FRANK VELTMAN

DEFAULTS IN UPDATE SEMANTICS

ABSTRACT. The aim of this paper is twofold: (i) to introduce the framework of update semantics and to explain what kind of phenomena may successfully be analysed in it; (ii) to give a detailed analysis of one such phenomenon: default reasoning.
KEY WORDS: dynamic semantics, defaults, epistemic modalities.

1. INTRODUCTION: THE FRAMEWORK OF UPDATE SEMANTICS

The standard definition of logical validity runs as follows: An argument is valid if its premises cannot all be true without its conclusion being true as well. Most logical theories developed so far have taken this definition of validity as their starting point. Consequently, the heart of these theories consists in a specification of truth conditions. The heart of the theories developed in this paper does not consist in a specification of truth conditions. The slogan ‘You know the meaning of a sentence if you know the conditions under which it is true’ is replaced by this one: ‘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’.[1] Thus, meaning becomes a dynamic notion: the meaning of a sentence is an operation on information states.

To define an update semantics for a language L, one has to specify a set Σ of relevant information states, and a function [ ] that assigns to each sentence ϕ an operation $[\phi]$ on Σ. The resulting triple $\langle L, \Sigma, [\ ]\rangle$ is called an update system. If σ is a state and ϕ a sentence, we write ‘$\sigma[\phi]$’ to denote the result of updating σ with ϕ. Since $[\phi]$ is the function and σ the argument, it would have been more in line with common practice to write ‘$[\phi](\sigma)$’, but postfix notation is more convenient for dealing with texts. Now we can write ‘$\sigma[\psi_1]...[\psi_n]$’ for the result of updating σ with the sequence of sentences $\psi_1,..., \psi_n$.

An important notion is the notion of *acceptance*. Let σ be any state and ϕ be any sentence. Consider the state $\sigma[\phi]$. This state will in most cases be different from σ, but every now and then it may happen that $\sigma[\phi]=\sigma$.

If so, the information conveyed by ϕ is already subsumed by σ. In such a case we write $\sigma \Vdash \phi$ and we say that ϕ *is accepted in σ.*

1.1 Constraints that Do Not Aways Hold

The phrase 'update semantics' might be misleading in that it suggests that all you have to do in order to update your information state with ϕ is to add the informational content of ϕ to the information you already have.

DEFINITION 1.1. An update system $\langle$ L, Σ, [ ] $\rangle$ is *additive* iff there exists a state **0**, the *minimal* state, in Σ and a binary operation + on Σ such that

(i) the operation + has all the properties of a join operation:

$$\mathbf{0} + \sigma = \sigma;$$
$$\sigma + \sigma = \sigma;$$
$$\sigma + \tau = \tau + \sigma;$$
$$(\rho + \sigma) + \tau = \rho + (\sigma + \tau).$$

(ii) for every sentence ϕ and state σ, $\sigma[\phi] = \sigma + \mathbf{0}[\phi]$.

Whenever (i) holds Σ is called an information lattice. If $\sigma + \tau = \tau$, we will write $\sigma \leq \tau$, and say that τ is *at least as strong* as σ.

As long as one is dealing with phenomena that can be captured by a classical update system, the dynamic approach has nothing to offer over and above the static approach. In such cases one can associate with every sentence ϕ of L a static meaning — $\mathbf{0}[\phi]$, representing 'the' informational content of ϕ — and define the dynamic meaning of ϕ in terms of it.

There are various constraints that must be fulfilled by an update system for it to be additive. For one thing, $\sigma[\phi]$ should be defined for every σ. The systems discussed in this paper have this property, but it is not difficult to think of phenomena that cannot be covered in this way. Take the case of a pronoun desperately looking for a referent:

'He is just joking.'

If it is not clear to whom the speaker is referring, the hearer will not know what to do with this statement. Or take the case of presupposition. The framework of update semantics offers a natural explanation of this notion:

ϕ *presupposes* ψ iff for every state σ, $\sigma[\phi]$ is defined only if $\sigma \Vdash \psi$

Clearly, this definition can only be instrumental in systems in which $\sigma[\phi]$ is sometimes undefined.[2]

Another necessary condition for an update system to be additive is this:

Idempotence: For every state σ and sentence ϕ, $\sigma[\phi] \Vdash \phi$.

At first sight this principle goes without saying. What would ‘updating your state with ϕ’ mean if not at least ‘changing your state in such a manner that you come to accept ϕ’? Still, there are sentences for which no successful update exists. Here paradoxical sentences like ‘This sentence is false’ are a case in point. As shown in Groeneveld[1994], the paradoxicality of this sentence resides in the fact that every time you try to accommodate the information it conveys, you have to change your mind.

A third constraint worth looking at is the principle of Persistence:

Persistence: If $\sigma \Vdash \phi$ and $\sigma \leq \tau$, then $\tau \Vdash \phi$.

The clearest examples of non-persistent sentences can be found among sentences in which modal qualifications like ‘presumably’, ‘probably’, ‘must’, ‘may’ or ‘might’ occur. Consider for example the next two sequences. Processing the first does not cause any problems, but processing the second does.

Somebody is knocking at the door... Maybe it's John... It's Mary.
Somebody is knocking at the door... Maybe it's John... It's Mary ... Maybe it's John

Explanation: it is quite normal for one's expectations to be overruled by the facts — that is what is going on in the first sequence. But once you know something, it is a bit silly to pretend that you still expect something else, which is what is going on in the second.

One of the advantages of the dynamic approach is that these differences can be accounted for. The set-up enables us to deal with sequences of sentences, whole texts. Let ϕ_1 = ‘Somebody is knocking at the door’, ϕ_2 = ‘Maybe it's John’, and ϕ_3= ‘It's Mary’. If we want, we can compare $\sigma[\phi_1][\phi_2][\phi_3]$ with $\sigma[\phi_1][\phi_2][\phi_3][\phi_2]$ for any state σ, and see if there are any differences.

There are two more important constraints:

Strengthening: $\sigma \leq \sigma[\phi]$

Monotony: If $\sigma \leq \tau$, then $\sigma[\phi] \leq \tau[\phi]$.

We will have more to say on these in due course. As for now, we note

PROPOSITION 1.2. An update system $\langle L, \Sigma, [\ ] \rangle$ is additive iff (i) Σ is an update lattice on which [ ] is total, and (ii) the principles of Idempotence, Persistence, Monotony and Strengthening hold.

1.2. Notions of validity

Various notions of logical validity suggest themselves. The notion that will concern us most is this:

- An argument is valid$_1$ iff updating the minimal state **0** with the premises $\psi_1,..., \psi_n$ in that order, yields an information state in which the conclusion ϕ is accepted. Formally:

$$\psi_1,..., \psi_n \Vdash_1 \phi \text{ iff } \mathbf{0}[\psi_1]...[\psi_n] \Vdash \phi.$$

A more general notion of validity is this one:

- An argument is valid$_2$ iff updating any information state σ with the premises $\psi_1,..., \psi_n$ in that order, yields an information state in which the conclusion ϕ is accepted. Formally:

$$\psi_1,..., \psi_n \Vdash_2 \phi \text{ iff for every } \sigma, \sigma[\psi_1]...[\psi_n] \Vdash \phi.$$

And the next notion is closest to the classical one:

- An argument is valid$_3$ iff one cannot accept all its premises without having to accept the conclusion as well. More formally:

$$\psi_1,..., \psi_n \Vdash_3 \phi \text{ iff } \sigma \Vdash \phi \text{ for every } \sigma \text{ such that } \sigma \Vdash \psi_1,..., \sigma \Vdash \psi_n.$$

PROPOSITION 1.3. In every additive update system the following holds:

$$\psi_1,..., \psi_n \Vdash_1 \phi \text{ iff } \psi_1,..., \psi_n \Vdash_2 \phi \text{ iff } \psi_1,..., \psi_n \Vdash_3 \phi.$$

In general the three notions do not coincide. Notice that validity$_3$ is monotonic: If an argument with premises $\psi_1,..., \psi_n$ and conclusion ϕ is valid$_3$, then it remains valid$_3$ if you add more premises to $\psi_1,..., \psi_n$. Validity$_2$ is at least left monotonic:

$$\text{If } \psi_1,..., \psi_n \Vdash_2 \phi, \text{ then } \chi, \psi_1,..., \psi_n \Vdash_2 \phi.$$

Validity$_1$ is neither right nor left monotonic. But it is easy to verify that this notion conforms to the following principle of *Sequential Monotony*:

If $\psi_1, ..., \psi_n \Vdash_1 \phi$ and $\psi_1, ..., \psi_n, \theta_1, ..., \theta_k \Vdash_1 \chi$, then $\psi_1, ..., \psi_n, \phi, \theta_1, ..., \theta_k \Vdash_1 \chi$.

Moreover, validity$_1$ complies with the following version of the principle of *Cut Elimination*, which we shall call *Sequential Cut*:

If $\psi_1, ..., \psi_n \Vdash_1 \phi$ and $\psi_1, ..., \psi_n, \phi, \theta_1, ..., \theta_k \Vdash_1 \chi$, then $\psi_1, ..., \psi_n, \theta_1, ..., \theta_k \Vdash_1 \chi$.

Given the principle of Idempotence, validity$_1$ is *Reflexive*.

$$\psi_1, ..., \psi_n, \phi \Vdash_1 \phi.$$

Sequential Monotony, Sequential Cut, and Reflexivity completely characterise the structural properties of the notion of validity$_1$ in update systems in which the principle of Idempotence holds. (See van Benthem[1991] for a way to prove this.)

1.3 Overview

In the next section a simple nonadditive update system is discussed. It models the dynamics of the epistemic possibility operator '*might*'. In addition some further terminology will be introduced. In particular, a distinction is made between additive *propositional updates* and non-classical *tests*.

In §3 a slightly more complex system is studied, covering the interplay between rules of the form '*Normally it is the case that...*' and the expectations they give rise to, which are expressed by sentences of the form '*Presumably it is the case that...*'. It will appear that rules are classical, just like ordinary descriptive sentences, although the kind of updates they give rise to are not propositional.

§ 4 is the heart of the paper. There the system developed in §3 is extended with restricted rules, i.e. sentences of the form '*If..., it is normally the case that...*'. I will show that the logical behaviour of these sentences can be explained by a simple coherence constraint which determines when a rule is acceptable, supplemented with an applicability criterion which explains why a rule is sometimes overruled by other rules.

Finally, in §5, we will see that the system developed in §4 is sufficiently rich to deal with most of the examples that are used as bench–mark problems in the literature.

Here are some examples to indicate the end result: Within the system developed in §4 and §5 the following argument form turns out to be valid$_1$:

premise 1: P's normally are R
premise 2: x is P

---

conclusion: Presumably, x is R

This argument remains valid$_1$ if one learns more about the object x, provided there is no evidence that the new information is relevant to the conclusion. So in the next case the inference still goes through.

premise 1: P's normally are R
premise 2: x is P
premise 3: x is Q

---

conclusion: Presumably, x is R

However, if on top of the premises 1, 2, and 3 the rule ‘Q's normally are not R’ is adopted, the argument is not valid$_1$ any more. If all one knows is

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

then it remains open whether one can presume that x is R. Clearly, the object x must be an exception to one of the rules, but there is no reason to expect it to be an exception to the one rule rather than to the other. Adding further default rules may make the balance tip. If, for instance, we add ‘Q's normally are P’ as a premise, we get the following valid$_1$ argument:

premise 1: Q's normally are P
premise 2: Q's normally are not R
premise 3: P's normally are R
premise 4: x is P
premise 5: x is Q
conclusion: Presumably, x is not R

In the presence of the principle ‘Q's normally are P’ the principle ‘Q's normally are not R’ takes precedence over the principle ‘P's normally are R’. (If a concrete example is wanted, read ‘x is P’ as ‘x is adult’, ‘x is Q’ as ‘x is a student’ and ‘x is R’ as ‘x is employed’).

None of the arguments above is valid$_2$ or valid$_3$. Both the definition of validity$_2$ and the definition of validity$_3$ contain a quantification over the set of states. Hence, in checking the validity$_2$ or validity$_3$ of an argument, one must reckon with the possibility that more is known than is stated in the premises. Conclusions drawn from default rules, however, are typically drawn ‘in the absence of any information to the contrary’; they may have to be withdrawn in the light of new information. Therefore, in evaluating a default argument it is important to know exactly which information is available. That is why I will concentrate on the notion of validity$_1$.

The dynamic set up and the notion of validity$_1$ that comes with it are the main features setting the theory developed in this paper apart from other default theories. Another difference between this theory and other theories is this: The fact that a conclusion has been drawn by default is made visible in the object language. It is not valid$_1$ to infer from ‘P's normally are R’ and ‘x is P’ that x *is* R; only that this is *presumably* so. Sentences starting with ‘presumably’ are non–persistent, so this qualification makes explicit the fact that the conclusion is defeasible. In other theories, a conclusion which is drawn by default inference is not marked; it is only at the meta–level that a defeasible conclusion gets a special status.

Finally, the research that led to this paper started off from the idea that questions of priority, which are likely to arise in the case of conflicting defaults should be decided at the level of semantics. Take the fact that the rule ‘Q's normally are not R’ can override the rule ‘P's normally are R’ in the presence of the rule ‘Q's normally are P’. (See the last example above). This is enforced by what these rules mean. It is not something to be stipulated over and above the semantics — as most theories would have it — but something to be explained by it.

2. A FIRST EXAMPLE: *MIGHT*

DEFINITION 2.1. Let **A** be a set consisting of finitely many *atomic sentences*. With **A** we associate two languages, $L_0^{\mathbf{A}}$ and $L_1^{\mathbf{A}}$. Both have **A** as their non-logical vocabulary. $L_0^{\mathbf{A}}$ has as its logical vocabulary one unary operator ¬, two binary operators ∧ and ∨, and two parentheses) and (. The sentences of $L_0^{\mathbf{A}}$ are just the ones one would expect for a language with such a vocabulary. $L_1^{\mathbf{A}}$ has in its logical vocabulary one additional unary operator *might.* A string ϕ of symbols is a sentence of $L_1^{\mathbf{A}}$ iff there is some sentence ψ of $L_0^{\mathbf{A}}$ such that either $\phi = \psi$ or $\phi = \mathit{might}\,\psi$.

Below, ‘p’, ‘q’, ‘r’, etc. are used as metavariables for atomic sentences. Different such metavariables refer to different atomic sentences. The symbols ‘ϕ’, ‘ψ’, and ‘χ’ are used as metavariables for arbitrary sentences.

The idea behind the analysis of ‘might’ is this: One has to agree to *might* ϕ if ϕ is consistent with ones knowledge — or rather with what one takes to be ones knowledge. Otherwise *might* ϕ is to be rejected.

In order to fix this idea into a mathematical model we need a way to represent an agent's knowledge. Below, a knowledge state [3] σ is given by a set of subsets of **A**. Intuitively, a subset *w* of **A** — or a possible world as we shall call it — will be an element of σ if, for all the agent in state σ knows, *w* might give a correct picture of the facts — given the agent's information, the possibility is not excluded that the atomic sentences in *w* are all true and the other false.

The powerset of **A** determines the space of *a priori* possibilities: if the agent happens to know nothing at all, any subset of **A** might picture reality correctly. As the agent's knowledge increases σ shrinks, until σ consists of a single subset of **A**. Then the agent's knowledge is complete. Thus, growth of knowledge is understood as a process of elimination.

DEFINITION 2.2. Let *W* be the powerset of the set **A** of atomic sentences.

(i) σ is an *information state* iff $\sigma \subseteq W$;

(ii) **0**, *the minimal state*, is the information state given by *W*;

1, *the absurd state*, is the information state given by the empty set;

(iii) For every two states σ and τ, $\sigma + \tau = \sigma \cap \tau$.

Note that $\sigma \leq \tau$ iff $\tau \subseteq \sigma$.

The notion of information state is language dependent: different sets of atomic sentences give rise to different sets of possible information states. The definition obscures this. It would be more accurate to speak of **A**-information states, and of the **A**-minimal state. I will occasionally use the latter terminology, in particular when we are ready to prove that in matters of logic it is not important to know exactly which language is at stake.

DEFINITION 2.3. Let **A** be given. For every sentence ϕ of $L_1^{\mathbf{A}}$ and state σ, $\sigma[\phi]$ is determined as follows:

$$\begin{array}{rl}
\text{atoms:} & \sigma[\mathrm{p}] = \sigma \cap \{w \in W \mid \mathrm{p} \in w\} \\
\neg: & \sigma[\neg\phi] = \sigma \sim \sigma[\phi] \\
\wedge: & \sigma[\phi \wedge \psi] = \sigma[\phi] \cap \sigma[\psi] \\
\vee: & \sigma[\phi \vee \psi] = \sigma[\phi] \cup \sigma[\psi] \\
\textit{might}: & \sigma[\textit{might}\,\phi] = \sigma \text{ if } \sigma[\phi] \neq \mathbf{1} \\
 & \sigma[\textit{might}\,\phi] = \mathbf{1} \text{ if } \sigma[\phi] = \mathbf{1}
\end{array}$$

The update clauses tell for each sentence ϕ and each state σ how σ changes when somebody in state σ accepts ϕ. If $\sigma[\phi] \neq \mathbf{1}$, ϕ is *acceptable in* σ. If $\sigma[\phi] = \mathbf{1}$, ϕ is *not acceptable in* σ and if $\sigma[\phi] = \sigma$, ϕ is *accepted* in σ. These notions are normative rather than descriptive: If $\sigma[\phi] = \mathbf{1}$, an agent in state σ *should* not accept ϕ. And if $\sigma[\phi] = \sigma$, an agent in state σ *has* to accept ϕ. An agent who refuses to do so is willingly or unwillingly breaking the conventions that govern the use of $\neg$, $\wedge$, $\vee$, *might*, etc.

It is also important to keep in mind that these notions have little or nothing to do with the notions of truth and falsity. It is very well possible that $\sigma[\mathrm{p}] = \mathbf{1}$, whereas in fact p is true or that $\sigma[\mathrm{p}] = \sigma$, whereas in fact p is false. Suppose that p is in fact true, and that $\sigma[\mathrm{p}] = \mathbf{1}$. Given the terminology introduced above, p is not acceptable for an agent in state σ. Does this mean that an agent in state σ must refuse to accept p, even when he or she is confronted with the facts? Of course not. The sentence p is not acceptable *in* state σ. So, the agent should *revise* σ in such a manner that p *becomes* acceptable. In definition 2.3 we are not dealing with revision: The update clauses do not tell for any sentence ϕ how a state σ in which ϕ is not acceptable must be revised so that ϕ can be accepted in the result. They stop at the point where it is clear that an inconsistency would arise if the information contained in ϕ would be incorporated in σ itself.

Note that for every sentence ϕ, $\mathbf{1}[\phi] = \mathbf{1}$. So, in the absurd state every sentence is accepted, but no sentence is acceptable. This explains how it can be that although we are not dealing with revision, the principle of Idempotence still goes through: Even if a sentence ϕ is not acceptable in σ — even if you *should* not accept ϕ — the result of updating σ with ϕ is an information state in which ϕ *is* accepted.

Although we are not dealing with belief revision, it may very well happen that a sentence is accepted at one stage, and rejected later. Revision is not the only possible source of non-persistence; testing is another. Here, sentences of the form *might* ϕ provide an example. As the definition says, all you can do when told that it might be the case that ϕ is to agree or to disagree. If ϕ is acceptable in your information state σ, you must accept *might* ϕ. And if ϕ is not acceptable in σ, neither is *might* ϕ. Clearly, then, sentences of the form *might* ϕ provide an invitation to perform a test on σ rather than to incorporate some new information in it. And the outcome of this test can be positive at first and negative later. In the minimal state you have to accept ‘It might be raining’, but as soon as you learn that it is not raining ‘It might be raining’ has to be rejected.

DEFINITION 2.4. A sequence of sentences $\psi_1, \ldots, \psi_n$ is *consistent* iff there is an information state σ such that $\sigma[\psi_1]\ldots[\psi_n] \neq \mathbf{1}$.

Again, since the set of information states varies with the non-logical vocabulary of the language in which $\psi_1, \ldots, \psi_n$ have been formulated, it would have been more accurate to speak of **A**-consistency. The next lemma and proposition show, however, that this prefix **A** can be omitted.

LEMMA 2.5. Let $\mathbf{A} \subseteq \mathbf{A}'$. With each **A**-state σ we associate an **A'**-state $\sigma^* = \{w \subseteq \mathbf{A}' \mid w \cap \mathbf{A} \in \sigma\}$. With each **A'**-state σ we associate an **A**-state $\sigma = \{w \subseteq \mathbf{A} \mid w = v \cap \mathbf{A}$ for some $v \in \sigma\}$.
Now, for every ϕ of L_1^A the following holds:
(i) if σ is an **A**-state, then $\sigma[\phi]^* = \sigma^*[\phi]$;
(ii) if σ, τ are **A**-states and $\sigma \neq \tau$, then $\sigma^* \neq \tau^*$;
(iii) if σ is an **A'**-state, then $\sigma[\phi]^\circ = \sigma^\circ[\phi]$;

(iv) if σ is an $\mathbf{A}'$-state, and $\phi\sigma[\phi] \neq \sigma$, then $\sigma^{\circ}[\phi] \neq \sigma^{\circ}$.

PROPOSITION 2.6. Let $p_1,\dots, p_k$ be the atomic sentences occurring in $\psi_1,\dots, \psi_n, \phi$. Suppose that $\{p_1,\dots, p_k\} \subseteq \mathbf{A}$ and $\{p_1,\dots, p_k\} \subseteq \mathbf{A}'$.
(i) The argument $\psi_1,\dots, \psi_n/\phi$ is $\mathbf{A}$-valid$_1$ iff it is $\mathbf{A}'$-valid$_1$;
(ii) $\psi_1;\dots;\psi_n$ is $\mathbf{A}$-consistent iff $\psi_1;\dots;\psi_n$ is $\mathbf{A}'$-consistent.

Suppose $p_1,\dots, p_k$ are the atoms in the argument $\psi_1,\dots, \psi_n/\phi$. Given proposition 2.6, we may rest assured that the answer to the question whether $\psi_1,\dots, \psi_n/\phi$ is valid is language independent, as it should be. Actually, in looking for the answer to this question we can always restrict ourselves to looking at the set of states generated by $\mathbf{A} = \{p_1,\dots, p_k\}$. Since there are only finitely many of these, the logic is decidable.

Henceforth I will omit the subscript ‘1’ in ‘validity$_1$’ and ‘$\Vdash_1$’. The next examples illustrate some of the points made in the preceding section.

EXAMPLES 2.7
(i) *might* $\neg$p, p is consistent;
p, *might* $\neg$p is not consistent.
(ii) Right-monotonicity fails: *might* $\neg$p $\Vdash$ *might* $\neg$p, but it is not the case that *might* $\neg$p, p $\Vdash$ *might* $\neg$p;
(iii) Left-monotonicity fails, too: $\Vdash$ *might* p, but it is not the case that $\neg$p $\Vdash$ *might* p.

A systematic study of the logical behaviour of *might* will have to be left to another occasion. What follows are some preliminary observations, which will play a role in the next sections.

LEMMA 2.8. Let σ and τ be information states and ϕ a sentence of $L_1^{\mathbf{A}}$.
(i) $\sigma \leq \sigma[\phi]$;
(ii) $\sigma[\phi][\phi] = \sigma[\phi]$;
(iii) if $\sigma \leq \tau$, then $\sigma[\phi] \leq \tau[\phi]$;
(iv) if ϕ a sentence of $L_0^{\mathbf{A}}$, the following holds:
if $\sigma \leq \tau$ and $\sigma \Vdash \phi$, then $\tau \Vdash \phi$.

The principles of Strengthening, Idempotence, Monotony and Persistence hold in $\langle L_0^{\mathbf{A}}, \Sigma, [\,]\rangle$. Hence, the system $\langle L_0^{\mathbf{A}}, \Sigma, [\,]\rangle$ is additive: we can associate with every sentence ϕ of $L_0^{\mathbf{A}}$ a static meaning, $\mathbf{0}[\phi]$. Updating any state σ with ϕ boils down to taking the intersection of σ and $\mathbf{0}[\phi]$. In the following, whenever we are dealing with a sentence ϕ of $L_0^{\mathbf{A}}$, I will refer to $\mathbf{0}[\phi]$ as *the proposition expressed by* ϕ, and write $\|\phi\|$ instead of $\mathbf{0}[\phi]$**.**

What would be the starting point in a static set up, can now be proved:

$$\|p\| = \{w\in W \mid p\in w\}$$
$$\|\neg\phi\| = W\sim \|\phi\|$$
$$\|\phi\wedge\psi\| = \|\phi\|\cap\|\psi\|$$
$$\|\phi\vee\psi\| = \|\phi\|\cup\|\psi\|$$

Given this, it will come as no surprise that for sentences of $L_0^{\mathbf{A}}$ we have that $\psi_1,..., \psi_n \Vdash \phi$ iff the argument $\psi_1,..., \psi_n / \phi$ is valid in classical logic

The system $\langle L_1^{\mathbf{A}}, \Sigma, [\ ]\rangle$ is not additive. Sentences of the form *might* ϕ are not persistent; they do not express a proposition; their informational content is not context independent. If you learn a sentence ϕ of $L_0^{\mathbf{A}}$, you learn that the real world is one of the worlds in which the proposition expressed by ϕ holds: the real world is a ϕ-world. But it would be nonsense to speak of the '*might* ϕ-worlds'. If ϕ might be true, this is not a property of the world but of your knowledge of the world.

3. RULES WITH EXCEPTIONS

In the previous section we studied a simple update process. The only information an agent could acquire was information about the actual facts. In this section we are interested in a slightly more complex process: Not only will the agents be able to learn which propositions *in fact* hold, but also which propositions *normally* hold. On top of that, they will be able to decide whether — in view of the information at hand — a given proposition *presumably* holds.

DEFINITION 3.1. Let $\mathbf{A}$ and $L_0^{\mathbf{A}}$ be as in § 2. The language $L_2^{\mathbf{A}}$ has $\mathbf{A}$ as its non-logical vocabulary, and in its logical vocabulary two additional unary operators: *normally*, and *presumably*. A string of symbols ϕ is a sentence of $L_2^{\mathbf{A}}$ iff there is a sentence ψ of $L_0^{\mathbf{A}}$ such that either $\phi = \psi$, or $\phi =$ *normally* ψ, or $\phi =$ *presumably* ψ.

Below, sentences of the form *normally* ϕ will be called *(default) rules*. To describe their impact on an agent's state of mind, we must give more structure to an information state than we did in the previous section. We want to capture two things: an agent's knowledge and an agent's expectations. And we want to do so in such a way that we can describe how an agent's expectations are adjusted as his or her knowledge increases. One way to do this is to think of a state σ as a pair $\langle \varepsilon, s\rangle$. Here s is a subset of the set of possible worlds, playing much the same role as it did in the previous section; it represents the agent's knowledge of the facts. The set ε represents the agent's knowledge of the rules.

DEFINITION 3.2. Let W be as before. Then ε is an (*expectation) pattern* on W iff ε is a reflexive and transitive relation on W.

The relation ε encodes the rules the agent is acquainted with. It does so in the following manner. Let P be the set of all propositions that a certain agent considers to be normally the case. Then $\langle w, v\rangle$ is an element of this agent's expectation pattern ε if every proposition in P that holds in v also holds in w. In other words, w conforms to all the rules in P that v conforms to, and perhaps to more.

Instead of '$\langle w, v\rangle \in \varepsilon$', we often write '$w \leq_{\varepsilon} v$'. If both $v \leq_{\varepsilon} w$ and $w \leq_{\varepsilon} v$, we write '$v \cong_{\varepsilon} w$'. Clearly, $\cong_{\varepsilon}$ is an equivalence relation. If $v \leq_{\varepsilon} w$ but not $w \leq_{\varepsilon} v$, we write '$v <_{\varepsilon} w$' and say that v is less exceptional than w.

DEFINITION 3.3. Let ε be a pattern on W;

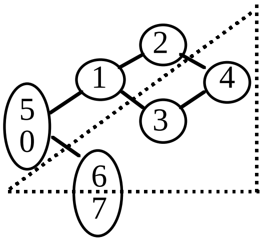
(i) w is a *normal world* in ε iff $w \in W$ and $w \leq_{\varepsilon} v$ for every $v \in W$;

(ii) $\mathbf{n}\varepsilon$ is the set of all normal worlds in ε;

(iii) ε is *coherent* iff $\mathbf{n}\varepsilon \neq \emptyset$.

Again, let P be the set of all propositions that a certain agent considers to be normally the case. Assume that $\emptyset \notin P$. (For a rule *normally* ϕ to be acceptable it is a necessary condition that the proposition expressed by ϕ holds at least in one world.) Given this, clause (iii) says that a pattern ε is coherent iff there is at least one possible world in which all the propositions in P hold. It seems reasonable to require that patterns be coherent in this sense. If it is not even conceivable that everything is normal, something is wrong. This does not mean, of course, that everything must in fact be normal, or that one must in all circumstances expect everything to be normal. It would not be very realistic to expect things to be more normal than the data leave room for.

Every now and then it is helpful to picture a state. The figure below pictures a state $\sigma = \langle \varepsilon, s \rangle$ pertaining to a language with three atoms.



If two worlds belong to the same $\cong_\varepsilon$-equivalence class, they are placed within the same circle or oval. So, the $\cong_\varepsilon$-equivalence classes are $\{w_1\}$, $\{w_2\}$, $\{w_3\}$, $\{w_4\}$, $\{w_0, w_5\}$, and $\{w_6, w_7\}$. If $w_i <_\varepsilon w_j$, the diagram contains a rightward path from the $\cong_\varepsilon$-equivalence class to which w_i belongs to the $\cong_\varepsilon$-equivalence class to which w_j belongs. We have for example that $w_0 <_\varepsilon w_3$, while it is neither the case that $w_2 \leq_\varepsilon w_3$, nor that $w_3 \leq_\varepsilon w_2$. The worlds constituting s are placed in an area with dashed borders; $s = \{w_3, w_4, w_6\}$. The normal worlds, w_5 and $w_{0,}$ do not belong to s. So, an agent who is in state σ knows that the actual world is not normal. Among the worlds that might be the actual world the worlds w_3 and w_6 take a special place: they are optimal in the sense of the next definition.

DEFINITION 3.4. 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) $\mathbf{m}_{\langle \varepsilon, s \rangle}$ is the set of all optimal worlds in $\langle \varepsilon, s \rangle$.

Default rules are of crucial importance when some decision must be made in circumstances where the facts of the matter are only partly known. In such a case one must reckon with several possibilities: for all an agent in state $\langle \varepsilon, s \rangle$ knows, each element of s might give a correct picture of the facts. Defaults serve to narrow down this range of possibilities: some elements of s are more normal than other. An agent in state $\langle \varepsilon, s \rangle$ will assume that the actual world conforms to as many standards of normality as possible; presumably, it is one of the *optimal* worlds. Worlds that are less than optimal become important when expectations have to be adjusted. As ones knowledge increases s shrinks, and the worlds that were optimal in s may disappear from s, and other worlds will become optimal.

DEFINITION 3.5. Let ε and ε' be patterns on W, and $e \subseteq W$.
(i) ε' is a *refinement* of ε iff $\varepsilon' \sqsubseteq \varepsilon$;

(ii) $\varepsilon\bullet e=\{\langle v, w\rangle \in \varepsilon \mid$ if $w\in e$, then $v\in e\}$; $\varepsilon\bullet e$ is the *refinement of ε with the proposition e.*

The refinement operation $\bullet$ is put to work when a new rule is learnt. Think of it as follows: Suppose $\langle v, w\rangle \in \varepsilon$. Then every rule which holds in w, also holds in v — at least in so far as the rules encoded in ε are concerned. Now a new rule comes in: *normally* ϕ. Two possibilities obtain:

(i) $\mathbf{n}\varepsilon\cap\|\phi\|\neq\varnothing$. There are normal worlds in which $\|\phi\|$ holds. Hence, the new rule is compatible with the rules encoded in ε; it is acceptable. If it is accepted, the new pattern will become $\varepsilon\bullet\|\phi\|$. That is, if $w\in\|\phi\|$ but $v\notin\|\phi\|$, the pair $\langle v, w\rangle$ has to be removed from ε. Given the new rule, it is no longer the case that v conforms to every rule that w conforms to.

(ii) $\mathbf{n}\varepsilon\cap\|\phi\|=\varnothing$. In this case the new rule is incompatible with the rules encoded in ε. Therefore it is not acceptable.

PROPOSITION 3.6.

(i) $(\varepsilon\bullet\varnothing)=\varepsilon$
$(\varepsilon\bullet W)=\varepsilon$

(ii) $(\varepsilon\bullet e)\bullet e=\varepsilon\bullet e$

(iii) If ε is a refinement of ε', and $'\varepsilon\bullet e=\varepsilon$,’ then $\varepsilon\bullet e=\varepsilon$

(iv) If ε is a refinement of ε', then $\bullet\varepsilon\, e$ is a refinement of $\varepsilon'\ \bullet e$.

Clauses (ii), (iii), and (iv) of this proposition are the basis for the proof that rules are idempotent, persistent and monotonous.

Let ε be a pattern. A proposition $e\subseteq W$ is said to be a *default in ε* iff $e\neq\varnothing$ and $(\varepsilon\bullet e)=\varepsilon$. The next proposition shows that this terminology fits in well with the explanation of the notion of a pattern given above.

PROPOSITION 3.7. Let ε be a pattern on W. Then for every $v, w\in W$, $w\leq_{\varepsilon}v$ iff $w\in e$ for every default e in ε such that $v\in e$.

I have not yet officially stated what an information state is.

DEFINITION 3.8. 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) $\varepsilon=\{\langle w, w\rangle\mid w\in W\}$ and $s=\varnothing$;

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

$\mathbf{1}$, *the absurd state*, is the state given by $\langle\{\langle w, w\rangle \mid w\in W\}, \varnothing\rangle$.

(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;

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

Note that $\langle \varepsilon, s\rangle \leq \langle \varepsilon', s'\rangle$ iff $s' \subseteq s$ and $\varepsilon' \subseteq \varepsilon$..

In the minimal state $\mathbf{0}$ no defaults are known: all worlds are equally normal.

There exist many pairs $\langle\varepsilon, s\rangle$, with the property that ε is incoherent, or $s = \varnothing$. Only one of these, the absurd state $\mathbf{1}$, has acquired official status as an information state — the idea being that the other incongruous states, being no less absurd, can be identified with $\mathbf{1}$.

DEFINITION 3.9. Let $\sigma = \langle \varepsilon, s\rangle$ be an information state. For every sentence ϕ of $L^{\mathbf{A}}_2$, $\sigma[\phi]$ is determined as follows:

- if ϕ is a sentence of $L^{\mathbf{A}}_0$, then
    - if $s \cap \| \phi \| = \varnothing$, $\sigma[\phi] = \mathbf{1}$;
    - otherwise, $\sigma[\phi] = \langle \varepsilon, s \cap \| \phi \|\rangle$.
- if ϕ = *normally* ψ, then
    - if $\mathbf{n}\varepsilon \cap \| \psi \| = \varnothing$, $\sigma[\phi] = \mathbf{1}$;
    - otherwise, $\sigma[\phi] = \langle \varepsilon \bullet \| \psi \|, s\rangle$.
- if ϕ = *presumably* ψ, then
    - if $\mathbf{m}_\sigma \cap \| \psi \| = \mathbf{m}_\sigma$, $\sigma[\phi] = \sigma$;
    - otherwise, $\sigma[\phi] = \mathbf{1}$.

The rule for *presumably* ϕ resembles the one for *might* ϕ in being an invitation to perform a test: If the proposition expressed by ϕ holds in all optimal worlds of σ, the sentence *presumably* ϕ must be accepted. Otherwise, *presumably* ϕ is not acceptable — not acceptable *in* σ, that is.

A sentence of the form *presumably* ϕ is not meant to convey new information. By asserting *presumably* ϕ, a speaker makes a kind of comment: 'Given the defaults and the facts that I am acquainted with it is to be expected that ϕ'. The addressee is supposed to determine whether on the basis of his or her own information ϕ is to be expected, too. If not so, a discussion will arise: 'Why do you think ϕ is to be expected?' the addressee will ask, and in the ensuing exchange of information both the speaker and the addressee may learn some new defaults or facts, so that in the end both will expect the same. (Admittedly, this is a somewhat idyllic picture).

EXAMPLES 3.10

(i) $\mathbf{0}\,[\mathit{normally}\,\mathrm{p}]\,[\neg\mathrm{p}] \neq \mathbf{1}$

$\mathbf{0}\,[\mathit{normally}\,\mathrm{p}]\,[\mathit{normally}\,\neg\mathrm{p}] = \mathbf{1}$

(ii) $\mathit{normally}\,\mathrm{p} \Vdash \mathit{presumably}\,\mathrm{p}$

$\mathit{normally}\,\mathrm{p}, \neg\mathrm{p} \nVdash \mathit{presumably}\,\mathrm{p}$

$\mathit{normally}\,\mathrm{p}, \neg\mathrm{p} \Vdash \mathit{normally}\,\mathrm{p}$

(iii) $\mathit{normally}\,\mathrm{p}, \mathrm{q} \Vdash \mathit{presumably}\,\mathrm{p}$

$\mathit{normally}\,\mathrm{p}, \mathrm{q}, \neg\mathrm{p} \nVdash \mathit{presumably}\,\mathrm{p}$

(iv) $\mathit{normally}\,\mathrm{p}, \mathit{normally}\,\mathrm{q} \Vdash \mathit{presumably}\,\mathrm{p}$

$\mathit{normally}\,\mathrm{p}, \mathit{normally}\,\mathrm{q}, \neg\mathrm{p} \nVdash \mathit{presumably}\,\mathrm{p}$

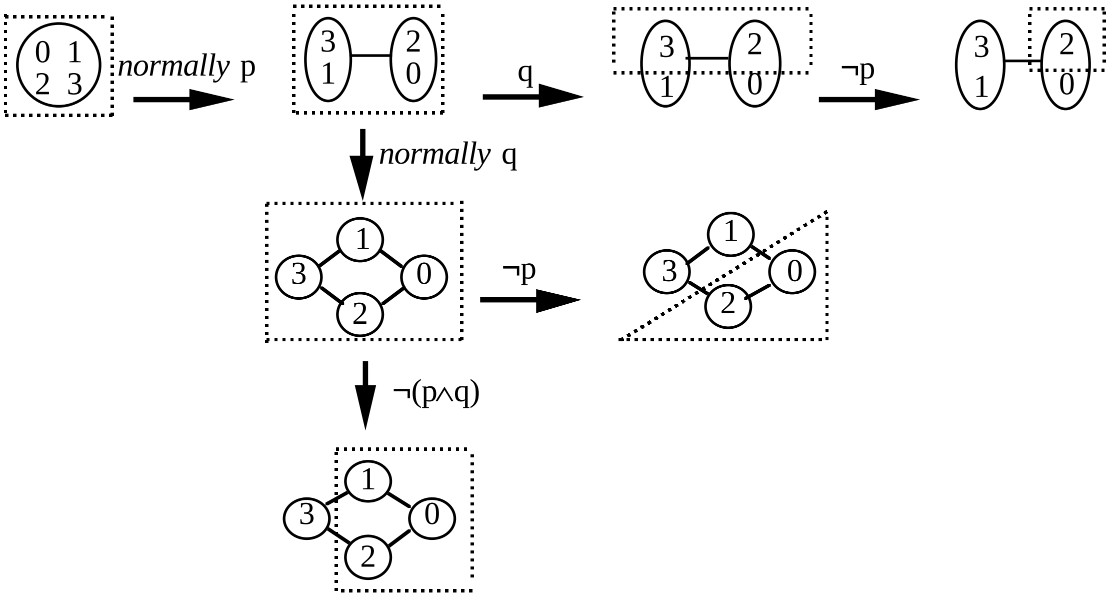
$\mathit{normally}\,\mathrm{p}, \mathit{normally}\,\mathrm{q}, \neg\mathrm{p} \Vdash \mathit{presumably}\,\mathrm{q}$

(v) $\mathit{normally}\,\mathrm{p}, \mathit{normally}\,\mathrm{q}, \neg(\mathrm{p} \wedge \mathrm{q}) \nVdash \mathit{presumably}\,\mathrm{p}$

$\mathit{normally}\,\mathrm{p}, \mathit{normally}\,\mathrm{q}, \neg(\mathrm{p} \wedge \mathrm{q}) \nVdash \mathit{presumably}\,\mathrm{q}$

The examples illustrate some important characteristics of the system. The first example under (i) shows that rules can have exceptions: An agent may first learn *normally* p — ‘normally it rains’ — and then discover that in fact it isn't raining. However, once an agent has accepted that it normally rains, the opposite rule ‘Normally it does not rain’ is unacceptable.

The states pertaining to the examples mentioned under (ii), (iii), (iv) and (v) are pictured below. $W = \{w_0, w_1, w_2, w_3\}$, where $w_0 = \varnothing$, $w_1 = \{\mathrm{p}\}$, $w_2 = \{\mathrm{q}\}$, and $w_3 = \{\mathrm{p,q}\}$. The first two examples mentioned under (ii) show that sentences of the form *presumably* ϕ are not persistent. If it is a rule that it normally rains, and if that is all you know, you may presume that it is raining now. But once you know that in fact it is not raining, it is silly to go on presuming that it is. Note that this does not mean you have to give up the rule in question. Today's weather may be exceptional, tomorrow's presumably will be normal again. [4] Even though the consequences one can draw from a rule are not persistent, the rule itself is[5].



The point of the examples in (iii) and (iv) is this: Having accepted a rule *normally* p you may expect p provided the other information you have is irrelevant to p — or at least not known to be relevant to p. So, if it is a rule that it normally rains, and all you know on top of that is that there is an easterly wind, you may presume that it is raining now. (In the next section we will see what happens when you learn that an easterly wind normally means that the weather is dry).

The examples in (iv) show that a sentence of the form *normally* ϕ says quite a bit more than just that ϕ holds in all normal worlds. It induces a general preference for worlds in which ϕ holds to worlds in which ϕ does not hold. Hence, if the real world has turned out to be exceptional in one respect, one can go on assuming it is normal in other respects.

As the examples in (v) illustrate, sometimes one gets in a predicament. If you prefer worlds in which p holds to worlds in which p doesn't hold, and worlds in which q holds to worlds in which q doesn't hold, then it is hard to choose if you cannot have both. Or to put it in terms of the next definition: the state $\mathbf{0}[\textit{normally}\,\mathrm{p}][\textit{normally}\,\mathrm{q}][\neg(\mathrm{p}\wedge\mathrm{q})]$ is ambiguous.

DEFINITION 3.11. Let $\langle\varepsilon, s\rangle$ be an information state.

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

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

I will not pursue a systematic study of *normally* and *presumably* here. However, the following seems to me essential.

LEMMA 3.12. Let ϕ be a sentence of $L_2^{\mathbf{A}}$ and let σ and τ be any states.

(i) $\sigma \leq \sigma[\phi]$;

(ii) $\sigma[\phi][\phi] = \sigma[\phi]$;

(iii) If $\phi \neq \mathit{presumably}\,\psi$ and $\sigma \leq \tau$, then $\sigma[\phi] \leq \tau[\phi]$;

(iv) If $\phi \neq \mathit{presumably}\,\psi$ and $\sigma \leq \tau$ and $\sigma \Vdash \phi$, then $\tau \Vdash \phi$.

We already saw that sentences of the form *presumably* ϕ are not persistent. That they are not monotonous either is due to the fact that the test for *presumably* ϕ may very well at first have a negative outcome, and a positive outcome later. Note, for example that $\mathbf{0} \leq \mathbf{0}$ [p], but it is not the case that $\mathbf{0}$ [*presumably* p] $\leq$ $\mathbf{0}$ [p][*presumably* p].

Note, however, that (iii) and (iv) of lemma 3.12 do hold for rules. We can assign to *normally* ϕ a static meaning, viz. $\mathbf{0}$ [*normally* ϕ], and think of the process of updating a state σ with *normally* ϕ as adding the information contained in $\mathbf{0}$ [*normally* ϕ] to σ. Not only purely descriptive sentences carry context independent information, but rules do so as well.

One way to gain some insight in the logical properties of the operator *normally* is to compare it with the alethic necessity operator. The next principles give a characterisation of the logical properties of the latter in a normal system of modal logic[6].

$$\mathit{necessarily}\,\phi \Vdash \phi$$
$$\mathit{necessarily}\,\phi,\ \mathit{necessarily}\,\psi \Vdash \mathit{necessarily}\,(\phi \wedge \psi)$$
$$\mathit{necessarily}\,\phi \Vdash \mathit{necessarily}\,(\phi \vee \psi)$$
$$\text{If } \Vdash \phi, \text{ then } \Vdash \mathit{necessarily}\,\phi$$

Only the second and the fourth of these principles remain valid — in our sense of the word — if we substitute *normally* for *necessarily*. We find:

$$\mathit{normally}\,\phi,\ \mathit{normally}\,\psi \Vdash \mathit{normally}\,(\phi \wedge \psi)$$
$$\text{If } \Vdash \phi, \text{ then } \Vdash \mathit{normally}\,\phi$$

We already know that the first principle does not hold for *normally*. What we have instead is the much weaker principle

$$\mathit{normally}\,\phi \Vdash \mathit{presumably}\,\phi\,.$$

The third principle fails, too. It is not generally so that

$$\textit{normally}\,\phi \Vdash \textit{normally}\,(\phi \vee \psi)$$

Perhaps the point is best brought out by an example. Compare:

— Normally it rains. It is not raining now. So, presumably it is snowing.
— Normally it rains or it snows. It is not raining now. So, presumably it is snowing.

Intuitively, the first line of thought is incorrect. Formally, it is invalid:

$$\textit{normally}\,p,\ \neg p \nVdash \textit{presumably}\ q$$

The second line of thought, however, seems correct. Formally we find:

$$\textit{normally}\,(p \vee q),\ \neg p \Vdash \textit{presumably}\ q$$

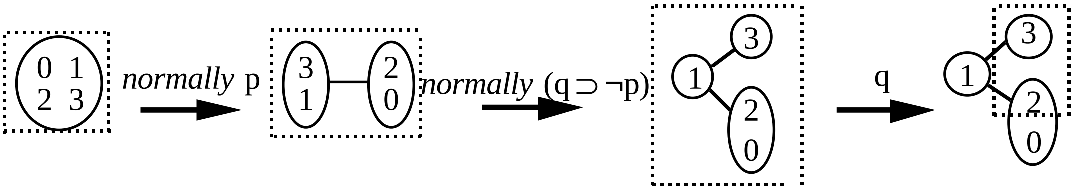
The example also shows why an agent might accept *normally* p, while refusing to accept *normally* $(p \vee q)$. The latter gives some indication as to what one can expect in case it is found that p happens to be false, the former does not. An agent may agree that p is normally the case but disagree that q rather than $\neg q$ is to be expected if p is false.[7]

4. RULES FOR EXCEPTIONS

The system devised above lacks expressive power. It works fine for general rules with accidental exceptions — ‘Normally it rains, but today it doesn't’ — but there is no room for non accidental exceptions: we cannot say when exceptional circumstances are to be expected and what one can expect when they obtain — ‘Normally it rains. But if there is an easterly wind, the weather is usually dry.’

Here is an example illustrating this. Suppose an agent in state **0** accepts the rule *normally* p — normally it rains. This induces an overall preference for worlds in which $\| p \|$ holds. Now, the agent wants to make an exception: If $\| q \|$ holds, $\| p \|$ normally does not hold — if there is an easterly wind, then normally it does not rain. The problem is that this exception cannot be made with the formula *normally*$(q \supset \neg p)$. The effect should be that in the domain of q–worlds the rule *normally* p is overridden, but things do not work out that way. The formula *normally*$(q \supset \neg p)$ induces another overall preference, this time for worlds in which the proposition

$\|q \supset \neg p\|$ holds. So, when it is learnt that in the actual world $\| q \|$ holds, an ambiguous situation arises: There are two optimal sets, one for the world that conforms to *normally* p, and the other for the world that conforms to *normally*$(q \supset \neg p)$. In the picture below $w_3 = \{p, q\}$, $w_2 = \{q\}$, $w_1 = \{p\}$ and $w_0 = \varnothing$.



One cannot equate ‘if q, then normally ¬p’ with *normally*$(q \supset \neg p)$. The binary operator ‘*if..., then normally* ... is not definable in terms of unary operator ‘*normally...*’.[8]

DEFINITION 4.1. Let **A** and $L_0^{\mathbf{A}}$ be as in § 2. The language $L_3^{\mathbf{A}}$ has **A** as its non-logical vocabulary, and in its logical vocabulary one additional binary operator $\rightsquigarrow$ and one additional unary operator *presumably*. A string of symbols ϕ is a sentence of $L_3^{\mathbf{A}}$ iff there are sentences ψ and χ of $L_0^{\mathbf{A}}$ such that $\phi = \psi$, or $\phi = \psi \rightsquigarrow \chi$, or $\phi =$ *presumably* ψ.

Read ‘$\phi \rightsquigarrow \psi$’ as ‘If ϕ, then normally ψ’. A sentence of the form ‘$\phi \rightsquigarrow \psi$’ is going to express that the proposition $\| \psi \|$ is a default in the domain of worlds given by ϕ. If this domain is a proper subset of the set of possible worlds, ‘$\phi \rightsquigarrow \psi$’ is called a *restricted* rule. General rules of the form *normally* ψ are reintroduced here as an abbreviation of $(\psi \vee \neg\psi) \rightsquigarrow \psi$.

DEFINITION 4.2.

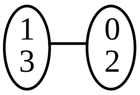
(i) Let W be as before. A *frame on W* is a function π assigning to every subset d of W a pattern πd on d.

(ii) Let π be a frame on W and $d, e \subseteq W$. The proposition e is a default in πd iff $d \cap e \neq \varnothing$ and $\pi d \bullet e = \pi d$.

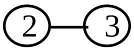
Whenever it is clear which frame is at stake we will say ‘e is a d-default’ rather than ‘e is a default in πd’.

For the example introduced above, the resulting frame π looks like this:

$\|$p$\|$ is a default in πW:



$\|\neg$p$\|$ is a default in $\pi\|$q$\|$



And for $d \neq W$ or $d \neq \|$q$\|$, $\pi d = d \times d$.

Given definition 4.2 every subset d of W can have its own pattern πd. So, now our agents can make as many exceptions as they wish. But of course, not anything goes. If they make too many exceptions, their expectation frames get incoherent.

4.1 Coherence

DEFINITION 4.3. Let π be a frame on W, and $d \subseteq W$.

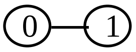
(i) w is a *normal* world in πd iff $w \in d$ and for every $d' \subseteq d$ such that $w \in d'$ it holds that $w \not\leq_{\pi d'} v$ for every $v \in d'$;

(ii) **n**πd is the set of all normal worlds in πd;

(iii) π is *coherent* iff for every non empty $d \subseteq W$, **n**$\pi d \neq \emptyset$.

Consider the frame depicted above. Given definition 4.3, **n**$\pi W = \{w_1\}$. So, despite the fact that w_3 conforms to the general rule *normally* p, w_3 does not count as a normal world in πW. Think of this as follows. By accepting $\|\neg$p$\|$ as a $\|$q$\|$-default, the agent has made an exception: the worlds in the domain $\|$q$\|$ are exempted from the general rule. So, to say that w_3 conforms to the general rule, as I did above, is misleading as it suggests that w_3 is subjected to this rule in the first place. But it is not. It is only subjected to the more specific rule q $\rightsquigarrow$ $\neg$p, to which it happens to be an exception The world w_3 is an exception to an exceptive clause, and we are not going to consider such an ‘exception to an exception’ as normal.

Here is a simple example of a frame that is not coherent. We are dealing with an agent who believes that it normally rains and who has made an exception for the case that there is an easterly wind: if there is an easterly wind, then normally it does not rain. On top of this the agent wants to make an exception for the case that there is no easterly wind: if there is no easterly wind, then normally it does not rain either. This is too much: the agent is making too many exceptions. Formally: the resulting frame π' is the same as the frame π depicted above except that now $\|\neg$p$\|$ is a rule in $\pi\|\neg$q$\|$. So $\pi'\{w_0, w_1\}$ looks like this:



But this means that $\mathbf{n}\pi W = \varnothing$. The frame π' is incoherent.

DEFINITION 4.4. 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 $\langle \iota, \varnothing \rangle$, where $\iota d = \{\langle w, w \rangle \mid w \in d\}$ for every $d \subseteq W$.

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

(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;
$\sigma + \sigma' = \mathbf{1}$, otherwise.

The differences between these definitions and the corresponding ones in the preceding section (see definition 3.8) are all due to the fact that we are not dealing with just one pattern, but with a frame of patterns.

Updating an information state with a new rule is a matter of refinement, just like before. If an agent in state $\sigma = \langle \pi, s \rangle$ decides to accept $\phi \rightsquigarrow \psi$, the pattern $\pi\|\phi\|$ will have to be refined with $\|\psi\|$. But of course, no agent should accept $\phi \rightsquigarrow \psi$ if the result of refining $\pi\|\phi\|$ with $\|\psi\|$ is incoherent.

DEFINITION 4.5.

(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 $d, e \subseteq W$. $\pi_{d \bullet e}$ is the refinement of π given by

(a) if $d' \neq d$, then $\pi_{d \bullet e} d' = \pi d'$;

(b) $\pi_{d \bullet e} d = \pi d \bullet e$.

The frame $\pi_{d \bullet e}$ is *the result of refining πd in π with e*.

DEFINITION 4.6. Let $\sigma = \langle \pi, s \rangle$ be an information state.

- $\sigma[\phi \rightsquigarrow \psi] = \mathbf{1}$ if $\|\phi\| \cap \|\psi\| = \varnothing$ or $\pi_{\|\phi\| \bullet \|\psi\|}$ is incoherent.
- Otherwise, $\sigma[\phi \rightsquigarrow \psi] = \langle \pi_{\|\phi\| \bullet \|\psi\|}, s \rangle$.

The case that $\|\phi\| \cap \|\psi\| = \varnothing$ is special: according to definition 4.2(ii), $\|\psi\|$ cannot be a default in $\pi\|\phi\|$ in this case. Still, according to proposition

3.6(i), $\pi\|\phi\|\bullet\|\psi\| = \pi\|\phi\|$. Hence, $\pi\|\phi\|\bullet\|\psi\|$ is coherent — a technical inconvenience.

PROPOSITION 4.7. Let π be coherent $d, e \subseteq W$. Suppose $d \cap e \neq \emptyset$.
$\pi_{d\bullet e}$ is coherent iff there is no $d' \supseteq d$ such that $\mathbf{n}\pi d' \subseteq d \sim e$.

Combining the definition and the proposition we get
Let $\sigma = \langle \pi, s \rangle$ be an information state. $\sigma[\phi \rightsquigarrow \psi]$ is determined as follows:

- If $\mathbf{n}\pi d \subseteq \|\phi\| \sim \|\psi\|$ for some $d \supseteq \|\phi\|$, then $\sigma[\phi \rightsquigarrow \psi] = \mathbf{1}$.
- Otherwise, $\sigma[\phi \rightsquigarrow \psi] = \langle \pi\|\phi\|\bullet\|\psi\|, s \rangle$.

EXAMPLES 4.8
(i) $\mathbf{0}$ [*normally* p] [q $\rightsquigarrow$ ¬p] $\neq \mathbf{1}$;
(ii) $\mathbf{0}$ [*normally* p] [q $\rightsquigarrow$ ¬p] [¬q $\rightsquigarrow$¬p] $= \mathbf{1}$;
(iii) $\mathbf{0}$ [*normally* p] [q $\rightsquigarrow$ ¬p] [*normally* q] $= \mathbf{1}$;
(iv) $\mathbf{0}$ [p $\rightsquigarrow$ q] [q $\rightsquigarrow$ p] [p$\rightsquigarrow$ r] [q $\rightsquigarrow$ ¬r] $= \mathbf{1}$.

(i) and (ii) were discussed above. (iii and (iv) are left as exercises.

4.2 Applicability

Let $\sigma = \langle \pi, s \rangle$ be a state. The frame π encodes the rules an agent in state σ is acquainted with and *s* his or hers knowledge of the facts. Now, what will an agent in state σ expect? In the previous section, where we were dealing with states consisting of just one pattern ε, this question was easy to answer: all we had to do was to sort out which of the worlds in *s* were optimal given the pattern ε. In this section things are more complicated. We are dealing with a number of patterns not all of which need have the same impact on *s*.

The crucial notion here is the notion of applicability: If you want to know what an agent in state $\langle \pi, s \rangle$ expects, you will have to sort out which of the rules encoded in π *apply* within *s*.

DEFINITION 4.9. 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 $w \in \mathbf{n}\pi d$ such that *w* complies with $\{e_1, \ldots, e_n\}$.

(Instead of saying 'the set $\{e_1,\dots,e_n\}$ applies within *s*', we often say '$e_1,\dots,e_n$ *jointly apply within s'*).

To see what is going on here, let us first look at the case that we are dealing with one default only. (In this case we say that the *d*–default *e,* rather than the singleton $\{e\}$ applies within *s*). The definition reduces to:

Let $\langle \pi, s\rangle$ be a coherent information state and *e* be a default in πd. The default *e* applies within *s* iff there is no $d' \supseteq s$ such that $\mathbf{n}\pi d' \subseteq d \sim e$.

An even more special case obtains if *s* is a subset of *d*. Then we say that the *d*- default *e* applies *to s* (rather than *within s*).

PROPOSITION 4.10. Let π be a coherent frame. Let *e* be a default in πd and suppose $s \subseteq d$. The default *e* applies to *s* iff there exists a coherent refinement π' of π such that for every domain d' with $s \subseteq d' \subseteq d$, *e* is a default in $\pi' d'$.

In other words, the *d*-default *e* applies to the subdomain *s* of *d* just in case *e* is an acceptable default in every domain between *s* and *d*. If there is some domain *d*' between *s* and *d* that cannot be coherently refined with *e*, then *e* does not apply to *s*.

EXAMPLES 4.11 For each of the following states $\sigma_i = \langle \pi_i, s_i\rangle$ we want to know which defaults apply to s_i.

(i) $\sigma_1 = \mathbf{0}$ [*normally* p] [q $\rightsquigarrow$ ¬p] [q];

(ii) $\sigma_2 = \mathbf{0}$ [*normally* p] [q $\rightsquigarrow$ ¬p] [q ∧ r];

(iii) $\sigma_3 = \mathbf{0}$ [*normally* p] [q $\rightsquigarrow$ ¬p] [(q ∧ r) $\rightsquigarrow$ p] [q ∧ r];

(iv) $\sigma_4 = \mathbf{0}$ [p $\rightsquigarrow$ r] [q $\rightsquigarrow$ (p ∧ ¬r)] [p ∧ q].

Here and in the following it may help if you read p as 'it rains', q as 'there is an easterly wind' and r as 'the temperature is below 15 °C'. Imagine that in each of these cases we are talking about a different country. All you know about the climate of this country is given by the rules mentioned. All you know about today's weather condition is given by the descriptive sentences mentioned. The question is: what else do you expect?

Example (i). We already know the frame π_1: ‖ p ‖ is a default in πW and ‖ ¬p ‖ is a default in π‖ q ‖. The agent's factual knowledge is given by $s_1 =$ ‖ q‖. Clearly, π‖ q ‖ cannot coherently be refined with ‖ p ‖. So, according to proposition 4.10, $\| p \|$ does not apply to s_1. It is overridden by the more specific $\| q \|$-default $\| \neg p \|$, which does apply to s_1.

Example (ii). For this example eight possibilities must be taken into account. Apart from that, the frame π_2 is much like π_1; its only interesting features are that $\| p \|$ is a default in *W*, and that $\| \neg p \|$ is a default in $\| q \|$. The agent's factual knowledge is given by $s_2 = \| q \wedge r \|$. When $\pi_2 \| q \|$ is refined with $\| p \|$, the result is incoherent. Since $s_2 \subseteq \| q \| \subseteq W$, it follows by proposition 4.10 that the *W*-default $\| p \|$ does not apply to s_2. The more specific $\| q \|$–default $\| \neg p \|$ does apply to s_2.

Example (iii). It is important to realise that we are working with a three place relation ‘the *d*-default *e* applies to *s*’. Often the first argument will be suppressed, but sometimes we cannot do so. This becomes evident when we compare the second example with the third. We saw above that in σ_2 the *W*-default $\| p \|$ does not apply to $\| q \wedge r \|$. There is nothing wrong, however, if an agent in addition to the rules *normally* p and $q \rightsquigarrow \neg p$ accepts the rule $(q \wedge r) \rightsquigarrow p$ — as an exceptive clause to an exceptive clause. But even after doing so, the *W*-default $\| p \|$ does not apply to $\| q \wedge r \|$. It is the more specific $\| q \wedge r \|$–default $\| p \|$ which does.

Examples (i)-(iii) show how the applicability criterion enforces that more specific rules take precedence over more general rules. However, as the next example shows, that is not the only thing enforced by it.

Example (iv). Neither of the rules $p \rightsquigarrow r$ and $q \rightsquigarrow (p \wedge \neg r)$ is more specific than the other. Yet, in the context given by $p \wedge q$ only the rule $q \rightsquigarrow (p \wedge \neg r)$ has to be taken into account, which is the main reason why an agent in state σ_4 is allowed to draw the following conclusion:

$p \rightsquigarrow r$	If it rains, normally the temperature is below 15°C.
$q \rightsquigarrow (p \wedge \neg r)$	If there is an easterly wind, then normally it rains, but the temperature is 15°C or higher.
$p \wedge q$	It is raining and there is an easterly wind
presumably $\neg r$	Presumably, the temperature 15°C or higher

The $\| p \|$-default $\| r \|$ does not apply to $s_4 = \| p \wedge q \|$, because $\| q \| \supseteq s_4$, while $\mathbf{n}\pi_4 \| q \| \subseteq \| p \| \sim \| r \|$. The $\| q \|$-default $\| p \wedge \neg r \|$ does apply to $\| p \wedge q \|$, because there is no $d \supseteq \| p \wedge q \|$ such that $\mathbf{n}\pi d' \nsubseteq q \| \sim \| p \wedge \neg r \|$.

Definition 4.9 pertains to sets of defaults rather than to single defaults. From the next example it will become clear why this is so.

EXAMPLES 4.11 (continued). For each of the states $\sigma_i = \langle \pi_i, s_i \rangle$ we want to know which defaults jointly apply within s_i.

(v) $\sigma_5 = \mathbf{0}\ [p \rightsquigarrow r]\ [q \rightsquigarrow \neg r]\ [p \wedge q]$;

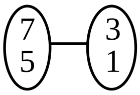
(vi) $\sigma_6 = \mathbf{0}\ [q \rightsquigarrow p]\ [p \rightsquigarrow r]\ [q]$;

Example (v). If it rains, the temperature is normally below 15°C. If there is an easterly wind the temperature is normally 15°C or higher. It's raining and there happens to be an easterly wind. What would the temperature be? The following analysis reveals why there is not much to be said here.

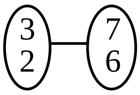
index	world
0	—
1	p
2	q
3	q, p
4	r
5	r, p
6	r, q
7	r, q, p

We are dealing with a set $W = \{w_0, ..., w_7\}$ of eight possible worlds described in the table on the left. The set $s_5 = \{w_3, w_7\}$.
π_5 is the following frame:
If $d \neq \{w_1, w_3, w_5, w_7\}$ and $d \neq \{w_2, w_3, w_6, w_7\}$, $\pi_5 d = d \times d$.

$\pi_5 \| p \|$ looks like this:



$\pi_5 \| q \|$ is this:



So, if $\{w_1, w_3, w_5, w_7\} \subseteq d$, $\mathbf{n}\pi d = d \sim \{w_1, w_3\}$; and if $\{w_2, w_3, w_6, w_7\} \subseteq d$, $\mathbf{n}\pi d = d \sim \{w_6, w_7\}$. Otherwise, $\mathbf{n}\pi d = d$.
The proposition $\| r \| = \{w_4, w_5, w_6, w_7\}$ is acceptable as a default in every domain between $s_5 = \{w_3, w_7\}$ and $\| p \| = \{w_1, w_3, w_5, w_7\}$. Hence, the $\| p \|$-default $\| r \|$ applies to s_5. Likewise we find that the $\| q \|$-default $\| \neg r \|$ applies to s_5. However, there is no coherent refinement π' of π_5 such that both $\| r \| = \{w_4, w_5, w_6, w_7\}$ and $\| \neg r \| = \{w_0, w_1, w_2, w_3\}$ are defaults in $\pi'\{\ w_3 w_7\}$. Which amounts to saying that the $\| p \|$-default $\| r \|$ and the $\| q \|$-default $\| \neg r \|$ do not *jointly* apply to s_5.

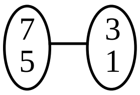
PROPOSITION 4.12. Let $\sigma = \langle \pi, s \rangle$ be a coherent information state and assume that $e_1, \ldots, e_n$ are defaults in $\pi d_1, \ldots, \pi d_n$ respectively. Suppose $s \subseteq d_i$ for every i $(1 \leq i \leq n)$. The defaults $e_1, \ldots, e_n$ jointly apply to s iff there exists a coherent refinement π' of π such that for every i it holds that e_i is a default in $\pi'd'$ for every domain d' such that $s \subseteq d' \subseteq d_i$.

The important thing to notice here is the order of the quantifiers: "there exists a coherent refinement such that for every i it holds that e_i is..." it says, rather than "for every i there exists a coherent refinement such that e_i is..." In the latter case each of the defaults $e_1,\ldots, e_n$ taken separately applies to s, but perhaps $e_1,\ldots, e_n$ do not jointly apply.

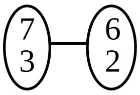
Let us now turn to a case in which not all rules the agent is acquainted with express defaults in a domain extending s.

Example (vi). We will find that $\mathrm{q} \rightsquigarrow \mathrm{p}$, $\mathrm{p} \rightsquigarrow \mathrm{r}$, $\mathrm{q} \Vdash$ *presumably* r.
The main reason why this is so is because in state $\mathbf{0}\,[\mathrm{q} \rightsquigarrow \mathrm{p}]\,[\mathrm{p} \rightsquigarrow \mathrm{r}]\,[\mathrm{q}]$ the $\|\mathrm{q}\|$-default $\|\mathrm{p}\|$ and the $\|\mathrm{p}\|$-default $\|\mathrm{r}\|$ jointly apply within $\|\mathrm{q}\|$.
Consider $W = \{w_0,..., w_7\}$ as described above under (v).
The set $s_6 = \{w_2, w_3, w_6, w_7\}$; π_6 is the following frame:
If $d \neq \{w_1, w_3, w_5, w_7\}$ and $d \neq \{w_2, w_3, w_6, w_7\}$, $\pi_6 d = d \times d$.

$\pi_6\|\mathrm{p}\|$ looks like this:



$\pi_6\|\mathrm{q}\|$ is this:



So, if $\{w_1, w_3, w_5, w_7\} \subseteq d$, $\mathbf{n}\pi d = d \sim \{w_1, w_3\}$, and if $\{w_2, w_3, w_6, w_7\} \subseteq d$, $\mathbf{n}\pi d = d \sim \{w_2, w_6\}$. Otherwise, $\mathbf{n}\pi d = d$.

The $\|\mathrm{q}\|$-default $\|\mathrm{p}\|$ and the $\|\mathrm{p}\|$-default $\|\mathrm{r}\|$ jointly apply within $\|\mathrm{q}\|$ if for every $d \supseteq \|\mathrm{q}\|$ there is some $w \in \mathbf{n}\pi d$ which complies with both. Since $w_7 \in \mathbf{n}\pi d$ for every $d \supseteq \|\mathrm{q}\|$, this is so. And since w_7 is the only world in $\|\mathrm{q}\|$ which complies with both these defaults, an agent in state σ_6 will expect the real world to be like w_7 rather than like w_2, w_3, or w_6. Which means that the agent will expect both p and r to be true.

By now the basic ideas behind definition 4.9 will be clear. First, *if* a set of defaults applies within a given context s, the effect will be that worlds not complying with these defaults do not count as normal s-worlds any more. Second, from the previous section we know that in a coherent frame the following holds: if a world is not normal in s, it is not normal in any domain extending s. So, *when* does a set of defaults apply within s? If for no domain d extending s, the set $\mathbf{n}\pi d$ of normal d-worlds consists entirely of worlds not complying with the defaults in question. Because otherwise, if the defaults did apply, the frame would get incoherent.

In the above I alluded several times to the next definition.

DEFINITION 4.13. 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 $e_1,\ldots, e_n$ 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,æ, e_n, 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 $\mathbf{m}_\sigma$.

(iii) $\sigma[$*presumably* $\psi]$ is determined as follows:

- If $\mathbf{m}_\sigma \cap \| \psi \| = \mathbf{m}_\sigma$, then $\sigma[$*presumably* $\psi] = \sigma$.
- Otherwise, $\sigma[$*presumably* $\psi] = \mathbf{1}$.

It is very well possible for there to be more than one maximal applicable set of defaults. If so, the state is called ambiguous.

PROPOSITION 4.14. Let $\sigma = \langle \pi, s \rangle$ be a coherent information state. Let each πd be given by $\pi d = (d \times d) \bullet (e_d)_1 \ldots \bullet (e_d)_m$.
Then w is optimal in σ iff $w \in s$ and w complies with a set of defaults D with the following properties:

(i) Each element of D is identical to some $(e_d)_i$;

(ii) D applies within s;

(iii) for every $(e_d)_i$ such that D‹$\{(e_d)_i\}$ applies within s, it holds that $(e_d)_i \in D$.

Suppose you have to sort out whether a certain argument of the form $\phi_1 \rightsquigarrow \psi_1,\ldots, \phi_n \rightsquigarrow \psi_n, \chi_1,\ldots, \chi_m$ / *presumably* θ is valid. What you have to do then is to determine the set of optimal worlds in the state $\sigma = \mathbf{0}\,[\phi_1 \rightsquigarrow \psi_1]\ldots[\phi_n \rightsquigarrow \psi_n]\,[\chi_1]\ldots[\chi_m]$. Definition 4.13 says that in order to do so you have to determine all maximal applicable sets of defaults in σ. Proposition 4.14 facilitates this work: you never have to take more defaults into account than the explicitly given defaults $\| \psi_1 \|, \ldots , \| \psi_n \|$ in their respective domains $\| \phi_1 \|, \ldots , \| \phi_n \|$. All you have to do is to determine the maximal subsets of $\{\| \psi_1 \|, \ldots , \| \psi_n \|\}$ applying within $\| \chi_1 \| \cap \ldots \cap \| \chi_m \|$. The set of optimal worlds is given by these.

Given proposition 4.14, it is easy to determine the set of optimal worlds in the states $\sigma_1, \ldots, \sigma_6$ figuring in example 4.11. Thus, we find:

(i) *normally* $p, q \leadsto \neg p, q \Vdash$ *presumably* $\neg p$.

(ii) *normally* $p, q \leadsto \neg p, q \wedge r \Vdash$ *presumably* $\neg p$

(iii) *normally* $p, q \leadsto \neg p, (q \wedge r) \leadsto p, q \wedge r \Vdash$ *presumably* p

(iv) $p \leadsto r,\ q \leadsto (p \wedge \neg r),\ p \wedge q \Vdash$ *presumably* $\neg r$

(v) $p \leadsto r,\ q \leadsto \neg r,\ p \wedge q \nVdash$ *presumably* r
$p \leadsto r,\ q \leadsto \neg r,\ p \wedge q \nVdash$ *presumably* $\neg r$

(vi) $q \leadsto p,\ p \leadsto r,\ q \Vdash$ *presumably* r.

5. COMPARISONS

So far, we have been thinking of the language $L_3^{\mathbf{A}}$ as a propositional language, but we can also give a predicate logical interpretation to it. Think of p, q, etc. as monadic predicates rather than atomic sentences. Each such predicate specifies a property and each well-formed expression of $L_0^{\mathbf{A}}$ specifies a Boolean combination of properties. Think of *W* as the set of possible *objects* rather than the set of possible worlds. A possible object $i \in W$ has the property expressed by the atom p if and only if $p \in i$. Note that different possible objects have different properties. Therefore it would be more precise to call the elements of *W* possible *types* of objects: in reality there can be more than one or no object fitting the description of a given possible object in *W*.

Like before, the set *s* in a state $\langle \pi, s \rangle$ represents the agent's knowledge, only now it is not the agent's knowledge about the real world, but about *some* real object. With a formula ϕ of $L_0^{\mathbf{A}}$ it is learnt that this object, which is not explicitly mentioned in ϕ, has the property expressed by ϕ.

A default in a pattern πd is a property now — a property that objects with the property *d* normally possess. Since ϕ-worlds (worlds in which the proposition expressed by ϕ holds) have become ϕ-objects (objects with the property expressed by ϕ), '$\phi \leadsto \psi$' can be read as 'ϕ-objects normally are ψ-objects' instead of 'ϕ-worlds normally are ψ-worlds'.

Let me repeat one of the things I said above: in reality there can be more than one or no object fitting the description of a given possible object. Expectation frames are conceptual frames. So, if the coherence condition requires that $\mathbf{n}\pi d \neq \emptyset$, this just means that it must be conceivable for an object in d to have all the properties that objects in d normally have. It does not mean that such an object must really exist. It may very well be that in reality no object fitting the description of any object in $\mathbf{n}\pi d$ can be found. It might be that each and every real bird lacks one or more of the properties that birds normally have, either by rule or by accident. It can be a fact that every bird is in some respect abnormal. But it cannot be a rule. If you want a system in which the sentence ‘Birds normally aren't normal’ is acceptable, you will have to look elsewhere.

Looking at the examples treated in the preceding section through predicate logical glasses, you will recognise some old acquaintances. Example 4.11(v), for instance, which is repeated below on the right hand side, can also serve as a formalisation of the well known Nixon Dilemma:

Quakers normally are pacifist	$p \leadsto r$
Republicans normally are not pacifist	$q \leadsto \neg r$
Nixon is both republican and Quaker	$p \wedge q$

As we saw, from these premises no conclusion, not even a tentative one, concerning Nixon's pacifism can be drawn.
Equally well known is the next example, which we did not discuss so far.

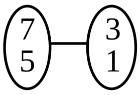
Adults normally are employed	$p \leadsto r$
Students normally are not employed	$q \leadsto \neg r$
Students normally are adults	$q \leadsto p$
John is a student	q
Presumably, John is adult and not employed	*presumably* $(p \wedge \neg r)$

This argument is valid. To see why, we have to determine the state

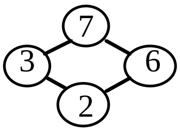
$$\mathbf{0}\ [p \leadsto r]\ [q \leadsto \neg r]\ [q \leadsto p]\ [q] = \sigma = \langle \pi, s \rangle.$$

Let W be defined as in example 4.11(v). Then $s = \{w_2, w_3, w_6, w_7\}$. For π we find: if $d \neq \{w_1, w_3, w_5, w_7\}$ and $\mathrm{d} \neq \{w_2, w_3, w_6, w_7\}$, $\pi d = d \times d$.

$\pi \| p \|$ can be depicted as:



And this is $\pi\|q\|$:



Since $\mathbf{n}\pi s = \{w_3\} \subseteq \|p\| \sim \|r\|$, the $\|p\|$-default $\|r\|$ does not apply within *s*. The other rules apply, which means that $\sigma \Vdash$ *presumably* $(p \wedge \neg r)$.

DEFINITION 5.1. Let $\sigma = \langle \pi, s \rangle$ be a state.

(i) The *factual information contained in* σ is given by

$\{\phi \mid \phi$ is a sentence of L_0^A such that $s \subseteq \|\phi\|\}$.

(ii) A set Δ of sentences of L_0^A is called an *extension* of the factual information contained in σ iff there exists some maximal applicable set *E* of defaults such that $\Delta = \{\phi \mid \{w \in s \mid w$ complies with $E\} \subseteq \|\phi\|\}$.

One way to compare the theory developed here with other theories, is to compare what they have to say about extensions. Note for example that we have:

$\phi_1 \rightsquigarrow \psi_1, \ldots, \phi_n \rightsquigarrow \psi_n, \chi_1, \ldots, \chi_m \Vdash$ *presumably* θ iff θ belongs to every extension of the factual information in $\mathbf{0}[\phi_1 \rightsquigarrow \psi_1] \ldots [\phi_n \rightsquigarrow \psi_n][\chi_1] \ldots [\chi_m]$.

In other words, the theory developed here belongs to the class of *sceptical* theories. It differs from other sceptical theories in that some sets of sentences that qualify as an extension in this sceptical theory will not do so in some of the other, or *vice versa*. Take the last example above. Within the framework of Reiter's default logic, this argument can be represented as the default theory $\langle D, W \rangle$, where

$D = \{(p: Mr / r), (q: M\neg r / \neg r), (q: Mp / p)\}$, and $W = \{q\}$

Given Reiter's definition of extension this default theory has two extensions: the deductive closure of $\{p, q, \neg r\}$, and the deductive closure of $\{p, q, r\}$. On our account, however, only the first of these counts.

The main shortcoming of Reiter's original theory is that it does not answer questions of priority. In many cases where conflicting rules are at stake, some take priority over other. In the above, I have tried to uncover the mechanisms behind this phenomenon. The resulting theory has much in common with the theories presented in Delgrande[1988] and Asher&Morreau[1990], which are built on the semantics of conditionals developed by David Lewis[1973].

In one respect the theory developed here is simpler than those of Delgrande and Asher&Morreau. In checking the validity of an argument, all three theories intend to look at the state of an agent who does not know more than what is given by the premises. Both Delgrande and Asher&Morreau try to give a direct definition of this state, whereas in the dynamic framework it is built up incrementally. In another respect, the theory developed here is more complex. Indeed, readers acquainted with the papers mentioned will have wondered why I did not choose selection functions[9] to represent an agent's knowledge of the rules. From a mathematical point of view, these are much simpler objects than expectation frames, and so far I have done nothing to show that it is really necessary to make things as complex as they are now.

There is a simpler version of the present theory in which selection functions are used as one of the components in an information state. In many cases this simpler version works just as well as the present one. Actually, so long as we restrict ourselves to cases in which for each domain at most one (non trivial) default has to be taken into account, both versions amount to the same thing. But as soon as we have more than one rule in the same domain differences obtain.

Students normally are adult
Students normally are not employed
John is a student
John is employed

Presumably, John is an adult

Students normally are adult
Students normally are not employed
Adults normally are employed
Adults normally know how to drive a car
Peter is a student

Presumably, Peter knows how to drive a car

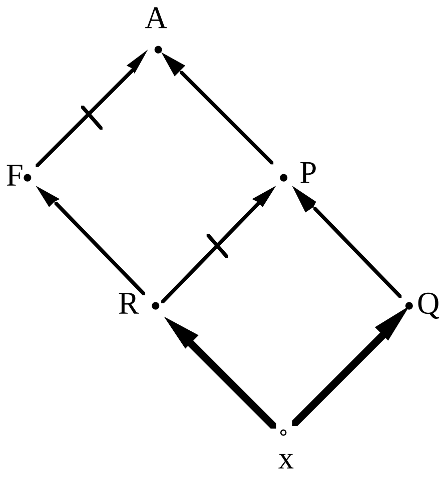
These are instances of a principle that is sometimes called the principle of Independence. If an object is exceptional in one respect, this does not necessarily mean it will be exceptional in other respects as well. Often you may rest assured that in other respects it will be normal. As the examples show, this holds not only if the object concerned happens to be an accidental exception to one of the rules you are acquainted with, but also if it is a non accidental exception. Given the premises of the left example, John happens to be an exception to the rule that students are not employed. So, John is not a normal student — not entirely normal at least. However, this is no reason to think that the rule that students normally are adults does not apply. You may still presume that John is an adult. You may do so given Reiter's theory, you may do so given the theory presented here, but you may not do so given Delgrande's or Asher&Morreau's theory. As for the example on the right, a formal analysis reveals that the optimal Peter — the Peter that conforms to as many applicable standards of normality as possible — is an adult who is a non accidental exception to the rule that adults are employed, but who knows how to drive a car anyway. The only other theories I know of that give the same outcome here are the theories of inheritance to which I shall turn below.

The principle of Independence comes out valid mainly because a pattern πd can be more than just a bipartition of d in normal and abnormal elements. I believe this principle embodies an essential feature of common sense reasoning. So, I cannot but conclude that selection functions are not the right kind of entities to model an agent's knowledge of the rules.

The expressive power of our formalism is limited. However, it is sufficiently rich to express everything expressible in a semantic network. The theory presented here supplies a semantics for multiple inheritance networks in which cyclic paths and complex predicates are allowed. It yields a decidable non-monotonic notion of logical consequence, viz. validity$_1$, which is comparable to the 'support'-relation in inheritance theory. It can be used as a basis for answering questions of soundness and completeness: Given an inference algorithm for a suitable[10] class of nets, is it the case that a net Γ belonging to this class supports a conclusion ϕ iff it is valid$_1$ to infer *presumably* ϕ from the rules and the facts that make up Γ?

For all algorithms I am acquainted with, the answer to this question is no. The algorithm for which the answer comes closest to yes is the one presented in Horty & Thomason & Touretzky [1987]. For the examples discussed so far this algorithm gives the same outcome as the theory presented here. Still, from our point of view, the algorithm is not sound. If it were, the next argument, would be valid in our sense of the word, but it is not

Quakers normally are pacifist
Republicans normally are not pacifist
Pacifists normally are anti-military
Republicans normally are football fans
Football fans normally are not anti military
John is both a Quaker and a republican
Presumably, John is not anti military



In our formalism this argument, which exemplifies the case of cascaded ambiguities, has the form:

$$q \rightsquigarrow p,\ p \rightsquigarrow a,\ r \rightsquigarrow \neg p,\ r \rightsquigarrow f,\ f \rightsquigarrow \neg a,\ q \wedge r \,/\, \textit{presumably}\ \neg a.$$

The state of somebody who has just learned all these premises is highly ambiguous. There turn out to be four optimal objects: $\{q, r, f\}$, $\{q, r, f, p\}$, $\{q, r, p, a\}$ and $\{q, r, f, p, a\}$. Therefore, it is neither valid to expect that x is anti-military, nor that x is not anti-military.

Here is an example showing that from our point of view, the algorithm of Horty *cum suis* is not complete either. A defeasible version of *Modus Tollens* is valid in our sense of the word, but the net representing the premises of the argument given below does not support its conclusion.

Adults normally have a driver's licence.	$p \rightsquigarrow q$
John does not have a driver's license.	$\neg q$
Presumably, John is not adult .	*presumably* $\neg p$

To see why this argument is valid, set $W = \{w_0, w_1, w_2, w_3\}$, where $w_0 = \varnothing$ $w_1 = \{p\}$, $w_2 = \{q\}$, and $w_3 = \{p, q\}$. Consider $\mathbf{0}\,[p \rightsquigarrow q]\,[\neg q]$. Of the worlds in $\|\neg q\|$, the world w_0 is complies with the $\|p\|$-default $\|q\|$. The world w_1, however, does not. And since for no domain d extending $\|\neg q\|$ it holds that $\mathbf{n}\pi d \subseteq \{w_1\}$, the $\|p\|$-default $\|q\|$ applies within $\|\neg q\|$. Hence, someone in state $\mathbf{0}\,[p \rightsquigarrow q]\,[\neg q]$ will expect the real world to be like w_0 rather than like w_1. And in w_0 the proposition $\|p\|$ does not hold.

It is instructive to compare the above argument with the following:

Students normally are adults	$p \leadsto q$
Adults normally are not students	$q \leadsto \neg p$
John is a student	p
Presumably, John is an adult	*presumably* q

Note that the inheritance net corresponding to this argument is cyclic. At first sight, the premises of the argument may seem ambiguous: by *Modus Ponens* one can infer *presumably* q, and by *Modus Tollens* one can infer *presumably* ¬q. However, a closer inspection of the state $\mathbf{0}\,[p \leadsto q]\,[q \leadsto \neg p]\,[p]$ reveals that *Modus Ponens* takes precedence over *Modus Tollens*: Let *W* be like above. The crucial point is that the $\|q\|$-default $\|\neg p\|$ does not apply within $\|p\|$ because $\mathbf{n}\pi\|p\| = \{w_3\}$ and $\{w_3\} \subseteq \|q\| \sim \|\neg p\|$. The $\|p\|$-default $\|q\|$, on the other hand, does apply within $\|p\|$. So, the real world will be like w_3 rather than like w_1, which means that $\mathbf{0}\,[p \leadsto q]\,[q \leadsto \neg p]\,[p] \Vdash$ *presumably* q.

By now it will be clear that the theory of defaults developed in this paper differs from all other theories not only in its explanations but also in its predictions. I will leave it to the readers who have missed their favourite examples to check these for themselves, and conclude this section by pointing out some more general peculiarities.

First a reminder: the logic generated by the validity notion that we have been using is not closed under substitution. For example, we saw in the previous section that the following argument is valid

$$q \leadsto p,\ p \leadsto r,\ q \Vdash \textit{presumably}\ r \qquad (*)$$

However, (*) is only valid for predicates that are independent — or at least not known to be dependent. If we substitute '¬q' for 'r', we find

$$q \leadsto p,\ p \leadsto \neg q,\ q \nVdash \textit{presumably}\ \neg q$$

As an introduction to the second point, consider the rules 'Students are normally adults' and 'Adults are normally employed'. Suppose these are the only rules you are acquainted with. Given (*), it is correct to infer for any student x (of whom you don't know more than this) that x is presumably employed. This does not mean, however, that it is correct to conclude that students normally are employed. That is:

$$q \rightsquigarrow p,\ p \rightsquigarrow r \nVdash q \rightsquigarrow r \qquad (**)$$

The *Hypothetical Syllogism* is not valid; we only have a defeasible version of it, exemplified by (*). The conclusion of (*) is defeasible. It will be defeated when for example you learn that students normally are not employed. The conclusion of (**), on the other hand, is not defeasible. Rules are persistent.

There are more examples of this kind. Ever so often we find that

$$\phi_1 \rightsquigarrow \psi_1,\ldots,\phi_n \rightsquigarrow \psi_n,\ \chi \Vdash \textit{presumably}\ \theta,$$

whereas

$$\phi_1 \rightsquigarrow \psi_1,\ldots,\phi_n \rightsquigarrow \psi_n \nVdash \chi \rightsquigarrow \theta.$$

For instance, as we saw, a defeasible form of *Modus Tollens* is valid:

$$p \rightsquigarrow q,\ \neg q \Vdash \textit{presumably}\ \neg p,$$

but *Contraposition* fails:

$$p \rightsquigarrow q \nVdash \neg q \rightsquigarrow \neg p.$$

We also have

$$p \rightsquigarrow q,\ p \wedge r \Vdash \textit{presumably}\ q,$$

but *Strengthening the Antecedent* is not allowed:

$$p \rightsquigarrow q \nVdash (p \wedge r) \rightsquigarrow q.$$

Well known principles of implication like the *Hypothetical Syllogism*, *Contraposition* and *Strengthening the Antecedent* fail for the default arrow $\rightsquigarrow$. So, naturally the question if there any left which do hold. If the arrow $\rightsquigarrow$ is not a strict implication, as the failure of these principles shows, is it then perhaps a variable *strict implication*? If it were, the next principles, which give a complete characterisation of the interplay of any variable strict implication with the classical connectives, would hold:

Conditional Identity (CI)[11]	:	$\phi \rightsquigarrow \psi \Vdash \phi \rightsquigarrow \phi$
Conjunction of Consequents (CC)	:	$\phi \rightsquigarrow \psi,\ \phi \rightsquigarrow \chi \Vdash \phi \rightsquigarrow (\psi \wedge \chi)$
Weakening the Consequent (CW)	:	$\phi \rightsquigarrow \psi \Vdash \phi \rightsquigarrow (\psi \vee \chi)$
Strengthening with a Consequent (ASC)	:	$\phi \rightsquigarrow \psi,\ \phi \rightsquigarrow \chi \Vdash (\phi \wedge \psi) \rightsquigarrow \chi$
Disjunction of Antecedents (AD)	:	$\phi \rightsquigarrow \chi,\ \psi > \chi \Vdash (\phi \vee \psi) \rightsquigarrow \chi$

It turns out that only the first two of these principles are valid. The remaining three are *almost* valid. For example, for any state σ the following holds:

$$\sigma[\phi \rightsquigarrow \psi][\phi \rightsquigarrow \neg(\psi \vee \chi)] = \mathbf{0};$$
$$\sigma[\phi \rightsquigarrow \psi][\phi \rightsquigarrow \chi][(\phi \wedge \psi) \rightsquigarrow \neg\chi] = \mathbf{0};$$
$$\sigma[\phi \rightsquigarrow \chi][\psi \rightsquigarrow \chi][(\phi \vee \psi) \rightsquigarrow \neg\chi] = \mathbf{0}.$$

For a principle like the *Hypothetical Syllogism*, something analogous does not hold. It is very well possible that

$$\sigma[\phi \rightsquigarrow \psi][\psi \rightsquigarrow \chi][\phi \rightsquigarrow \neg\chi] \neq \mathbf{0}.$$

Here is another specification of '*almost* valid': Let Δ be any sequence of rules. Then we have the following:

$$\Delta,\ \phi \rightsquigarrow \psi,\ \phi \ \Vdash \textit{presumably}(\psi \vee \chi);$$
$$\Delta,\ \phi \rightsquigarrow \psi,\ \phi \rightsquigarrow \chi,\ \phi \wedge \psi \ \Vdash \textit{presumably}\,\chi;$$
$$\Delta,\ \phi \rightsquigarrow \chi,\ \psi \rightsquigarrow \chi,\ \phi \vee \psi \ \Vdash \textit{presumably}\,\chi.$$

These are defeasible versions of CW, ASC and AD, but they have a special property: their conclusions can only be defeated by *factual* information. So, here, too, there is a big difference with a principle like the *Hypothetical Syllogism*, since

$$\phi \rightsquigarrow \neg\chi,\ \phi \rightsquigarrow \psi,\ \psi \rightsquigarrow \chi,\ \phi \ \not\Vdash \textit{presumably}\,\chi.$$

I have not been able to find a good intuitive explanation why the principles ASC and AD should not hold. Only for the case of CW have I an argument showing that something would be wrong if this principle were valid. Indeed, it is perfectly alright that

$$\phi \rightsquigarrow \psi \ \not\Vdash \phi \rightsquigarrow (\psi \vee \chi)$$

Here, I can repeat what I wrote near the end of the previous section. As the next examples show, a sentence of the form $\phi \rightsquigarrow (\psi \vee \chi)$ is in certain respects stronger than $\phi \rightsquigarrow \psi$.

— Tigers normally have four legs. Shere Khan is a tiger. Shere Khan does not have four legs. So, presumably Shere Khan has five legs.

— Tigers normally have four or five legs. Shere Khan is a tiger. Shere Khan does not have four legs. So, presumably Shere Khan has five legs.

The second argument is valid, the first is not. The rule ‘Tigers normally have four or five legs’ indicates what one can expect in case one encounters a tiger that does not have four legs; the rule ‘Tigers normally have four legs’ does not. No wonder an agent might be willing to accept the latter without wanting to accept the former.

6 CONCLUDING REMARKS

The aim of this paper has been twofold: (i) to introduce the framework of update semantics, and to explain what kind of semantic phenomena may be successfully analysed within it; and (ii) to give an analysis of one such phenomenon: default reasoning.

Within the framework of update semantics default reasoning is not considered a special kind of reasoning with ordinary sentences, but an ordinary kind of reasoning with a special kind of sentences. It is just as valid to conclude ‘Presumably x is B’ from ’x is A’ and ‘A's are normally B’ as it is to conclude ‘x is B’ from ‘x is A’ and the ‘All A's are B’. One does not have to set ones mind to a different mode of reasoning to get the former. In both cases the same validity notion is at stake, which for ordinary descriptive sentences yields the same monotonic logic as the classical notion. However, as soon as the language is enriched with sentences that express default rules and operators like ‘presumably’ the logic gets non-monotonic, because sentences starting with ‘presumably’ are special — they are non-persistent.

The specific theory of defaults developed in the preceding sections is not the only possible one within the framework of update semantics. Indeed, one would hope that somebody will come up with a more elegant formalisation of the same intuitive ideas. Still, I think that these intuitive ideas, culminating in the coherence criterion and the applicability criterion, are sound, and I take the fact that the theory behaves a lot better than other theories in predicting the capricious logical behaviour of defaults to be evidence in favour of this position.

I hope that the ideas set out in this paper will be helpful not only to logicians interested in defaults, but also to linguists interested in the semantics of generic sentences. I realise, however, that what I offer here is at best one missing piece in a giant puzzle — nobody knows how many pieces are still missing, let alone how they fit together. I have given a logical analysis of one particular kind of generic sentence, viz. sentences of the form ‘P's normally are Q’. And whatever merits this analysis may have, it does not say anything about the relation between this particular kind of generic sentence and other kinds. It does not explain why a sentence of the form

(i) P's normally are Q

so often conveys the same information as (ii)-(iv):

(ii) the P is Q
(iii) P's are Q
(iv) a P is Q

It does not even explain why such sentences are often equivalent to:

(v) Normally P's are Q

In the AI-literature, these sentence forms are often used interchangeably. And, indeed, there are many instances where all of them seem to have the same impact. Compare for example:

(i)' Tigers normally have four legs
(ii)' The tiger has four legs
(iii)' Tigers have four legs
(iv)' A tiger has four legs
(v)' Normally tigers have four legs

But linguists, much more so than logicians, have always been aware of the differences between these sentence forms. If sentences of the form (i)-(v) really were always equivalent, we could say:

(i)" Tigers normally are extinct

and mean the same as we would mean with (ii)" or (iii)"

(ii)" The tiger is extinct

Likewise, if (i) and (ii) really were equivalent, the sentence

(iii)"' Tigers eat people

would imply

(i)"' Tigers normally eat people

And what to think of the next sentence?

(iv)"" A tiger is available

Whatever this means, it is not equivalent to

(i)"" Tigers normally are available

which in its turn differs widely in meaning from

(v)"" Normally tigers are available

This is just a sample from the long list of problems surrounding generic sentences[12]. Since Carlsson[1977] it is clear that part of the solution lies in a proper subcategorization of predicates, some being exclusively predicable of kinds, other primarily of individuals, and still other primarily of temporal stages of individuals. But so far there is no theory explaining when a generic sentence can get a default reading, and how such a reading comes about. This paper does not offer such a theory either — at best it explains what a default reading amounts to.[13]

REFERENCES

Asher, N. and M. Morreau: 1990, ‘Commonsense entailment: A modal theory of non-monotonic reasoning’, in J. van Eijck (ed.), *Logics in AI*, Proceedings of JELIA '90, Springer Lecture Notes in Computer Science **478**, 1-30.

Beaver, D.: 1991, ‘The kinematics of presupposition’, *Proceedings of the 8th A' dam Colloquium.*

Carlson, G.: 1977, *Reference to Kinds in English*, Ph.D dissertation, University of Massachusetts, Amherst.

Delgrande, J.: 1988, ‘An approach to default reasoning based on a first-order conditional logic: Revised report’, *Artificial Intelligence* **36**, 63-90.

Gärdenfors, P.: 1984, ‘The dynamics of belief as a basis for logic’, *British Journal for the Philosophy of Science* **35**, 1-10.

Groenendijk, J. and M. Stokhof: 1991, ‘Dynamic predicate logic’, *Linguistics and Philosophy* **14**, 39-101.

Groeneveld, W.: 1989, ‘Dynamic Semantics and Circular Propositions’, *Journal of Philosophical Logic* **23**, 267-306.

Heim, I. R.: 1982, *The Semantics of Definite and Indefinite Noun Phrases*, Ph. D. Dissertation, University of Massachusetts, Amherst.

Horty, J., R. Thomason, and D. Touretzky: 1987, *A Skeptical Theory of Inheritance in Non- Monotonic Nets*, report CMU-CS-87-175, Carnegie Mellon University, iii+52 pp.

Kamp, J.A.*W*.: 1981, ‘A theory of truth and semantic representation’, in J. Groenendijk, T.M.V. Janssen, and M. Stokhof (eds.), *Formal Methods in the Study of Language*, Mathematical Centre Tracts 135, Amsterdam, 277-322.

Krifka, M.: 1987, ‘*An Outline of Genericity*’, SNS-Bericht 87-25, Seminar für Natürlich-Sprachliche Systeme, Universität Tübingen.

Lewis, D.: 1973, *Counterfactuals*, Basil Blackwell, Oxford.

Reiter, R.: 1980, ‘A logic for default reasoning’, in *Artificial Intelligence* **13**, 81-132.

Stalnaker, R.: 1974, ‘Pragmatic presuppositions’, in Munitz, M. and P. Unger (eds.), *Semantics and philosophy*, University Press, New York, 197-213.

van Benthem, J.F.A.K.: 1991, *Language in Action*, Elsevier Science Publishers (North-Holland), Amsterdam.

Zeevat, H: 1992, ‘Presupposition and accomodation in update semantics’, *Journal of Semantics* **9**, pp 379-412.

[1] This notion of meaning underlies much recent work in formal semantics. Its origin can be traced back to Robert Stalnaker's work on presupposition and assertion. (See for instance Stalnaker[1974]). It took further shape in the work of Hans Kamp and Irene Heim on anaphora, and in Peter Gärdenfors's work on the dynamics of belief (See for example Kamp[1981], Heim[1982], and Gärdenfors[1984]). The most direct inspiration for the present paper came from the work of Jeroen Groenendijk and Martin Stokhof on Dynamic Predicate Logic. (See Groenendijk, J. and M. Stokhof[1991]).

[2] See Beaver[1991] and Zeevat[1992] for more elaborated views.

[3] I use the phrases ‘knowledge’ and ‘knowledge state’ where the reader might prefer ‘beliefs’ and ‘belief state’. Actually, I want the information states σ to represent something in between: if σ is the state of a given agent, it should stand for what the agent regards as his or her knowledge. Things the agent would qualify as mere beliefs do not count. But it might very well be that something the agent takes as known, is in fact false.

[4] It is not possible to formalise this example within in the present system, because the set *s* in an information state $<\varepsilon, s>$models the agent‘s knowledge of the ‘actual’ situation. It would be more general to work with states $<\varepsilon, f>$, where ε is a pattern on *W* (just like above) and *f* is a function which assigns to every point of time *t* a subset $f(t)$ of *W* , representing the agent’s knowledge of the situation at time *t*. In so doing, we could also formally deal with an agent’s expectations about tomorrow‘s weather.

[5] In trying to get to grips with the definition of an information state, the reader may have wondered why the pattern ε in a state $<\varepsilon, s>$ is taken to be a pattern on *W* rather than *s*. Could not $<\varepsilon \cap (s x s)>$ do the job that is now done by $<\varepsilon, s>$? The answer to this question is no: under the alternative definition rules would no longer be persistent.

[6] We restrict our attention here to a language in which the necessity operator only occurs as the outermost operator of a sentence. It is not difficult to extend the theory in such a manner that not only default rules but also *strict* rules can be understood by our agents. Here is the basic idea: A state σ is a pair $<\varepsilon,\ s>$ just like before, only now ε is a pattern on a subset *V* of *W* of rather than on *W* itself. As the next update clause shows, *V* is determined by the strict rules the agent is acquainted with: a world *w* is an element of *V* just in case every proposition that the agent considers necessary holds in *w*.

- if $\mathbf{n}\varepsilon \cap \| \phi \| = \emptyset$ or $s \cap \| \phi \| = \emptyset$, $\sigma[\mathit{necessarily}\, \phi] = \mathbf{1}$;
- otherwise, $\sigma[\mathit{necessarily}\, \phi] = < \varepsilon \mid (V \cap \| \psi \|)\ s \cap \| \psi \|>$.

[7] I cannot prove that it is impossible for there to be an update system for which the following would hold:

(i) *normally* p $\Vdash$ *normally* (p $\vee$ q);

(ii) *normally* p, $\neg$p $\nVdash$*presumably* q;

(iii) *normally* (p $\vee$ q), $\neg$p $\Vdash$ *presumably*q.

However, if you want such a system you will have to give up *Sequential Cut*. (To see why, suppose (i) holds. Given Sequential Cut and (ii), it follows that *normally* p, *normally* (p $\vee$ q), $\neg$p $\nVdash$*presumably* q. But this is almost as bad as not having (iii).

[8] Things go also wrong if one equates ‘if q, then normally ¬p’ with ‘q ⊃ *normally* ¬p’. This sentence would at best bring the agents in a state in which they believe that either q happens to be false in the actual world, or p is normally not the case. (NB: Officially q ⊃*normally* ¬p is not a sentence of L_2^A).

[9] A selection function is a function f that assigns to each subset *d* of *W*, a subset $f(d)$ of *d*. Intuitively, $f(d)$ contains the normal elements of *d*.

[10] Here ‘suitable’ means ‘everything that can be said in the net language, can be said in the language L_3^A ’. I am going to be rather sloppy in distinguishing between the two.

[11] In most conditional logics CI holds unrestictedly: $\Vdash \phi \rightsquigarrow \phi$. I had to make one proviso: $\phi \rightsquigarrow \phi$ is only valid for non-contradictory ϕ. (In absurd circumstances nothing is normal)

[12] See Krifka[1987] for a mind boggling overview.

[13] This paper had many drafts. The first was published in Report 2.5.A of the ESPRIT Basic Research Action 3175, DYANA, Centre for Cognitive Science, Edinburgh, 1990.