by supplying an algorithm for inheritance networks. Arguments that can be expressed in both frameworks are valid on the circumscriptive account if and only if the inheritance algorithm has a positive outcome.
Keywords: circumscription, defaults, nonmonotonic logic, inheritance nets

1 Introduction

Discussions often end before the issues that started them have been resolved. In the eighties and nineties of the previous century default reasoning was a hot topic in the field of logic & AI. The result of this discussion was not one single theory that met with general agreement, but a collection of alternative theories, each with its merits, but none entirely satisfactory. This paper aims to give a new impetus to this discussion.

The issue is the logical behavior of sentences of the form

S's are normally P

Such sentences function as default rules: when you are confronted with an object with property S, and you have no evidence to the contrary, you are legitimized to assume that this object has property P.

The 'evidence to the contrary' can vary. Sometimes it simply consists in the empirical observation that the object concerned is in fact an exception to the rule. On other occasions the evidence may be more indirect. Consider:

premise 1	*A's are normally E*
premise 2	*S's are normally not E*
premise 3	*S's are normally A*
premise 4	*c is A and c is S*
by default	*c is not E*

This is a case of conflicting defaults.[1] At first sight one might be tempted to draw both the conclusion that *c is E* (from premises 1 and 4) and that *c is not E* (from premises 2 and 4) , and maybe on second thought to draw neither. But the third premise states that objects with the property S normally have the property A as well. So, apparently, *normal* S's are *exceptional* A's, as the rule that *A's are normally E* does not hold for them. In other words, only the S-defaults apply to c. So, presumably, c is not E.

Default reasoning has been formalized in various ways, and within each of the existing theoretical frameworks a number of strategies have been proposed to deal with conflicting defaults — many of them rather *ad hoc*. In the following we will focus on two of these frameworks, Circumscription, and Inheritance Nets[2], and implement a new, principled strategy to deal with conflicting rules in each of these.

2 Naive Circumscription

Within the circumscriptive approach a sentence of the form *S's are normally P* is represented by a formula of the form

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

Here $Ab_{SxPx}\, x$ is a one place predicate. The subscript '$SxPx$' serves as an index, indicating the rule concerned. If an object a satisfies the formula $Ab_{SxPx}\, x$, this means that a is an *abnormal* object with respect to this rule.

More generally, let $\mathcal{L}_0$ be a language of *monadic* first order logic. With each pair $\langle\varphi(x), \psi(x)\rangle$[3], we associate a new one-place predicate $Ab_{\varphi(x)\psi(x)}$, thus obtaining the first order language $\mathcal{L}$.

[1] If a concrete example is wanted, substitute '*adult*' for A, '*employed*' for E, and '*student*' for S.

[2] See [1],[2],[3]

[3] Notation: we write $\varphi(x)$ to denote a formula φ of $\mathcal{L}_0$ in which (at most) the variable x occurs freely

A *default rule* is a formula of $\mathcal{L}$ of the form

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

Here, $\varphi(x)$ and $\psi(x)$ must be formulas of $\mathcal{L}_0$ that are quantifier-free and in which no individual constant occurs. The formula $\varphi(x)$ is called the *antecedent* of the rule, $Ab_{\varphi(x)\psi(x)}\, x$ is its *abnormality clause*, and $\psi(x)$ its *consequent*. Again, the index $\varphi(x)\psi(x)$ is there just to indicate that it concerns the abnormality predicate of the rule with antecedent $\varphi(x)$ and consequent $\psi(x)$. When it is clear which variable is at stake we will write $Ab_{\varphi\psi}$ rather than $Ab_{\varphi(x)\psi(x)}$. And often we will shorten '$\forall x((\varphi(x) \wedge \neg Ab_{\varphi\psi}\, x) \rightarrow \psi(x))$' further to

$$\forall x(\varphi(x) \rightsquigarrow \psi(x))$$

Since it is clear from the antecedent and the consequent of a default rule what the abnormality clause is, this should not cause confusion.[4]

In ordinary logic, for an argument to be valid, the conclusion must be true in *all* models in which the premises are true. The basic idea underlying circumscription is that not all models of the premises matter but only the most normal ones — only the ones in which the extension of the abnormality predicates is minimal given the information at hand. Formally:

Definition 2.1

(i) Let $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I} \rangle$ and $\mathfrak{A}' = \langle \mathcal{A}', \mathcal{I}' \rangle$ be two models with the following properties:

(a) $\mathcal{A} = \mathcal{A}'$

(b) for all individual constants c, $\mathcal{I}(c) = \mathcal{I}'(c)$

[4]Some readers may not like the fact that in this set up the formulas $\forall x(Sx \rightsquigarrow Px)$ and $\forall y(Sy \rightsquigarrow Py)$ are not logically equivalent, because they contain different abnormality predicates. We could remedy this defect by introducing the same abnormality predicate $Ab_{\varphi(\cdot)\psi(\cdot)}$ for all pairs $\langle \varphi(x), \psi(x) \rangle$, independent of the free variable x occurring in $\varphi(x)$ and $\psi(x)$. Here '$\cdot$' refers to a symbol that does not belong to the vocabulary of $\mathcal{L}_0$, and by $\varphi(\cdot)$, we mean the expression that one obtains from $\varphi(x)$ by replacing each free occurrence of x by an occurrence of $\cdot$.

Some readers may insist that on top of this we should enforce that whenever $\varphi(x)$ is logical equivalent to $\chi(x)$, and $\psi(x)$ to $\theta(x)$, $\forall x(\varphi(x) \rightsquigarrow \psi(x))$ gets equivalent to $\forall x(\chi(x) \rightsquigarrow \theta(x))$. This can be done by stipulating that we are only interested in models that assign the same extension to $Ab_{\varphi(\cdot)\psi(\cdot)}$ and $Ab_{\chi(\cdot)\theta(\cdot)}$ if $\varphi(x)$ is logical equivalent to $\chi(x)$ and $\psi(x)$ to $\theta(x)$. However, for our purposes, we can keep things simple.

(c) *for all predicates* $Ab_{\varphi\psi}$, $\mathcal{I}(Ab_{\varphi\psi}) \subseteq \mathcal{I}'(Ab_{\varphi\psi})$

Then $\mathfrak{A}$ is at least as normal as $\mathfrak{A}'$. *If* $\mathfrak{A}$ *is at least as normal as* $\mathfrak{A}'$, *but* $\mathfrak{A}'$ *is not at least as normal as* $\mathfrak{A}$, *then* $\mathfrak{A}$ is more normal than $\mathfrak{A}'$.

(ii) *Let* $\mathfrak{C}$ *be a class of models. Then* $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I} \rangle$ *is an* optimal *model in* $\mathfrak{C}$ *iff* $\mathfrak{A} \in \mathfrak{C}$ *and there is no model in* $\mathfrak{C}$ *that is more normal than* $\mathfrak{A}$.

(iii) *Let* Δ *be a set of sentences. Then* $\Delta \models_c \varphi$ *iff* φ *is true in all optimal models of* Δ.

If $\Delta \models_c \varphi$, we say that φ follows by circumscription from Δ. Here is an example of an argument for which this is so.

premise 1	*Adults normally have a bank account*
premise 2	*Adults normally have a driver's license*
premise 3	*John is an adult*
premise 4	*John does not have a driver's license*
by default	*John is an adult with a bank account*

This can be formalized as

premise 1	$\forall x((Ax \wedge \neg Ab_{AB}\, x) \rightarrow Bx)$
premise 2	$\forall x((Ax \wedge \neg Ab_{AD}\, x) \rightarrow Dx)$
premise 3	Aj
premise 4	$\neg Dj$
by circumscription	Bj

This example illustrates why the abnormality predicates have a double index referring to both the antecedent and the consequent of the rule, rather than a single one referring to just the antecedent. It is not sufficient to distinguish between normal and abnormal A's, and formalize a sentence like *Adults normally have a bank account* as $\forall x((Ax \wedge \neg Ab_A\, x) \rightarrow Bx)$. The distinction has to be more fine grained. An object with the property A can be a normal A in some respects and an abnormal A in other. Even though John is an abnormal adult in not having a driver's license, he is a normal adult in having a bank account, or at least we want to be able to conclude by default that he is. If we had formalized the argument in the following way, we would not have gotten very far.

premise 1	$\forall x((Ax \wedge \neg Ab_A\, x) \rightarrow Bx)$
premise 2	$\forall x((Ax \wedge \neg Ab_A\, x) \rightarrow Dx)$
premise 3	Aj
premise 4	$\neg Dj$

Let us now look at the case of conflicting defaults introduced at the end of section 1. The formalized version looks like this:

premise 1	$\forall x(Ax \rightsquigarrow Ex)$
premise 2	$\forall x(Sx \rightsquigarrow \neg Ex)$
premise 3	$\forall x(Sx \rightsquigarrow Ax)$
premise 4	$Ac \wedge Sc$
by circumscription	$\neg Ea$

Unfortunately, in this simple set up the conclusion $\neg Ea$ does not follow from the premises. We find two kinds of optimal models: in some the sentences $\neg Ab_{SA}\, c$, $\neg Ab_{S\neg E}\, c$, and $Ab_{AE}\, c$ hold, which is fine, but in the other the sentences $\neg Ab_{SA}\, c$, $Ab_{S\neg E}\, c$, and $\neg Ab_{AE}\, c$ are true.

Recall that in the informal discussion of this example it was suggested that the three default rules involved together imply that objects with property S are *exceptional* A's; normal A's have the property E, but normal S's don't, even though normal S's do have property A.

In the next section we will see how one can enforce that in all models in which these three defaults hold, also the formula $\forall x(Sx \rightarrow Ab_{AE}\, x)$ wil be true. Once we have this, the only optimal models will be models in which $\neg Ab_{SA}\, c$, $\neg Ab_{S\neg E}\, c$, and $\neg Ab_{AE}\, c$ are true. Which means that the conclusion follows.

3 Exemption and Inheritance

In the following, we will distinguish two kinds of rules, rules that allow for exceptions and rules that do not allow for exceptions. So far we only talked about the first kind, but we also want to discuss the second kind. In order to do so, sentences of the form $\forall x(\varphi(x) \rightarrow \psi(x))$ can get a special status as *strict rules*. These strict rules are to be distinguished from universal sentences that are only accidentally true, and they will be treated differently.[5]

[5] It is tempting to introduce a necessity operator in the object language to distinguish rules form accidental statements, but we resist this temptation, and only make the distinction at a meta-level.

The general set up will be this: Let Σ be a set of default and strict rules and Π be a set of sentences. Think of $I = \langle \Sigma, \Pi \rangle$ as the *information* of some agent at some time, where Σ is the set of rules the agent is acquainted with, and Π the agent's factual information. We 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. The universe $\mathcal{U}_I$ must satisfy some additional *constraint* that will be discussed below. $\mathcal{F}_I$ consists of all models in $\mathcal{U}_I$ that are models of Π.

In this set up we can define validity as follows:

$$\Sigma, \Pi \models_d \varphi \textit{ iff for all optimal models } \mathfrak{A} \in \mathcal{F}_I, \mathfrak{A} \models \varphi$$

Read '$\Sigma, \Pi \models_d \varphi$' as '$\varphi$ follows by default from Σ and Π'.

Before we can turn to a discussion of the constraint, we need to introduce some technical notions .

Definition 3.1

(i) 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* $Ab_{\varphi\psi}x$*.*

(ii) Let Σ *be a set of rules, and let* d *be some element of the domain of some model* $\mathfrak{A}$ *of* Σ*. Then* d complies with Σ *(in* $\mathfrak{A}$*) iff* d *complies with all the default rules in* Σ*.*

So, if an object satisfying $\varphi(x)$ complies with $\forall x(\varphi(x) \rightsquigarrow \psi(x))$, it will also satisfy $\psi(x)$. But notice that the definition allows for the following situations:

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

We will present examples later on. For now, just take 'comply' as a technical term.

3.1 The exemption principle

The constraint we will impose on $\mathcal{U}_I$ is motivated by the following minimal requirement.

> If the *only* factual information about some object is that it has property P, it must be valid to infer by default that this object complies with all the default rules for objects with property P.

What would be the use of these rules if they would not at least allow this inference?

It may seem easy to satisfy this requirement, but it is not.

Definition 3.2

(i) *An exemption clause is a formula of the form* $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$*, for* Δ *a set of default rules.*[6]

(ii) *Let* Σ *be a set of rules.* $\Sigma^{\varphi(x)}$ *is the set of rules in* Σ *with antecedent* $\varphi(x)$*.*

(iii) *The exemption clause* $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$ *is an exemption clause* for Σ *iff* $\Sigma \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}} Ab_\delta\, x)$.

To see how these definitions work, consider again

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

Here $\Sigma^{Sx} = \{\forall x(Sx \rightsquigarrow Ax), \forall x(Sx \rightsquigarrow \neg Ex)\}$. Let $\Delta = \{\forall x(Ax \rightsquigarrow Ex)\}$. Clearly, there is no model such that some object in its domain satisfies Sx and complies with $\Delta \cup \Sigma^{Sx}$. So,

$$\Sigma \models \forall x(Sx \to \bigvee_{\delta\in\Delta\cup\Sigma^{Sx}} Ab_\delta\, x)$$

By *(iii)* above this means that $\forall x(Sx \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$, i.e. $\forall x(Sx \to Ab_{AE} x)$, is an exemption clause for Σ, the idea being that objects with property S are, so to speak, *exempted* from the rule that A's are normally E.

[6] Where δ is a default rule, Ab_δ is the abnormality clause of δ. By definition, if $\Delta = \emptyset$, $\bigvee_{\delta\in\Delta} Ab_\delta\, x = \bot$.

The word 'exempted' suggests that default rules are some kind of normative rules. Indeed, often it is helpful to think of them that way. The use of the word 'normally', already suggests that we are dealing with a kind of norms here. To count as a normal S, S's must be A, and to count as a normal A, A's must be E, but here an exception is made for the S's. S's must be A, but they do not have to be E, they are not subjected to this rule. Actually, they must be not E.

In the following definition it is made explicit for any set of rules Σ which kinds of objects are exempted from which rules in Σ.

Definition 3.3 *Let Σ be a set of rules, and let Π be an arbitrary set of formulas.*

(i) The exemption extension $\Sigma^{\in}$ *of Σ is given by*

$$\Sigma^{\in} = \bigcup_{n \in \omega} \Sigma^{\in}_n$$

where $\Sigma^{\in}_0 = \Sigma$ and $\Sigma^{\in}_{n+1} = \Sigma^{\in}_n \cup \{\varphi \mid \varphi$ is an exemption clause for $\Sigma^{\in}_n\}$

(ii) The state generated by $I = \langle \Sigma, \Pi \rangle$ is the state $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ given by

(a) $\mathfrak{A} \in \mathcal{U}_I$ iff $\mathfrak{A}$ is a model of $\Sigma^{\in}$;

(b) $\mathcal{F}_I$ consists of all models in $\mathcal{U}_I$ that are models of Π.

Notice that $\Sigma^{\in}$ has the following property, which we will call the *Exemption Principle*.

$$\text{If } \Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta \cup \Sigma^{\varphi(x)}} Ab_\delta\, x), \text{ then } \Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta} Ab_\delta\, x)$$

In fact $\Sigma^{\in}$ is the weakest extension of Σ with this property.

Proposition 3.4 (Minimal requirement)
Suppose $\forall x(\varphi(x) \rightsquigarrow \psi(x)) \in \Sigma$. Then $\Sigma, \{\varphi(c)\} \models_d \psi(c)$.

Proof: Let $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ be the state generated by $I = \langle \Sigma, \{\varphi(c)\} \rangle$. It suffices to show that every optimal model in $\mathcal{F}_I$ has the property that the object named c complies with $\Sigma^{\varphi(x)}$. If $\mathcal{F}_I = \emptyset$, this holds trivially. Suppose $\mathcal{F}_I \neq \emptyset$.

Consider any model $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I} \rangle$ in $\mathcal{F}_I$ in which the object $\mathcal{I}(c)$ does not comply with $\Sigma^{\varphi(x)}$. We will show that $\mathfrak{A}$ is not optimal.

Let Δ be the set of defaults in $\Sigma^{\in}$ with which $\mathcal{I}(c)$ complies. Apparently, $\Sigma^{\in} \not\models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta} Ab_\delta\, x)$. By the exemption principle this means that $\Sigma^{\in} \not\models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta \cup \Sigma^{\varphi(x)}} Ab_\delta\, x)$. Hence, there exists a model $\mathfrak{A}' = \langle \mathcal{A}', \mathcal{I}' \rangle$ in $\mathcal{U}_I$ such that some element d_0 in $\mathcal{A}'$ satisfies $(\varphi(x) \wedge \neg \bigvee_{\delta \in \Delta \cup \Sigma^{\varphi(x)}} Ab_\delta\, x)$.
Now, let $\mathfrak{A}'' = \langle \mathcal{A}'', \mathcal{I}'' \rangle$ be defined as follows:
– $\mathcal{A}'' = \mathcal{A}$;
– For individual constants a, $\mathcal{I}''(a) = \mathcal{I}(a)$;
– For P an ordinary predicate or an abnormality predicate,
  if $d \neq \mathcal{I}(c)$, then $d \in \mathcal{I}''(P)$ iff $d \in \mathcal{I}(P)$, and
  if $d = \mathcal{I}(c)$, then $d \in \mathcal{I}''(P)$ iff $d_0 \in \mathcal{I}'(P)$.
Consider any quantifier-free formula $\theta(x)$ in which no individual constant occurs. Clearly, if $d \neq \mathcal{I}''(c)$, then d satisfies $\theta(x)$ in $\mathfrak{A}''$ iff d satisfies $\theta(x)$ in $\mathfrak{A}$, while $\mathcal{I}''(c)$ satisfies $\theta(x)$ in $\mathfrak{A}''$ iff $I'(c)$ satisfies $\theta(x)$ in $\mathfrak{A}'$.

Given that all sentences of $\Sigma^{\in}$ are of the form $\forall x \theta(x)$ with θ as described, $\mathfrak{A}''$ will be a model of $\Sigma^{\in}$. And clearly, $\mathfrak{A}''$ is more normal than $\mathfrak{A}$. Therefore $\mathfrak{A}$ is not optimal.

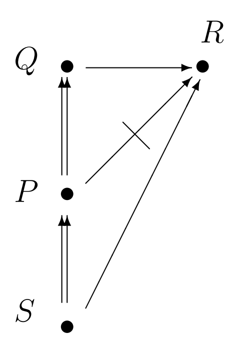
3.2 The Inheritance Property

On the face of it the exemption principle is not very strong. But it is amazing to see its consequences. One is the *inheritance property*, which in its simplest form runs as follows:

> Let Σ be a set of rules. Suppose that $\Sigma^{\in} \models \forall x(\varphi(x) \rightsquigarrow \psi(x))$ and $\Sigma^{\in} \models \forall x(\psi(x) \to Ab_{\chi\theta}\, x)$. Then $\Sigma^{\in} \models \forall x(\varphi(x) \to Ab_{\chi\theta}\, x)$

To see how this works, consider the theory Σ consisting of the following five rules

$\forall x(Qx \rightsquigarrow Rx)$
$\forall x(Px \to Qx)$
$\forall x(Px \rightsquigarrow \neg Rx)$
$\forall x(Sx \to Px)$
$\forall x(Sx \rightsquigarrow Rx)$



Consider the first three rules, and notice that the exemption principle enforces that $\forall x(Px \to Ab_{_{QR}}\, x) \in \Sigma^{\in}$.[7] Now, the inheritance principle yields that $\forall x(Sx \to Ab_{_{QR}}\, x) \in \Sigma^{\in}$. By applying the the exemption principle to the last three rules we also have $\forall x(Sx \to Ab_{_{P\neg R}}\, x) \in \Sigma^{\in}$.

So, all S's are P's but all S's are exceptional P's because they normally have the property R whereas P's normally do not have property R. The P's are exceptional Q's because Q's normally do have the property R. Now, does this make the S's normal Q 's? No! The S's neither count as normal P's nor as normal Q's. Exceptions to exceptions are not normal. The S's are doubly exceptional Q's rather than normal Q's. (Would you call a flying penguin a normal bird?)

The example illustrates the fact that it is possible for an object not to comply with a rule whereas both the antecedent and the consequent of the rule hold for it. Objects with the property S do not comply with the rule $\forall x(Qx \rightsquigarrow Rx)$, but in optimal circumstances they will have both the properties Q and R.

We will now prove a general form of the inheritance property.

Proposition 3.5 (Inheritance Property)
Let $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ be the state correlated with with the information $I = \langle \Sigma, \Pi \rangle$. Let $\Delta \subseteq \Sigma$ be a set of default rules.

Suppose

$$\text{(a) } \Sigma^{\in} \models \forall x(\varphi(x) \rightsquigarrow \psi(x)) \text{ and (b) } \Sigma^{\in} \models \forall x(\psi(x) \to \bigvee_{\delta\in\Delta} Ab_{\delta}\, x)$$

Then

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_{\delta}\, x)$$

Proof: By first-order logic alone, it is trivially true that

$$\Sigma^{\in} \models \forall x(\varphi(x) \to (\psi(x) \vee \neg\psi(x)))$$

[7] If a proof is wanted: Take $\Sigma = \{\forall x(Qx \rightsquigarrow Rx), \forall x(Px \to Qx), \forall x(Px \rightsquigarrow \neg Rx)\}$ and $\Delta = \{\forall x(Qx \rightsquigarrow Rx)\}$. Note that $\Sigma \models \forall x(Px \to \bigvee_{\delta\in\Delta\cup\Sigma^{Px}} Ab_{\delta}\, x)$, or simply put $\Sigma \models \forall x(Px \to (Ab_{_{QR}} \vee Ab_{_{P\neg R}}\, x))$; apply the exemption principle to find that $\Sigma^{\in} \models \forall x(Px \to Ab_{_{QR}} x)$.

Given (a), objects that satisfy $\varphi(x)$ and $\neg\psi(x)$ will also satisfy $Ab_{\varphi\psi}\, x$. Thus, the above statement remains true when we replace $\neg\psi(x)$ with $Ab_{\varphi\psi}\, x$. Similarly, given (b) we can replace $\psi(x)$ in the formula above with $\bigvee_{\delta\in\Delta} Ab_\delta\, x$ while keeping the statement true. This gives us

$$\Sigma^{\in} \models \forall x(\varphi(x) \to (\bigvee_{\delta\in\Delta} Ab_\delta\, x \vee Ab_{\varphi\psi}\, x))$$

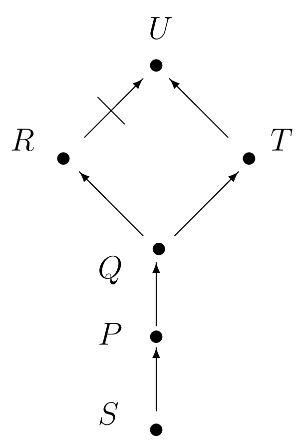
Given the exemption principle this means

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$$

3.3 Some more examples

(i) Using the inheritance principle it is easy to see why the following argument is valid.

$$\begin{array}{ll} & \forall x(Rx \rightsquigarrow \neg Ux) \\ & \forall x(Tx \rightsquigarrow Ux) \\ & \forall x(Qx \rightsquigarrow Tx) \\ & \forall x(Qx \rightsquigarrow Rx) \\ & \forall x(Px \rightsquigarrow Qx) \\ & \forall x(Sx \rightsquigarrow Px) \\ & Sc \\ \hline \therefore & Rc \wedge Tc \end{array}$$



Looking at the first four rules, we see that the exemption principle enforces that $\forall x(Qx \to (Ab_{R\neg U}\, x \ \vee\ Ab_{TU}\, x)) \in \Sigma^{\in}$. By applying the Inheritance Principle twice we see that $\forall x(Sx \to (Ab_{R\neg U}\, x \ \vee\ Ab_{TU}\, x)) \in \Sigma^{\in}$. So in all relevant models either $Ab_{R\neg U}\, c$ or $Ab_{TU}\, c$ is true. From this it follows that in all optimal models $\neg Ab_{S\neg P}$, $\neg Ab_{PQ}$, $\neg Ab_{QR}$, and $\neg Ab_{QT}$ are true, which enables us to conclude by default that Pc, Qc, Rc and Tc.

Notice that on the naive account from section 2 none of these can be concluded. It would not even be possible to make the first step upwards from Sc to Pc. Here we can not only make this first step but also a second to Qc and further up to Rc and Tc. Only when we hit a direct conflict do we need to stop. By having the upper abnormalities propagate downward, we do not have to take into account potential abnormalities at the lower levels.

(ii) Both *Defeasible Modus Ponens* and *Defeasible Modus Tollens* are valid.[8]

$$\begin{array}{ll} & \forall x(Sx \rightsquigarrow Px) \\ & Sc \\ \hline \therefore & Pc \end{array} \qquad\qquad \begin{array}{ll} & \forall x(Sx \rightsquigarrow Px) \\ & \neg Pc \\ \hline \therefore & \neg Sc \end{array}$$

The latter shows that an object need not have property S to count as an object that complies with the rule $\forall x(Sx \rightsquigarrow Px)$. Intuitively, if the object c had property S, it would be an abnormal S. So, assuming that the object c is normal and *complies* with the rule, it will not have property S.

Now, consider the following premises

$$\begin{array}{ll} \textit{premise 1} & \forall x(Sx \rightsquigarrow Px) \\ \textit{premise 2} & \forall x(Px \rightsquigarrow \neg Sx) \\ \textit{premise 3} & Sc \end{array}$$

At first sight one might be tempted to conclude Pc by *Defeasible Modus Ponens* and $\neg Pc$ by *Defeasible Modus Tollens*, but in fact the exemption principle enforces that $\forall x(Sx \rightarrow Ab_{P\neg S}\, x) \in \Sigma^{\in}$. This means that the only default conclusion to be drawn is Pc.

The reason we bring this up is because several authors have questioned the validity of *Defeasible Modus Tollens* with putative counterexamples like the following:

$$\begin{array}{ll} \textit{premise 1} & \textit{Men normally don't have a beard} \\ \textit{premise 2} & \textit{John has a beard} \\ \hline \textit{by default} & \textit{John is not a man} \end{array}$$

However, all this example shows is that one has to be very careful in providing 'intuitive' counterexamples when dealing with default arguments. One must

[8] There is a huge difference between this kind of *Modus Tollens* (From $\forall x(Sx \rightsquigarrow Px)$ and $\neg Pa$ it follows (by default) that $\neg Sa$) and *Contraposition* (From $\forall x(Sx \rightsquigarrow Px)$ it follows that $\forall x(\neg Px \rightsquigarrow \neg Sx)$). For a discussion, see [4].

be sure that the premises faithfully represent *all* one knows about the matter at issue.

In this case we know in fact more than the premises state. For instance, people with a beard normally are men. (This is why the conclusion sounds weird in the first place.)

Now, if we state this explicitly as a third premise we get:

premise 1	*People with a beard normally are men*
premise 2	*Men normally don't have a beard*
premise 3	*John has a beard*

And as we saw, Defeasible modus ponens beats defeasible modus tollens, so the only conclusion to be drawn is that *John is a man*.

3.4 Coherence

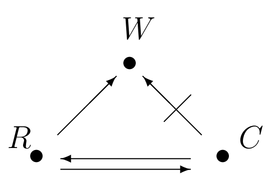
Every set Σ of default rules is consistent.[9] This does not mean that from a logical point of view every such set is okay. Here are some examples.

Consider

$$\Sigma = \{\forall x(Sx \rightsquigarrow Px), \forall x(Sx \rightsquigarrow \neg Px)\}$$

Clearly, a theory of this form is of no use. Note that $\Sigma \models \forall x(Sx \to (Ab_{SP} \vee Ab_{S\neg P}))$. We can apply the exemption principle (take $\Delta = \emptyset$ and $\varphi(x) = Sx$) to find that $\forall x(Sx \to \bot)$ is an exemption clause for Σ. So, $\Sigma^{\in} \models \neg\exists x Sx$.

A more complicated example is this one:



$$\begin{array}{l} \forall x(Rx \rightsquigarrow Cx) \\ \forall x(Cx \rightsquigarrow Rx) \\ \forall x(Rx \rightsquigarrow Wx) \\ \forall x(Cx \rightsquigarrow \neg Wx) \end{array}$$

'Rainy days normally are cold', 'Cold days normally are rainy', 'On rainy days the wind is normally west', 'On cold days the wind is normally not west'. Something is wrong with this theory. The exemption principle does not allow such days: $\Sigma^{\in} \models \neg\exists x Rx$. Proof: note first that $\forall x(Cx \to Ab_{RW}x) \in \Sigma^{\in}$.

[9]To see why this is so, consider a model in which all objects are abnormal in all respects.

By the inheritance property it follows that $\forall x(Rx \to Ab_{RW}) \in \Sigma^{\in}$. Applying the exemption principle once more yields $\forall x(Rx \to \bot) \in \Sigma^{\in}$.

A third example is given by

$$\Sigma = \{\forall x(Sx \rightsquigarrow Px), \forall x((Sx \wedge Qx) \rightsquigarrow \neg Px), \forall x((Sx \wedge \neg Qx) \rightsquigarrow \neg Px)\}$$

Again, this does not sound like an acceptable theory. Too many exceptions are being made. $\Sigma^{\in}$ does not allow this. Note that $\Sigma \models \forall x((Sx \wedge Qx) \to (Ab_{(Sx \wedge Qx)\neg Px}x) \vee Ab_{SxPx}x))$. Hence, by the exemption principle $\forall x((Sx \wedge Qx) \to Ab_{SxPx}x) \in \Sigma^{\in}$; similarly, $\forall x((Sx \wedge \neg Qx) \to Ab_{SP}x) \in \Sigma^{\in}$. Hence, $\Sigma^{\in} \models \forall x(Sx \to Ab_{SP}x)$. But then $\forall x(Sx \to \bot) \in \Sigma^{\in}$.

The above leads to the following definition:

Definition 3.6 *A set of rules Σ is coherent iff for every $\varphi(x)$ which is the antecedent of some rule in Σ, $\Sigma^{\in} \cup \{\exists x\varphi(x)\}$ is consistent.*

A set of rules is incoherent if it is logically impossible to satisfy the minimal requirement. In such a case there is some property such that no object with this property can comply with all the rules for objects with this property. Given the exemption principle, no such objects are allowed.

As will become clear in the next section, for inheritance nets we can give an exact syntactic characterization of the sets of rules that are incoherent.

4 Networks

Inheritance networks are, simply put, the kind of directed graphs we have used to illustrate some of the examples in the previous sections. Thus, an inheritance network is a directed graph where the arrows represent default rules, nodes represent properties and specifically marked arrows are used for negative rules and for strict rules. More formally:

Definition 4.1 *An* Inheritance Network *is a pair $\langle V, \Sigma \rangle$, where each element of Σ is a combination of an ordered pair of elements of V and a* polarity *which may be positive, negative, strict positive or strict negative.*
Elements of Σ are referred to as arrows going from the first element of the ordered pair to the second. We will generally refer to an arrow from u to v as uv if positive, uv^- if negative, uv^* if strict positive and uv^{*-} if strict negative.

Nodes will generally represent properties, but may also represent objects or individuals, provided they are only connected to other nodes by strict arrows. In many examples, there is a single node representing an individual, and strict arrows from it to nodes representing properties indicate that that individual has those properties. The definition we use does not distinguish between nodes representing individuals and nodes representing properties: the difference is purely a matter of interpretation.
For making inferences in these networks, the notion of a *path* is crucial.

Definition 4.2 *Let $\langle V, \Sigma \rangle$ be an Inheritance Network, with $a, b \in V$.*

(i) A positive path *from a to b is a subset $\{\alpha_1, \ldots, \alpha_n\} \subseteq \Sigma$ such that there exist $v_1, \ldots, v_{n-1} \in V$ such that:*

- *α_1 is a positive (or strict positive) arrow from a to v_1*
- *α_i is a positive (or strict positive) arrow from v_{i-1} to v_i, where $1 < i < n$*
- *α_n is a positive (or strict positive) arrow from v_{n-1} to b*

Furthermore, the empty set is considered a positive path from any $v \in V$ to itself.

(ii) $X \subseteq \Sigma$ is a negative path *from a to b if there are X_1, X_2, a', b', α such that*

- *$X = X_1 \cup \{\alpha\} \cup X_2$*
- *X_1 is a positive path from a to a'*
- *X_2 is a positive path from b to b'*
- *α is a negative (or strict negative) arrow from a' to b', or from b' to a'*[10]

If there exists a positive (negative) path from a to b, this serves as *prima facie* evidence that objects with property a have (do not have) property b. Of course, in interesting examples we have prima facie evidence for both b and not b, which brings us to the next key notion: the *conflicting set*.

[10]Note that it's possible that $a = a', b = b'$ and X_1 and X_2 are empty.

Definition 4.3 *Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, a subset $X \subseteq \Sigma$ is a* conflicting set relative to *a iff there is some $b \in V$ such that X contains both a positive and a negative path from a to b.*
Such an X is a minimal conflicting set *if every proper subset of X is not a conflicting set relative to a.*

In the above definition, note that 'minimal' does mean having the least possible number of elements. Rather, it simply means that nothing more can be taken out without losing the property.

4.1 Making inferences in inheritance nets

Let $\langle V, \Sigma \rangle$ be an inheritance network, and $u, v \in V$. In the following we will write $u \asymp v$ to indicate that there is a positive path from u to v and a positive path from v to u.

Definition 4.4 *Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, let*

$$Ess_\Sigma(a) = \{uv \in \Sigma \mid u \asymp a\} \cup \{uv^- \in \Sigma \mid u \asymp a\} \cup \{\alpha \in \Sigma \mid \alpha \textit{ is a strict arrow}\}$$

For a given property a, the set $Ess_\Sigma(a)$ contains the rules that are essential for a, i.e. all rules from which the objects with property a cannot be exempted. No object can be exempted from any strict rule; the objects with property a cannot be exempted from any rule for objects with property a, and more generally, the objects with property a cannot be exempted from any rule for objects with a property b that is "default equivalent" to a.

Definition 4.5 *Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, let*

$$d(a) := \{X - Ess_\Sigma(a) \mid X \text{ is a minimal conflicting set relative to } a\}$$

Note that $d(a)$ is not a set of arrows but rather a set of sets of arrows. The intuition is that the objects with property a are exempted from at least one rule in every set in $d(a)$.

The inheritance property comes in by letting the d function propagate backwards along positive paths, collecting d-sets in the D function defined below.

Definition 4.6 *Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, let*

$$D(a) := \bigcup \{d(b) \mid \text{there is a positive path from } a \text{ to } b\}$$

Thus, $D(a)$ is the union of all $d(b)$ for b's to which there is a positive path from a. Its elements are sets of arrows, just like the elements of each $d(b)$ are.

We are now close to defining the consequence relation for networks. This will be done in terms of *exception sets*, potential sets of default rules (that is, arrows) to which an exception must be made.

Definition 4.7 *Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, $X \subseteq \Sigma$ is an* acceptable exception set of a *iff for all $Y \in D(a)$ there is some $\alpha \in X$ such that $\alpha \in Y$.*

Such an X is a minimal exception set *if every proper subset of X is not an acceptable exception set of a.*

Each minimal exception set represents a way to make as few exceptions as possible. A given conclusion b now follows from a in a network if b can be reached from a under each of these ways.

Definition 4.8 *Let $\langle V, \Sigma \rangle$ be an inheritance network. Let $a, b \in V$.*

- *$a \vdash_\Sigma b$ iff for every minimal exception set X of a there is a positive path Y from a to b such that $X \cap Y = \emptyset$.*
- *$a \vdash_\Sigma \neg b$ iff at least one of the following is true:*
  - *(i) For every minimal exception set X of a there is a negative path Y from a to b such that $X \cap Y = \emptyset$.*
  - *(ii) No minimal exception set X of a is also an acceptable exception set of b.*

We did not prepare the reader for the second clause of negative entailment. It is there for the special case in which there is no path from a to b.[11] In such a case it may happen that objects with property b are so abnormal that one can safely assume that the object under consideration does not have property b. When we know nothing about an object, we like to assume that it is normal in all respects. Thus if objects with property b are never normal in all respects, like a penguin which is either a non-flying bird or an even more abnormal flying penguin, we assume that objects we do not know

[11] If there is a path from a to b every minimal exception set of a is an acceptable exception set for b.

anything about do not have property b. This is certainly how it works in the circumscription semantics. (Note that if b does not force exceptions to be made then any set is an acceptable exception set of b. Thus only 'exceptional' b's are affected.)
We do not cover arguments from complete ignorance here, but the above also holds if we do know something about the object but what we know (in this case, that it has property a) is completely unrelated to b. So for example, if we combine a Nixon Diamond and a Tweety Triangle into a single inheritance network (without adding any extra arrows), this clause lets us conclude that Nixon is presumably not a penguin and vice versa.
Determining exactly when a and b are not sufficiently related is non-trivial. There are situations where a and b are both connected to some third node c, yet still distinct enough that we should allow $a \vdash_\Sigma \neg b$ to follow. The key here is that if a necessarily creates the same abnormalities b does, then someone who already accepts a cannot reject b on the basis of those abnormalities. This is what is stated by the condition that the minimal exception sets for a are all acceptable exception sets for b. We will see in the Appendix that this condition is the correct one for the sake of making the completeness proof work.

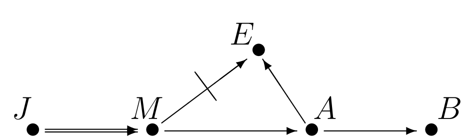
4.2 Examples

4.2.1

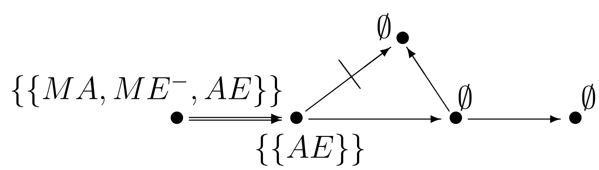
As a first example, we consider the following desirable inference.

premise 1	*Adults normally have a bank account*
premise 2	*Master students are normally adults*
premise 3	*Master students are normally not employed*
premise 4	*Adults are normally employed*
premise 5	J*ohn is a master student*
by default	*John is an adult with a bank account, but he is not employed*

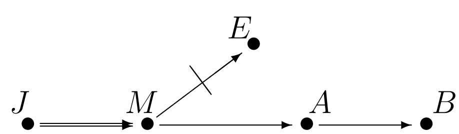
Rendered as an inheritance network, this looks as follows.



Our first step is to determine the d function. Since there are no conflicting sets relative to A, B, and E, we have $d(A) = d(B) = d(E) = \emptyset$. The conflicting sets relative to master student are $\{MA, ME^{-}, AE\}$ and $\{AB, MA, ME^{-}, AE\}$. Only the first of these is minimal. Since $Ess_{\Sigma}(M) = \{MA, ME^{-}, JM\}$, we obtain $d(M) = \{\{AE\}\}$.
Similarly, there is a single minimal conflicting set relative to J: the set $\{MA, ME^{-}, AE, JM\}$. We have $Ess_{\Sigma}(J) = \{JM\}$, so $d(J) = \{\{MA, ME^{-}, AE\}\}$.



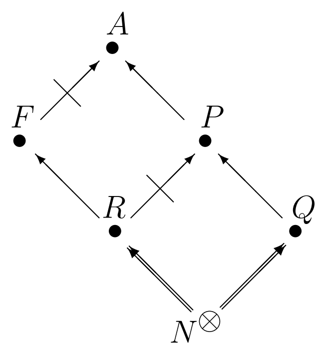
We can now determine $D(J)$. Since there is a positive path from J to every other node, $D(J)$ is the union of all the d's. Only two are non-trivial, so $D(J) = \{\{MA, ME^{-}, AE\}, \{AE\}\}$.
Since $\{AE\} \in D(J)$, every acceptable exception set for John will contain arrow AE. Since $\{AE\}$ is itself an acceptable exception set, this makes it the only minimal exception set. Thus, a conclusion is acceptable iff there is a path from J to it that does not use arrow AE. That is, if there is a path in the following network.



Therefore as desired we obtain $J \vdash_{\Sigma} \neg E$, $J \vdash_{\Sigma} A$, $J \vdash_{\Sigma} B$.

4.2.2 The Double Diamond

The following network is a well-known extension of the Nixon Diamond, generally referred to as the Double Diamond.



premise 1	*Nixon is a Republican and a Quaker*
premise 2	*Quakers are normally Pacifist*
premise 3	*Republicans are normally not Pacifist*
premise 4	*Republicans are normally Football fans*
premise 5	*Pacifists are normally Anti-military*
premise 6	*Football fans are normally not Anti-military*

The questions whether Nixon is Anti-military. In traditional pre-emption based approaches (notably [3]), the positive path from N to A is disabled by the negative path from N to P, so that $\neg A$ may be concluded. This outcome is considered counterintuitive since the negative path to A is itself disabled by its positive counterpart. This has led to paths like that being referred to as *zombie paths.*[5] Since our own approach is not based on this kind of pre-emption, we can do a bit better here.

The first thing to notice is that there are no pairs of conflicting paths starting at P, F, A, R, or Q. Therefore all of them have empty d, and $D(N) = d(N)$.

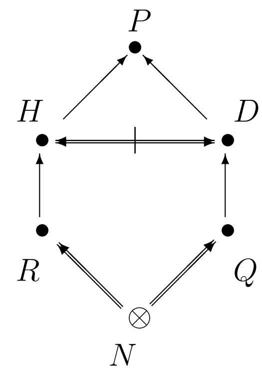
We subsequently find that $D(N) = \{\{QP, RP^-\}, \{QP, RF, PA, FA^-\}\}$. (Details left to the reader.) It is important to keep in mind that "minimal exception set" does not mean "exception set with the smallest amount of elements", meaning that $\{QP\}$ is not the only minimal exception set (relative to N) here. The others are $\{RP^-, RF\}, \{RP^-, PA\}$ and $\{RP^-, FA^-\}$.

We trivially obtain $N \vdash_\Sigma R$, $N \vdash_\Sigma Q$. But as to the other proprties, nothing can be concluded. While this seems natural enough for P and A, some people might see it as counterintuitive for F. However, it should be noted that there is both a positive and a negative path from N to F.

4.2.3 A floating conclusion

The next example is much discussed in the literature on inheritance nets.

premise 1	*Nixon is a Republican and a Quaker*
premise 2	*Quakers are normally Doves*
premise 3	*Republicans are normally Hawks*
premise 4	*Nobody is both a hawk and a dove*
premise 5	*Hawks normally are politically motivated*
premise 6	*Doves normally are politically motivated*

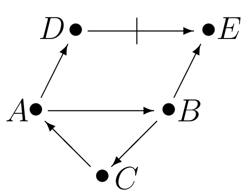


According to the theory presented here, the answer to the question is 'Yes'.[12] It is easy to see that $D(N) = d(N)$. Furthermore, $d(N) = \{\{RH, QD\}\}$ (left to the reader). This means there are two minimal exception sets for N, namely $\{RH\}$ and $\{QD\}$.

The exception set $\{RH\}$ does not contain any element of the rightmost path from N to P, and the exception set $\{QD\}$ does not contain any element of the leftmost path from N to P. Thus, for each minimal exception set there is a positive path from N to P which does not contain any element of that set. Therefore $N \vdash_\Sigma P$.

4.2.4 Closed loops

The algorithm we will present below can also handle inheritance nets with cyclic paths. For an example of how this works, consider the following



premise 1	A's are normally B
premise 2	B's are normally C
premise 3	C's are normally A
premise 4	A's are normally D
premise 5	D's are normally not E
premise 6	B's are normally E
premise 7	x is A
by default	x is E

This example overlaps a small loop with part of a Nixon diamond. At first glance then, one might expect $d(A) = \{\{DE^-, BE\}\}$. However, this is not

[12]This is what most people working in this field want. Horty ([6]) provides a counterexample, but it concerns normative rules rather than defaults. See also Prakken

the case.
Since all points of the loop must be taken into account, we have $Ess_\Sigma(A) = \{AD, AB, BE, BC, CA\}$. Therefore the conflicting set $X = \{AB, AD, DE^-, BE\}$ leads to the inclusion of not $\{DE^-, BE\}$ but rather $X - Ess_\Sigma(A) = \{DE^-\}$ in $d(A)$. Thus, E may be validly concluded when starting at A, B or C.

4.3 An algorithm and examples

The way inheritance works in this system makes a backward-induction approach ideal. Consider the following pseudo-code algorithm to determine d and D across a network.

for $i = 1$ to n **do**
  for each positive path X starting at x_i **do**
    for each negative path Y starting at x_i **do**
      if X and Y have the same endpoint **then**
        $d(x_i) := d(x_i) \cup \{X \cup Y - Ess_\Sigma(x_i)\}$
      end if
    end for
  end for
end for
for $i = 1$ to n **do**
  for $X \in d(x_n)$ **do**
    $D(x_i) := D(x_i) \cup \{X\}$
  end for
  for $j = i + 1$ to n **do**
    if $x_j x_i \in \Sigma$ or $x_j x_i^* \in \Sigma$ **then**
      $D(x_j) := D(x_j) \cup D(x_i)$
    end if
  end for
end for
for $i = 1$ to n **do**
  for $X \in D(x_i)$ **do**
    if $\exists Y \in D(x_i) : Y \subset X$ **then**
      $D(x_i) := D(x_i) - \{X\}$
    end if
  end for

end for

The above will work so long as the nodes have have already been put in backward-induction order; that is, so long as for $i < j$ there is never a positive arrow from i to j. In cases where such an ordering is impossible (ie when the network contains positive loops), the correct results can still be obtained by simply rerunning the parts for D until the results stop changing.[13]

The algorithm is polynomial-time relative to n, Σ and P, where P is the number of paths there are. While we know $|\Sigma| \leq n^2$, P of course cannot be guaranteed to be less than exponential in Σ.

Since it is based on pairs of paths, we know that $|d(x_i)| < 0.5P^2$ for any i. The inheritance thereby puts $|D(x_i)|$ in the order of nP^2. In the absence of a way to reduce this figure, this means the most intensive part of the algorithm is the part where non-minimal elements are removed from D. Indeed, this is why it is generally more efficient to do this at the end (as we do here), rather than on-the-fly inside another loop.

While determining D is the bulk of the work when trying to determine whether $a \vdash_\Sigma b$ or whether the first option for $a \vdash_\Sigma \neg b$ holds, more is needed to check for the second option for $a \vdash_\Sigma \neg b$. Recall that under this item, $a \vdash_\Sigma \neg b$ is true if no minimal exception set X of a is an acceptable exception set of b. Instead of constructing every such X, we will check this for every *choice set* of $D(a)$. A choice set of $D(a)$ is a set $X \subseteq D(a)$ constructed by choosing for each $Y \subseteq D(a)$ one element $y \in Y$ to put in X. Each minimal exception set is contained in such a choice set (left to the reader) and each such choice set contains a minimal exception set (since it is itself an acceptable exception set), so it follows that this has the same result as checking all minimal exception sets.

for each choice set X of $D(a)$ **do**
  Acceptable:=true
  for all $Y \in D(b)$ **do**
    if $X \cap Y = \emptyset$ **then**
      Acceptable:=false
    end if
  end for
  if Acceptable=true **then**
    return $a \not\vdash_\Sigma \neg b$

[13]Specifically constructed perverse examples can necessitate any amount of runs up to n, but generally only a couple should be needed.

end if
end for
return $a \vdash_\Sigma \neg b$

In this algorithm, for each choice set X of $D(a)$, we first assume that X is an acceptable exception set of b and then check if there is a reason to revise this. If it is indeed an acceptable exception set of b then we conclude that $a \not\vdash_\Sigma \neg b$ (we assume the other option for $a \vdash_\Sigma \neg b$ has already been ruled out) and halt the algorithm. Otherwise we move on to the next choice set X. If no choice set X is an acceptable exception set of b, then we conclude that $a \vdash_\Sigma \neg b$.

Of course, the time it takes to create all choice sets is exponential in $|D(a)|$, so one may wish to be careful about when to choose to use this second algorithm.

4.4 Completeness

Networks are a natural way to illustrate most examples even when working in a circumscriptive framework, so it will come as no surprise that the inheritance networks from this chapter can be interpreted in terms of the system we introduced before. However, what is far from trivial is the interpretation can be done in such a way that all the results coincide; that is, that the network-based approach is sound and complete (as to what it can express) relative to the other framework.

We provide the (rather straightforward) translation and the formal statement here. For the extensive proof, see Appendix A.

Definition 4.9 *Let* $N = \langle V, \Sigma \rangle$ *be an inheritance network, with* $V = \{v_1, \ldots, v_n\}$. *We associate with every* $v_i \in V$ *a predicate* P_i, *and with every arrow* α *a rule* α *given by*

$$\begin{aligned} v_i v_j^{\uparrow} &= \forall x(P_i x \rightsquigarrow P_j x) \\ v_i {v_j}^{-\uparrow} &= \forall x(P_i x \rightsquigarrow \neg P_j x) \\ v_i v_j^{*\uparrow} &= \forall x(P_i x \rightarrow P_j x) \\ v_i v_j^{*-\uparrow} &= \forall x(P_i x \rightarrow \neg P_j x) \end{aligned}$$

We will call $\Sigma^{\uparrow} = \{\alpha \mid \alpha \in \Sigma\}$ *the* lift *of* N.

Note that since networks do not distinguish between individuals and properties, the lift will convert to predicates any individuals used in example networks. A premise like "John is an Adult", which in the circumscription framework could be represented as Aj, would be represented in an inheritance network as a strict arrow from J to A, the lift of which would be $\forall x(Jx \to Ax)$.

Theorem 4.10 ***(Soundness-Completeness Theorem)*** *Let $N = \langle V, \Sigma \rangle$ and $\Sigma^{\uparrow}$ be as in the definition. Suppose $\Sigma^{\uparrow}$ is coherent. Then $v_i \vdash_{\Sigma} v_j$ if and only if $\Sigma^{\uparrow}, \{P_i c\} \models_d P_j c$, and $v_i \vdash_{\Sigma} \neg v_j$ if and only if $\Sigma^{\uparrow}, \{P_i c\} \models_d \neg P_j c$.*

Since the above theorem only works in the case of coherence, it is desirable to have a comparable network-based notion. This is where the following theorem comes in. Again, the proof can be found in Appendix A.

Theorem 4.11 *Let $\Sigma^{\uparrow}$ be the lift of the inheritance network $\langle V, \Sigma \rangle$. Then $\Sigma^{\uparrow}$ is coherent if and only if there is no $v \in V$ with $\emptyset \in d(v)$.*

Broadly speaking, the latter is the case if two equivalent points yield unresolvably different conclusions about a third. This is made explicit by the following definition and proposition, also proven in the Appendix.

Definition 4.12 *The vertex x* semi-strictly implies *(*semi-strictly refutes*) y if there is a positive (negative) path from x to y where every arrow after the first is strict.*

Proposition 4.13 *Let $\langle V, \Sigma \rangle$ be an inheritance network. If $\emptyset \in d(x)$, then there are some z and some $y \approx x, y' \approx x$ such that y semi-strictly implies z and y' semi-strictly refutes z.*

5 Conclusion

In the above we have studied the logical properties of defaults, or more particularly of sentences of the form *S's are normally P*. We have shown that their capricious logical behavior can be wholly explained on the basis of one simple underlying principle that determines in cases of conflicting defaults which objects are *exempted* from which rules. We have developed the theory both semantically (within a circumscriptive theory) and syntactically (using inheritance nets). In the appendix we will prove a completeness theorem showing that arguments that can be expressed in both systems are valid on the one account iff they are valid on the other.

Despite the length of this paper, we have only taken the first steps developing these systems. Undoubtedly, a more systematic *model theoretic* study of the circumscriptive part will result in a more elegant proof of the completeness theorem. We also think that on the *algorithmic* side further investigations may yield simplifications. For example, things get a lot less complicated (and complex) if the nets do not have cycles. Finally, a study like this should be complemented by a study which answers the question under which conditions a set of default rules can be safely adopted as a guiding line for taking decisions. Maybe this is a question for methodologists rather than for logicians, but the answer is important to everybody interested in common sense reasoning.

Bibliography

References

[1] J. McCarthy, Circumscription: A form of non-monotonic reasoning, in: M. L. Ginsberg (Ed.), Readings in Nonmonotonic Reasoning, Kaufmann, Los Altos, CA, 1987, pp. 145–151.

[2] J. McCarthy, Applications of circumscription to formalizing common sense knowledge, in: V. Lifschitz (Ed.), Formalizing Common Sense: Papers by John McCarthy, Ablex Publishing Corporation, Norwood, New Jersey, 1990, pp. 198–225.
URL `citeseer.ist.psu.edu/mccarthy86applications.html`

[3] J. F. Horty, R. M. Thomason, D. S. Touretzy, A skeptical theory of inheritance in nonmonotonic semantic networks, Artificial Intelligence 42 (1990) 311–348.

[4] M. Caminada, On the issue of contraposition of defeasible rules, Frontiers in Artificial Intelligence and Applications 172 (2008) 109.

[5] D. Makinson, K. Schlechta, Floating conclusions and zombie paths: two deep difficulties in the "directly skeptical" approach to defeasible inheritance nets, Artificial Intelligence 48 (1991) 199–209.

[6] J. F. Horty, Skepticism and floating conclusions, Artificial Intelligence 135 (12) (2002) 55 – 72. doi:http://dx.doi.org/10.1016/S0004-3702(01)00160-6.

A Completeness of Networks relative to the Semantics

When defining the d and D functions we already suggested that they amount to implementing the inheritance property and a weak version of the exemption principle. Before starting with the completeness proof proper, we will first make this explicit and prove that when working with inheritance networks the combination of this weak version of the exemption principle and the inheritance property is equivalent to the regular exemption principle.

A.1 New constraints, same consequences

Definition A.1 *Let Σ be a set of rules. The formulas φ and ψ are* equivalent in Σ *iff there are $\varphi_1, \ldots, \varphi_m$, $\psi_1, \ldots, \psi_n$ such that $\varphi_m = \psi = \psi_1$, $\psi_n = \varphi = \varphi_1$ and for all $1 \leq i < m$, $1 \leq j < n$*

$$\begin{array}{rcl} \Sigma & \models & \forall x(\varphi_i(x) \rightsquigarrow \varphi_{i+1}(x)) \\ \Sigma & \models & \forall x(\psi_j(x) \rightsquigarrow \psi_{j+1}(x)) \end{array}$$

We denote this as $\varphi \approx_\Sigma \psi$, or simply $\varphi \approx \psi$ is no confusion is possible.

Definition A.2

(i) The clause $\forall x(\varphi(x) \rightarrow \bigvee_{\delta\in\Delta} Ab_\delta x)$ is an expanded exemption clause for Σ iff there are $\psi_1 \approx \psi_2 \approx \ldots \approx \psi_n \approx \varphi$ such that

$$\Sigma \models \forall x(\varphi(x) \rightarrow \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_1(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

.

(ii) The expanded weak exemption extension Σ^W of Σ is given by

$$\Sigma^W = \Sigma \cup \{\varphi \mid \varphi \textit{ is an expanded exemption clause for } \Sigma\}$$

Definition A.3

(i) The clause $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$ is an inherited clause for Σ iff there is some ψ such that $\forall x(\varphi(x) \rightsquigarrow \psi(x)) \in \Sigma$ and $\Sigma \models \forall x(\psi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$.

(ii) The inheritance extension Σ^I of Σ is given by

$$\Sigma^I = \bigcup_{n\in\omega} \Sigma^I_n$$

where $\Sigma^I_0 = \Sigma$ and $\Sigma^I_{n+1} = \Sigma^I_n \cup \{\varphi \mid \varphi \textit{ is an inherited clause for } \Sigma^I_n a\}$

Theorem A.4 $\Sigma^\in \models \Sigma^{WI}$

Proof: We first prove that $\Sigma^\in \models \Sigma^W$. Let $\theta \in \Sigma^W$. We may assume that $\Sigma \not\models \theta$ (otherwise $\Sigma^\in \models \theta$ follows immediately). Therefore θ is of the form

$$\theta = \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$$

with

$$\Sigma \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_1(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

for some $\psi_1 \approx \psi_2 \approx \ldots \approx \psi_n \approx \varphi$.
Since $\psi_1 \approx \varphi$, (repeated) use of the inheritance property lets us conclude

$$\Sigma^\in \models \forall x(\psi_1(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_1(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

By taking $\Delta' = \Delta \cup \Sigma^{\varphi(x)}$, we may use the exemption principle to conclude

$$\Sigma^\in \models \forall x(\psi_1(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_2(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

Now by (repeatedly) using the inheritance property again we arrive at

$$\Sigma^\in \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_2(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

The same process can be repeated for all ψ_i, leaving us with

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta \cup \Sigma^{\varphi(x)}} Ab_\delta x)$$

from which it follows through the exemption principle that

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta} Ab_\delta x)$$

This proves that $\Sigma^{\in} \models \Sigma^{W}$. Therefore $\Sigma^{\in I} \models \Sigma^{WI}$. Since the exemption principle implies the inheritance property, $\Sigma^{\in} \models \Sigma^{\in I}$, and thus $\Sigma^{\in} \models \Sigma^{WI}$.

How about $\Sigma^{WI} \models \Sigma^{\in}$? We doubt this holds for every Σ, but it does hold for the special case that Σ is the lift of an inheritance network. Before we turn to the proof of this statement some more observations are needed.

The rules and exemption clauses figuring in the sets $(\Sigma^{\uparrow})^{WI}$ have a very specific syntactic form, which gives us a lot of freedom when we construct models of such sets. For example, all the sentences concerned are universal, so every $(\Sigma^{\uparrow})^{WI}$ is preserved under submodels. Note also that if the only difference between two models $\mathcal{A}$ and $\mathcal{A}'$ is that $\mathcal{A}'$ has more abnormalities than $\mathcal{A}$, then $\mathcal{A}'$ will be a model of $(\Sigma^{\uparrow})^{WI}$ if $\mathcal{A}$ is. This also holds if for some some predicates P_i that do not occur in the consequent of any rule in $(\Sigma^{\uparrow})^{WI}$, the extension of P_i in $\mathcal{A}'$ is a subset of the extension of P_i in $\mathcal{A}$. More precisely:

Lemma A.5 *Let $\Sigma^{\uparrow}$ be the lift of an inheritance network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_m\}$. Let Γ consist of sentences of the form $\forall x(Q_j x \to \bigvee_{\delta \in \Delta_j} Ab_\delta x)$. Let $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I}\rangle$ and $\mathfrak{A}' = \langle \mathcal{A}', \mathcal{I}'\rangle$ be two models with the following properties:*

(a) $\mathfrak{A} \models \Sigma^{\uparrow} \cup \Gamma$;

(b) $\mathcal{A} = \mathcal{A}'$;

(c) for all individual constants c, $\mathcal{I}(c) = \mathcal{I}'(c)$;

(d) for all predicates P_i, the following holds:

- *(da) If P_i does not occur in the consequent of any rule in $\Sigma^{\uparrow}$, then $\mathcal{I}'(P_i) \subseteq \mathcal{I}(P_i)$;*

(db) Otherwise, $\mathcal{I}'(P_i) = \mathcal{I}(P_i)$*;*

(e) for all predicates $Ab_{P_iP_j}$, $\mathcal{I}(Ab_{P_iP_j}) \subseteq \mathcal{I}'(Ab_{P_iP_j})$*;*

Then $\mathfrak{A}' \models \Sigma^{\uparrow} \cup \Gamma$.

Proof: Left to the reader.

On the way to the completeness theorem, we are often looking for correspondences between notions that play a role in inheritance nets on the one hand and notions in the circumscription framework on the other. One such notion is the notion of a *path*.

Note that if in a network $\langle V, \Sigma\rangle$ there is a positive path from v_i to v_j, then $\Sigma^{\uparrow} \models \forall x((P_i x \wedge \bigwedge_{\alpha\in X} \neg Ab_\alpha x) \rightarrow P_j x)$.[14] For coherent theories the converse is also true. This follows immediately from the following more general proposition.

Lemma A.6 *Let* $\Sigma^{\uparrow}$ *be the lift of an inheritance network* $\langle V, \Sigma\rangle$*, with* $V = \{v_1, \ldots, v_m\}$*. Let* Γ *consist of sentences of the form* $\forall x(Q_j x \rightarrow \bigvee_{\delta\in\Delta_j} Ab_\delta x)$*. Let* $\varphi(x)$ *be a quantifier-free formula in which all predicates are abnormality predicates, and such that* $\Sigma^{\uparrow}\cup\Gamma\cup\{\exists x(P_i x\wedge\varphi(x))\}$ *is consistent. If* $\Sigma^{\uparrow}\cup\Gamma \models \forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$*, then there is a positive path from* v_i *to* v_j*.*

Proof: We proceed with an unusual induction, one on the number of distinct consequents of rules in $\Sigma^{\uparrow}$.
Case n=0: If $\emptyset\cup\Gamma \models \forall x((P_i x\wedge\varphi(x)) \rightarrow P_j x)$ and $\emptyset\cup\Gamma\cup\{\exists x(P_i x\wedge\varphi(x))\}$ is consistent, then $i = j$. So, there is a path from v_i to v_j.
Induction Hypothesis: The theorem is true if the number of distinct consequents occurring in the rules of $\Sigma^{\uparrow}$ is at most n.
Case n+1: Let $\Sigma^{\uparrow}$ have $n+1$ such consequents.

We first show that there is at least one l such that $\Sigma^{\uparrow}$ contains the rule $\forall x((P_l x \wedge \neg Ab_{P_lP_j} x) \rightarrow P_j x)$.

Suppose there is no such l. Given the fact that both $\Sigma^{\uparrow}$ and Γ consist of universal sentences we can construct a model $\mathfrak{A}$ of $\Sigma^{\uparrow}\cup\Gamma\cup\{\forall x(P_i x\wedge\varphi(x))\}$. Since $\Sigma^{\uparrow}\cup\Gamma \models \forall x((P_i x\wedge\varphi(x)) \rightarrow P_j x)$, $\forall x P_j x$ is true in $\mathfrak{A}$. Now, notice that if we change the interpretation of P_j in $\mathfrak{A}$, while leaving the interpretation of all other predicates the same, the resulting model $\mathfrak{A}'$ will still be a model

[14] We are a bit sloppy here. We should have written '$Ab_{\alpha^{\uparrow}}$' instead of 'Ab_α', because it concerns the abnormality predicate of the lift $\alpha^{\uparrow}$ of the arrow α.

of Γ, $\forall x(P_i x \wedge \varphi(x))$, and also $\Sigma^{\uparrow}$, the latter by lemma A.5. However, $\mathfrak{A}'$ is not a model of $\forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$ any more. This contradicts the fact that $\Sigma^{\uparrow} \cup \Gamma \models \forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$.

Now, let L be the set of l for which $\Sigma^{\uparrow}$ contains the rule $\forall x(P_l x \wedge \neg Ab_{P_l P_j} x \rightarrow P_j x)$. Let $\Sigma^{\uparrow}_{-j}$ be $\Sigma^{\uparrow}$ with all rules in which P_j is the consequent removed. The next claim is that for at least one $l \in L$, $\Sigma^{\uparrow}_{-j} \cup \Gamma \models \forall x(P_i x \wedge \varphi(x) \rightarrow P_l x)$.

To prove this let $\mathfrak{A}_{-j}$ be a model of $\Sigma^{\uparrow}_{-j} \cup \Gamma \cup \{\forall x(P_i x \wedge \varphi(x))\}$. Suppose the claim does not hold. Then we can change the interpretation of P_l for all $l \in L$ in such a manner that $\forall x \neg P_l x$ gets true for all $l \in L$ and such that $\forall x \neg P_j x$ gets true, while leaving the interpretation of all other predicates the same, without affecting the truth of the sentences in $\Sigma^{\uparrow}_{-j} \cup \Gamma \cup \{\forall x(P_i x \wedge \varphi(x))\}$. (Again by Lemma A.5.)

The model would then trivially make true all default rules with $P_l x$ in the antecedent for any $l \in L$, and therefore be a model of $\Sigma^{\uparrow} \cup \Gamma$. However, it would make $\forall x(P_i x \wedge \varphi(x) \rightarrow P_j x)$ false, contradicting the fact that $\Sigma^{\uparrow} \cup \Gamma \models \forall x(P_i x \wedge \varphi(x) \rightarrow P_j x)$.

So we find some l such that $\Sigma^{\uparrow}$ contains a rule of the form $\forall x(P_l x \wedge \neg Ab_{P_{l'} P_j} x \rightarrow P_j x)$ and $\Sigma^{\uparrow}_{-j} \cup \Gamma \models \forall x(P_i x \wedge \varphi(x) \rightarrow P_l x)$. Note that $\Sigma^{\uparrow}_{-j}$ has n distinct consequents of default rules in it. Thus by the induction hypothesis there is a positive path from v_i to v_l. The rule $\forall x(P_{l'} x \wedge \neg Ab_{P_{l'} P_j} x \rightarrow P_j x)$ corresponds to an arrow from $v_{l'}$ to v_j, extending the path to one from v_i to v_j.

Theorem A.7 *If $\Sigma^{\uparrow}$ is the lift of an inheritance network $\langle V, \Sigma \rangle$ and is coherent, then $(\Sigma^{\uparrow})^{WI} \models (\Sigma^{\uparrow})^{WI\in}$*

Proof: It suffices to show that $(\Sigma^{\uparrow})^{WI}$ satisfies the exemption principle. So, let θ, θ' be any clauses of the form below:

$$\theta = \forall x(P_i x \rightarrow \bigvee_{\delta \in \Delta} Ab_\delta x)$$

$$\theta' = \forall x(P_i x \rightarrow \bigvee_{\delta \in \Delta \cup (\Sigma^{\uparrow})^{P_i x}} Ab_\delta x)$$

We have to prove that whenever such a θ' is implied by $(\Sigma^{\uparrow})^{WI}$, so is θ.

Suppose $(\Sigma^{\uparrow})^{WI} \models \theta'$. Note first that if $\Sigma^{\uparrow} \models \theta'$, then $(\Sigma^{\uparrow})^{W} \models \theta$ by construction (because $\varphi \approx \varphi$), and we're done. So, the interesting case is when $\Sigma^{\uparrow} \not\models \theta'$.

Set $(\Sigma^{\uparrow})^{WI} = \Sigma^{\uparrow} \cup \{\phi_1, \phi_2, \ldots\}$, where

$$\phi_j = \forall x(Q_j x \to \bigvee_{\delta \in \Delta_j} Ab_\delta x)$$

For any k, let $(\Sigma^{\uparrow})^{WI}_k = \Sigma^{\uparrow} \cup \{\phi_1, \ldots, \phi_k\}$.

Now, there is some n such that $(\Sigma^{\uparrow})^{WI}_{n-1} \not\models \theta'$, while $(\Sigma^{\uparrow})^{WI}_n \models \theta'$. Two important things are true about ϕ_n.

Claim 1: There is a path from the node corresponding to P_i to the node corresponding to Q_n.

To show this, consider a model $\mathfrak{A}_1$ of $(\Sigma^{\uparrow})^{WI}_{n-1}$ where θ' is not true. Define $\mu(x) = \bigwedge_{\delta \in \Delta \cup (\Sigma^{\uparrow})^{P_i x}} \neg Ab_\delta x$. Thus, $\mathfrak{A}_1 \models \exists x(P_i x \wedge \mu(x))$, demonstrating that $(\Sigma^{\uparrow})^{WI}_{n-1} \cup \{\exists x(P_i x \wedge \mu(x))\}$ is consistent.

If the claimed path does not exist, contraposition of Lemma A.6 tells us it cannot be the case that $(\Sigma^{\uparrow})^{WI}_{n-1} \models \forall x((P_i x \wedge \mu(x)) \to Q_n x)$. Thus there is a model $\mathfrak{A}_2$ of $(\Sigma^{\uparrow})^{WI}_{n-1}$ with some element d satisfying $P_i x \wedge \neg Q_n x \wedge \mu(x)$. Restrict $\mathfrak{A}_2$ to d to get $\mathfrak{A}_3$. Since $\mathfrak{A}_3 \models \forall x \neg Q_n x$, trivially $\mathfrak{A}_3 \models \phi_n$. Therefore $\mathfrak{A}_3 \models (\Sigma^{\uparrow})^{WI}_n$. However, $\mathfrak{A}_3 \not\models \theta'$, contradicting the choice of n. This contradiction proves the claimed path must exist.

Claim 2: $\Delta_n \subseteq \Delta \cup (\Sigma^{\uparrow})^{P_i x}$.

For this, let $\mathfrak{A}_4$ be the restriction of $\mathfrak{A}_1$, above, to elements satisfying $P_i x \wedge \mu(x)$. Construct $\mathfrak{A}_5$ from $\mathfrak{A}_4$ by making $\forall x Ab_\delta x$ true for all $\delta \in \Delta_n - (\Delta \cup (\Sigma^{\uparrow})^{P_i x})$. Note that $\mathfrak{A}_5$ is still a model of $(\Sigma^{\uparrow})^{WI}_{n-1}$. Also, $\mathfrak{A}_5 \models \neg\theta'$. However, if the claim is false then trivially $\mathfrak{A}_5 \models \phi_n$ and hence $\mathfrak{A}_5 \models (\Sigma^{\uparrow})^{WI}_n$. By contradiction, the claim must be true.

Having proven these claims, we now distinguish two cases, depending on where ϕ_n was added.

Case I: $\phi_n \in (\Sigma^{\uparrow})^W$.

If $\phi_n \in (\Sigma^{\uparrow})^W$, then there are $Q'_1 \approx \ldots \approx Q'_u \approx Q_n$ such that

$$\Sigma^{\uparrow} \models \forall x(Q_n x \to \bigvee_{\delta \in \Delta_n \cup (\Sigma^{\uparrow})^{Q_n} \cup (\Sigma^{\uparrow})^{Q'_1} \cup \ldots \cup (\Sigma^{\uparrow})^{Q'_u}} Ab_\delta x).$$

Given that $\Delta_n \subseteq \Delta \cup (\Sigma^{\uparrow})^{P_i x}$, this implies

$$\Sigma^{\uparrow} \models \forall x(Q_n x \to \bigvee_{\delta \in \Delta \cup (\Sigma^{\uparrow})^{P_i} \cup (\Sigma^{\uparrow})^{Q_n} \cup (\Sigma^{\uparrow})^{Q'_1} \cup \ldots \cup (\Sigma^{\uparrow})^{Q'_u}} Ab_\delta x).$$

This leaves two possibilities. If $Q_n \approx P_i$ then $(\Sigma^{\uparrow})^W \models \forall x(Q_n x \to \bigvee_{\delta\in\Delta} Ab_\delta x)$, which implies $\theta \in (\Sigma^{\uparrow})^{WI}$.
If it is not the case that $Q_n \approx P_i$, the above can be simplified to

$$\Sigma^{\uparrow} \models \forall x(Q_n x \to \bigvee_{\delta\in\Delta\cup(\Sigma^{\uparrow})^{Q_n}\cup(\Sigma^{\uparrow})^{Q'_1}\cup\ldots\cup(\Sigma^{\uparrow})^{Q'_u}} Ab_\delta x).$$

To prove this, let χ be the simplified formula and χ' the unsimplified one. Suppose there is no path from the node corresponding to Q_n to the node corresponding to P_i and yet $\Sigma^{\uparrow} \not\models \chi$. Define $\mu(x)$ as follows:

$$\mu(x) = \left(\bigwedge_{\delta\in\Delta\cup(\Sigma^{\uparrow})^{Q_n}\cup(\Sigma^{\uparrow})^{Q'_1}\cup\ldots\cup(\Sigma^{\uparrow})^{Q'_u}} \neg Ab_\delta x)\right) \wedge \left(\bigvee_{\delta\in(\Sigma^{\uparrow})^{P_i}} Ab_\delta x\right).$$

Then $\Sigma^{\uparrow}\cup\{\exists x(Q_n x\wedge\mu(x))\}$ is consistent. (Since $\Sigma^{\uparrow}$ is coherent, $\Sigma^{\uparrow}\cup\{\exists x Q_n x\}$ is consistent. Therefore this follows directly from $\Sigma^{\uparrow} \models \chi'$, $\Sigma^{\uparrow} \not\models \chi$.) Since there is no path from the node corresponding to Q_n to the node corresponding to P_i, contraposition of Lemma A.6 tells us it cannot be the case that $\Sigma^{\uparrow} \models \forall x((Q_n x\wedge\mu(x)) \to P_i x)$. Therefore there is a model of $\Sigma^{\uparrow}$ with some element d satisfying $Q_n x \wedge \neg P_i x \wedge \mu(x)$.
Adjust this model such that for no δ in $(\Sigma^{\uparrow})^{P_i}$ d satisfies $Ab_\delta x$. Since d does not satisfy $P_i x$, this adjusted model is still a model of $\Sigma^{\uparrow}$ (if there is no path as above). However, this model does not make χ' true. This contradiction proves that if such a path does not exist then $\Sigma^{\uparrow} \models \chi$.

Given that $\Sigma^{\uparrow} \models \chi$, it follows that $\forall x(Q_n x \to \bigvee_{\delta\in\Delta} Ab_\delta x) \in (\Sigma^{\uparrow})^W$. Since there is a path from the node corresponding to P_i to the one corresponding to Q_n, this in turn leads to $\theta \in (\Sigma^{\uparrow})^{WI}$.
Case II: $\phi_n \in (\Sigma^{\uparrow})^{WI} - (\Sigma^{\uparrow})^W$.
In this case there is some Q' such that there is a positive path from the node corresponding to Q_n to the node corresponding to Q' and

$$(\Sigma^{\uparrow})^W \models \forall x(Q' x \to \bigvee_{\delta\in\Delta_n} Ab_\delta x).$$

Recall that we have established $\Delta_n \subseteq \Delta \cup (\Sigma^{\uparrow})^{P_i}$. Thus the above implies

$$(\Sigma^{\uparrow})^W \models \forall x(Q' x \to \bigvee_{\delta\in\Delta\cup(\Sigma^{\uparrow})^{P_i}} Ab_\delta x).$$

Now pick m such that $(\Sigma^{\uparrow})^{WI}_{m}$ implies the above formula and $(\Sigma^{\uparrow})^{WI}_{m-1}$ does not. Since $(\Sigma^{\uparrow})^{W}$ does so, we may assume that $\phi_m \in (\Sigma^{\uparrow})^{W}$. Therefore by the same arguments as above (the ones used for the case that $\phi_n \in (\Sigma^{\uparrow})^{W}$) it follows that $\theta \in (\Sigma^{\uparrow})^{WI}$.

The above theorems give us $\Sigma^{\in} \models \Sigma^{WI}$ and $\Sigma^{WI} \models \Sigma^{WI\in}$. Since it is trivially true that $\Sigma^{WI\in} \models \Sigma^{\in}$, this means $\Sigma^{\in}$ and Σ^{WI} have the same models.

What is perhaps easier to see but still important to prove is that the alternative constraints leading to Σ^{WI} correctly model what happens in constructing the D function. The following Lemma and Proposition cover this part.

Lemma A.8 *Let $\Sigma^{\uparrow}$ be the lift of some network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_n\}$. Then $X \subseteq \Sigma$ is a conflicting set relative to v_i if and only if*

$$\Sigma^{\uparrow} \models \forall x \left(P_i x \rightarrow \bigvee_{\alpha \in X} Ab_{\alpha} x \right)$$

(Note that the above means the formula is true on every model of $\Sigma^{\uparrow}$, even those which are not models of $(\Sigma^{\uparrow})^{WI}$.)

Proof: Suppose $X \subseteq \Sigma$ is a conflicting set relative to v_i. Suppose towards contradiction that there is a model $\mathfrak{A}$ of $\Sigma^{\uparrow}$ such that

$$\mathfrak{A} \models \exists x \left(P_i x \wedge \bigwedge_{\alpha \in X} \neg Ab_{\alpha} x \right)$$

Since X is a conflicting set relative to v_i, there is some v_j such that X contains both a positive and a negative path to v_j. Therefore by repeated modus ponens (as well as modus tollens, possibly) it follows that both $P_j x$ and $\neg P_j x$. Contradiction.

For the other direction, suppose X is not a conflicting set relative to v_i. Let $\mathfrak{A}$ be a model where $\forall x P_i x$ and $\forall x \neg Ab_{\alpha} x$ for all $\alpha \in X$ hold, with the rest of the predicates having their truth-value determined by applying the rules in $\Sigma^{\uparrow}$. Since there are no logical relations between the predicates other than those provided by $\Sigma^{\uparrow}$, this can be done while letting $\mathfrak{A}$ be a consistent model of $\Sigma^{\uparrow}$. But the relevant formula is now false on $\mathfrak{A}$. Hence, $\Sigma^{\uparrow}$ does not entail it.

Proposition A.9 *Let $\Sigma^{\uparrow}$ be the lift of an inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let*

$$\phi = \forall x \left(P_i x \to \bigvee_{\alpha \in X} Ab_\alpha x \right)$$

If $(\Sigma^{\uparrow})^{WI} \models \phi$, then $Y \in D(v_i)$ for some $Y \subseteq X$. Conversely, if $X \in D(v_i)$ then $(\Sigma^{\uparrow})^{WI} \models \phi$.

PROOF: Suppose $(\Sigma^{\uparrow})^{WI} \models \phi$. By the construction of $(\Sigma^{\uparrow})^{WI}$, there must be some k such that there is a positive path from v_i to v_k and

$$(\Sigma^{\uparrow})^{W} \models \forall x \left(P_k x \to \bigvee_{\alpha \in X} Ab_\alpha x \right)$$

By the construction of $(\Sigma^{\uparrow})^{W}$, it follows that

$$\Sigma^{\uparrow} \models \forall x \left(P_k x \to \bigvee_{\alpha \in X \cup Ess_\Sigma(v_k)} Ab_\alpha x \right)$$

By Lemma A.8, this means that $X \cup Ess_\Sigma(v_k)$ is a conflicting set relative to v_k. Therefore $Y \in d(v_k)$ and hence $Y \in D(v_i)$, where $Y = X - Ess_\Sigma(v_k) \subseteq X$.

For the converse, suppose $X \in D(v_i)$. Then there is some v_j such that there is a positive path from v_i to v_j and $X \in d(v_j)$. Therefore $X \cup Ess_\Sigma(v_j)$ is a conflicting set relative to v_j. By Lemma A.8,

$$\Sigma^{\uparrow} \models \forall x \left(P_j x \to \bigvee_{\alpha \in (X \cup Min_\Sigma(v_j))} Ab_\alpha x \right)$$

By construction of $(\Sigma^{\uparrow})^{W}$,

$$(\Sigma^{\uparrow})^{W} \models \forall x \left(P_j x \to \bigvee_{\alpha \in X} Ab_\alpha x \right)$$

And therefore by construction $(\Sigma^{\uparrow})^{WI} \models \phi$.

A.2 Completeness Proof

Knowing (via Σ^{WI}) how the D function and $\Sigma^{\in}$ are related is an important step on our way to completeness, but we are far from done. At this point it may not be entirely clear to what thing on the inheritance network side the models in the sets $\mathcal{F}$ of the states $\langle \mathcal{U}, \mathcal{F} \rangle$ correspond. The bulk of the completeness proof lies in showing that they correspond to acceptable exception sets, with optimal models corresponding to minimal exception sets. The correspondence can be interpreted through the notion defined below.

Definition A.10 *Let $N = \langle V, \Sigma \rangle$ be an inheritance network. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the state generated by the lift of N. Let $\mathfrak{A}$ be a model in $\mathcal{F}$ whose domain contains an element referred to by the constant c. Let $X \subseteq \Sigma$ be an exception set.*

We say that $\mathfrak{A}$ models *the exception set X for c if $\mathfrak{A} \models Ab_{\alpha^\uparrow} c$ if and only if $\alpha \in X$.*

Now we first show that $\mathcal{F}$ consists of those models in $\mathcal{U}$ which correspond to an acceptable exception set of v_i.

Proposition A.11 *Let $\Sigma^\uparrow$ be the lift of the inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^\uparrow, \{P_i c\} \rangle$. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the information state generated by I.*
For $\mathfrak{A} \in \mathcal{U}$, we have $\mathfrak{A} \in \mathcal{F}$ if and only if $\mathfrak{A}$ models an acceptable exception set of v_i for c and makes $P_i c$ true.

PROOF: Let $\mathfrak{A} \in \mathcal{U}$. Then $\mathfrak{A} \in \mathcal{F}$ if and only if $\mathfrak{A} \models P_i c$. This makes the right-to-left direction trivial, so now assume $\mathfrak{A} \models P_i c$.
Choose $X \subseteq \Sigma$ such that $\mathfrak{A}$ models X for c. (By definition there is exactly one way to do this.) The only thing left to show is that X is an acceptable exception set of v_i. Let $Y \in D(v_i)$. We must show that $\exists \delta \in Y : \delta \in X$.

By Proposition A.9,

$$(\Sigma^\uparrow)^{WI} \models \forall x \left(P_i x \to \bigvee_{\alpha \in Y} Ab_{\alpha^\uparrow} x \right)$$

Therefore $\mathfrak{A} \models \bigvee_{\alpha \in Y} Ab_{\alpha^\uparrow} c$. Hence, there is some $\alpha \in Y$ such that $\mathfrak{A} \models Ab_\alpha c$. Since $\mathfrak{A}$ models X for c, this implies implies $\delta \in X$.

Next we show that every minimal exception set is in fact represented by at least one model.

Theorem A.12 *Let $\Sigma^{\uparrow}$ be the lift of the inheritance network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^{\uparrow}, \{P_i c\}\rangle$. Let $\langle \mathcal{U}, \mathcal{F}\rangle$ be the information state generated by I.*
For every minimal exception set X relative to v_i there is a model in $\mathcal{F}$ which models X for c.

PROOF: Let X be a minimal exception set relative to v_i. Construct $\mathfrak{A}$ as follows:

- For the domain, take the same domain as that of some other model in $\mathcal{F}$.
- Let $\mathfrak{A} \models P_i c$ and let $\mathfrak{A}$ model X for c.
- For all P_j, let $\mathfrak{A} \models P_j c$ if and only if there is a positive path from v_i to v_j that does not contain an element of X.
- For all y other than c in its domain and for all P_j, let $\mathfrak{A} \models \neg P_j y$.

We need to show that $\mathfrak{A} \in \mathcal{U}$. (The previous proposition then implies $\mathfrak{A} \in \mathcal{F}$.) For this it suffices to show that $\mathfrak{A} \models (\Sigma^{\uparrow})^{WI}$. (Since $(\Sigma^{\uparrow})^{\in}$ and $(\Sigma^{\uparrow})^{WI}$ have the same models, $\mathcal{U}$ consists exactly of all models of $(\Sigma^{\uparrow})^{WI}$.)
For elements other than c, the predicate assignments are trivially consistent with all rules and exemption clauses in $(\Sigma^{\uparrow})^{WI}$. For c, we first look at the rules in $\Sigma^{\uparrow}$.
Rules in $\Sigma^{\uparrow}$: So let $\phi \in \Sigma^{\uparrow}$, where

$$\phi = \forall x((P_j x \wedge \neg Ab_{P_j P_k} x) \rightarrow P_k x)$$

We may assume that $\mathfrak{A} \models P_j c \wedge \neg Ab_{P_j P_k} c$. (Otherwise c is trivially consistent with the rule.) Thus there is a positive path from v_i to v_j that does not contain an element of X, and the arrow from v_j to v_k is not in X. Therefore there is also such a path from v_i to v_k, and thus $P_k c$.
For negative rules, again take $\phi \in \Sigma^{\uparrow}$ but now with

$$\phi = \forall x((P_j x \wedge \neg Ab_{P_j P_k} x) \rightarrow \neg P_k x).$$

Again we may assume that $\mathfrak{A} \models P_j c \wedge \neg Ab_{P_j P_k} c$. Thus there is a negative path from v_i to v_k containing no element of X. Suppose there is also a positive path from v_i to v_k, and let Y be the union of these two paths. Then Y is a conflicting set relative to v_i. Since X is a minimal exception set relative to v_i, some $\alpha \in Y$ must be in X. Since the negative path had no such overlap, this α must be part of the positive path.

As we've shown that every such positive path contains an element of X, it follows by construction that $\mathfrak{A} \models \neg P_k c$. Therefore the valuation for c is consistent with this rule.

Exemption clauses in $(\Sigma^{\uparrow})^{WI}$**:** Suppose $\theta \in (\Sigma^{\uparrow})^{WI}$, where

$$\theta = \forall x (P_j x \rightarrow \bigvee_{\alpha \in \Delta} Ab_{\alpha^{\uparrow}} x)$$

By Proposition A.9, $Y \in D(v_j)$ for some $Y \subseteq \Delta$. We may assume that $P_j c$. Therefore there is a positive path from v_i to v_j, and thus $Y \in D(v_i)$. Since X is a minimal exception set relative to v_i, it follows that there is some $\alpha' \in Y$ for which $\alpha' \in X$. By construction, $\mathfrak{A} \models \neg Ab_{\alpha'} c$, and therefore c is consistent with θ.

Finally, we show that minimal exception sets correspond to optimal models.

Theorem A.13 *Let $\Sigma^{\uparrow}$ be the lift of the inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^{\uparrow}, \{P_i c\} \rangle$. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the information state generated by I.*

Every optimal model of $\mathcal{F}$ models an minimal exception set of v_i for c, and every minimal exception set of v_i has a model (for c) which is optimal in $\mathcal{F}$.

Proof: For the first part, let $\mathfrak{A}$ be optimal in $\mathcal{F}$. Per Proposition A.11, $\mathfrak{A}$ models some acceptable exception set X of v_i for c. Assume towards contradiction that X is not a minimal exception set of v_i, and that $X' \subset X$ is. Per Theorem A.12, there is a model $\mathfrak{A}' \in \mathcal{F}$ which models X'.

Now construct model $\mathfrak{A}''$ to be exactly like $\mathfrak{A}$ except that when evaluating predicates (including abnormality predicates) applied to c, it uses the same evaluation as $\mathfrak{A}'$.[15] Now the abnormality predicates made true by $\mathfrak{A}''$ are a strict subset of those made true by $\mathfrak{A}$. Thus it is strictly more normal than $\mathfrak{A}$, which is therefore not optimal.

For the second part, let X be a minimal exception set of v_i. By Theorem A.12, there are models in $\mathcal{F}$ which model X for c. Pick $\mathfrak{A}$ to be a model

[15] Showing that $\mathfrak{A}'' \in \mathcal{F}$ is fairly trivial and left to the reader.

which is optimal amongst those models. Suppose $\mathfrak{B} \in \mathcal{F}$ is at least as normal as $\mathfrak{A}$.
By Proposition A.11, $\mathfrak{B}$ models some acceptable exception set Y of v_i. Since $\mathfrak{B}$ is at least as normal as $\mathfrak{A}$, we have $Y \subseteq X$. Since X is minimal, this means $Y = X$. As we picked $\mathfrak{A}$ to be optimal amongst those that model X, this means $\mathfrak{A}$ is at least as normal as $\mathfrak{B}$.
Thus, $\mathfrak{A}$ is an optimal model.

Having proven the correspondence between optimal models and minimal exception sets, the last step in the completeness proof is to go from these models to the allowable inferences as defined in Definition 4.8. After doing this in the next theorem, the result we are after follows almost as a corollary.

Theorem A.14 *Let $\Sigma^{\uparrow}$ be the lift of the inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^{\uparrow}, \{P_i c\} \rangle$. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the information state generated by I. Let X be a minimal exception set of v_i.*
Then:

1. *If there is a positive path from v_i to v_j which doesn't contain any element of X, then every $\mathfrak{A} \in \mathcal{F}$ which models X makes $P_j c$ true.*

2. *If there is a negative path from v_i to v_j which doesn't contain any element of X, then every $\mathfrak{A} \in \mathcal{F}$ which models X makes $\neg P_j c$ true.*

3. *If X is not an acceptable exception set of v_j, then every $\mathfrak{A} \in \mathfrak{F}$ which models X makes $\neg P_j c$ true.*

4. *If every $\mathfrak{A} \in \mathcal{F}$ which models X makes $P_j c$ true, then there is a positive path from v_i to v_j which doesn't contain any element of X.*

5. *If every $\mathfrak{A} \in \mathcal{F}$ which models X makes $\neg P_j c$ true, then either there is a negative path from v_i to v_j which doesn't contain any element of X or X is not an acceptable exception set of v_j.*

PROOF: Point 1 and 2 are trivial by repeated modus ponens/tollens. Point 3 is almost as easy: If X is not an acceptable exception set of v_j, then there is some $Y \in D(v_j)$ such that $X \cap Y = \emptyset$. Since $Y \in D(v_j)$, $(\Sigma^{\uparrow})^{WI} \models \forall x (P_j x \to \bigvee_{\alpha \in Y} Ab_\alpha x)$ (Proposition A.9). Suppose $\mathfrak{A} \in \mathfrak{F}$ models X. Since $X \cap Y = \emptyset$, $\mathfrak{A}$ does not make $\bigvee_{\alpha \in Y} Ab_\alpha c$ true. Therefore $\mathfrak{A} \models \neg P_j c$.

For point 4, suppose every $\mathfrak{A} \in \mathcal{F}$ which models X makes $P_j c$ true. Construct $\mathfrak{B}$ as follows:

- For the domain, take the same domain as that of some other model in $\mathcal{F}$.
- Let $\mathfrak{B} \models P_i c$ and let $\mathfrak{B}$ model X for c.
- For all P_j, let $\mathfrak{B} \models P_j c$ if and only if there is a positive path from v_i to v_j that does not contain an element of X.
- For all y other than c in its domain and for all P_j, let $\mathfrak{B} \models \neg P_j y$.

We have shown in the proof of Theorem A.12 that $\mathfrak{B} \in \mathcal{F}$. Thus, by construction there is a positive path from v_i to v_j that does not contain an element of X.
For point 5, suppose every $\mathfrak{A} \in \mathcal{F}$ which models X makes $\neg P_j c$ true. Now construct $\mathfrak{B}'$ to be as $\mathfrak{B}$ except that $\mathfrak{B}' \models P_j c$. Then $\mathfrak{B}'$ is not in $\mathcal{F}$, and more specifically $\mathfrak{B}' \not\models (\Sigma^{\uparrow})^{WI}$. Pick $\phi \in (\Sigma^{\uparrow})^{WI}$ such that $\mathfrak{B}' \models \neg\phi$. A number of cases arise, depending on ϕ.

a $\phi = \forall x(P_k x \wedge \neg Ab_\phi x \to \neg P_j x)$ for some k, with $\mathfrak{B}' \models P_k c \wedge \neg Ab_\phi c$. In this case, there is a negative path from v_i to v_j (via v_k) that does not contain an element of X.

b $\phi = \forall x(P_j x \wedge \neg Ab_\phi x \to \neg P_k x)$ for some k, with $\mathfrak{B}' \models P_k c \wedge \neg Ab_\phi c$. In this case too, there is a negative path from v_i to v_j (via v_k using modus tollens at the end) that does not contain an element of X.

c $\phi = \forall x(P_j x \to \bigvee_{\delta\in\Delta} Ab_\delta x)$ for some Δ, with $\mathfrak{B}' \models \neg \bigvee_{\delta\in\Delta} Ab_\delta c$. Then it follows that $X \cap \Delta = \emptyset$, and therefore by Proposition A.11, $Y \in D(v_j)$ for some $Y \subseteq \Delta$. Since X contains no element of Δ, it contains no element of this Y. Therefore X is not an acceptable exception set of v_j.

d $\phi = \forall x(P_j x \wedge \neg Ab_\phi \to P_k x)$ for some k, with $\mathfrak{B}' \models \neg P_k c \wedge \neg Ab_\phi c$. In this case, change the model one step further, making $P_k c$ true. As the new model still cannot be in $\mathfrak{F}$, find a new ϕ' it now contradicts.
If this ϕ' is like in case a or b, then there is still a negative path, which is just one step longer. (Recall that a negative path can go through any amount of positive arrows 'in the wrong direction' at the end.) If it is like case c, then the Y which is found is also part of v_j. If it is itself like case

d, then we continue to proceed in the same way.
Since no amount of making predicates true will make the model part of $\mathfrak{F}$, going on long enough will lead to a ϕ' of one of the first three forms. The only potential complication in this induction is the possibility that we are led to a formula like type a or b where P_k is true merely because of a change we made to the model. In this case there is a negative path from v_j to itself of which no element is in X. Since this path is a contradicting set relative to v_j, it follows that X is not an acceptable exception set of v_j.

Theorem A.15 ***(Soundness-Completeness Theorem)*** *Suppose $\Sigma^\uparrow$ is coherent. Then $v_i \vdash_\Sigma v_j$ if and only if $\Sigma^\uparrow, \{P_i c\} \models_d P_j c$, and $v_i \vdash_\Sigma \neg v_j$ if and only if $\Sigma^\uparrow, \{P_i c\} \models_d \neg P_j c$.*

PROOF: Let $\langle \mathcal{U}, \mathcal{F} \rangle$ correspond to $\langle \Sigma^\uparrow, \{P_i c\}\rangle$.

- By definition $v_i \vdash_\Sigma v_j$ holds if and only if for every minimal exception set X of v_i, there is a positive path Y from v_i to v_j with $X \cap Y = \emptyset$. Likewise, $v_i \vdash_\Sigma \neg v_j$ holds iff either for every such X there is a negative path Y like that, or no such X is an acceptable exception set of v_j.

- By Theorem A.14, this is true iff each $\mathfrak{A} \in \mathcal{F}$ which models a minimal exception set of v_i makes $P_j c$ true ($\neg P_j c$ for the negative case).

- By Theorem A.13, this is true iff each optimal model in $\mathcal{F}$ makes $P_j c$ ($\neg P_j c$) true.

- By definition this is true iff $\Sigma^\uparrow, \{P_i c\} \models_d P_j c$ ($\neg P_j c$).

A.3 Coherence

Theorem A.16 *Let $\langle V, \Sigma \rangle$ be an inheritance network with $V = \{v_1, \ldots, v_n\}$. Then $\Sigma^\uparrow$ is incoherent if and only if there is some v_i such that $\emptyset \in D(v_i)$.*

PROOF: $\Sigma^\uparrow$ is incoherent if and only if there is some P_i such that $\Sigma^{\uparrow WI} \cup \{\exists x P_i x\}$ is inconsistent. This is if and only if $(\Sigma^\uparrow)^{WI} \models \forall x \neg P_i x$ for some P_i. By the convention on empty disjunctions, $\forall x \neg P_i x$ is equivalent to $\forall x(P_i x \rightarrow \bigvee_\emptyset Ab_\alpha x)$. Therefore the last step follows from Proposition A.9.

Proposition A.17 *Let $\langle V, \Sigma \rangle$ be an inheritance network without strict arrows. If $\emptyset \in d(x)$, then there are some z and some $y \approx x, y' \approx x$ such that Σ contains a positive arrow from y to z and a negative arrow from y' to z.*

PROOF: Suppose $\emptyset \in d(x)$. Then there is some minimal conflicting set $X \subseteq Ess_\Sigma(x)$. We may assume without loss of generality that X is the union of a positive path $\{xy_1, y_1y_2, \ldots, y_mz\}$ and a negative path $\{xy'_1, y'_1y'_2, \ldots, y'_nz^-\}$. Since $y_mz \in X$, it follows that $y_mz \in Ess_\Sigma(x)$. Therefore $x \approx y_m$. Analogously, $x \approx y'_n$.

Definition A.18 *The vertex x* semi-strictly implies *(*semi-strictly refutes*) y if there is a positive (negative) path from x to y where every arrow after the first is strict.*

Proposition A.19 *Let $\langle V, \Sigma\rangle$ be an inheritance network. If $\emptyset \in d(x)$, then there are some z and some $y \approx x, y' \approx x$ such that y semi-strictly implies z and y' semi-strictly refutes z.*

PROOF: Suppose $\emptyset \in d(x)$. Then there is some minimal conflicting set $X \subseteq Ess_\Sigma(x)$. We may assume without loss of generality that X is the union of a positive path $\{xy_1, y_1y_2, \ldots, y_nz\}$ and a negative path $\{xy'_1, y'_1y'_2, \ldots, y'_nz^-\}$ (where some of these may actually be strict).
Pick the smallest i for which y_i strictly implies z.[16] Since $y_{i-1}y_i \in X$, it follows that $y_{i-1}y_i \in Ess_\Sigma(x)$. But by construction $y_{i-1}y_i$ is not strict. Therefore $y_{i-1} \approx x$.
Analogously, $y'_{j-1} \approx x$ when we pick the smallest j for which y'_j strictly refutes z. (If no y'_j does so, pick $j = n+1$ instead.) Now let $y = y_{i-1}$, $y' = y'_{j-1}$. By construction, y semi-strictly implies z and y' semi-strictly refutes it.

[16]For y_{i-1} to exist we must assume x does not semi-strictly imply z, but this is safe because if it does then we can pick $y = x$ and skip the next couple of steps in the proof.