Clarity in Non-Monotonic Logic

MSc Thesis (*Afstudeerscriptie*)

written by

Harald Bastiaanse
(born September 21st, 1983 in Purmerend, The Netherlands)

under the supervision of **Prof Dr Johan van Benthem** and **Prof Dr Frank Veltman**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
July 5, 2007	Prof Dr Johan van Benthem
	Prof Dr Frank Veltman
	Fenrong Liu
	Prof Dr Dick de Jongh
	Dr Eric Pacuit
	Dr Sujata Ghosh

Institute for Logic, Language and Computation

Contents

1 Introduction

Most books whose titles include the word "mathematician" end up leveling a lawn chair.
-John Waggoner in USA Today

The above quote illustrates how most people aren't prone to spending their time reading texts that are unclear to them. Of course, people who engage in various highly symbolic formal endeavors on a professional basis are more tolerant of works that are hard to get through. Or are they?
It would seem so at first, and surely plenty of painstakingly hard-to-read works are being produced in all of the various fields. But in the end they too are capable of renouncing overly complex works at least as strongly, as the following quote from a well-known mathematician reveals.

[But] I am not concerned with ballistics or aerodynamics, [...]. That [...] is repulsively ugly and intolerably dull; -G. H. Hardy in *Mathematics in Wartime*

Consider also that in any form of art, the vast majority of what is produced is utter trash that ought never to reach the public[1] (the differences lying in the extent to which this trash reaches the public anyway). Thus it is besides the point to look at what is being produced (after all, we do not suppose that artists are at all tolerant towards trash). Rather, what is important is the following: are these unclear pieces actually being read?
Outlook hazy. It is telling that historically mathematics and even the sciences have often made great leaps forward by switching to new formalisms and paradigms in terms of which the phenomena under study were easier to express and comprehend. Evidently unclarity can stifle a field, or at least prevent readily available results from being picked up on when needed.
In this thesis we will ascertain as much by looking into the field of non-monotonic logic. We take a look at a system for default rules that has great empirical adequacy but also no lack of complexity, and that has thus remained basically unnoticed for a long time.[2]
And exactly as our little theory suggests, a fruitful application presents itself immediately after we've rephrased the system to increase its clarity...

[1] Which, incidentally, makes it mistaken to reject something as not being an art form merely because the vast majority of it isn't worthy of being called art.

[2] Interestingly, we will see that even celebrity is no mitigator for unclarity, as the system appeared in an (otherwise) extremely popular paper.

1.1 Overview

In the next chapter we will embed the system for default rules from Defaults in Update Semantics into DEUL, starting out with a relatively straightforward embedding of its third chapter before moving on to a non-trivial reduction-embedding that removes the need to use frames at the same time. The extent to which clarity is improved at that point is debateable, but that's only halfway into the thesis.

In the chapter after that we modalize circumscription, creating the Modal Circumscription Family framework, then proceed to lift the embedding from the previous chapter onto this, cutting out a few needlessly complex parts in the process. The resulting hybrid of Update Semantics, DEUL and Circumscription is then discussed along with various intuitions underlying it, and we then proceed to test it against a number of stock examples to demonstrate its workings and its empirical adequacy.

1.2 Practical matters

This thesis is designed to be (theoretically) readable even by someone with no prior experience with any of the papers mentioned herein[3]. The various appendices give small formal introductions to the various systems, reissuing all the relevant definitions. Those unfamiliar with the terms "non-monotonic logic" and "default rules" may look them up as well.

Furthermore, the technical parts of the thesis are consistently put into separate sections so that the casual and/or lazy reader can elect to skip them, temporarily or altogether. (Casual joking besides, all of these allow for easy referencing for the advanced reader.)

I've strived for a clear categorization of the content, which also means the reader should be prepared for some page-flipping. I am quite serious when I redirect some readers to the appendix at the start of the technical section, recommend going to the "Notes" sections for intuitive explanations of certain technical parts, or recommend looking up an example in the Examples section.

This is to keep out what would otherwise be long-winded distractions interfering with the clear readability of the technical sections[4], and conversely to keep out what would otherwise be obfuscatory formalism interfering with the clear readability of the non-technical sections.

[3] However, it is not specifically designed to be *interesting* for such people...

[4] Which already have too much text in them for my personal tastes, though I don't expect everyone to agree on this particular matter.

2 Clarifying Defaults in Update Semantics

2.1 Introduction

The paper *Defaults in Update Semantics*[11] is a case in point regarding the importance of clarity and the effect of the lack thereof. It starts out simple enough and gets progressively harder to grasp as the pages go by. Meanwhile, the impact of parts of it has been inversely proportional to their distance from the front page. The main ideas from the introduction have been picked up on soon, especially the slogan

> You know the meaning of a sentence if you know the change it brings about in the information state of anyone who accepts the news conveyed by it.

The straightforward general system for update semantics and its relatively easy application to correctly deal with the "might" operator have also been of some influence, especially in linguistics. Things go quickly downhill from there, though. The third chapter, dealing with necessarily and presumably, has gotten nowhere near as much attention. And I have yet to lay my eyes on even a single paper that gives more than a token nod to the fourth chapter which deals with conditional defaults.

Now, obviously many scholars felt that the first two chapters provided more than enough food for thought, or else these would never have been so popular either. And it did get plenty of looks[5], so perhaps it came at the wrong time or in the wrong place.
Regardless, the fact that the later chapters never got much attention is unfortunate for two reasons. For one thing, as the title Defaults in Update Semantics suggests the paper was never meant to be merely about dynamic meaning, update semantics and the might operator. Its main content was meant to be the issue of handling defaults, which only starts to be addressed in chapter 3.
More importantly, this system for defaults actually works quite well, giving the intuitively correct answer for many examples on which more popular systems fail. (We will not go into these here; a number of important examples dealt with successfully can be found in the next chapter, as well as in the original paper.)

It would seem like a good idea, then, to draw renewed attention and understanding to these chapters by clearing them up. Our main approach will be to embed them into a more clear and

[5] being the third most quoted article ever to appear in the *Journal of Philosophical Logic*.

popular system, complemented for the conditional part by a simultaneous bid to remove the concept of "frames" in the process. As an added bonus, this will also provide use with an analysis as to wherein the difficulty and unclarity lies, which will be useful to us later on.

So how shall we proceed? The patterns of possible worlds standing in a normality relation suggest a modal logic, but it also becomes quickly apparent that the various operators cannot be captured with mere modalities. As the slogan indicates meaning is dynamic in the update semantics system; its formulas are model-change operators. We seek a dynamic modal logic with model-changing operators, then. As of 2005, we are in luck.

The Dynamic-Epistemic Upgrade Logic DEUL from the 2005 paper[6] *Dynamic Logic of Preference Upgrade* [10] is just what we need: a (multi-)modal logic with model-change operators (It even self-embeds into modal logic through reduction axioms.). As noted earlier, for readers unfamiliar with DEUL a quick recap has been included (Appendix C).

[6]Though still due to be published this year.

2.2 Technical

Readers unfamiliar with DEUL should refer to Appendix C at this point as should readers who are familiar with it but confused by the slightly different notation used here.
For easier checking, the relevant definitions from Defaults in Update Semantics are repeated in a separate appendix as well, though readers wholly unfamiliar with it may be better off merely glancing over the technical parts of that and this section.
(At one point, explanation was intentionally extremely sparse in this section. While there are parts for which at this point I cannot say that this is the case, I do leave in my remark that In sofar as what we're doing here isn't self-explanatory, one should refer to the notes section.)

2.2.1 The @ operator

Before we can start the embedding proper, we first enrich the target system with a new update and associated operator. Reduction axioms will self-embed this enriched system so that this causes no issues.
(Some information about the intuitions behind this operator is to be found in the notes section. It basically changes the extension of p to φ, and might be thought of as meaning-change or somesuch.)

Definition 2.1 *(update)*

$$\begin{aligned}(X, A, B, V)_{p@\varphi} &:= (X, A, B, V_{p@\varphi}) \\ V_{p@\varphi}(p) &:= \{x \in X | (X, A, B, V), x \models \varphi\} \\ V_{p@\varphi}(q) &:= V(q) \text{ (for } q \neq p)\end{aligned}$$

Definition 2.2 *(trivial semantic enrichment)*

$$M, x \models \langle p@\varphi\rangle\psi \Leftrightarrow M_{p@\varphi}, x \models \psi$$

Lemma 2.3 *(reduction axioms)*

$$\begin{aligned}
\langle p@\varphi\rangle p &\leftrightarrow \varphi \\
\langle p@\varphi\rangle q &\leftrightarrow q \text{ (for } q \neq p) \\
\langle p@\varphi\rangle \neg\psi &\leftrightarrow \neg\langle p@\varphi\rangle\psi \\
\langle p@\varphi\rangle(\phi \wedge \psi) &\leftrightarrow \langle p@\varphi\rangle\phi \wedge \langle p@\varphi\rangle\psi \\
\langle p@\varphi\rangle \Diamond\psi &\leftrightarrow \Diamond\langle p@\varphi\rangle\psi \\
\langle p@\varphi\rangle \Diamond^{\leq}\psi &\leftrightarrow \Diamond^{\leq}\langle p@\varphi\rangle\psi \\
\langle p@\varphi\rangle E\psi &\leftrightarrow E\langle p@\varphi\rangle\psi
\end{aligned}$$

(Proof of the soundness of these reduction axioms is left to the reader, as are the things to conclude from this.)

2.2.2 The easy case

This section embeds only the simple system from chapter 3 of [11]. For the system with conditional defaults from chapter 4, see the next section.

Definition 2.4 *Let $\sigma = \langle \epsilon, s \rangle$ be an information state on W, per [11]3.8. Then*

$$M_\sigma := (X, \{\}, \{\leq\}, V)$$

, where

$$\begin{aligned} X &:= W \\ \leq &:= \{(w,v)|(v,w) \in \epsilon\} \\ V(p) &:= \{w \in X | p \in w\}(p \in \mathbb{A}) \\ V(1) &:= s(1 \notin \mathbb{A}) \end{aligned}$$

The proposition letter 1 should be new, not already in the original language. Note also that the relationship is reversed from Defaults in Update Semantics: Here the larger worlds are the more normal ones, which is more in line with the system we're translating into (especially with regard to the # update).
The not very well-filled second slot is where the epistemic relation would go in DEUL, something which we have no use for at this point but will use in the advanced section later on. [7]
The correspondence between information states and derived models is one-to-one except for the absurd information state (of which the original system accepts only one). The next theorem is a statement to this effect, but first an equivalence relation to accommodate the exception.

Definition 2.5 *M is* absurd *iff $M \models U\neg 1$.*
$M \cong M'$ iff $M = M'$ or M, M' are both absurd.

Many of the following formulas results seem to depend on coherence. However, note that, per definition [11]3.8, every information state is either coherent or the absurd state **1**, which makes things a lot easier.
It might also be interesting to note that while incoherence in the original system can also result by

[7] DEUL purists might want a placeholder $\sim$ to fill the slot. This shouldn't otherwise make any appreciable difference.

performing an update which would make make all worlds non-normal, we only deal with s being empty here. We get away with this by doing things per update: we translate formulas into updates in such a way that each update that would cause incoherence will make $|1|$ empty.

As a matter of notational convenience, we talk about [*actually* ψ] in cases where the original would simply put [ψ]. This helps distinguish between the formula ψ and the factual update [ψ] a bit more clearly, and to refer to this update a bit more easily.

Theorem 2.6 *(Update Correspondence)*

$$\begin{aligned} M_{\sigma[\text{normally } \psi]} &\cong (M_\sigma)_{(1@(1\wedge E\Box^{\leq}\Diamond^{\leq}\psi));\#\psi} \\ M_{\sigma[\text{presumably } \psi]} &\cong (M_\sigma)_{1@(1\wedge\langle!1\rangle U\Diamond^{\leq}\Box^{\leq}\psi)} \\ M_{\sigma[\text{actually } \psi]} &\cong (M_\sigma)_{1@(1\wedge\psi)} \\ M_{\sigma[\text{might } \psi]} &\cong (M_\sigma)_{1@(1\wedge E(1\wedge\psi))} \end{aligned}$$

(All of these updates possibly restrict but never enlarge $|1|$, so that if $\sigma = \mathbf{1}$ we trivially get an absurd model which is therefore equivalent to M_σ.)
In the *might* case, there are two possibilities: if some s-world satisfies ψ, then $E(1 \wedge \psi)$ is true everywhere, making $1@(1 \wedge E(1 \wedge \psi))$ a trivial update. If no s-world satisfies ψ, then $E(1 \wedge \psi)$ is false everywhere, so that $1@(1 \wedge E(1 \wedge \psi))$ results in an absurd model.
Similarly the *presumably* ψ update is trivial if and only if $\langle!1\rangle U\Diamond^{\leq}\Box^{\leq}\psi$ is true at all 1-worlds, which is to say that all optimal[8] s-worlds make ψ true, and results an absurd model otherwise.
(To see that $U\Diamond^{\leq}\Box^{\leq}\psi$ means all optimal worlds satisfy ψ: If all optimal worlds satisfy ψ use that any world (U) can see ($\Diamond^{\leq}$) an optimal world, and any optimal world sees only ($\Box^{\leq}$) optimal worlds, which satisfy ψ. If $U\Diamond^{\leq}\Box^{\leq}\psi$ is true use that any (U) optimal world sees ($\Diamond^{\leq}$) only worlds that can see it, hence must satisfy ψ since at least one of those worlds satisfies $\Box^{\leq}\psi$.)
The *normally* ψ update consists of in the first place a test (which is passed iff $E\Box^{\leq}\Diamond^{\leq}\psi$ is true (everywhere) which is iff some optimal world satisfies ψ (note that $E\Box^{\leq}\Diamond^{\leq}\psi$ is the dual to $U\Diamond^{\leq}\Box^{\leq}\psi$); and which results in an absurd state otherwise) and if that test is succesful the update $\#\psi$ which makes ψ-worlds more normal than $\neg\psi$-worlds. (Strictly speaking the $\#\psi$ update follows regardless

[8]Here and in the rest of this proof I use "optimal" in the sense of being $\leq$-maximal, rather than in the sense of Definition [11]3.4

but since it cannot undo the absurdity of the model this is irrelevant for our purposes.)
Finally, the *actually* ψ update simply restricts $|1|$ (ie, s) to $|1 \wedge \psi|$ ($s \cap |\psi|$) (and results in an absurd model if this makes $|1|$ (s) empty). □

The translation of the formulas is two-fold. The formula segments produced by translation function T_1 are very much like the updates above. Using T_1 basically allows us to replace all updates in the original by parts of DEUL-formulas, as we will see Corollary 2.8.
The translation function T_2 allows us to deal with acceptance, the translations it provides effectively stating that the corresponding updates don't change the model[9]. We will get at this with Theorem 2.9.

Definition 2.7 *(Translation functions)*

$$\begin{aligned}
T_1(\text{normally } \varphi) &:= \langle 1@1 \wedge E\Box^{\leq}\Diamond^{\leq}\varphi\rangle\langle\#\varphi\rangle \\
T_1(\text{presumably } \varphi) &:= \langle 1@1 \wedge \langle !1\rangle U\Diamond^{\leq}\Box^{\leq}\varphi\rangle \\
T_1(\text{actually } \varphi) &:= \langle 1@1 \wedge \varphi\rangle \\
T_1(\text{might } \varphi) &:= \langle 1@1 \wedge E(1 \wedge \varphi)\rangle \\
T_2(\text{normally } \varphi) &:= (U\neg 1) \vee (E\Box^{\leq}\Diamond^{\leq}\varphi \wedge U(\varphi \rightarrow \Box^{\leq}\varphi)) \\
T_2(\text{presumably } \varphi) &:= U(1 \rightarrow \langle !1\rangle\Diamond^{\leq}\Box^{\leq}\varphi) \\
T_2(\text{actually } \varphi) &:= U(1 \rightarrow \varphi) \\
T_2(\text{might } \varphi) &:= (U\neg 1) \vee E(1 \wedge \varphi)
\end{aligned}$$

Corollary 2.8 *(Update Correspondence for T_1)*

$$M_\sigma, w \models T_1(\psi)T_2(\phi) \Leftrightarrow M_{\sigma[\psi]}, w \models T_2(\phi)$$

This is immediate from Theorem 2.6. It states that the model derived from M_σ through the update in $T_1(\psi)$ is cannot be distinguished through T_2-formulas from the model $M_{\sigma[\psi]}$, but Theorem 2.6 already states that these models are equivalent and we know that the T_2-formulas cannot distinguish between models that are equivalent in that sense. □

[9]Except for vacuous truth for absurd models: we already know that absurd models always remain absurd and specifically don't want to distinguish between absurd models.

Theorem 2.9 *(Acceptance)*

$$\sigma \Vdash \phi \Leftrightarrow \exists w : M_\sigma, w \models T_2(\phi)$$

The proof can be done in two steps. The first and easiest is to establish that the model of the absurd state accepts any T_2-formula. This follows easily from the definition of T_2, as M_1, w satisfies $U\neg 1$.

For the rest of the proof, note that there is a one-to-one correspondence between non-absurd information states σ and accompanying models M_σ; therefore if $\sigma \neq 1$ then because of Theorem 2.6 we have $\sigma[\phi] = \sigma$ iff M_σ is unchanged by the corresponding update. For the cases *presumably*, *actually*, and *might* this trivially corresponds to $T_2(\phi)$.
For the *normally* case, $E\Box^{\leq}\Diamond^{\leq}\varphi$ corresponds to the first update changing nothing, while $U(\varphi \rightarrow \Box^{\leq}\varphi)$ corresponds to $\#\varphi$ (that update removes just those arrows which are from φ-worlds to $\neg\varphi$-worlds, whose nonexistence is exactly what is expressed by $U(\varphi \rightarrow \Box^{\leq}\varphi)$). □

Finally, we obtain the end result of a translation of the three notions of entailment.

Corollary 2.10 *(Entailment)*

$$\begin{aligned}
\psi_1, \ldots, \psi_n \Vdash_1 \phi &\Leftrightarrow \exists w : M_0, w \models T_1(\psi_1) \ldots T_1(\psi_n) T_2(\phi) \\
\psi_1, \ldots, \psi_n \Vdash_2 \phi &\Leftrightarrow \forall\sigma \exists w : M_\sigma, w \models T_1(\psi_1) \ldots T_1(\psi_n) T_2(\phi) \\
\psi_1, \ldots, \psi_n \Vdash_3 \phi &\Leftrightarrow \forall\sigma \exists w : M_\sigma, w \models T_2(\psi_1), \ldots, M_\sigma, w \models T_2(\psi_n) \Rightarrow M_\sigma, w \models T_2(\phi)
\end{aligned}$$

From Theorem 2.9 it follows that

$$\psi_1, \ldots, \psi_n \Vdash_1 \phi \Leftrightarrow \exists w : M_{0[\psi_1]\ldots[\psi_n]}, w \models T_2(\phi)$$

. Using Corollary 2.8 we can take the next step:

$$\psi_1, \ldots, \psi_n \Vdash_1 \phi \Leftrightarrow \exists w : M_{0[\psi_1]\ldots[\psi_{n-1}]}, w \models T_1(\psi_2) T_2(\phi)$$

, and the rest is by induction.
The $\Vdash_2$ case is analogous, while the $\Vdash_3$ case follows from Theorem 2.9 directly. In the first two cases it is of some importance to pick a w where all the updates can take place, but any 1-world always suffices for that purpose.

2.2.3 Conditional defaults without frames

In the next definition we will make use of an encoding of the rules, assigning a natural number k to each one and referring to such things as "the k-th rule" and "the k-th slot" as if part of some list.

There are two things about these rules which we will need to refer to later on and for which we hence introduce special predicates. The first is their antecedent, for which we will make use of dom_k. The update introducing the k-th rule will include an @ operation making dom_k true exactly there where it needs to be.

The second is their being in force, for which we use ap_k. The update introducing the rule will ensure that any world where the rule is in force will satisfy the corresponding material implication. However, we work ap_k into our model in a different way, so that the truth of the material implication in no way implies that the rule is actually in force. This uncertainty about whether rules are in force is captured by the $\sim$ relation, which equates worlds which agree on all information other than which rules are in force[10]. The appropriate update will also make it preferable for a rule to be in force, by updating with $\#ap_k$.

Definition 2.11 *We let*

$$M_0 := (X_0, \{\sim\}, \{\leq\}, V)$$

, where

$$\begin{aligned} X_0 &:= P(\mathbb{A} \cup \{ap_k | k \in \mathbb{N}\}) \\ \sim &:= \{(x,y) | x - \{ap_k | k \in \mathbb{N}\} = y - \{ap_k | k \in \mathbb{N}\}\} \\ \leq &:= \{(x,y) | x, y \in X_0\} \\ V(p) &:= \{x \in X_0 | p \in x\} \\ V(1) &:= X_0 \\ V(dom_k) &:= \emptyset \end{aligned}$$

We won't generally have specific translations of the type M_σ this time around: instead we'll be using equivalence classes of models representing σ, defined further below. This is because we will

[10]...and so could work as an epistemic relation for an agent who knows only this factual information, though here intuition takes a back seat to us having a good use for an equivalence relation and an open slot to fill.

put less emphasis on σ itself and more on how it is derived from 0. (We will not be dealing with σ's that cannot be derived from 0 through the defined formulas.)

Definition 2.12 *(Default-related notions)*
Let $N \subseteq \mathbb{N}$, $k \in \mathbb{N}$. Then:

$$\begin{aligned}
Comp(N) &:= \bigwedge_{k \in N} (dom_k \to ap_k) \\
Normal_k(\phi) &:= \phi \wedge \bigwedge_{l=1}^{k-1} (U(dom_l \to \phi) \to ap_l) \\
App_k(N, \phi) &:= \bigwedge_{\psi \in D(k)} (U(\phi \to \psi) \to E(Normal_k(\psi) \wedge Comp(N))) \\
D(k) &:= \{dom_1, \ldots, dom_k, dom_1 \vee dom_2, \ldots, dom_{k-1} \vee dom_k, \ldots, dom_1 \vee \ldots \vee dom_k\}
\end{aligned}$$

Here N is a subset of $\mathbb{N}$, and therefore encodes a set of default rules. $Comp(N)$ states that the world evaluated at complies with this set, $App_k(N, \phi)$ that this set applies within $|\phi|$ (where ϕ is any formula; though intuitively conceiving of ϕ as a subset may be more helpful).
$Normal_k(\phi)$ states that the world evaluated at is normal with regards to ϕ, presupposing that there are no rules with index k or larger. For this simplified definition we use that we only consider frames derivable from 0: then only subsets that are the domain of a rule are assigned a non-trivial pattern by the frame, and maximality there corresponds with complying with relevant rules; thus a world is normal wrt ϕ if it complies with each rule whose domain is within ϕ.
Note that $App_k(N, \phi)$ also presupposes that there are no rules with index k or larger. Also note that $Comp(N)$ and $Normal_k(\phi)$ don't apply to $\sim$-equivalent worlds that don't make any ap_n true (say). This makes other definitions a bit tricky, but is more workable than the alternative.

Correct correspondence for these formulas is as follows: for $Comp(N)$ it is trivial, special considerations about ap_n notwithstanding. For $Normal_k(\phi)$ definition 4.3[11] makes a world of d normal with regards to d iff it is maximal in $\pi d'$ for every $d' \subseteq d$. But, under the assumption that the current information state is the result of consecutive updates to $\mathbf{0}$, the only d' for which $\pi d'$ is non-trivial are those that are the domain of some default rule. Furthermore, for those d' where $\pi d'$ *is* non-trivial maximality simply corresponds to complying with all default rules that have d' as their domain. Hence normality with regards to d corresponds to complying with every default

rule of which the domain is within d (note that $U(dom_l \to \phi)$ expresses this subset relation).
For $App_k(N, \phi)$, definition 4.9[11] makes a set of rules apply within s iff for every $d \supseteq s$ there is a normal d-world (normal with regards to d) which complies with them. This is what our formula states, except that it checks only unions of non-trivial subframes. The reason we can get away with this is not that normality with regards to trivial subframes is trivial (which it isn't), but rather that with the way normality is defined it is relevant only which domains are contained in a subframe. (It might seem that we should want to include more to prevent automatic trivial applicability if s isn't contained in any disjunction of domains. However, in such a situation trivial applicability is the correct result since any $d \supseteq s$ will have worlds that are not in any domain and are therefore both trivially normal and trivially in compliance with all rules.)
In actual fact for all but the most construed of examples it suffices to check only the subframes that are themselves domains, but unfortunately there are such examples (one of them is

$$0[A \vee B \rightsquigarrow X][B \vee C \rightsquigarrow Y][A \rightsquigarrow \neg X][C \rightsquigarrow \neg Y][B \rightsquigarrow \neg(X \wedge Y)]$$

here there are normal $|A \vee B|$- and $|B \vee C|$-worlds that comply with the last three rules, but there is no normal $|A \vee B \vee C|$-world that does so).

(Special considerations required the ordering of the next pieces to be a bit peculiar.)
Our translation functions and the things we do with them are similar to the approach in the previous section, with the new T_0 function giving the appropriate model updates themselves. Thus the explanations given there can be of some use for undestanding these here (and don't get repeated here).
Keep the encoding of defaults using numbers in mind and read sets $N \subseteq \mathbb{N}$ as sets of rules.

Definition 2.13 *(Translation functions)*

$$
\begin{aligned}
T_0(\varphi \rightsquigarrow \psi, k) &:= 1@(1 \wedge E(\varphi \wedge \psi)); !((\varphi \wedge ap_k) \rightarrow \psi); dom_k@\varphi; \#ap_k \\
App?_k &:= 1@(1 \wedge \bigwedge_{N \subseteq \mathbb{N}_1^{k-1}} ((\bigwedge_{n \in N} ap_n) \rightarrow App_k(N, 1))) \\
T_0(\text{presumably } \varphi, k) &:= 1@(1 \wedge E\langle App?_k\rangle\langle !1\rangle U \Diamond^{\leq}\Box^{\leq}\varphi) \\
T_0(\text{actually } \varphi, k) &:= 1@(1 \wedge \varphi) \\
T_0(\text{might } \varphi, k) &:= 1@(1 \wedge E(1 \wedge \varphi)) \\
T_1(\varphi \rightsquigarrow \psi, k) &:= \langle 1@(1 \wedge E(\varphi \wedge \psi))\rangle\langle !((\varphi \wedge ap_k) \rightarrow \psi)\rangle\langle dom_k@\varphi\rangle\langle \#ap_k\rangle \\
T_1(\text{presumably } \varphi, k) &:= \langle 1@(1 \wedge E\langle App?_k\rangle\langle !1\rangle U \Diamond^{\leq}\Box^{\leq}\varphi)\rangle \\
T_1(\text{actually } \varphi, k) &:= \langle 1@(1 \wedge \varphi)\rangle \\
T_1(\text{might } \varphi, k) &:= \langle 1@(1 \wedge E(1 \wedge \varphi))\rangle \\
T_2(\varphi \rightsquigarrow \psi, k) &:= U\neg 1 \vee E \bigvee_{N \subseteq \mathbb{N}_1^{k-1}} ((\bigwedge_{n \in N} U(dom_n \leftrightarrow \varphi)) \wedge (\bigwedge_{n \in \mathbb{N}_1^{k-1} - N} \neg U(dom_n \leftrightarrow \varphi)) \wedge \\
&\quad \langle ! \bigwedge_{n \in \mathbb{N}_1^{k-1} - N} \neg ap_n\rangle\langle !\varphi\rangle(\psi \rightarrow \Diamond\Box^{\leq}\psi)) \\
T_2(\text{presumably } \varphi, k) &:= U1 \vee E\langle App?_k\rangle\langle !1\rangle U \Diamond^{\leq}\Box^{\leq}\varphi \\
T_2(\text{actually } \varphi, k) &:= U(1 \rightarrow \varphi) \\
T_2(\text{might } \varphi, k) &:= U\neg 1 \vee E(1 \wedge \varphi)
\end{aligned}
$$

The T_0 translations are model upgrades.

The k variable serves a few purposes. The conditional default is encoded into the k-th slot by the T_0 translation, while in the other translations it is presumed that slots k and further have not yet been used.

The $T_0(\varphi \rightsquigarrow \psi, k)$ is read as follows: first we check that $\varphi \wedge \psi$ is possible at all (and get in an absurd model if not). Then we get the encoding right, first by making sure that ap_k guarantees the material implication, then by making dom_k true at exactly the domain (ie: φ). Finally, we make worlds where the rule is in force (ie: where ap_k) superior.

The $App?_k$ update is meant to capture the pragmatics properly. It restricts 1 to those worlds that only make true sets of ap_n for which the corresponding set of rules applies within s[11]. (The

[11] As dealt with by App_k, explained earlier

intuition: you should never enforce combinations of rules that aren't jointly applicable.) This means that after this update optimality within 1 corresponds to complying with a maximal-applicable set of rules, and therefore coincides exactly with optimality in the sense of [11] definition $4.13_{ii)}$ (using its Proposition 4.14).

The T_2 translation for the conditional default reads as follows: first establish N as the set of rules whose domain is $|\varphi|$. Then, after restricting to only these rules and to worlds where φ holds, it becomes true that every ψ-world is $\sim$-equivalent to a world which is superior to all not-ψ worlds. (We have to take this workaround since the 'original' worlds that don't make any *ap* true are always all equivalent.) This is the correct translation since in the original system the update changes (only!) $\pi|\varphi|$ to make all ψ-worlds superior to all $\neg\psi$-worlds and since we get the equivalent of $\pi|\varphi|$ back by disregarding all rules whose domain is not exactly $|\varphi|$.

Not that unlike the other T_2 translations we cannot read this one as saying that our model remains the same under the appropriate update (if only because the conditional will always get encoded into the next slot). Rather, we are forced to take the detour interpreting it as the information state corresponding to the current model being unchanged by the formula being translated.

Theorem 2.14 *(Update Correspondence 1)*

$$(0[\phi_1]\dots[\phi_n] \Vdash \psi) \Leftrightarrow ((M_0)_{T_0(\phi_1,1);\dots;T_0(\phi_n,n)} \models T_2(\psi, n+1))$$

The *actually* and *might* cases are as before. The *presumably* case is as before except for having to correctly capture the pragmatics, with that part explained at length above. We've also already gone through all the necessary explanation to show that the update conditional implication behaves correctly at this point.

Definition 2.15 *(Equivalence and related notions)*

$$((M_0)_{T_0(\phi_1,1);\dots;T_0(\phi_n,n)} \cong (M_0)_{T_0(\psi_1,1);\dots;T_0(\psi_n,n)}) \Leftrightarrow (0[\phi_1]\dots[\phi_n] = 0[\psi_1]\dots[\psi_n])$$

$$k(M) = \max\{n \in \mathbb{N} | U(ap_n \rightarrow \Box^{\leq} ap_n)\}$$

$$|M_\sigma| = \{(M_0)_{T_0(\phi_1,1);\dots;T_0(\phi_n,n)} | 0[\phi_1]\dots[\phi_n] = \sigma\}$$

Models are equivalent if they are the result of series of updates which (applied to 0) would lead to the same information state. Extending this, $|M_\sigma|$ is the equivalence class of all models which are

the result of a series of updates which (applied to 0) would lead to information state σ.
$k(M)$ is the largest n for which a default rule has been encoded into slot n.

Lemma 2.16 *(Success of equivalence)*

$$\begin{aligned} M \cong M' &\Rightarrow \forall\varphi(M \models T_2(\varphi, k(M)+1) \Leftrightarrow M' \models T_2(\varphi, k(M')+1)) \\ M \in |M_\sigma| &\Rightarrow \forall\varphi(M \models T_2(\varphi, k(M)+1) \Leftrightarrow \sigma \Vdash \varphi) \end{aligned}$$

If two models are equivalent then they make translations of the same formulas true, provided the translations take into account all slots used. A model in $|M_\sigma|$ makes the appropriate translation of a formula true iff σ makes the formula true.
In both these cases, the implications can be reversed if we assume that the models are derived from appropriate updates on M_0.

Lemma 2.17 *(Update Correspondence 2)*

$$M \in |M_\sigma| \Leftrightarrow M_{T_0(\varphi, k(M)+1)} \in |M_{\sigma[\varphi]}|$$

A model is in $|M_\sigma|$ iff the result of updating it with the appropriate translation of φ is in $|M_{\sigma[\varphi]}|$.

2.3 Notes

2.3.1 The @ operator

In the most basic sense this operation is about changing the extensions of predicates. It is a useful tool to keep track of subsets such as s, a situation where updates that work through link-cutting or world-elimination would be inappropriate and result in a much less elegant solution. Any subset of worlds which is sufficiently non-arbitrary (ie any subset that can be described using modal formulas) can be properly captured and updated by assigning to it a new predicate and using the @ operator, a fact of which we make extensive use in the more advanced section.
Interestingly, in some contexts it might instead be conceived of as meaning revision, with $p@\phi$ changing assigning to p the (current) meaning of ϕ. This facilitates reasoning about meaning and allows one to discuss such principles as $\langle p@\phi\rangle\psi \rightarrow ((p \leftrightarrow \phi) \rightarrow \psi)$ (if changing the meaning of p to ϕ makes ψ true then it is already true if this is a vacuous change) and $\vdash \phi \Rightarrow \vdash \langle p@\psi\rangle\phi$ (if ϕ is a tautology then it doesn't depend on the meaning of p).

2.3.2 Eliminating the frames

Frames would seem like an essential aspect of Defaults in Update Semantics, and hence doing an embedding which gets rid of frames at the same time would seem like an ill-fated endeavour. In actuality it works and results in improved clarity (if not improved size). Some explanation is obviously needed, though.
The first thing to note is that in the original system the only non-trivial parts of the frame are those that correspond to the domain of a default rule. Therefore at least theoretically all of the necessary information is encoded within those rules. We proceed, then, to encode these rules using the @ operation. We let dom_k be the domain of the k-th rule. The ap_k predicate will distinguish between worlds where the rule is in force and worlds where it isn't. At the moment it suffices to say that worlds where ap_k is true will satisfy the appropriate material implication (or have been deleted by the update with the rule) whereas worlds where it is false may or may not do so.
Using this we adopt a somewhat different relation between worlds. Whereas originally satisfying rules makes worlds superior, here it is rules actually being in force that makes worlds superior. For optimal-world purposes this makes no practical difference, but it allows us to "turn a rule off" through a (hypothetical) update with $\neg ap_k$, allowing us to reason about which worlds are

optimal after disregarding certain rules. This ability to disregard rules is an essential part of the pragmatics and indeed should be an essential part of any pragmatics for default rules that aims to be empirically adequate.
(For how the pragmatics is now correctly dealt with see the technical section. For more on how the pragmatics works intuitively speaking, see the notes in chapter 4.)

2.3.3 Success from the perspective of DEUL

That the various updates can all be translated into DEUL isn't the most interesting aspect of our success. What's much more interesting is that the meta-notions of compliance with a set of rules, normality with regards to a domain and application of a set of rules within a domain all translate into DEUL-formulas as well. Moreover, they translate into unexpectedly easy formulas that should be easily expanded by a computer and for simple examples (which most interesting ones are) can in fact be dealt with manually without too much trouble.
And all of that is merely about the destination. In this case, the journey has perhaps been even more enlightening to the modal logician. The new @ operator looks to be a powerful tool for almost any embedding into DEUL, as well as being possibly interesting in other ways. Indeed, some unexpected uses may still be looming on the horizon since we've only applied it to a few special cases.

2.3.4 Improvements in terms of clarity

For being the main purpose of this embedding, there have perhaps but perhaps not been as many successes here as we might have hoped. Though for the most part clear and understandable, some of the translated updates are perhaps a bit on the long side. Thus whether the embedding has improved clarity can be a matter of some debate. One thing that certainly does benefit ease of use is the fact that checking whether a certain combination of rules applies -which used to involve an arduous inspection of many domains- now reduces to checking whether a single formula is true (and to a lesser extent compliance and normality should also be easier to work with).
In terms of the sheer size to be dealt with results are also contradictory. An important side-goal of the embedding was to eliminate the need to utilize frames, which have the size of the powerset of the initial set of possible worlds. While there has been success in this regard, it has come at a high price: the need to incorporate the "in force"-predicates ap_k means that the number of possible

worlds we work with is increased exponentially in the number of default rules (technically we are working with an inifinite model, but there's never more than a finite part of it that is relevant). So in terms of size we will typically not be much better off.

Where insightfulness and ease of use are concerned, though, the fact that we are working with just a single large model rather than will a lot of separate submodels should help tremendously. Also, a certain amount of size and complexity is simply the price that has to be paid for the empirical adequacy the system provides. The essential ability to turn rules off necessitates an extra predicate per rule to deal with this.

3 Modal Circumscription Families

Clarity is a strange beastie. We left the previous chapter with the conclusion that a certain amount of size and complexity are the price to be paid for the empirical adequacy provided by the update semantics system for defaults. This was primarily due to the need to use special predicates about whether certain rules apply or not. But what if there existed a system that isn't especially lacking in clarity or popularity that already had such predicates? Such a system might have the pragmatics from the previous chapter added to it for free!

There is such a system.

The systems referred to under the header Circumscription (classical texts include [8][7][5]) revolve around encoding the assumption that the extensions of certain predicates are as small as possible. It is common practice in the circumscription literature to use 'abnormality predicates' that indicate that an object is abnormal with regards to a certain rule (compare: that a certain rule is(n't) being enforced with regards to it).

In the next section we talk about the desirability of adding a good empirically adequate pragmatics to circumscription. But we aren't there yet: abnormality predicates notwithstanding this is not a trivial addition. When starting out the previous chapter, we noted that Defaults in Update Semantics suggested a dynamic, modal approach. Circumscription is neither modal nor dynamic, at least not in any of the classic approaches. Thus, our first step is to borrow from DEUL and develop a new approach to circumscription that *is* modal and dynamic: that of Modal Circumscription Families.

Once that is taken care of, we proceed to lift the embedding from the previous chapter onto this new system. In the process we also make some small changes, cutting out certain unneeded parts of the pragmatics that are also particularly unwieldy.

The structure of this chapter is largely the same as that of the previous chapter. First we look at some reasons for the conversion to be interesting. Next we go through the conversion itself with just a bit of text explaining the various moves and formulas. The conversion is again done in two stages, and followed by a section that explains the various non-technical aspects of the conversion at length and goes into how to use and understand the new system. After that we put the system to work, dealing with a few stock examples to display the ease of working with it as well as its empirical adequacy. We then round the chapter off with some concluding remarks.

3.1 Motivation

3.1.1 The Problem

There is an important open issue which has haunted circumscription (and indeed most systems for non-monotonic reasoning) since its conception. It is mostly an issue of empirical inadequacy: in a number of (very simple) examples our intuitions suggest a conclusion which does not follow through circumscription alone. Usually the desired conclusions can still be reached through a small enrichment of the system (for which many methods have been suggested throughout the years) and the addition of one or two pieces of information to the example.
One of the earliest and simplest such examples is the "birds fly" example, which is "generally considered to be the rock-bottom benchmark problem of any nonmonotonic logic"[12]. It has a number of incarnations, typically resembling:

(1) Birds fly.
(2) Penguins are birds.
(3) Penguins don't fly.
(4) Tweety is a penguin.[12][13]

Here the conclusion we would intuitively want to draw is that Tweety is a non-flying bird. However, using vanilla circumscription leaves open the option that Tweety is instead a flying penguin or a non-bird. Possible enrichments to the system to deal with this problem have been around for a long time and include amongst others circumscription policies and prioritized circumscription (eg [7]). Solutions revolve around the intuition that penguins are to be an exception to the rule that birds fly, creating means for (3) to overpower or temporarily disable (1). However, they fail to address what it is about this example that makes use want to make an exception to rule (1) rather than (2) or (3).
The enrichments of the system are hardly problematic, but the need to add more information to the examples is. For each explicit example the added information may seem intuitively obvious, but typically no attempt is made to explain the additions as warranted by certain principles. Indeed,

[12] Obviously the classic circumscriptionists are no big Looney Tunes fans...

[13] Normally "penguins are birds" is taken to be a strict rule rather than a default rule in this example. That actually makes things easier, amongst other things because it already rules out one violation. (If "penguins don't fly" is also taken as a hard rule, the problem disappears, along with the appeal of the example.) See the notes on specificity for more on this. The notes also have a section on how to do hard rules.

often we encounter no real justification for them at all. This is especially worrying since in a system that tries to capture common sense reasoning it is unacceptable to justify arbitrary additions merely because they should follow intuitively. Hence Asher and Morreau [1] describe such as committing oneself to the Hypothesis of the Ghost in the Machine, given as follows:

> Hypothesis of the Ghost in the Machine
> That specific information takes precedence over general information is not to be accounted for by the semantics of generic statements itself. Rather, it is due to the intervention of a power which is extraneous to the semantic machinery, but which guides this machinery to have this effect (by ordering the defaults, deciding the priorities of predicates to be minimized, or whatever).

To this day an empirically adequate means of dealing with these examples in a strictly systematic fashion has not been found. Even now (this quote from [6]) we can find in the literature such remarks as

> It is distressing that there is no obvious means to discover a circumscription policy that will entail the correct conclusions. [McCarthy, 1986] gives a partial solution in terms of policies, where the set of predicates to be minimized and varied is explicitly listed, but in no way helps us decide how to construct such policies.

The pattern is all too familiar: a method to implement additional information to get at the correct conclusion is easily found, but no systematic method is given to decide what information to add (in this case: what policy to use).

3.1.2 Enter Defaults in Update Semantics

There are a number of reasons to seek to solve the aforementioned issue by combining the system from Defaults in Update Semantics with circumscription. The most important reason is that this system manages to deal with the "birds fly" example and a number of other problematic cases in a way that is both empirically adequate and wholly systematic. It has a systematic way to determine which default rules should apply under which conditions, which is exactly what we have need of.

3.2 Technical

(This technical section *is* (somewhat) low on explanation: again, insofar as what we are doing isn't self-explanatory, refer to the notes in the next section.)

3.2.1 Introducing Modal Circumscription Families

Definition 3.1 *A* modal circumscription family *is a pair*

$$\Omega = (O, \leq)$$

such that O is a set of first-order predicate logic models, all of which share the same domain $D(\Omega)$ and set of predicates $A(\Omega)$, and $\leq$ is a transitive-reflexive relation on O.
$\Omega_0(D, A)$ is the specific modal circumscription family where O consists of all models with domain D and set of predicates A and $\leq = \{(M, M)|M \in O\}$.

Definition 3.2 *(Updates)*
For φ a first-order formula with no free variables,

$$(O, \leq)[\varphi] = (O', \leq \cap (O' \times O')), \text{ where } O' = \{M \in O | M \models \varphi\}$$

For P in $A(\Omega)$, with V_M referring to the valuation of M,

$$(O, \leq)[Circ(P)] = (O, \leq \cup \leq'), \text{ where}$$

$$\leq' = \{(M_i, M_j)|V_{M_i}(P) \supseteq V_{M_j}(P), \forall Q \in (A(\Omega) - \{P\}) : V_{M_i}(Q) = V_{M_j}(Q)\}$$

For same, and $Q_1, \ldots, Q_n$ in $A(\Omega)$,

$$(O, \leq)[Circvar(P, \{Q_1, \ldots, Q_n\})] = (O, \leq \cup \leq'), \text{ where}$$

$$\leq' = \{(M_i, M_j)|V_{M_i}(P) \supseteq V_{M_j}(P), \forall Q \in (A(\Omega) - \{P, Q_1, \ldots, Q_n\}) : V_{M_i}(Q) = V_{M_j}(Q)\}$$

Definition 3.3 *(Semantics)*

$$
\begin{aligned}
\Omega, M \models \phi &\Leftrightarrow M \models \phi \text{ for } \phi \text{ predicate-logical with no free variable} \\
\Omega, M \models \neg\varphi &\Leftrightarrow \Omega, M \not\models \varphi \\
\Omega, M \models \varphi \wedge \psi &\Leftrightarrow \Omega, M \models \varphi \text{ and } \Omega, M \models \psi \\
\Omega, M \models \Diamond\varphi &\Leftrightarrow \exists M' : M \leq M' \text{ and } \Omega, M' \models \varphi \\
\Omega, M \models E\varphi &\Leftrightarrow \exists M' : \Omega, M' \models \varphi \\
\Omega, M \models \langle X \rangle \varphi &\Leftrightarrow \Omega[X], M \models \varphi \text{ where } [X] \text{ is some update} \\
\Omega \Vdash \phi &\Leftrightarrow \forall M \in O : \Omega, M \models \Diamond\Box\phi
\end{aligned}
$$

(The other symbols (such as $\Box$) are dealt with in the usual fashion.)

Definition 3.4 *(Inference)*
For $\phi_1 \ldots \phi_n$ first-order predicate logic formulas without free variables and/or of the form $Circ(P)$ or $Circvar(P, \{Q_1, \ldots, Q_n\})$, and for φ a first-order predicate logic formula, and if D is the set of constants used in these formulas and A the set of predicates used there, we say

$$\phi_1 \ldots \phi_n \Vdash \varphi \Leftrightarrow \Omega_0(D, A)[\phi_1] \ldots [\phi_n] \Vdash \varphi$$

3.2.2 Adding default rules and pragmatics

Definition 3.5 *A* modal circumscription family with rules *is a quadruple*

$$\Omega = (O, \leq, R, s, a)$$

such that:
- O is a set of first-order predicate logic models, all of which share the same domain $D(\Omega)$ and set of predicates $A(\Omega)$
- $\leq$ is a transitive-reflexive relation on O
- a is a subset of $A(\Omega)$
- R is a finite set whose elements are of the form (φ, ψ, P), with φ and ψ being first-order predicate logic formulas and P a predicate in a
- s is a subset of O
$\Omega_0(D, A, a)$ is the specific modal circumscription family with rules where O consists of all models with domain D and set of predicates A, $\leq = \{(M, M)|M \in O\}$, R is empty, and appropriate a.

The default rules are encoded in R. Basically, interpret (φ, ψ, P) as $\varphi \wedge \neg P \rightarrow \psi$. The set s is as in Defaults in Update Semantics. The P part of each element of R is the abnormality predicate used for that rule (in this system we use manually assigned abnormality predicates.
We will need to distinguish between abnormality predicates and other predicates. This is what the set a is for: basically a is the set of (available) abnormality predicates, all rules having to have a predicate from a as their abnormality predicate.

Definition 3.6 *(Updates)*
For φ a first-order formula with no free variables,

$$(O, \leq, R, s, a)[\varphi] = (O, \leq, R, \{M \in s|M \models \varphi\}, a)$$

For φ a first-order formula with no free variables,

$$(O, \leq, R, s, a)[!\varphi] = (O', \leq \cap (O' \times O'), R, s \cap O', a), \text{ where } O' = \{M \in O|M \models \varphi\}$$

For φ and ψ first-order formulas with one free variable, and P a one-place predicate,

$$(O, \leq, R, s, a)[\varphi \rightsquigarrow \psi, P] = (O, \leq \cup \leq', R \cup \{(\varphi, \psi, P)\}, s, a)[!\forall x : \varphi(x) \wedge \neg Px \rightarrow \psi(x)], \text{ where}$$

$$\leq' = \{(M_i, M_j)|V_{M_i}(P) \supseteq V_{M_j}(P), \forall Q \in (a - \{P\}) : V_{M_i}(Q) = V_{M_j}(Q)\}$$

$$(O, \leq, R, s, a)[s] = (s, \leq \cap (s \times s), R, s, a)$$

$[\varphi]$ is soft update with φ; $[!\varphi]$ is hard update with φ; $[\varphi \rightsquigarrow \psi, P]$ is the update adding the default rule that if φ then normally ψ (P being the abnormality predicate); $[s]$ is a restriction to s, which is used to take the optimal $s-$world.
Note that $a-\{P\}$ could be rewritten as $A-((A-a)\cup\{P\})$, and hence corresponds to circumscribing P varying all non-abnormality predicates.

Definition 3.7 *(Pragmatics)*
Let $\Omega = (O, \leq, R, s, a)$, $x \in D(\Omega)$, $Y \subseteq D(\Omega)$, $r \subseteq R$. *Then:*

$$\begin{aligned}
Normal(x,r) &:= \bigwedge_{(\varphi,\psi,P)\in r} (\varphi(x) \rightarrow \neg Px) \\
Expected_\Omega(Y,x) &:= \bigwedge_{(\varphi,\psi,P)\in R:|\varphi(x)|\subseteq Y} \neg Px \\
App_\Omega(r,x) &:= \bigwedge_{(\varphi,\psi,P)\in R:|\varphi(x)|\supseteq s} E(\varphi(x) \wedge Expected_\Omega(|\varphi(x)|,x) \wedge Normal(x,r)) \\
Prag_\Omega &:= \bigwedge_{x\in D(\Omega)} \bigwedge_{r\subseteq R} ((\bigwedge_{(\varphi,\psi,P)\in r} \neg Px) \rightarrow App_\Omega(r,x))
\end{aligned}$$

(Intuitive readings mentioned here are strictly for legibility purposes. Consult the notes for a real and much better discussion of the intuitions involved. The explanations here are brief and based on "exceptions propagate downward" (also explained in the notes).)
$Normal(x,r)$ states that -at the current model- x complies with all the rules in $r \subseteq R$, which is to say that for those rules for which it satisfies the antecedent, it does not satisfy the appropriate abnormality clause.
Intuitively it means that x is normal with regards to all the $r-$rules that might apply to it.
$Expected_\Omega(Y,x)$ states that -in circumscription family Ω, at the current model- x is as expected with regards to the set of rules of more specific antecedent than Y. Intuitively, since exceptions propagate down and not up we must expect x to be normal with regards to a rule if our information is less specific than it's antecedent.
$App_\Omega(r,x)$ states that the set of rules r jointly applies to x within s. It is read as saying that for every rule where the antecedent is true of x across s, the accompanying information state allows for a valuation that is as expected with regards to x while being normal with regards to all $r-$rules.
For the intuition, suppose that for some more general information state this is not the case. Then, since we must expect x to be as expected in that state, we must expect it to not be normal with regards to all $r-$rules; ie to be an exception. But exceptions propagate downwards, so we must

then expect this to be the case in the information state s as well. So then in information state s we expect these rules to not all apply: they don't jointly apply in s. (One might think we should check more information states here, as we did at this point in the previous chapter. However, this time we are not restricted by the need to emulate any specific system and we can let other considerations prevail. Cutting this corner significantly benefits clarity and has no appreciable adverse effect on empirical adequacy. In the unlikely event someone manages to come up with a natural example that gives rise to an intuitive answer that we can obtain only with the addition, that would be the moment to reconsider.)

Finally, $Prag_\Omega$ will be made true in those models where whenever a set of rules doesn't jointly apply to x within s, x is in fact abnormal for at least one of them. This simply ensures that our intuitions are actually borne out.

Definition 3.8 *(Slightly changed Semantics)*

$$\Omega \Vdash \phi \quad \Leftrightarrow \quad \forall M \in O : \Omega[!Prag_\Omega][s], M \models U\Diamond\Box\phi$$

We now take the optimal worlds only from among those satisfying the pragmatic considerations and known facts. The semantics is otherwise the same as earlier.

3.3 Notes

The Modal Circumscription Family framework combines aspects of DEUL, Circumscription, and Update Semantics into a new whole. In the first step, a modal framework is created which can host circumscription and give the same results as should be expected from regular circumscription yet which is inherently modal and sufficiently similar to DEUL to allow the embedding from the previous chapter to work. This is by and large the easy step.
In the second step, the framework is expanded to incorporate a set of default rules, a set of still-possible worlds, and a set of predicates designated as abnormality predicates. Then update with default rules is defined in a circumscriptive way. Finally, suitable formulas for dealing with the pragmatics are defined based on the formulas used in the advanced embedding in the previous chapter (though with some renaming for our convenience) and the semantics is slightly adjusted.

3.3.1 Modal Circumscription Families

In the classical literature circumscription works on the syntactic level, replacing a specific sentence with the circumscription of certain parts of it with regards to it. The system introduced here works differently: a family of all possible valuations is used, and the circumscription operation is applied not a sentence but to this family, making those possible valuations that give the predicate to be circumscribed a smaller extension preferable to those that give it a larger extension. Subsequent predictions are then based on those possible valuations which are optimal with regards to this ordering.
The conclusions that may be drawn are the same, but there are some distinct benefits to this approach. The main benefit is flexibility: facts can be added after circumscribing without any problems, and predicates can be circumscribed in any order (though one must still take special care if some predicates are to be varied when circumscribing others). There is also a matter of elegance, as syntactically circumscribing against conjunctions of relevant facts can quickly become unwieldy. Finally, the 'possible valuation'-rendering makes circumscription modal and brings it in line with the intuitions of such approaches as Defaults in Update Semantics, allowing us to do what we are doing here.

3.3.2 Circumscription in the latter approach

While not immediately obvious at first, the "with Rules" version of the system still has circumscription as an essential ingredient. As hinted at in the technical section, if we rewrite $(a - \{P\})$ as $A(\Omega) - ((A(\Omega) - a) \cup \{P\})$ a comparison between Definition 3.2 and Definition 3.6 reveals that the ordering in the latter corresponds to circumscribing the abnormality predicates varying all the other predicates, not unlike the approaches to default rules in the classic circumscription literature.

3.3.3 Exceptions, 'exceptionality' and reversible exceptions

The most obvious change made in this embedding is the inclusion of special predicates indicating which rules are in effect. This allows exceptions to be made and reasoned about more easily. The main concept behind the pragmatics of application can now be interpreted as being the notion of *exceptionality*, manifested through two principles: more specific (factual) information can only make things more exceptional, and conversely less specific information can only make things less exceptional. Or in other words, exceptions propagate down while rules propagate up.
If we conclude from our current information that x is an exception to a given rule, then gaining even more information cannot reverse this conclusion. This may seem inadequate at first: what if x is an exception to the rule that would make it an exception to the first rule? An example of this might be a penguin which is inside an airplane and therefore does fly (in a somewhat broad interpretation of the word). The idea here -which is somewhat borne out in the example- is that exceptions to exceptions are even more exceptional, not less so. As far as the original rule is concerned their adhering to it is purely accidental, their truth rests solely on the rule that makes them a counter-exception. We can see this in the example: the penguin flies because it is on an airplane, not because it is a bird; therefore in a sense it is perfectly acceptable to say that it is still an exception to the rule that birds fly.
The way exceptionality is used here is similar to the concept of *specificity*, discussed below.

3.3.4 Specificity and the formulas

The main driving force behind the notions of application can also be interpreted as a special variant on the principle of *specificity*. A principle which roughly states that more specific rules should take precedence over those which are about less specific circumstances, this principle is often lauded but rarely realized in any non-monotonic system.

Specificity is dealt with directly through the $Expected_\Omega(s', x)$ formula: in any information state we must expect all rules that are about more specific states to be in force. Under this reading, the $App_\Omega(r, x)$ formula can then be seen as saying that applicability of a set of rules is about it being possible for them to apply whilst also adhering to specificity (for a select group of information states).

As a consequence of all this, we get specificity built into the system to a high degree. For example, in situations of the form:

> If A then normally C
>
> If A and B then normally not-C
>
> Ax and Bx

it will follow that not-Cx.[14]

This example notwithstanding, note that the kind of specificity we incorporate into the system is much more subtle and advanced than merely giving the more specific rule precedence. For we do not look at whether rules are more specific than other rules, but rather at whether they are more specific than the available information.

This is the kind of specificity that is the single pragmatic condition we enforce on the information states, deeming a set of rules non-applicable if their being applied would contradict specificity in some information state. Unlike simplistic specificity, this allows us to succesfully deal with such examples as the non-strict Birds fly variant and the easterly wind problem[15], where no rule is more specific than another.

3.3.5 What about incoherence?

Going back and comparing this system with the original Defaults in Update Semantics yields an interesting observation. That system makes extensive use of the notion of *incoherence*, a notion which is wholly absent here. It is interesting to see how this difference is reflected in the conclusions the systems give rise to.

Incoherence occurs as the consequence of updating with information that contradicts the information already available. When updating with factual information, this can be taken rather literally.

[14] Particularly, the strict version of the birds fly example can be rendered like this using A for "Bird", C for "Flies" and B for "Penguin". Note that the non-strict version needs the more advanced notion, as there "Penguin" is not more specific than "Bird".

[15] See the examples section.

Our counterparts to such incoherent states are modal circumscription families with empty s. In such families, the update $[s]$ eliminates all models. Therefore from the way the semantics is defined every formula is trivially entailed and can bee concluded in these cases.
When updating with a rule, incoherence is about making too many exceptions. As it turns out, the old notion of incoherence corresponds neatly to falsity of the $Expected_\Omega(s', x)$ formula. (This is hardly coincidental by the way, as the $Expected_\Omega(s', x)$ formula corresponds to the original notion of normality, with coherence being directly linked to the existence of normal worlds.) So we should investigate cases where the $Expected_\Omega(s', x)$ formula is false. An easy example that comes to mind is the following:

If A, then normally C
If A and B, then normally not-C
If A and not-B, then normally not-C
Ax

The truth of Ax makes x an exception to one of these three rules, regardless of whether Bx is true. Hence there is a direct contradiction between Ax and $Expected_\Omega(|Ax|, x)$, which automatically makes the $App_\Omega(r, x)$ formula false for any non-empty r. In this state no rules can be meaningfully applied to x at all. So in an interesting reversal we get that incoherence makes it impossible to conclude anything (other than facts already available), instead of automatically making everything true.[16]

3.3.6 Dealing with hard rules

Incorporating hard rules into the system can be easily done in at least three straightforward ways. One way is to use the same announcement as for a regular rule but to leave out the abnormality predicate (ie to use an update of the form $[!\forall x : \phi(x) \rightarrow \psi(x)]$). Another way is to do a factual (soft) update with this exceptionless rule. A third way is to update with the rule as normal, then do a factual update stating that the appropriate abnormality predicate is always false $[\forall x : \neg Ab(x)]$. Of these last two methods it might be said that they produce the effect that the information that

[16]Though if this result is deemed undesirable it should be easy enough to adapt the $Prag_\Omega$ formula in such a way that incoherence results in all models being eliminated by the $!Prag_\Omega$ update and therefore also results in every formula following trivially.

the rule is hard is itself soft (though hard enough to be incontrovertible). An advantage they have is that they don't involve a real alteration of the system itself.

3.4 Examples

(Only a handful of examples are given here (all adapted from [11]): in case these examples are not yet sufficiently convincing the reader is invited to verify that the system does well on all applicable benchmark examples from [4] and [12]. Also, be advised that this section will get fairly technical fairly quickly; it is certainly much more technical than the notes section.)

Dealing with stock examples about some rules being overpowered by others is mostly very easy: just check the $App_\Omega(r, x)$ formula for the rule that is going to not apply (with r containing just that rule) and show that it does indeed not apply. After concluding that $App_\Omega(r, x)$ is false for all models in Ω, it follows that x satisfies the abnormality predicate for that rule in all models of $\Omega[!Prag]$. The rule has then been "turned off", and you can leave the (typically) trivial rest for the reader.

Technically speaking one should do more, such as check that the other rules do apply. However, this perhaps shouldn't be part of our recommended approach as anecdotal evidence suggests that in practice no-one will do it anyway...

Other examples will need to be done on more of a case-by-case basis. We will start out by doing one of those entirely, specifically the well-known "Nixon diamond".

3.4.1 Nixon diamond

A well-known stock example, not the most difficult one but a good introduction.

Consider the family

$$\begin{aligned}\Omega \;=\;& \Omega_0(\{Nixon\}, \{Quaker, Republican, Pacifist, Ab_1, Ab_2\}, \{Ab_1, Ab_2\}) \\ & [Quaker(x) \rightsquigarrow Pacifist(x), Ab_1][Republican(x) \rightsquigarrow \neg Pacifist(x), Ab_2] \\ & [Quaker(Nixon) \wedge Republican(Nixon)]\end{aligned}$$

The updates will already ensure that models where $Quaker(Nixon) \wedge Republican(Nixon)$ holds do not have both $\neg Ab_1(Nixon)$ and $\neg Ab_2(Nixon)$, so it only remains to show that these can each occur separately. To show that the first rule applies, we inspect $App_\Omega(\{(Quaker(x), Pacifist(x), Ab_1)\}, Nixon)$, which is equivalent to

$$\begin{aligned}&E(Quaker(Nixon) \wedge \neg Ab_1(Nixon) \wedge (Quaker(Nixon) \rightarrow \neg Ab_1(Nixon)) \wedge \\ &E(Republican(Nixon) \wedge \neg Ab_2(Nixon) \wedge (Quaker(Nixon) \rightarrow \neg Ab_1(Nixon))\end{aligned}$$

The first term is true because there is a possible model that makes Nixon a quaker and a pacifist and not abnormal with regards to the first rule. Likewise, the second terms is true because there is a possible model that makes Nixon a republican non-pacifist and not abnormal with regards to either rule. Note that this second possible model will make Nixon a non-quaker and will therefore be a non-actual model: the formula explicitly ranges over non-actual models as well and we deliberately don't eliminate non-actual models until after going through the pragmatics.
Showing that the second rule also applies is entirely analogous. Hence we have now shown that of the models with $Quaker(Nixon) \wedge Republican(Nixon)$, both those with $\neg Ab_1(Nixon)$, $Ab_2(Nixon)$ and $Pacifist(Nixon)$ and those with $Ab_1(Nixon)$, $\neg Ab_2(Nixon)$ and $\neg Pacifist(Nixon)$ will survive the $!Prag_\Omega$ update.
Since models where $Quaker(Nixon) \wedge Republican(Nixon)$ holds do not have both $\neg Ab_1(Nixon)$ and $\neg Ab_2(Nixon)$, both kinds will be optimal in $\Omega[!Prag_\Omega][s]$ (as any model that would be strictly superior will have been eliminated by $[!Prag_\Omega][s]$). Thus neither $U\Diamond\Box Pacifist(Nixon)$ nor $U\Diamond\Box\neg Pacifist(Nixon)$ will hold at (any model of) $\Omega[!Prag_\Omega][s]$, and thus neither $\Omega \Vdash Pacifist(Nixon)$ nor $\Omega \Vdash \neg Pacifist(Nixon)$ holds.

3.4.2 Birds fly

The quintessential stock example. Consider the family

$$\begin{aligned}\Omega \;=\; & \Omega_0(\{Tweety\}, \{Bird, Penguin, Flies, Ab_1, Ab_2, Ab_3\}, \{Ab_1, Ab_2, Ab_3\})[Bird(x) \rightsquigarrow Flies(x), Ab_1] \\ & [Penguin(x) \rightsquigarrow Bird(x), Ab_2][Penguin(x) \rightsquigarrow \neg Flies(x), Ab_3][Penguin(Tweety)]\end{aligned}$$

Then we have that $App_\Omega(\{(Bird(x), Flies(x), Ab_1)\}, Tweety)$ is a conjunction containing the term

$$E(Penguin(Tweety) \wedge (\neg Ab_2(Tweety) \wedge \neg Ab_3(Tweety)) \wedge (Bird(Tweety) \rightarrow \neg Ab_1(Tweety)))$$

But any world that would witness this term would have $Bird(x)$ (since $\neg Ab_2(x)$) and therefore also have $\neg Ab_1(Tweety)$. Then, having both $Bird(Tweety), \neg Ab_1(Tweety)$ and $Penguin(Tweety), \neg Ab_3(Tweety)$ it would have both $Flies(Tweety)$ and $\neg Flies(Tweety)$. Since this is absurd, we conclude that $\neg App_\Omega(\{(Bird(x), Flies(x), Ab_1)\}, Tweety)$, ie the first rule does not apply to Tweety in Ω.
After updating with $[!Prag_\Omega]$, all worlds will have $Ab_1(Tweety)$. Since with the first rule taken out there is no longer any conflict, the conclusions that $Bird(Tweety)$ and $\neg Flies(Tweety)$ can then

be easily obtained.[17]

3.4.3 Modus tollens

A good test for non-arbitraryness. Consider the family

$$\Omega = \Omega_0(\{Tweety\}, \{Bird, Flies, Ab_1\}, \{Ab_1\})[Bird(x) \rightsquigarrow Flies(x), Ab_1][\neg Flies(Tweety)]$$

Since we don't have that $s \subseteq |Bird(Tweety)|$, we get that $App_\Omega(\{(Ax, Dx, Ab_1)\}, Tweety)$ is a trivially true empty conjunction. Therefore the $[!Prag_\Omega]$ update doesn't eliminate any world.
Now all optimal models satisfy $\neg Ab_1(Tweety)$ and therefore $\neg Flies(Tweety) \rightarrow \neg Bird(Tweety)$ (contraposition), so we have $\Omega \Vdash \neg Bird(Tweety)$.
Note that even if we consequently update with "Penguins are birds" and "Penguins don't fly" (starting out with a larger Ω_0 of course), the App_Ω formulas will still be trivially true and the conclusion will still be that Tweety is not a bird (and not a penguin, either): without additional information about Tweety, the idea that Tweety is actually a penguin is too exceptional to consider (if this seems contrary to intuition, I'll counter with the argument that the fact that penguins are suddenly being mentioned tends to be seen as constituting such additional information, even though it really isn't). Only if we also add the information that $Penguin(Tweety)$ do we get back at the result from the previous example.

3.4.4 Easterly wind

An exceptionally tricky example, which not a single other system seems to get right.

> If it rains, then normally it's cold.
> If there is an easterly wind, then normally it rains but isn't cold.
> It rains and there is an easterly wind.

Let R, C, W stand for it Rains, it's Cold, there is an easterly Wind, respectively. To add an object to the mix, let a stand for Amsterdam where we will let all of this take place.

[17] The optimal worlds are those with $\neg Ab_2(Tweety)$ and $\neg Ab_3(Tweety)$, which in turn makes all of them satisfy $Bird(Tweety)$ and $\neg Flies(Tweety)$

Now consider the family

$$\begin{aligned}\Omega &= \Omega_0(\{a\},\{R,C,W,Ab_1,Ab_2\},\{Ab_1,Ab_2\})[Rx \rightsquigarrow Cx, Ab_1] \\ &\quad [Wx \rightsquigarrow (Rx \wedge \neg Cx), Ab_2][Ra \wedge Wa]\end{aligned}$$

Now, the first rule doesn't apply to Amsterdam, because the term $E(Wa \wedge \neg Ab_2 \wedge (Rx \rightarrow \neg Ab_1))$ is false in Ω (as it implies $E(Wa \wedge Ra \wedge \neg Ca \wedge (Ra \rightsquigarrow Ca)))$. The second rule is applicable, notably the more interesting term $E(Ra \wedge \neg Ab_1 \wedge (Wa \rightarrow \neg Ab_2))$ is true by virtue of a certain possible world without Wa.
From there on after taking the usual steps it follows that $\neg Ca$; if it rains and there is an easterly wind, it's not cold.

3.4.5 Circular defaults

The following is a typical worst-case example of circular defaults.
Let B, P, t stand for *Bird*, *Penguin* and *Tweety*, respectively. Consider the family

$$\Omega = \Omega_0(\{t\},\{B,P,Ab_1,Ab_2\},\{Ab_1,Ab_2\})[Px \rightsquigarrow Bx, Ab_1][Bx \rightsquigarrow \neg Px, Ab_2][Pt]$$

We have that $App_\Omega(\{(By, Py, Ab_2)\}, t)$ is equivalent to

$$E(Pt \wedge \neg Ab_1 t \wedge (Bt \rightarrow \neg Ab_2 t))$$

which under the circumstances implies the absurd $E(Pt \wedge Bt \wedge (Bt \rightarrow \neg Pt))$.
Hence the second rule does not apply to Tweety, and from there on with the usual steps it follows that $\Omega \Vdash Bt$.
Note that if the information were instead Bt we would have gotten the then intuitively correct conclusion $\neg Pt$ in a similar fashion.

3.4.6 Dealing with additional information

Not something we would expect to be a problem, but still it's a good thing to see this non-expectation work out.
Let B, P, F, G stand for *Bird*, *Penguin*, *Flies*, *lays eGgs* respectively. Consider the family

$$\begin{aligned}\Omega &= \Omega_0(\{t\},\{A,P,F,E,Ab_1,Ab_2,Ab_3,Ab_4\},\{Ab_1,Ab_2,Ab_3,Ab_4\})\ldots \\ &\quad \ldots[Bx \rightsquigarrow Fx, Ab_1][Px \rightsquigarrow \neg Fx, Ab_2][Px \rightsquigarrow Bx, Ab_3][Ax \rightsquigarrow Gx, Ab_4][Pt]\end{aligned}$$

As before we obtain Ab_1t and $\Omega \Vdash Bt \wedge \neg Ft$. Amongst the models satisfying $Ab_1t \wedge \neg Ab_2t \wedge \neg Ab_3t$, some also satisfy $\neg Ab_4t$ and some don't. Those that do are preferred to those that don't, and therefore the optimal models make $\neg Ab_4t$ true.
Hence we can conclude $\Omega \Vdash Gt$.

3.5 Further notes

As we've seen, the combined system achieves the goal of transferring the empirical adequacy of the update semantics system onto circumscription, allowing us to deal systematically and correctly with the "Birds fly" example and many others (e.g. almost all applicable examples from [4] and [12]). There are a few other things still worth talking about:

3.5.1 Inherent Predicate Logic

Technically speaking Defaults in Update Semantics has only a propositional system, and only creates the illusion of supporting examples ascribing properties to objects by having only one object per example. But circumscription and the combined system *are* predicate-logical. It is interesting to note that we obtain correct results on examples involving multiple objects and even on examples involving quantified conclusions (for example, $\Omega_0[Bx \rightsquigarrow Fx, Ab]$ defeasibly supports the conclusion $\forall x : Bx \rightarrow Fx$, ie before learning of penguins we expect that all birds fly) or quantified properties ($(\Omega_0[\forall y : xLy \rightsquigarrow \forall y : xHy, Ab][\forall y : aLy] \Vdash \forall y : aHy$, ie if the largest object is usually the heaviest and a is the largest object, we expect that a is indeed the heaviest).

3.5.2 Clarity

The embedding from the previous chapter was not an overwhelming success in terms of clearing things up, but by not being limited by the need for an exact embedding we were in a much better starting position this time around. Indeed, this is where we take the failures of the previous chapter and turn them into successes, as our previous experience has taught us what we should keep and what to throw out.

As mentioned in the technical sections, the first is the excessive checking of information states in the pragmatics, which we have unproblematically reduced to only one instance per rule.

The second is the system for generalized conclusions. Defaults in Update Semantics had a means by which to conclude default statements, but it was problematic from the start because the conclusions needed to be undefeasible. This is obviously inappropriate for a non-monotonic logic, and in practice led to such statements being concludable only in highly trivial situations (not to mention to the single most unappealing part of the embedding).

In taking the predicate-logical turn we have now obtained a more appropriate alternative, in the

form of defeasible universally quantified conclusions. This is not without issues of its own (since penguins always violate some rule we will end up effectively minimizing the amount of penguins; not varying the penguin predicate would solve this but is the single most undesirable solution as it puts us back at square one), but does give rise to all the desirable conclusions. Hence it is still a significant improvement over the status quo.
Finally, there is the remark that started out this chapter. The most significant issue of the previous chapter, that the pragmatics necessitated an extra predicate per rule, has turned into a non-issue by seeking out the right partner system.

All of that notwithstanding, the analysis wouldn't be complete without a (quick) look at how the clarity of circumscription is affected. Here it's worth noting that the Modal Circumscription Family framework allows us to easily circumscribe one variable at the time and to intersperse information and circumscription orders almost arbitrarily, resulting in a more incremental system. This doesn't appear to have anything to do with clarity... until you go ahead and circumscribe multiple variables in a conjunction of four or more sentences -as needs to be done even with the "Birds Fly" example. The way the relation between models is updated might take some getting used to, but ultimately the resulting relation and the parlance used in the definition should be rather familiar, to the circumscriptionist at least. Keeping our conclusion on the modest side, at least the modal framework certainly isn't less clear by any significant amount.

4 Conclusions

To create your "conclusions" section, take your abstract and put it in the past tense.

-unknown

What can an advocate of clarity have to say about his own piece without tacitly suggesting it wasn't clear enough to begin with? In this light, perhaps I should have taken an optimistic rather than pessimistic starting position, and should have started out not with the above quote but with the following quote about self-explanatoryness[18]:

We don't have time to explain it to ourselves, so why don't you explain it to us?

Well then: Clarity in Non-Monotonic Reasoning, what are the results of our brief foray into this subject? Whether we view the underlying psychological reasons in a cynical or optimistic fashion. the fact remains that a lack of clarity can make a system unpopular regardless of merit, and thereby withhold important results from our view. The system for conditional defaults from Defaults in Update Semantics is a good example: it had much merit in terms of empirical adequacy but went almost entirely unnoticed.

Now, improving its clarity through an embedding was only partly successful, but this should not come as too much of a surprise or disappointment: if the unclarity were entirely due to bad phrasing, *that* would be surprising! Rather, the way to look at it is that at least as interesting as the clarity gained is the meta-clarity we've gained about what the problematic parts are.

This is true not just because the meta-clarity allowed us to refine our way out of the problem with great precision. Much more interestingly, after putting the system in its most natural form it immediately became clear that the most natural and potentially fruitful application of it would be to combine it with circumscription. While already knowing something of value was to be found, it is only because of our clarification efforts that we've got to see where and how to apply it for optimal effect.

The result was a great success: the DEUL-inspired modal framework for circumscription is clear and of some interesting in it's own right, and like we anticipated the price of adding the pragmatics to it is quite low[19]. Furthermore the system generates the intuitively correct results for almost any

[18] taken from a popular computer game (*Sam & Max Season 1*)

[19] We even retain the classical approach of circumscribing abnormality predicates while varying the rest.

example, so at this point little more can be said about it without going beyond the scope of this section. Hence[20] this is as good a point as any on which to conclude our brief stroll.

[20]and since I prefer not to use a cliché like "It works well and conveniently leaves room for further research.",

A Recap: Nonmonotonic logic

Classical logics are *monotonic*, which means that gaining new information can never be sufficient cause to force you to retract a conclusion. However, this property may not always be desirable. For example, if we learn that Tweety is a bird we may want to conclude that he can fly, but if we then learn that he is a penguin we may want to retract that conclusion.
Basically, nonmonotonic logic is a field encompassing any logic that doesn't have the property of monotonicity. It can be seen as closely related to the field of belief revision, especially since these defeasible conclusions are usually not considered to have the status of knowledge. It is mostly concerned with examples such as these and our intuitions about them.

A.1 Default reasoning

Default reasoning is about *default rules*: rules which are true "by default" but which allow for exceptions (such as "Birds normally can fly."). Since the knowledge that something is in fact an exception can cancel out an earlier expectation, it comes as no surprise that default reasoning is one of the main objects of investigation in nonmonotonic logic.

B Recap: Defaults in Update Semantics

This section contains all the relevant definitions from Defaults in Update Semantics, which should allow the reader to verify the technical parts of the embedding. For various reasons this section stays as close to the original as possible in terms of wording, presentation and even order of appearance of the definitions. This also means that consuming this section without at least passing familiarity with the original is ill-advised.
If this section seems to be lacking in clarity and presentation, that is part of the point of this thesis. In terms of clarity it is also interesting to note that if we take the explanatory text and proofs out of the embedding, this section is in fact larger than the sum of the embedding and the system embedded into, (recap in the next section). But quips like these aside, let's get to it:

As the general framework is introduced earlier on, the notions of acceptance and validity don't actually appear as a definition unto themselves but rather with the introductory chapter dealing with a more general framework. The formal definitions are as follows, though they will only make sense after reading the definitions further below.
Here σ is an *information state*, $\sigma[\phi]$ is *the result of updating* σ *with the formula* ϕ, $\sigma \Vdash \phi$ means σ *accepts* ϕ, and $\psi_1, \ldots, \psi_n \Vdash_i \phi$ means ϕ is a *valid* conclusion to draw from $\psi_1, \ldots, \psi_n$, for various notions of validity.

Definition B.1

$$
\begin{aligned}
\sigma \Vdash \phi \quad &\text{iff} \quad \sigma[\phi] = \sigma \\
\psi_1, \ldots, \psi_n \Vdash_1 \phi \quad &\text{iff} \quad \boldsymbol{0}[\psi_1] \ldots [\psi_n] \Vdash \phi \\
\psi_1, \ldots, \psi_n \Vdash_2 \phi \quad &\text{iff} \quad \text{for every } \sigma, \sigma[\psi_1] \ldots [\psi_n] \Vdash \phi \\
\psi_1, \ldots, \psi_n \Vdash_2 \phi \quad &\text{iff} \quad \sigma \Vdash \phi \text{ for every } \sigma \text{ such that } \sigma \Vdash \psi_1, \ldots, \sigma \Vdash \psi_n
\end{aligned}
$$

For convenience, the following table matches up definitions from Defaults in Update Semantics with the same definitions from this appendix.

DIUS	App.	DIUS	App.
3.1	B.2	4.1	B.10
3.2	B.3	4.2	B.11
3.3	B.4	4.3	B.12
3.4	B.5	4.4	B.13
3.5	B.6	4.5	B.14
3.8	B.7	4.6	B.15
3.9	B.8	4.9	B.16
3.11	B.9	4.13	B.17

B.1 Rules with exceptions [*presumably* and *normally*]

Definition B.2 *Let* $\boldsymbol{A}$ *be a set consisting of finitely many* atomic sentences. *With* $\boldsymbol{A}$ *we associate two languages,* $L_0^{\boldsymbol{A}}$ *and* $L_2^{\boldsymbol{A}}$. *Both have* $\boldsymbol{A}$ *as their non-logical vocabulary.* $L_0^{\boldsymbol{A}}$ *has as its logical vocabulary one unary operator* $\neg$, *two binary operators* $\wedge$ *and* $\vee$, *and two parenthese*) *and* (. *The sentences of* $L_0^{\boldsymbol{A}}$ *are just the ones one would expect for a language with such a vocabulary.* $L_2^{\boldsymbol{A}}$ *has in its logical vocabulary two additional unary operators:* normally, *and* presumably. *A string of symbols* ϕ *is a sentence of* $L_2^{\boldsymbol{A}}$ *iff there is a sentence* ψ *of* $L_0^{\boldsymbol{A}}$ *such that either* $\phi = \psi$, *or* $\phi =$ normally ψ, *or* $\phi =$ presumably ψ.

Definition B.3 *Let* W *be the powerset of the set* $\boldsymbol{A}$ *of atomic sentences. Then* ε *is an* (expectation) pattern *on* W *iff* ε *is a reflexive and transitive relation on* W.

Definition B.4 *Let* ε *be a pattern on* W;
(i) w *is a* normal world *in* ε *iff* $w \in W$ *and* $w \leq_\varepsilon v$ *for every* $v \in W$;
(ii) $\boldsymbol{n}\varepsilon$ *is the set of all normal worlds in* ε;
(iii) ε *is* coherent *iff* $\boldsymbol{n}\varepsilon \neq \emptyset$.

Definition B.5 *Let* ε *be a pattern on* W, *and* $s \subseteq W$.
(i) w *is* optimal in $\langle \varepsilon, s \rangle$ *iff* $w \in s$ *and there is no* $v \in s$ *such that* $v <_\varepsilon w$;
(ii) $\boldsymbol{m}_{\langle \varepsilon, s \rangle}$ *is the set of all optimal worlds in* $\langle \varepsilon, s \rangle$.

Definition B.6 *Let* ε *and* ε' *be patterns on* W, *and* $e \subseteq W$.
(i) ε' *is a* refinement *of* ε *iff* $\varepsilon' \subseteq \varepsilon$;
(ii) $\varepsilon \cdot e = \{\langle v, w \rangle \in \varepsilon | \text{if } w \in e, \text{ then } v \in e\}$; $\varepsilon \cdot e$ *is* the refinement of ε with the proposition e.

Definition B.7 *Let W be as before.*

(i) σ *is an* information state *iff* $\sigma = \langle \varepsilon, s\rangle$ *and one of the following conditions is fulfilled:*

(a) ε *is a coherent pattern on* W *and* s *is a non-empty subset of* W*;*

(b) $\varepsilon = \{\langle w, w\rangle | w \in W\}$ *and* $s = \emptyset$*;*

(ii) $\boldsymbol{0}$*,* the minimal state*, is the state given by* $\langle W \times W, W\rangle$*;*

$\boldsymbol{1}$*,* the absurd state*, is the state given by* $\langle \{\langle w, w\rangle | w \in W\}, \emptyset\rangle$*;*

(iii) Let $\sigma = \langle \varepsilon, s\rangle$ *and* $\sigma' = \langle \varepsilon', s'\rangle$ *be states.*

$\sigma + \sigma' = \langle \varepsilon \cap \varepsilon', s \cap s'\rangle$*, if* $\langle \varepsilon \cap \varepsilon', s \cap s'\rangle$ *is coherent;*

$\sigma + \sigma' = \boldsymbol{1}$*, otherwise.*

Definition B.8 *Let* $\sigma = \langle \varepsilon, s\rangle$ *be an information state. For every sentence* ϕ *of* $L_2^{\boldsymbol{A}}$*,* $\sigma[\phi]$ *is determined as follows:*

if ϕ *is a sentence of* $L_0^{\boldsymbol{A}}$*, then*

· *if* $s \cap \|\phi\| = \emptyset$*,* $\sigma[\phi] = \boldsymbol{1}$*;*

· *otherwise,* $\sigma[\phi] = \langle \varepsilon, s \cap \|\phi\|\rangle$*.*

if $\phi =$ normally ψ*, then*

· *if* $\boldsymbol{n}\varepsilon \cap \|\phi\| = \emptyset$*,* $\sigma[\phi] = \boldsymbol{1}$*;*

· *otherwise,* $\sigma[\phi] = \langle \varepsilon \cdot \|\phi\|, s\rangle$*.*

if $\phi =$ presumably ψ*, then*

· *if* $\boldsymbol{m}_\sigma \varepsilon \cap \|\phi\| = \boldsymbol{m}_\sigma$*,* $\sigma[\phi] = \sigma$*;*

· *otherwise,* $\sigma[\phi] = \boldsymbol{1}$*.*

Definition B.9 *Let* $\langle \varepsilon, s\rangle$ *be an information state.*

(i) $\boldsymbol{m}$ *is an* optimal set in $\langle \varepsilon, s\rangle$ *iff there is some optimal world* w *in* $\langle \varepsilon, s\rangle$ *such that* $\boldsymbol{m} = \{v \in s | v \cong_\epsilon w\}$*;*

(ii) $\langle \varepsilon, s\rangle$ *is ambiguous if there is more than one optimal set in* $\langle \varepsilon, s\rangle$*.*

B.2 Rules for exceptions [conditional defaults]

Definition B.10 *Let* $\boldsymbol{A}$ *and* $L_0^{\boldsymbol{A}}$ *be as earlier. The language* $L_3^{\boldsymbol{A}}$ *has* $\boldsymbol{A}$ *as its non-logical vocabulary, and in its logical vocabulary one additional binary operator* $\rightsquigarrow$ *and one additional unary operator* presumable*. A string of symbols* ϕ *is a sentence of* $L_3^{\boldsymbol{A}}$ *iff there are sentences* ψ *and* χ *of* $L_0^{\boldsymbol{A}}$ *such that* $\phi = \psi$*, or* $\phi = \psi \rightsquigarrow \chi$*, or* $\phi =$ presumably ψ*.*

Definition B.11 *(i) Let W be as before. A* frame on W *is a function π assigning to every subset d of W a pattern πd on d.*
(ii) Let π be a frame on W and $d, e \subseteq W$. The proposition e is a default in πd iff $d \cap e \neq \emptyset$ and $\pi d \cdot e = \pi d$.

Definition B.12 *Let π be a frame on W, and $d \subseteq W$.*
(i) w is a normal *world in πd iff $w \in d$ and for every $d' \subseteq d$ such that $w \in d'$ it holds that $w \leq_{\pi d'} v$ for every $v \in d'$;*
(ii) $\boldsymbol{n}\pi d$ is the set of all normal worlds in πd;
(iii) π is coherent *iff for every non-empty $d \subseteq W$, $\boldsymbol{n}\pi d \neq \emptyset$.*

Definition B.13 *Let W be as before.*

(i) *σ is an* information state *iff $\sigma = \langle \pi, s \rangle$ and one of the following conditions is fulfilled:*
(a) π is a coherent frame on W, and s is a non-empty subset of W;
(b) π is the frame $\langle \iota, \emptyset \rangle$, where $\iota d = \{\langle w, w \rangle | w \in d\}$ for every $d \subseteq W$.

(ii) *$\boldsymbol{0} = \langle \upsilon, W \rangle$, where $\upsilon d = d \times d$ for every $d \subseteq W$.*
$\boldsymbol{1} = \langle \iota, \emptyset \rangle$.

(iii) *Let $\sigma = \langle \pi, s \rangle$ and $\sigma' = \langle \pi', s' \rangle$ be states.*
Let π'' be the frame such that for every d, $\pi'' d = \pi d \cap \pi' d$. Then
$\sigma + \sigma' = \langle \pi'', s \cap s' \rangle$, if $\langle \pi'', s \cap s' \rangle$ is coherent;
$\sigma + \sigma' = \boldsymbol{1}$, otherwise.

Definition B.14

(i) *Let π and π' be frames, both based on W.*
The frame π is a refinement *of π' iff $\pi d \subseteq \pi' d$ for every $d \subseteq W$.*

(ii) *Let π be a frame and $d, e \subseteq W$. $\pi_{d \cdot e}$ is the refinement of π given by*
(a) if $d' = d$, then $\pi_{d \cdot e} d' = \pi d'$;
(b) $\pi_{d \cdot e} d = \pi d \cdot e$.
The frame $\pi_{d \cdot e}$ is the result of refining πd in π with e.

Definition B.15 *Let $\sigma = \langle \pi, s \rangle$ be an information state.*
· *$\sigma[\phi \rightsquigarrow \psi] = \boldsymbol{1}$ if $\|\phi\| \cap \|\psi\| = \emptyset$ or $\pi_{\|\phi\| \cdot \|\psi\|}$ is incoherent.*
· *Otherwise, $\sigma[\phi \rightsquigarrow \psi] = \langle \pi_{\|\phi\| \cdot \|\psi\|}, s \rangle$.*

Definition B.16 *Let $\sigma = \langle \pi, s \rangle$ be a coherent information state and assume that $e_1, \ldots, e_n$ are defaults in $\pi d_1, \ldots, \pi d_n$ respectively.*
(i) A world w complies with $\{e_1, \ldots, e_n\}$ *iff $w \in e_i$ for every i such that $w \in d_i$ $(1 \leq i \leq n)$.*
(ii) The set of defaults $\{e_1, \ldots, e_n\}$ applies within *s iff for every $d \supseteq s$ there is some $w \in \boldsymbol{n}\pi d$ such that w complies with $\{e_1, \ldots, e_n\}$. In this case, we also say that $e_1, \ldots, e_n$* jointly apply within *s.*

Definition B.17 *Let $\sigma = \langle \pi, s \rangle$ be a coherent information state and assume that $e_1, \ldots, e_n$ are defaults in $\pi d_1, \ldots, \pi d_n$.*

(i) *Then* $\{e_1, \ldots, e_n\}$ *is* a maximal applicable set in *σ iff $e_1, \ldots, e_2$ jointly apply within s, and for every e_{n+1} and d_{n+1} such that e_{n+1} is a default in πd_{n+1}, and $e_1, \ldots, e_{n+1}$ jointly apply within s it holds that $e_{n+1} = e_i$ and $d_{n+1} = d_i$ for some $i \leq n$.*

(ii) *A world w is* optimal in *σ iff $w \in s$ and w complies with a maximal applicable set of defaults. The set of optimal worlds is denoted by $\boldsymbol{m}_\sigma$.*

(iii) *σ*[presumably *ψ*] *is determined as follows:*
· *If $\boldsymbol{m}_\sigma \cap \|\psi\| = \boldsymbol{m}_\sigma$, then σ*[presumably *ψ*] *$= \sigma$.*
· *Otherwise, σ*[presumably *ψ*] *$= \boldsymbol{1}$*

C Recap: The Dynamic-Epistemic Upgrade Logic

The Dynamic-Epistemic Upgrade Logic DEUL is introduced in the paper Dynamic Logic for Preference Upgrade[10]. We will use the version with world-eliminating announcement $!\varphi$ and use slightly different notation.
Outside this appendix we will also omit the subscripts. These are used in DEUL to deal with multi-agent situations, but for our purposes we can do fine effectively having only one of them.

Definition C.1 *(Language) Take a set of propositional variables P and a set of agents I, with p ranging over P and i over I. The* ***dynamic epistemic preference language*** *is given by:*

$$\begin{aligned} \varphi &::= \bot | p | \neg\varphi | \varphi \wedge \psi | \Box_i \varphi | \Box_i^{\leq} \varphi | U\varphi | \langle\pi\rangle\varphi \\ \pi &::= !\varphi | \#\varphi \end{aligned}$$

(Furthermore we will use $\vee$, $\Diamond_i$, $\Diamond_i^{\leq}$ and E being the duals of $\wedge$, $\Box_i$, $\Box_i^{\leq}$ and U, respectively.)

Definition C.2 *(Models) An* ***epistemic preference model*** *is a tuple $M = (S, \{\sim_i | i \in I\}, \{\leq_i | i \in I\}, V)$, with S a set of possible worlds, $\sim_i$ the usual equivalence relation of epistemic accessibility for agent i, and V a valuation for proposition letters. Moreover, $\leq_i$ is a reflexive and transitive relation over the worlds.*

Definition C.3 *(Model Upgrades) Let $M = (S, \{\sim_i | i \in I\}, \{\leq_i | i \in I\}, V)$. Then the* ***upgraded models*** *$M_{!\varphi}$ and $M_{\#\varphi}$ are defined as follows:*

$$\begin{aligned} M_{!\varphi} &:= (S', \{\sim_i | i \in I\} \cap (S' \times S'), \{\leq_i | i \in I\} \cap (S' \times S'), V'), \text{ where} \\ S' &= \{w \in S | M, w \models \varphi\} \\ V'(p) &= V(p) \cap S' \end{aligned}$$

$$\begin{aligned} M_{\#\varphi} &:= (S, \{\sim_i | i \in I\}, \{\leq_i^* | i \in I\}, V), \text{ where} \\ \leq_i^* &= \leq_i - \{(s,t) | M, s \models \varphi \text{ and } M, t \models \neg\varphi\} \end{aligned}$$

Definition C.4 *(Semantics) Given an epistemic preference model $M = (S, \{\sim_i | i \in I\}, \{\leq_i | i \in I\}, V)$, and a world $s \in S$, we define $M, s \models \varphi$ (****formula φ is true in M at s****) by induction on*

φ:

$$M, s \models p \quad \text{iff} \quad s \in V(p) \tag{1}$$

$$M, s \models \neg\varphi \quad \text{iff} \quad \text{not } M, s \models \varphi \tag{2}$$

$$M, s \models \varphi \wedge \psi \quad \text{iff} \quad M, s \models \varphi \text{and} M, s \models \psi \tag{3}$$

$$M, s \models \Box_i\varphi \quad \text{iff} \quad \text{for all } t : s \sim_i t \text{ implies } M, t \models \varphi \tag{4}$$

$$M, s \models \Box_i^{\leq}\varphi \quad \text{iff} \quad \text{for all } t : s \leq_i t \text{ implies } M, t \models \varphi \tag{5}$$

$$M, s \models U\varphi \quad \text{iff} \quad \text{for all } t : M, t \models \varphi \tag{6}$$

$$M, s \models \langle\pi\rangle\varphi \quad \text{iff} \quad M_\pi, s \models \varphi \tag{7}$$

(With convention that if s is eliminated by the update π then it is not the case that $M_\pi, s \models \varphi$.)

One of the interesting things about $DEUL$ is the fact it has *reduction axioms* eliminating the model-upgrade modalities, making it self-embed into regular modal logic. For more on the importance of these and what to do with them, consult [10].

Theorem C.5 *(Soundness) The following formulas are valid:*

$$\begin{aligned}
\langle!\varphi\rangle p &\leftrightarrow \varphi \wedge p \\
\langle!\varphi\rangle\neg\psi &\leftrightarrow \varphi \wedge \neg\langle!\varphi\rangle\psi \\
\langle!\varphi\rangle(\psi \wedge \phi) &\leftrightarrow \langle!\varphi\rangle\psi \wedge \langle!\varphi\rangle\phi \\
\langle!\varphi\rangle\Diamond_i\psi &\leftrightarrow \varphi \wedge \Diamond_i\langle!\varphi\rangle\psi \\
\langle!\varphi\rangle\Diamond_i^{\leq}\psi &\leftrightarrow \varphi \wedge \Diamond_i^{\leq}\langle!\varphi\rangle\psi \\
\langle!\varphi\rangle E\psi &\leftrightarrow \varphi \wedge E\langle!\varphi\rangle\psi
\end{aligned}$$

$$\begin{aligned}
\langle\#\varphi\rangle p &\leftrightarrow p \\
\langle\#\varphi\rangle\neg\psi &\leftrightarrow \neg\langle\#\varphi\rangle\psi \\
\langle\#\varphi\rangle(\psi \wedge \phi) &\leftrightarrow \langle\#\varphi\rangle\psi \wedge \langle\#\varphi\rangle\phi \\
\langle!\varphi\rangle\Diamond_i\psi &\leftrightarrow \Diamond_i\langle\#\varphi\rangle\psi \\
\langle\#\varphi\rangle\Diamond_i^{\leq}\psi &\leftrightarrow (\neg\varphi \wedge \Diamond_i^{\leq}\langle\#\varphi\rangle\psi) \vee (\Diamond_i^{\leq}(\varphi \wedge \langle!\varphi\rangle\psi)) \\
\langle!\varphi\rangle E\psi &\leftrightarrow E\langle!\varphi\rangle\psi
\end{aligned}$$

D Recap: Circumscription

This definition taken directly from Lifschitz ([5]), the simplest form of circumscription is as follows: First off, for any predicates of the same arity,

$$\begin{aligned} P = Q \quad & \textit{stands for} \quad \forall x(P(x) \equiv Q(x)) \\ P \leq Q \quad & \textit{stands for} \quad \forall x(P(x) \supseteq Q(x)) \\ P < Q \quad & \textit{stands for} \quad (P \leq Q) \wedge \neg(P = Q) \end{aligned}$$

Now if $A(P)$ is a sentence containing the predicate constant P, *the circumscription of* P *in* $A(P)$ is simply the following second-order sentence:

$$A(P) \wedge \neg\exists p[A(p) \wedge p < P]$$

This added clause in this sentence can be seen as meaning that you cannot reduce the extension of P while retaining the truth of $A(P)$. However, it is problematic in only talking about reducing the extension of P while leaving all other predicates the same.
Hence there is also such a thing as *the circumscription of* P *in* $A(P)$ *with varied* $Z_1, \ldots, Z_n$, which is the following second-order sentence:

$$A(P, Z_1, \ldots, Z_n) \wedge \neg\exists p, z_1, \ldots, z_n[A(p, z_1, \ldots, z_n) \wedge p < P]$$

A circumscription can be further circumscribed in, and varying a predicate that was circumscribed before can be used as a prioritizing tool.
Examples involving default rules are typically handled using abnormality predicates which are circumscribed against while varying the other predicates.

References

[1] N. Asher and M. Morreau. Commonsense entailment: A modal theory of nonmonotonic reasoning. In J. van Eijck, editor, *Logics in AI: Proc. of the European Workshop JELIA'90*, pages 1–30. Springer, Berlin, Heidelberg, 1991.

[2] M. L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, Los Altos, CA, 1987.

[3] J. Groenendijk and M. Stokhof. Two theories of dynamic semantics. In J. van Eijck, editor, *Logics in AI: Proc. of the European Workshop JELIA'90*, pages 55–64. Springer, Berlin, Heidelberg, 1991.

[4] V. Lifschitz. Benchmark problems for formal nonmonotonic reasoning, version 2.00. In M. Reinfrank, J. de Kleer, M. L. Ginsberg, and E. Sandewall, editors, *Non-Monotonic Reasoning: 2nd International Workshop, Grassau, Germany*, pages 202–219. Springer, Berlin, Heidelberg, 1989.

[5] Vladimir Lifschitz. Circumscription. In Dov Gabbay, Christopher J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, pages 298–352. Oxford University Press, 1994.

[6] Aarati Martino. *Formalizing Elaboration Tolerance*. PhD thesis, Stanford University, August 2003.

[7] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. In M. L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 153–166. Kaufmann, Los Altos, CA, 1987.

[8] J. McCarthy. Circumscription: A form of non-monotonic reasoning. In M. L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 145–151. Kaufmann, Los Altos, CA, 1987.

[9] Johan van Benthem. Situation calculus meets modal logic.

[10] Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17(2):157–182, 2007. forthcoming.

[11] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[12] Gerard Vreeswijk. Interpolation of benchmark problems in defeasible reasoning. In *WOCFAI*, pages 453–468, 1995.