Institute for Language, Logic and Information

DEFAULTS IN UPDATE SEMANTICS

Frank Veltman

ITLI Prepublication Series
for Logic, Semantics and Philosophy of Language LP-91-02

University of Amsterdam

The ITLI Prepublication Series

1986

86-01 The Institute of Language, Logic and Information

86-02 Peter van Emde Boas A Semantical Model for Integration and Modularization of Rules

86-03 Johan van Benthem Categorial Grammar and Lambda Calculus

86-04 Reinhard Muskens A Relational Formulation of the Theory of Types

86-05 Kenneth A. Bowen, Dick de Jongh Some Complete Logics for Branched Time, Part I Well-founded Time,

86-06 Johan van Benthem Logical Syntax Forward looking Operators

1987 87-01 Jeroen Groenendijk, Martin Stokhof Type shifting Rules and the Semantics of Interrogatives

87-02 Renate Bartsch Frame Representations and Discourse Representations

87-03 Jan Willem Klop, Roel de Vrijer Unique Normal Forms for Lambda Calculus with Surjective Pairing

87-04 Johan van Benthem Polyadic quantifiers

87-05 Víctor Sánchez Valencia Traditional Logicians and de Morgan's Example

87-06 Eleonore Oversteegen Temporal Adverbials in the Two Track Theory of Time

87-07 Johan van Benthem Categorial Grammar and Type Theory

87-08 Renate Bartsch The Construction of Properties under Perspectives

87-09 Herman Hendriks Type Change in Semantics: The Scope of Quantification and Coordination

1988 LP-88-01 Michiel van Lambalgen *Logic, Semantics and Philosophy of Language:* Algorithmic Information Theory

LP-88-02 Yde Venema Expressiveness and Completeness of an Interval Tense Logic

LP-88-03 Year Report 1987

LP-88-04 Reinhard Muskens Going partial in Montague Grammar

LP-88-05 Johan van Benthem Logical Constants across Varying Types

LP-88-06 Johan van Benthem Semantic Parallels in Natural Language and Computation

LP-88-07 Renate Bartsch Tenses, Aspects, and their Scopes in Discourse

LP-88-08 Jeroen Groenendijk, Martin Stokhof Context and Information in Dynamic Semantics

LP-88-09 Theo M.V. Janssen A mathematical model for the CAT framework of Eurotra

LP-88-10 Anneke Kleppe A Blissymbolics Translation Program

ML-88-01 Jaap van Oosten *Mathematical Logic and Foundations:* Lifschitz' Realizabiility

ML-88-02 M.D.G. Swaen The Arithmetical Fragment of Martin Löf's Type Theories with weak Σ-elimination

ML-88-03 Dick de Jongh, Frank Veltman Provability Logics for Relative Interpretability

ML-88-04 A.S. Troelstra On the Early History of Intuitionistic Logic

ML-88-05 A.S. Troelstra Remarks on Intuitionism and the Philosophy of Mathematics

CT-88-01 Ming Li, Paul M.B.Vitanyi *Computation and Complexity Theory:* Two Decades of Applied Kolmogorov Complexity

CT-88-02 Michiel H.M. Smid General Lower Bounds for the Partitioning of Range Trees

CT-88-03 Michiel H.M. Smid, Mark H. Overmars Maintaining Multiple Representations of
Leen Torenvliet, Peter van Emde Boas Dynamic Data Structures

CT-88-04 Dick de Jongh, Lex Hendriks Computations in Fragments of Intuitionistic Propositional Logic
Gerard R. Renardel de Lavalette

CT-88-05 Peter van Emde Boas Machine Models and Simulations (revised version)

CT-88-06 Michiel H.M. Smid A Data Structure for the Union-find Problem having good Single-Operation Complexity

CT-88-07 Johan van Benthem Time, Logic and Computation

CT-88-08 Michiel H.M. Smid, Mark H. Overmars Multiple Representations of Dynamic Data Structures
Leen Torenvliet, Peter van Emde Boas

CT-88-09 Theo M.V. Janssen Towards a Universal Parsing Algorithm for Functional Grammar

CT-88-10 Edith Spaan, Leen Torenvliet, Peter van Emde Boas Nondeterminism, Fairness and a Fundamental Analogy

CT-88-11 Sieger van Denneheuvel, Peter van Emde Boas Towards implementing RL

X-88-01 Marc Jumelet *Other prepublications:* On Solovay's Completeness Theorem

1989 LP-89-01 Johan van Benthem *Logic, Semantics and Philosophy of Language:* The Fine-Structure of Categorial Semantics

LP-89-02 Jeroen Groenendijk, Martin Stokhof Dynamic Predicate Logic, towards a compositional,
non-representational semantics of discourse

LP-89-03 Yde Venema Two-dimensional Modal Logics for Relation Algebras and Temporal Logic of Intervals

LP-89-04 Johan van Benthem Language in Action

LP-89-05 Johan van Benthem Modal Logic as a Theory of Information

LP-89-06 Andreja Prijatelj Intensional Lambek Calculi: Theory and Application

LP-89-07 Heinrich Wansing The Adequacy Problem for Sequential Propositional Logic

LP-89-08 Víctor Sánchez Valencia Peirce's Propositional Logic: From Algebra to Graphs

LP-89-09 Zhisheng Huang Dependency of Belief in Distributed Systems

ML-89-01 Dick de Jongh, Albert Visser *Mathematical Logic and Foundations:* Explicit Fixed Points for Interpretability Logic

ML-89-02 Roel de Vrijer Extending the Lambda Calculus with Surjective Pairing is conservative

ML-89-03 Dick de Jongh, Franco Montagna Rosser Orderings and Free Variables

ML-89-04 Dick de Jongh, Marc Jumelet, Franco Montagna On the Proof of Solovay's Theorem

ML-89-05 Rineke Verbrugge Σ-completeness and Bounded Arithmetic

ML-89-06 Michiel van Lambalgen The Axiomatization of Randomness

ML-89-07 Dirk Roorda Elementary Inductive Definitions in HA: from Strictly Positive towards Monotone

ML-89-08 Dirk Roorda Investigations into Classical Linear Logic

ML-89-09 Alessandra Carbone Provable Fixed points in $I\Delta_0+\Omega_1$

CT-89-01 Michiel H.M. Smid *Computation and Complexity Theory:* Dynamic Deferred Data Structures

CT-89-02 Peter van Emde Boas Machine Models and Simulations

CT-89-03 Ming Li, Herman Neuféglise, Leen Torenvliet, Peter van Emde Boas On Space Efficient Simulations

CT-89-04 Harry Buhrman, Leen Torenvliet A Comparison of Reductions on Nondeterministic Space

CT-89-05 Pieter H. Hartel, Michiel H.M. Smid A Parallel Functional Implementation of Range Queries
Leen Torenvliet, Willem G. Vree

CT-89-06 H.W. Lenstra, Jr. Finding Isomorphisms between Finite Fields

CT-89-07 Ming Li, Paul M.B. Vitanyi A Theory of Learning Simple Concepts under Simple Distributions and
Average Case Complexity for the Universal Distribution (Prel. Version)

CT-89-08 Harry Buhrman, Steven Homer Honest Reductions, Completeness and
Leen Torenvliet Nondeterminstic Complexity Classes

CT-89-09 Harry Buhrman, Edith Spaan, Leen Torenvliet On Adaptive Resource Bounded Computations

CT-89-10 Sieger van Denneheuvel The Rule Language RL/1

CT-89-11 Zhisheng Huang, Sieger van Denneheuvel Towards Functional Classification of Recursive Query Processing
Peter van Emde Boas

X-89-01 Marianne Kalsbeek *Other Prepublications:* An Orey Sentence for Predicative Arithmetic

X-89-02 G. Wagemakers New Foundations: a Survey of Quine's Set Theory

X-89-03 A.S. Troelstra Index of the Heyting Nachlass

X-89-04 Jeroen Groenendijk, Martin Stokhof Dynamic Montague Grammar, a first sketch

X-89-05 Maarten de Rijke The Modal Theory of Inequality

X-89-06 Peter van Emde Boas Een Relationele Semantiek voor Conceptueel Modelleren: Het RL-project

1990 SEE INSIDE BACK COVER

Instituut voor Taal, Logica en Informatie
Institute for Language, Logic and Information

Faculteit der Wiskunde en Informatica
(Department of Mathematics and Computer Science)
Plantage Muidergracht 24
1018TV Amsterdam

Faculteit der Wijsbegeerte
(Department of Philosophy)
Nieuwe Doelenstraat 15
1012CP Amsterdam

DEFAULTS IN UPDATE SEMANTICS

Frank Veltman
Department of Philosophy
University of Amsterdam

ITLI Prepublications
for Logic, Semantics and Philosophy of Language
ISSN 0924-2082

Received June 1991

DEFAULTS IN UPDATE SEMANTICS

Frank Veltman
Department of Philosophy
University of Amsterdam

June 1991

§ 1 THE FRAMEWORK

The standard definition of logical validity runs as follows: An argument is valid if its premises çannot all be true without its conclusion being true as well. By far the most logical theories developed so far have taken this definition of validity as their starting point. Consequently, the heart of such theories consists in a specification of truth conditions.

The heart of the theory presented below does not consist in a specification of truth conditions but of update conditions. According to this theory the slogan 'You know the meaning of a sentence if you know the conditions under which it is true' should be 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] In this way, meaning becomes a dynamic notion: the meaning $[\phi]$ of a sentence ϕ is an operation on information states.

Let σ be an information state and ϕ a sentence with meaning $[\phi]$. We write $\sigma[\phi]$ for the information state that results when σ is updated with ϕ.[2] In most cases $\sigma[\phi]$ will be different from σ, but every now and then it may happen that $\sigma[\phi] = \sigma$. Then the information conveyed by ϕ is already subsumed by σ: if σ is updated with ϕ, the resulting information state turns out to be σ again. In such a case, i.e. when $\sigma[\phi] = \sigma$, we write $\sigma \Vdash \phi$ and say that ϕ *is accepted in* σ.

Two explications of logical validity suggest themselves:

- An argument is valid$_1$ if one cannot accept all its premises without having to accept the conclusion as well. More formally:
  $\psi_1, ..., \psi_n \vDash_1 \phi$ iff $\sigma \Vdash \phi$ for every σ such that $\sigma \Vdash \psi_i$ for every $1 \leq i \leq n$.

[1] This conception 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[1979]). 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]).

[2] Since σ is the argument and $[\phi]$ the function, it would have been more in line with common practice to write '$[\phi](\sigma)$' instead of '$\sigma[\phi]$'. The present notation is more convenient for dealing with texts. Now we can write '$\sigma[\psi_1]...[\psi_n]$' —instead of '$[\psi_n](...([\psi_1](\sigma))...)$'— for the result of updating σ with the sequence of sentences $\psi_1,..., \psi_n$.

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

$$\psi_1,\ldots,\psi_n \models_2 \phi \text{ iff for every } \sigma,\ \sigma[\psi_1]\ldots[\psi_n] \Vdash \phi.$$

It is easy to see that arguments that are valid$_2$ are also valid$_1$. The converse does not hold.

1.1 PROPOSITION

Consider the following principle:

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

Given this principle the following are equivalent:

(i) *Stability:* If $\sigma \Vdash \phi$, then $\sigma[\psi] \Vdash \phi$;

(ii) If $\psi_1,\ldots,\psi_n \models_1 \phi$, then $\psi_1,\ldots,\psi_n \models_2 \phi$;

(iii) *Right-monotonicity:* If $\psi_1,\ldots,\psi_n \models_2 \phi$, then $\psi_1,\ldots,\psi_n, \chi \models_2 \phi$.

In the following we will take the principle of *Idempotence* for granted[1] — what would 'updating your information with ϕ' mean if not at least 'changing your information in such a manner that you come to accept ϕ'? But we will not assume the principle of *Stability*. There are sentences that are not stable. They will often at first be accepted on the basis of limited information, only to be rejected as more information comes at hand.

The clearest examples of unstable sentences are to be found among sentences in which modal qualifications like 'presumably', 'probably', 'must', 'may' or 'might' occur. That such sentences are not stable is shown by the next two sequences of sentences. Processing the first sequence does not cause any problems, but processing the second sequence does.

— Somebody is knocking at the door... Presumably, it's John... It's Mary.

— Somebody is knocking at the door... It's Mary ... Presumably, it's John.

These two sequences consist of the same sentences. Only the order differs. Nevertheless the first sequence makes sense, whereas the second does not.

[1] Actually, there are many sentences of natural language for which the Principle of Idempotence is questionable. Sentences with anaphoric pronouns and paradoxical sentences like the Liar can serve as examples. The formal languages discussed in this paper do not contain such sentences. See Groenendijk&Stokhof[1991] for more information on anaphora, and Groeneveld[1989] for a dynamic analysis of the Liar Paradox.

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 if you already 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 = 'Presumably, it's John', and ϕ_3 = 'It's Mary'. If we want, we can compare $\sigma\,[\phi_1]\,[\phi_2]\,[\phi_3]$ with $\sigma\,[\phi_1]\,[\phi_3]\,[\phi_2]$ for any information state σ, and see if there are any differences.

Proposition 1.1 says that in the absence of *Stability* the two explications of validity will give rise to different logics. The first notion of validity is monotonic. If an argument with premises $\psi_1,...,\psi_n$ and conclusion ϕ is valid$_1$, then it remains valid$_1$ if you add more premises to $\psi_1,...,\psi_n$. Given proposition 1.1, the second notion will be non-monotonic. Note however that the second notion is at least left-monotonic:

$$\text{If } \psi_1, ..., \psi_n \vDash_2 \phi, \text{ then } \chi, \psi_1, ..., \psi_n \vDash_2 \phi.$$

What fails is right-monotonicity:

$$\text{If } \psi_1, ..., \psi_n \vDash_2 \phi, \text{ then } \psi_1, ..., \psi_n, \chi \vDash_2 \phi.$$

There is a third explication of validity, one with which we will be much concerned when we come to discuss defaults:

- An argument is valid$_3$ iff updating the minimal information state '**1**' with the premises $\psi_1,....,.\psi_n$ in that order, yields an information state σ that affirms the conclusion ϕ. Formally:

$$\psi_1, ..., \psi_n \vDash_3 \phi \text{ iff } \mathbf{1}\,[\psi_1]...[\psi_n] \Vdash \phi.$$

Of course this definition presupposes that there is such a thing as the minimal information state.

Like validity$_1$, validity$_3$ gives rise to a stronger logic than validity$_2$. If $\psi_1, ... ,\psi_n \vDash_2 \phi$, then $\psi_1, ... ,\psi_n \vDash_3 \phi$, but the converse does not hold. Like validity$_2$, validity$_3$ is not right-monotonic. But validity$_3$ will turn out to be not left-monotonic either. In § 4 and §5 of this paper I will present a logic for default principles according to which the following argument form is valid$_3$:

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

This argument remains valid$_3$ if one learns more about the object x, so long as 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 and x is Q
conclusion:	Presumably, x is R

However, if on top of the premises 1 and 2, the rule 'Q's normally are not R' is adopted, the resulting argument is not valid$_3$ any more. That is, if all we know is

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

then it remains open whether we can presume that x is a 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 principles may make the balance tip. If, for instance, we add 'Q's normally are P' as a premise, we get the following valid$_3$ 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 and 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').

A first comment: One respect in which the default theory presented below differs from current theories is that the distinction between defeasible and indefeasible conclusions is made manifest at the level of the object language: It is not valid$_3$ to conclude from ‘P's normally are R’ and ‘x is P’ that x *is* R; only that this is *presumably* so. This qualification makes explicit the fact that a defeasible conclusion has been drawn. In other theories, one may well infer ‘x is R’ from the premises, but it is considered a special kind of inference. Where these other theories consider default reasoning as a special kind of reasoning with ordinary sentences, I prefer to think of it as an ordinary kind of reasoning with a special kind of sentences.

A second comment: The research that lead to this paper started off as an attempt to give a dynamic twist to the theory developed in Delgrande[1988], who in turn took Lewis[1973] as his starting point. One thing the present theory has in common with Delgrande’s is that questions of priority, which are likely to arise in the case of conflicting defaults (cf. the last two examples above), are decided at the level of semantics. That ‘Q's normally are not R’, overrides ‘P's normally are R’ in the presence of ‘Q's normally are P’ is enforced by what these principles *mean.* It is not something that has to be stipulated over and above the semantics — as most theories would have it — but something explained by it.

Thirdly: None of the arguments discussed above will turn out be valid$_1$ or valid$_2$. Both the definition of validity$_1$ and the definition of validity$_2$ contain a quantification over all possible information states. Therefore, we must always reckon with the possibility that it is *known* that the object under consideration is abnormal in the relevant respects.

Final remark: It is easy to verify that both validity$_3$ and validity$_2$ conform to the principle of *Cautious Monotonicity*:

If $\psi_1,\ldots,\psi_n \models_i \phi$ and $\psi_1,\ldots,\psi_n \models_i \chi$, then $\psi_1,\ldots,\psi_n,\chi \models_i \phi$ (i=2,3)

Moreover, both comply with the principle of *Lemma Generation*:

If $\psi_1,\ldots,\psi_n \models_i \chi$ and $\psi_1,\ldots,\psi_n,\chi \models_i \phi$, then $\psi_1,\ldots,\psi_n \models_i \phi$ (i=2,3)

Nowadays Restricted Monotonicity, Lemma Generation and *Reflexivity*:

$$\phi \models_i \phi \qquad (i=2,3)$$

are often considered as the minimal conditions that any reasonable consequence relation must satisfy. (See for example Makinson[1989]).

§ 2 A FIRST EXAMPLE: *MIGHT*

2.1 DEFINITION

Let $\mathcal{A}$ be a set consisting of finitely many *atomic sentences*. With $\mathcal{A}$ we associate two languages, $L_0^{\mathcal{A}}$ and $L_1^{\mathcal{A}}$. Both have $\mathcal{A}$ as their nonlogical vocabulary.
$L_0^{\mathcal{A}}$ has as its logical vocabulary one unary operator $\neg$, two binary operators $\wedge$ and $\vee$, and two parentheses) and (. The sentences of $L_0^{\mathcal{A}}$ are just the ones one would expect for a language with such a vocabulary.
$L_1^{\mathcal{A}}$ has in its logical vocabulary one additional unary operator *might*, and an interpunction sign denoted by ';'. A string ϕ of symbols is a sentence of $L_1^{\mathcal{A}}$ iff there is some sentence ψ of $L_0^{\mathcal{A}}$ such that either $\phi = \psi$ or $\phi = \mathit{might}\,\psi$.
The set of texts of $L_1^{\mathcal{A}}$ is the smallest set satisfying the following conditions:
(i) if ϕ is a sentence of $L_1^{\mathcal{A}}$, then ϕ is a text of $L_1^{\mathcal{A}}$;
(ii) if ν is a text of $L_1^{\mathcal{A}}$ and ϕ a sentence of $L_1^{\mathcal{A}}$, then $\nu\,;\phi$ is a text of $L_1^{\mathcal{A}}$.

Below, 'p', 'q', 'r', etc. are used as metavariables for atomic sentences. Different such metavariables refer to different atomic sentences. The greek letters 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[1] σ is given by a set of subsets of $\mathcal{A}$. Intuitively, a subset i of $\mathcal{A}$ will be an element of σ if — for all the agent in state σ knows — this subset *might* give a correct picture of reality. More precisely, for all the agent in state σ knows, the possibility is not excluded that the atomic sentences in i are all true and the other false. The powerset of $\mathcal{A}$ determines the space of *a priori* possibilities: if the agent happens to know nothing at all, then any subset of $\mathcal{A}$ might picture reality correctly. As the agent's knowledge increases σ shrinks, until σ consists of a single subset of $\mathcal{A}$.

[1] I use the phrases 'knowledge' and 'knowledge state' where some readers might prefer 'beliefs' and 'belief state'. In fact, 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 knowledge. Things of which the agent would say that he merely believes them do not count. But it could very well be that something the agent takes as known, is in fact false.

Then the agent's knowledge is complete. Thus, the growth of knowledge is understood as a process of elimination.

For an agent who knows nothing at all, any subset of $\mathcal{A}$ might picture reality correctly — any such set represents a 'way the world might be'. For lack of a better term, I will henceforth refer to the subsets of $\mathcal{A}$ as possible worlds[1].

2.2 DEFINITION

Let $\mathcal{A}$ be the set of atomic sentences and W the powerset of $\mathcal{A}$. Then

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

(ii) Let σ and τ be information states; σ *is at least as strong as* τ iff $\sigma \subseteq \tau$;

(iii) **1**, *the minimal state*, is the information state given by W;
0, *the absurd state*, is the information state given by the empty set.

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 $\mathcal{A}$-information states, and of the $\mathcal{A}$-minimal state. Below, 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.

2.3 DEFINITION

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

$$
\begin{array}{ll}
\text{atoms:} & \sigma[p] = \sigma \cap \{w \in W \mid p \in w\} \\
\neg: & \sigma[\neg\phi] = \sigma \sim \sigma[\phi] \\
\wedge: & \sigma[\phi \wedge \psi] = \sigma[\phi] \cap \sigma[\psi] \\
\vee: & \sigma[\phi \vee \psi] = \sigma[\phi] \cup \sigma[\psi] \\
\textit{might}: & \sigma[\textit{might}\,\phi] = \sigma \text{ if } \sigma[\phi] \neq \mathbf{0} \\
 & \sigma[\textit{might}\,\phi] = \mathbf{0} \text{ if } \sigma[\phi] = \mathbf{0} \\
\text{texts:} & \sigma[\nu;\phi] = \sigma[\nu][\phi]
\end{array}
$$

The update clauses tell for each sentence ϕ and every information state σ how σ changes when somebody in state σ accepts ϕ. If $\sigma[\phi] \neq \mathbf{0}$, we say that ϕ is *ac-*

[1] This is a bit misleading, since the phrase 'possible world' suggests that we are not talking about a way things might be, but about something that is that way. Cf Stalnaker [1976].

ceptable in σ. If $\sigma[\phi] = \mathbf{0}$, ϕ is *not acceptable in* σ and if $\sigma[\phi] = \sigma$, ϕ is *accepted* in σ. These notions are meant to be normative rather than descriptive: If $\sigma[\phi] = \mathbf{0}$, 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 the operators $\neg$, $\wedge$, $\vee$, *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 $\sigma[p] = \mathbf{0}$, whereas in fact p is true or that $\sigma[p] = \sigma$, whereas in fact p is false.

Suppose that p is in fact true and that $\sigma[p] = \mathbf{0}$. 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 must *revise* σ in such a manner that p *becomes* acceptable.

In the definition above 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, $\mathbf{0}[\phi] = \mathbf{0}$. 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: The Principle of Stability fails. The point is that revision is not the only possible source of unstability; testing is another. Here, sententences 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 information state you have to accept ‘It might be raining’, but as soon as you learn that it isn’t raining ‘It might be raining’ has to be rejected.

2.4 DEFINITION
A text ν is *consistent* iff there is an information state σ such that $\sigma[\nu]\neq\mathbf{0}$.

Again, since the set of information states varies with the nonlogical vocabulary of the language in which the text ν has been formulated, it would have been more accurate if we had introduced a notion of $\mathcal{A}$-consistency. The next lemma and proposition show however that this prefix can be omitted.

2.5 LEMMA
Let $\mathcal{A}\subseteq\mathcal{A}'$. With each $\mathcal{A}$-state σ we associate an $\mathcal{A}'$-state σ^* given by

$$\sigma^*=\{j\subseteq\mathcal{A}' \mid j\cap\mathcal{A}\in\sigma\}.$$

With each $\mathcal{A}'$-state σ we associate an $\mathcal{A}$-state σ° given by

$$\sigma^\circ=\{j\subseteq\mathcal{A} \mid j=i\cap\mathcal{A} \text{ for some } i\in\sigma\}.$$

Now, for every ϕ of $L_1^{\mathcal{A}}$ the following holds:
ia) if σ is an $\mathcal{A}$-state, then $\sigma[\phi]^*=\sigma^*[\phi]$;
ib) if σ, τ are $\mathcal{A}$-states and $\sigma\neq\tau$, then $\sigma^*\neq\tau^*$;
iia) if σ is an $\mathcal{A}'$-state, then $\sigma[\phi]^\circ=\sigma^\circ[\phi]$;
iib) if σ is an $\mathcal{A}'$-state, and $\sigma[\phi]\neq\sigma$, then $\sigma^\circ[\phi]\neq\sigma^\circ$.

It is not generally so that if σ and τ are $\mathcal{A}'$-states and $\sigma\neq\tau$, also $\sigma^\circ\neq\tau^\circ$. Fortunately, the property given under iib) is sufficiently strong to prove the next proposition.

2.6 PROPOSITION
Let $p_1,\dots,p_k$ be the atomic sentences occurring in $\psi_1,\dots,\psi_n$, ϕ. Suppose that $\{p_1,\dots,p_k\}\subseteq\mathcal{A}$ and $\{p_1,\dots,p_k\}\subseteq\mathcal{A}'$.
(i) The argument $\psi_1,\dots,\psi_n / \therefore\phi$ is $\mathcal{A}$-valid$_i$ iff it is $\mathcal{A}'$-valid$_i$ (i=1,2,3);
(ii) $\psi_1;\dots;\psi_n$ is $\mathcal{A}$-consistent iff $\psi_1;\dots;\psi_n$ is $\mathcal{A}'$-consistent.

Suppose $p_1,\dots,p_k$ are the atomic sentences occurring in the argument $\psi_1,\dots,\psi_n / \therefore\phi$. Given proposition 2.6, we may rest assured that the answer to the question whether $\psi_1,\dots,\psi_n / \therefore\phi$ is valid$_i$ (i=1,2,3), is language independent, as it should be. Actually, in looking for the answer we can always restrict ourselves to looking at the set of states generated by $\mathcal{A} = \{p_1, \ldots, p_k\}$. Since there are only finitely many of these, the logics generated by the three validity notions are decidable.

The next examples illustrate the points made in the preceding section.

2.7 EXAMPLES

(i) *might* $\neg$p ; p is consistent;
p ; *might* $\neg$p is not consistent.
(Compare this with the first example in § 1).

(ii) *might* $\neg$p $\models_2$ *might* $\neg$p, but *might* $\neg$p, p $\not\models_2$ *might* $\neg$p;
might $\neg$p $\models_3$*might* $\neg$p, but *might* $\neg$p, p $\not\models_3$ *might* $\neg$p.
In other words, neither validity$_2$ nor validity$_3$ are right-monotonic.

(iii) $\models_3$ *might*p, but $\neg$p $\not\models_3$ *might*p.
Validity$_3$ is not left-monotonic.
Note in passing that the logic generated by the third validity notion is not closed under substitution: $\models_3$ *might*p, but $\not\models_3$ *might*(p $\wedge$ $\neg$p).

A systematic study of the behaviour of *might* under the three validity notions will have to be left to another occassion. What follows are some preliminary observations, most of which will play a role in the next sections.

2.8 PROPOSITION

Let σ and τ be information states, and ϕ a sentence of $L_1^{\mathcal{A}}$.

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

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

(iii) If $\sigma \subseteq \tau$, then $\sigma[\phi] \subseteq \tau[\phi]$;

(iv) If ϕ is a sentence of $L_0^{\mathcal{A}}$, the following holds:
if $\sigma \subseteq \tau$, and $\tau[\phi] = \tau$, then $\sigma[\phi] = \sigma$.

We have already seen that (iv) does not hold for $L_1^{\mathcal{A}}$.
Consider the structure $\langle P(\mathrm{W}), \cap, \mathbf{1} \rangle$. This is a a semilattice with a zero-element. (Intersection serves as the join of this semilattice , and **1** serves as the zero-element). Clause (i), (ii) and (iii) of proposition 2.8 say that the meaning $[\phi]$ of every sentence ϕ is an idempotent monotonically increasing operation

on this semilattice. Clause (iv) adds that for sentences of L_0^A the operation is stable. Taking this together we find:[1]

2.9 PROPOSITION

Let ϕ be a sentence of L_0^A. For every σ, $\sigma[\phi] = \sigma \cap \mathbf{1}[\phi]$.

PROOF

That $\sigma[\phi] \subseteq \sigma \cap \mathbf{1}[\phi]$, follows almost immediately from (i) and (iii) of 2.8.
The converse is proved as follows:
Note first that $\sigma \cap \mathbf{1}[\phi] \subseteq \sigma$. Hence by (iii) of 2.8,

$$(\sigma \cap \mathbf{1}[\phi])[\phi] \subseteq \sigma[\phi]. \quad (*)$$

We also have that $\sigma \cap \mathbf{1}[\phi] \subseteq \mathbf{1}[\phi]$. By (ii) of 2.8, $\mathbf{1}[\phi][\phi] = \mathbf{1}[\phi]$. So, by (iv)

$$(\sigma \cap \mathbf{1}[\phi])[\phi] = \sigma \cap \mathbf{1}[\phi]. \ (**)$$

From (*) and (**) it follows immediately that $\sigma \cap \mathbf{1}[\phi] \subseteq \sigma[\phi]$.

This proposition—or rather its proof—shows that in many cases the dynamic approach has little to offer over and above the static approach. As soon as the following conditions are fulfilled:

- the set of information states has the structure of a semilattice with a zero-element;
- for every sentence ϕ, the dynamic meaning $[\phi]$ of ϕ is an operation with the properties mentioned in 2.8;

one can associate with every sentence ϕ a static meaning —*the* information contained in ϕ— in such a manner that updating a state σ with ϕ boils down to 'adding' the information contained in ϕ to the information contained in σ. In such a case, one might as well take this static notion of meaning as basic and define the dynamic notion in terms of it.

In the next sections, whenever we are dealing with a sentence ϕ of L_0^A, I will refer to the information contained in ϕ as *the proposition expressed by* ϕ, and write $[\![\phi]\!]$ instead of $\mathbf{1}[\phi]$.

What in the static approach is taken as the starting point, here can be proved:

$$\begin{aligned}
[\![p]\!] &= \{i \in W \mid p \in i\} \\
[\![\neg\phi]\!] &= W \sim [\![\phi]\!] \\
[\![\phi \wedge \psi]\!] &= [\![\phi]\!] \cap [\![\psi]\!] \\
[\![\phi \vee \psi]\!] &= [\![\phi]\!] \cup [\![\psi]\!]
\end{aligned}$$

[1] This result was inspired by similar observations in van Benthem[1989].

Using this result and the fact that for stable sentences the three validity notions coincide, one readily proves the next proposition.

2.10 PROPOSITION

Let $\psi_1, ..., \psi_n$, ϕ be sentences of L_0^A. Then

$\psi_1, ..., \psi_n \models_i \phi$ iff $\psi_1, ..., \psi_n$ / ∴ ϕ is valid in classical logic. (i=1,2,3)

Proposition 2.9 does not hold for L_1^A. Sentences of the form *might* ϕ do not express a proposition; their informational content is not contextindependent. If you learn a sentence ϕ of L_0^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.

LOOSE ENDS

In L_1^A *might* can only occur as the outermost operator of a sentence. Suppose we would change our grammar in such a way that *might* could occur everywhere, while leaving everything else as it is.

(i) It is easy to see that in that case *Idempotence* would no longer hold. (Consider the sentence p ∧ *might* ¬p).

(ii) Is there perhaps a convincing argument in favour of the position that *Idempotence* should no longer hold?

(iii) Is there perhaps an elegant way to restore *Idempotence*? After all, it could be that the update conditions given for ¬, ∧ and ∨ are too L_0^A-specific to work well for sentences that do not express a proposition.

(iv) Or should we just forbid *might* to occur everywhere? (After all, *might* expresses consistency, a meta-notion).

I really don't know which position to take here. All I know is that things get rather complicated if one opts for (iii) and starts out writing clauses like:

- $\sigma[\phi \wedge \psi]$ is the weakest τ as strong as σ such that $\tau \Vdash \phi$ and $\tau \Vdash \psi$
- $\sigma[\neg\phi]$ is the weakest τ as strong as σ such that $\tau[\phi] = \mathbf{0}$.

For one thing, there will not always be a *unique* weakest τ with the properties required (Consider the sentence ¬(*might* p ∧ *might* ¬p)).

§ 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.

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

Below, sentences of the form *normally* ϕ are called (default) rules. To describe their impact on an agent's state of mind, we need to 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 manner 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 an information 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. When a new rule is learnt, ε changes.

3.2 DEFINITION
Let W be as before. ε is an *expectation pattern* on W iff ε is a preordering of W:
(i) ε is reflexive: for every $i \in W$, $\langle i, i \rangle \in \varepsilon$;
(ii) ε is transitive: for every $i, j, k \in W$, if $\langle i, j \rangle \in \varepsilon$ and $\langle j, k \rangle \in \varepsilon$, then $\langle i, k \rangle \in \varepsilon$.

The relation ε encodes the rules the agent is acquainted with. It does so in the following manner: Let $\boldsymbol{P}$ be the set of all propositions that the agent considers normally to be the case. Now, intuitively, the pair $\langle i, j \rangle$ will be an element of

ε if every proposition in ***P*** that holds in j also holds in i. In other words, i conforms to all the rules that j conforms to, and perhaps to more. Given this interpretation, it will be clear that ε is reflexive and transitive.
If both $\langle j, i \rangle \in \varepsilon$ and $\langle i, j \rangle \in \varepsilon$, we write '$i \equiv_{\varepsilon} j$'. Clearly, $\equiv_{\varepsilon}$ is an equivalence relation.

3.3 DEFINITION

Let ε be a pattern on W;

(i) i is a *normal world* in ε iff $i \in W$ and $\langle i, j \rangle \in \varepsilon$ for every $j \in W$;

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

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

Clause (iii) says that a pattern ε is coherent if there is at least one possible world in which every proposition considered normally to be the case *is* the case. 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.

3.4 DEFINITION

Let ε be a pattern on W, and $s \subseteq W$.

(i) i is *optimal in* $\langle \varepsilon, s \rangle$ iff $i \in s$ and for every $j \in s$ such that $\langle j, i \rangle \in \varepsilon$, it holds that $\langle i, j \rangle \in \varepsilon$.

(ii) $\mathbf{m}_{<\varepsilon, s>}$ is the set of all optimal worlds in $\langle \varepsilon, s \rangle$.

Defaults 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. Default rules 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 more and more possibilities are eliminated. The actual world may turn out to be less normal than expected; the optimal worlds may disappear. When this happens, the best among the less than optimal worlds take over their role.

3.5 DEFINITION

Let ε and ε' be patterns on W, and $e \subseteq W$.

(i) ε' is a *refinement* of ε iff $\varepsilon' \subseteq \varepsilon$;

(ii) $\varepsilon \circ e = \{\langle i,j \rangle \in \varepsilon$: if $j \in e$, then $i \in e\}$; $\varepsilon \circ e$ is the *refinement of* ε *with* e.

It is left to the reader to check that $\varepsilon \circ e$ is a pattern.

The refinement operation is put to work when a new default rule is learnt. Think of this operation as follows: Suppose $\langle i,j \rangle \in \varepsilon$. Then i conforms to all the rules that j conforms to — that is to say: to all the rules learnt so far. Now a new default rule comes in: *normally* ϕ. Suppose $\mathbf{n}\varepsilon \cap [\![\phi]\!] \neq \emptyset$: the new rule is compatible with the old rules, and therefore it is acceptable. If it is accepted, the new pattern will become $\varepsilon \circ [\![\phi]\!]$. That is, if it happens to be the case that $i \notin [\![\phi]\!]$ and $j \in [\![\phi]\!]$, one has to remove the pair $\langle i,j \rangle$ from ε: given the new rule, it is no longer the case that i conforms to all the rules that j conforms to.

Let ε be a pattern. In the sequel, a proposition e is said to be a *(default) rule in* ε iff $e \neq \emptyset$ and $(\varepsilon \circ e) = \varepsilon$. The next proposition shows that this way of speaking fits in well with the informal explanation of the notion of a pattern given above.

3.6 PROPOSITION

Let ε be a pattern on W. The following holds for every $i, j \in W$:

$$\langle i,j \rangle \in \varepsilon \text{ iff } i \in e \text{ for every rule } e \text{ in } \varepsilon \text{ such that } j \in e.$$

PROOF

From left to right things are obvious. For the other direction, suppose $\langle i,j \rangle \notin \varepsilon$. Consider the proposition $e = \{ k \in W \mid \langle i,k \rangle \notin \varepsilon\}$.
Note that $i \notin e$ and $j \in e$. It is easy to check that e is a rule in ε.

The next proposition lists some basic properties of the refinement operation.

3.7 PROPOSITION

(i) $(\varepsilon \circ \emptyset)=\varepsilon$;

$(\varepsilon \circ W)=\varepsilon$;

(ii) $(\varepsilon \circ e) \circ e=\varepsilon \circ e$;

(iii) $(\varepsilon \circ e) \circ e'=(\varepsilon \circ e') \circ e$;

(iv) $\varepsilon \subseteq \varepsilon'$ iff for every e it holds that if $(\varepsilon' \circ e)=\varepsilon'$, then $(\varepsilon \circ e)=\varepsilon$.

The properties listed in proposition 3.7 have some important consequences for the system as a whole. From (ii) and (iii) it follows that the refinement operation is permutation invariant. So, if an agent has to learn a number of rules, then the order in which these are learnt does not make a difference to the result. According to (iv), a rule remains a rule as a pattern is further refined. This implies that sentences of the form *normally*ϕ are stable.

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

3.8 DEFINITION

Let W be as before. Then σ is an *information state* iff $\sigma=\langle\varepsilon, s\rangle$ and one of the following conditions is fulfilled:

(i) ε is a coherent pattern on W and s is a nonempty subset of W;

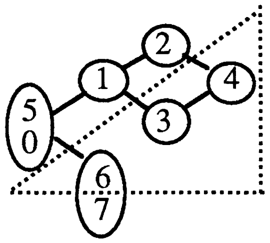
(ii) $\varepsilon=\{\langle i,i\rangle \mid i \in W\}$ and $s=\emptyset$.

If ε is incoherent, or if s is empty, something is wrong. In the above definition both kinds of incongruity are lumped together as only $\langle\{\langle i,i\rangle \mid i \in W\}, \emptyset\rangle$ has acquired official status as an information state.

$\mathbf{0}=\langle\{\langle i,i\rangle \mid i \in W\}, \emptyset\rangle$ is the *absurd* information state.

$\mathbf{1}=\langle W \mathrm{x} W, W\rangle$ is the *minimal* information state.

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



If two worlds belong to the same $\equiv_\varepsilon$-equivalence class, they are placed within the same circle or oval. So, the $\equiv_\varepsilon$-equivalence classes are $\{w_0, w_5\}$, $\{w_6, w_7\}$, $\{w_1\}$, $\{w_2\}$, $\{w_3\}$, and $\{w_4\}$. If $\langle w_i, w_j\rangle \in \varepsilon$ but $\langle w_j, w_i\rangle \notin \varepsilon$, the diagram contains a rightward path from the $\equiv_\varepsilon$-equivalence class to which w_i belongs to the $\equiv_\varepsilon$-equivalence class to which w_j belongs. We have for example that $\langle w_0, w_3\rangle \in \varepsilon$ but $\langle w_3, w_0\rangle \notin \varepsilon$, while it is neither the case that $\langle w_3, w_7\rangle \in \varepsilon$, nor that $\langle w_7, w_3\rangle \in \varepsilon$. The worlds constituting s are placed in an area with dashed borders: $s = \{w_3, w_4, w_6\}$. The normal worlds are w_5 and w_0; the worlds w_6 and w_3 are optimal.

3.9 DEFINITION

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

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

- if $s \cap [\![\phi]\!] = \emptyset$, $\sigma[\phi] = \mathbf{0}$;
- otherwise, $\sigma[\phi] = \langle \varepsilon, s \cap [\![\phi]\!]\rangle$.

NB: We already know from § 2 what $[\![\phi]\!]$ is.

if ϕ = *normally* ψ, then

- if $\mathbf{n}\varepsilon \cap [\![\psi]\!] = \emptyset$, $\sigma[\phi] = \mathbf{0}$;
- otherwise, $\sigma[\phi] = \langle \varepsilon \circ [\![\psi]\!], s\rangle$.

if ϕ = *presumably* ψ, then

- if $\mathbf{m}_\sigma \cap [\![\psi]\!] = \mathbf{m}_\sigma$, $\sigma[\phi] = \sigma$;
- otherwise, $\sigma[\phi] = \mathbf{0}$.

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 rules 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 rules or facts, so that in the end both will expect the same. (Admittedly, this is a somewhat idyllic picture).

3.10 EXAMPLES

Let $\mathcal{A} = \{p,q\}$

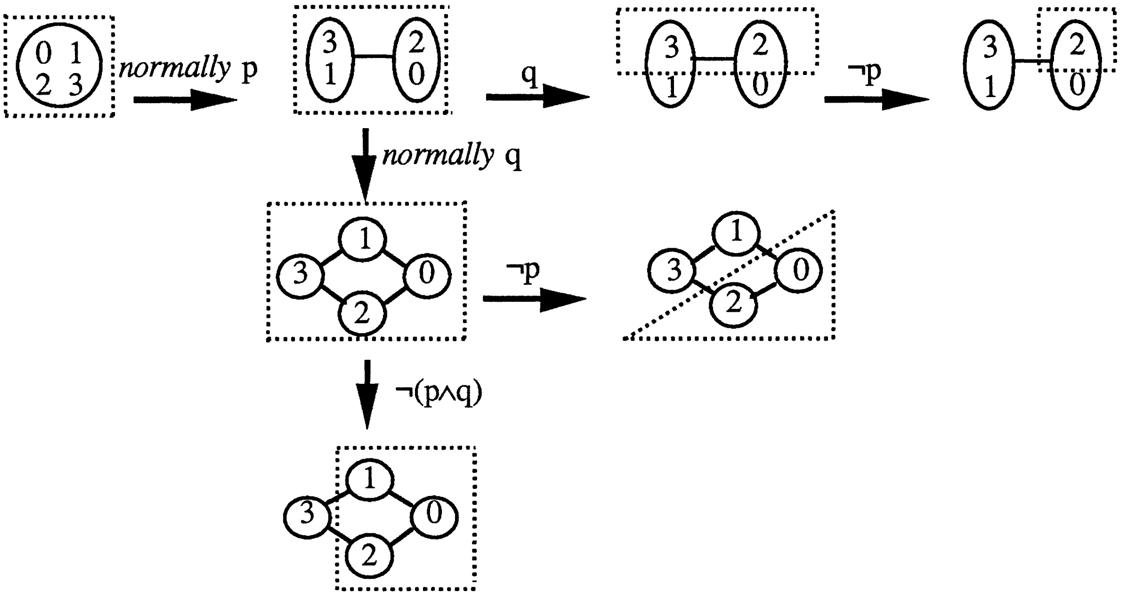
(i) $\mathbf{1}[\textit{normally}\, p][\neg p] \neq \mathbf{0}$
$\mathbf{1}[\textit{normally}\, p][\textit{normally}\, \neg p] = \mathbf{0}$

(ii) $\mathbf{1}[\textit{normally}\, p] \Vdash \textit{presumably}\, p$
$\mathbf{1}[\textit{normally}\, p][q] \Vdash \textit{presumably}\, p$
$\mathbf{1}[\textit{normally}\, p][q][\neg p] \nVdash \textit{presumably}\, p$

(iii) $\mathbf{1}[\textit{normally}\, p][\textit{normally}\, q] \Vdash \textit{presumably}\, p$
$\mathbf{1}[\textit{normally}\, p][\textit{normally}\, q][\neg p] \nVdash \textit{presumably}\, p$
$\mathbf{1}[\textit{normally}\, p][\textit{normally}\, q][\neg p] \Vdash \textit{presumably}\, q$

(iv) $\mathbf{1}[\textit{normally}\, p][\textit{normally}\, q][\neg (p \wedge q)] \nVdash \textit{presumably}\, p$
$\mathbf{1}[\textit{normally}\, p][\textit{normally}\, q][\neg (p \wedge q)] \nVdash \textit{presumably}\, q$

The information states pertaining to the examples mentioned under (ii), (iii) and (iv) are pictured below. $W = \{w_0, w_1, w_2, w_3\}$, where $w_3 = \{p,q\}$, $w_2 = \{q\}$, $w_1 = \{p\}$ and $w_0 = \emptyset$.



The examples illustrate some important characteristics of the system. The second example shows that sentences of the form *presumably* ϕ are not in general stable. Even if it is a rule that *normally* ϕ, it is sometimes wrong to expect that

ϕ. Only if the incoming information is irrelevant to ϕ —or at least not known to be relevant to ϕ — may one expect ϕ.

The third example shows 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 fourth example illustrates, sometimes one gets in a predicament. If one prefers worlds in which ϕ holds to worlds in which ϕ doesn't hold, and worlds in which ψ holds to worlds in which ψ doesn't hold, then it is hard to choose if the only choice one has got is between worlds in which ϕ holds but ψ doesn't hold, and worlds in which ψ holds but ϕ doesn't hold. In the next definition we introduce some more terminology to characterize this situation.

3.11 DEFINITION

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 i in $\langle \varepsilon, s \rangle$ such that $\mathbf{m} = \{j \in s\colon i \equiv_\varepsilon j\}$

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

The state **1** [*normally* p] [*normally* q] [$\neg(p \wedge q)$] is ambiguous. The other states discussed above are unambiguous.[1]

I will not pursue a systematic study of the operators *normally* and *presumably* here. But there are a few things that seem to me essential for a proper understanding of what's going on.

[1] It is tempting to continue this definition 3.10 as follows:

(iii) The *factual information contained in* $\langle \varepsilon, s \rangle$ is given by $\{\phi \mid \phi$ is a sentence of $L_0^{\mathcal{A}}$ that is accepted in s$\}$

(iv) A set Δ of sentences of $L_0^{\mathcal{A}}$ is called a *(default) extension* of the factual information contained in $\langle \varepsilon, s \rangle$ iff there is some optimal set **m** in $\langle \varepsilon, s \rangle$ such that $\Delta = \{\phi \mid \phi$ is a sentence of $L_0^{\mathcal{A}}$ that is accepted in $\mathbf{m}\}$

If one wants to compare the theories developed in this and the next section with other theories, then one way to go is to compare what each of them has to say about extensions. Note for example that we have:

$\sigma \Vdash$ *presumably* ϕ iff ϕ belongs to every default extension of the factual information contained in σ.

In other words, the theory developed here belongs to the class of *sceptical* theories.

3.12 DEFINITION

Let $\sigma = \langle \varepsilon , s \rangle$ and $\sigma' = \langle \varepsilon' , s' \rangle$ be information states.

(i) *σ is at least as strong as* σ' iff $s \subseteq s'$ and $\varepsilon \subseteq \varepsilon'$.

(ii) $\sigma \cap \sigma' = \langle \varepsilon \cap \varepsilon', s \cap s' \rangle$, if $\langle \varepsilon \cap \varepsilon', s \cap s' \rangle$ is coherent;
$\sigma \cap \sigma' = \mathbf{0}$, otherwise.

Note that $\sigma \cap \sigma'$ is the weakest τ that is at least as strong as σ and σ'.

Let Σ be the set of all information states for the language $L_2^{\mathcal{A}}$. Consider the structure $\langle \Sigma, \cap , \mathbf{1} \rangle$. This is a a semilattice with a zero-element (with $\cap$ serving as the join, and $\mathbf{1}$ as the zero-element).[1] For every ϕ, $[\phi]$ is an operation on this semilattice, which turns out to have the following structural properties:

3.13 PROPOSITION

Let ϕ be a sentence of $L_2^{\mathcal{A}}$ and σ and τ be any information states.

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

(ii) $\sigma[\phi]$ is at least as strong as σ;

(iii) If $\phi \neq$ *presumably* ψ the following holds:
if σ is at least as strong as τ, then $\sigma[\phi]$ is at least as strong as $\tau[\phi]$;

(iv) If $\phi \neq$ *presumably* ψ the following holds:
if σ is at least as strong as τ and $\tau[\phi] = \tau$ and , then $\sigma[\phi] = \sigma$.

We already saw that (iv) does not generally hold. That (iii) sometimes fails 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. Consider for example $\mathbf{1}[p]$, a state at least as strong as $\mathbf{1}$: $\mathbf{1}[p]$[*presumably* p] $= \mathbf{1}[p]$, but $\mathbf{1}$[*presumably* p] $= \mathbf{0}$. More important, however, is the fact that (i)-(iv) of proposition 3.13 do hold for sentences in which *presumably* does not occur. As we saw in the previous section, this means that we can assign to such sentences ϕ a ‘static’ meaning, viz. $\mathbf{1}[\phi]$, and think of the process of updating an information state σ with ϕ as adding the information contained in $\mathbf{1}[\phi]$ to the information contained in σ:

If $\phi \neq$ *presumably* ψ, then for every σ, $\sigma[\phi] = \sigma \cap \mathbf{1}[\phi]$.

[1] Unlike the semilattice studied in the previous section, this semilattice is not part of a boolean algebra. It is not generally so that one can find for given $\sigma, \sigma' \in \Sigma$, a unique strongest τ that is at least as weak as σ and σ'. (Let alone that Σ is fitted with complements).

In other words, not only purely descriptive sentences carry context independent information but rules do so, too. If $[\![\phi]\!] \neq \emptyset$, the information contained in the rule *normally* ϕ can be identified with the pattern $(W \mathrm{x} W) \circ [\![\phi]\!]$. What *normally* ϕ says is that ϕ-worlds—if at least there are any— are to be prefered to other worlds.

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 characterization of the logical properties of the latter in a normal system of modal logic[1].

(I) $\quad \textit{necessarily}\phi \models \phi$

(II) $\quad \textit{necessarily}\,\phi, \textit{necessarily}\,\psi \models \textit{necessarily}(\phi \wedge \psi)$

(III) $\quad \textit{necessarily}\,\phi \models \textit{necessarily}\,(\phi \vee \psi)$

(IV) $\quad$ If $\models \phi$, then $\models \textit{necessarily}\,\phi$

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

$$\textit{normally}\,\phi, \textit{normally}\,\psi \models_i \textit{normally}\,(\phi \wedge \psi) \quad (i=1,2,3)$$

$$\text{If } \models_i \phi, \text{ then } \models_i \textit{normally}\,\phi \quad (i=1,2,3)$$

The third principle fails. It is not generally so — not even if i=3 — that

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

Perhaps the point is best brought out by an example. Compare the following sequences of sentences.

— Normally John is at home. He is not at home now. So, presumably he is at school.

— Normally Fred is at home or at school. He is not at home now. So, presumably he is at school.

[1] We are restricting our attention here to a language in which the necessity operator can only occur 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: An *information state* σ is a pair $\langle \varepsilon , s \rangle$ just like before, only this time ε is a pattern on a subset f of W of rather than on W itself. As the next update clause shows, the extension of f is determined by the strict rules the agent is acquainted with: a world i is an element of f just in case every proposition that the agent considers as necessary holds in i.

- if $\mathbf{n}\varepsilon \cap [\![\phi]\!] = \emptyset$ or $s \cap [\![\phi]\!] = \emptyset$, $\sigma[\textit{necessarily}\,\phi] = \mathbf{0}$;
- otherwise, $\sigma[\textit{necessarily}\,\phi] = \langle \varepsilon | (f \cap [\![\psi]\!]), s \cap [\![\psi]\!] \rangle$

So, when a new strict rule *necessarily* ϕ is learnt both the pattern ε and the set s are *restricted* to the worlds in which the proposition expressed by ϕ holds.

Intuitively, the first line of thought is incorrect. Formally, it is invalid$_3$:

$$\mathbf{1}\,[\textit{normally}\ \mathrm{p}]\ [\neg \mathrm{p}] \nVdash \textit{presumably}\ \mathrm{q}$$

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

$$\mathbf{1}\,[\textit{normally}\ (\mathrm{p} \vee \mathrm{q})]\ [\neg \mathrm{p}] \Vdash \textit{presumably}\ \mathrm{q}$$

What this example shows is that a rule like 'Normally John is at home or at school' is in some respects stronger than the rule 'Normally John is at home'. The former gives us some indication as to what we can expect in case we have found that John is not at home, the latter doesn't. No wonder, then, that one can accept the latter rule without accepting the former.[1]

Finally, some remarks in connection with the first principle. We know already that this principle does not hold for *normally*. What we have instead is the much weaker principle:

$$\textit{normally}\ \phi \models \textit{presumably}\ \phi.$$

This principle is only valid$_3$. It is neither valid$_1$ nor valid$_2$. In many information states in which *normally* ϕ is accepted, one is not entitled to infer *presumably* ϕ. Often, this presumption is overruled by the facts.

The only way to rescue this weakened form the instantiation principle is to add a new operator to our language—an operator that enables one to express that no facts have been learnt that come between the premise *normally* ϕ on the one hand and the conclusion *presumably* ϕ on the other. Its syntax and semantics are given by the following rules:

- if ϕ is a sentence of $L_0^{\mathcal{A}}$, then *may* ϕ is a sentence of $L_2^{\mathcal{A}}$.
- if $\phi = \textit{may}\ \psi$, then
  - if $\mathbf{m} \cap [\![\psi]\!] \neq \emptyset$ for every optimal set $\mathbf{m}$ in $\sigma, \sigma[\phi] = \sigma$
  - otherwise, $\sigma[\phi] = \mathbf{0}$

According to this analysis *may*ϕ is somewhat stronger than *might*ϕ. All you say with *might*ϕ is that ϕ is consistent with what you know. With *may*ϕ you say that ϕ is consistent with what you expect — if at least your expectations are unambiguous, in which case the test for *may* ψ boils down to this:

[1] I cannot prove that that it is impossible for there to be a system in which (i) *normally* p $\models$ *normally* (p $\vee$ q); (ii) *normally* (p $\vee$ q), $\neg$p $\models$ *presumably* q; but (iii) *normally* p, $\neg$p $\not\models$ *presumably* q. Still, if you want such a system you will have to give up either (a) the principle of Lemma Generation; or (b) the idea that rules are stable. For, by (b) it follows from (i) that *normally* p, $\neg$p $\models$ *normally* (p $\vee$ q). Given (a) and (iii) this means that *normally* p, *normally* (p $\vee$ q), $\neg$p $\not\models$ *presumably* q. But this is almost as bad as not having (ii).

- if $\mathbf{m}_\sigma \cap [\![\psi]\!] \neq \emptyset$, $\sigma[may\,\psi] = \sigma$
- otherwise, $\sigma[may\,\psi] = \mathbf{0}$

If your expectations are ambiguous, things are more complicated. Then for the test on *may*ϕ to have a positive outcome, ϕ has to be consistent with every possible disambiguation of your expectations.

Given this interpretation of *may* we find:

$$normally\,\phi,\ may\,\phi \models_i presumably\ \phi.\ (i = 1,2)$$

The observations needed for the proof are given in the next lemma.

3.14 LEMMA

Let $\sigma = \langle \varepsilon, s \rangle$ be an information state. Suppose $\sigma \Vdash normally\,\phi$. Then for every i,j $\in$ W the following holds:

$$\text{If } i \equiv_\varepsilon j, \text{ then } i \in [\![\phi]\!] \text{ iff } j \in [\![\phi]\!].$$

LOOSE ENDS

Compare the next two sentences:

(i) Normally it does not rain in April.

(ii) It is not the case that it normally rains in April.

Both (i) and (ii) are grammatically correct, and they differ in meaning: (ii) is weaker than (i). One can continue (ii) with (iii), thus denying (i):

(iii) Neither is it the case that normally it does not rain in April.

The speakers of the language L_2^A can say things like (i), but they cannot say things like (ii). So, why not enable them to do so?

Here is one way to go about: We add to the language L_2^A a new one place operator $\sim$, called *denial.* This is the semantic rule that comes with it:

- $\sigma[\sim\phi] = \sigma$ if $\sigma[\phi] \neq \sigma$
- otherwise, $\sigma[\sim\phi] = \mathbf{0}$

In other words, $\sim\phi$ invites a test, and this test has a positive outcome just in case σ is not accepted ϕ.

It will be clear how this is supposed to work: if our agents want to say something like (ii), they can use a sentence of the form $\sim$*normally* ϕ; and if they want say something like (iii) they can do so with a sentence of the form $\sim$*normally* $\neg\phi$. However, failing an answer to the question as to when exactly the English 'not' has the meaning of $\neg$ and when it is just a denial $\sim$, this manœuvre can only be qualified as *ad hoc*. It would be much more elegant if we could get what we want by extending the update clause for sentences of the form $\neg\phi$ to the case that $\phi = \textit{normally}\ \psi$. Unfortunately, I see no natural way to do so. (Consider for example this generalization, which is based on the idea that 'not ϕ' means so much as 'it is absurd to accept ϕ':

$\sigma[\neg\phi]$ is the weakest τ as strong as σ such that $\tau[\phi] = \mathbf{0}$.

For descriptive ϕ this definition works fine, but there are no states σ such that both $\sigma \Vdash \neg\textit{normally}\,\mathrm{p}$ and $\sigma \Vdash \neg\textit{normally}\,\neg\mathrm{p}$ except the absurd one.)

This is just one of the problems that arise when one wants to give a meaning to formulas in which *normally* and *presumably* do not occur as outermost operators. To give another example, what meaning should we give to a formula of the form $\textit{presumably}\,\phi \vee \textit{presumably}\,\psi$? It is not easy to think of a context in which somebody seriously says:

(iv) Presumably it is raining, or presumably it is snowing;

if at least this is not to be understood as a roundabout way to say:

(v) Presumably it is raining or snowing.

Perhaps (iv) can be seriously said by a speaker who does not anything about the weather but this:

(vi) Normally it rains, or normally it snows.

But then, would you be prepared to accept such a disjunction of rules — even in circumstances in which it is not clear which of the disjuncts you are supposed to choose? (And if so, how to account for this formally?)
And what to think of:

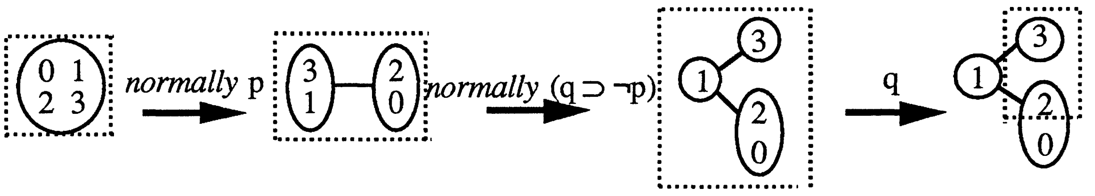
(vii) Normally it is presumably raining.

Is this grammatically correct at all? (If not so, why not?)
Sentences like this are not only a problem for the present theory, but also for the average speaker of English. One might hope that within a dynamic framework, where a distinction can be drawn between test operators and other operators, an explanation for this can be found. Perhaps a case can be made for the thesis that by their very nature test operators can only occur in certain positions. But note that it is not true that they can only occur as outermost operators. For one thing, it is perfectly alright to put them in front of the consequent of a conditional sentence. There is nothing wrong with 'If it is not raining , then presumably it is snowing'.

§ 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 nonaccidental exceptions: we cannot say when exceptional circumstances are to be expected and what one can expect when they obtain—'But if it's below 0°C, normally it will not rain. In that case it normally snows.'

Let me illustrate this with a formal example. Suppose an agent in state **1** accepts the rule *normally* p. As we saw, this induces a general preference for worlds in which ⟦p⟧ holds. Now, we want to *make* an exception: if ⟦q⟧ holds, ⟦p⟧ normally does not hold. And we want to do so in such a manner that in the event it is learnt that in fact ⟦q⟧ does hold, the agent will expect ⟦p⟧ *not* to hold. The point is that we cannot make this exception with a formula of the form *normally*($q \supset \neg p$). The effect of accepting this formula should be that in the domain of q-worlds the rule *normally* p is overridden, but the semantics for *normally* does not work out that way. The formula *normally*($q \supset \neg p$) induces another general preference, this time for worlds in which the proposition ⟦$q \supset \neg p$⟧ holds. So, when it is learnt that ⟦q⟧ holds an ambiguous situation arises: There are two optimal sets, one for the worlds that conform to the one general rule *normally* p, and the other for the worlds that conform to the other general rule *normally*($q \supset \neg p$). In the picture below the numbers refer to the same worlds as in the example on page 18).



One cannot equate 'q normally implies ¬p' with *normally*($q \supset \neg p$); the binary operator '*...normally implies...*' is not definable in terms of the unary operator *normally*; 'q normally implies ¬p' says that the proposition expressed by ¬p is a rule in the domain of q-worlds. Within this domain of q-worlds the rule *normally* p, which is a rule in the domain of all worlds, can be suspended.

4.1 DEFINITION

Let $\mathcal{A}$ and $L_0^{\mathcal{A}}$ be as in § 2. The language $L_3^{\mathcal{A}}$ has $\mathcal{A}$ as its nonlogical vocabulary, and in its logical vocabulary one additional unary operator *presumably*, and one additional binary operator $\rightsquigarrow$.

A string of symbols ϕ is a sentence of $L_3^{\mathcal{A}}$ iff there are sentences ψ and χ of $L_0^{\mathcal{A}}$ such that $\phi = \psi$, or $\phi = presumably\ \psi$, or $\phi = \psi \rightsquigarrow \chi$.

Read '$\phi \rightsquigarrow \psi$' as '$\phi$ normally implies ψ'.
We write '*normally* ϕ' to abbreviate '$(\phi \vee \neg\phi) \rightsquigarrow \phi$'.

It must be possible for a proposition to be a rule in a given domain without it automatically being a rule in all its subdomains. This means we need to be able to assign different patterns to different subsets of W.

4.2 DEFINITION

Let W be as before. π is an *(expectation) frame* on W iff π is a function that assigns to every subset d of W a pattern $\pi(d)$ on d.

4.3 DEFINITION

Let π be a frame on W; suppose $d, e \subseteq W$.
e is a *(default) rule* in $\pi(d)$ iff (i) $\pi(d) \circ e = \pi(d)$; and (ii) $d = \emptyset$ or $d \cap e \neq \emptyset$.

Suppose $d \cap e = \emptyset$. Then by the definition of $\circ$, $\pi(d) \circ e = \pi(d)$. But how could it be normal for the proposition e to hold in the domain d, if e holds nowhere in the domain d? The proposition e cannot be a rule in $\pi(d)$ in this case—unless $d = \emptyset$: in absurd circumstances nothing is abnormal.

Definition 4.2 allows for the possibility that every subset d of W—every domain—has its own rules. So, now we can make as many exceptions as we want to. But of course not anything goes. If we make too many exceptions, our expectation frame gets incoherent.

4.4 DEFINITION

Let π be a frame on W; suppose $d \subseteq W$ and let $i \in d$.

(i) i is a *normal world in* d iff $i \in \mathbf{n}\pi(d')$ for every $d' \subseteq d$ such that $i \in d'$;

(ii) $\nu_\pi(d)$ is the set of all normal worlds in d;

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

Below I will use the the term 'coherent' in two ways: A coherent *frame* is a frame with the property laid down in (iii). By a coherent *pattern* I will mean a pattern that is coherent in sense of definition 3.3: a pattern ε such that $\mathbf{n}\varepsilon \neq \emptyset$.

4.5 PROPOSITION
Let π be a frame on W. The following are equivalent.
(i) π is coherent;
(ii) If $d \neq \emptyset$, there is a coherent pattern ε on d such that for every $d' \subseteq d$,
$\varepsilon | d' \subseteq \pi(d')$.

For a frame to be coherent every single pattern in the frame must be coherent: for every nonempty domain d, $\mathbf{n}\pi(d)$ must be nonempty. But this is not sufficient. Clause (ii) of proposition 4.5 says that the patterns $\pi(d')$ on the subdomains of a domain d must all fit together in one coherent pattern ε on d.
What is the rationale for this constraint? A pattern is a kind of preference relation. Each rule in $\pi(d')$ provides some criterion for calling some of the worlds in d' more normal—in some respect—than the other worlds in d'. In establishing their preferences, our agents may employ different criteria in different domains. But not anything goes: once they have employed a certain criterion in a domain d', they must be prepared to employ the very same criterion in the d'-part of any domain larger than d'—in passing from a smaller domain to a larger domain no criterium can get lost: in a larger domain one can make at least as many distinctions. Actually, our agents must be prepared to do this cumulatively, so that in the end for every domeain d, all criteria employed in any subdomain d' of d have been taken into account. If there is some domain d for which this cannot be done without the resulting pattern on d getting incoherent, the frame is incoherent.
Note that (ii) doesn't say that the pattern $\pi(d)$ itself must have the property that its restriction $\pi(d)| d'$ to any of its subdomains d' is a refinement of $\pi(d')$. The coherence constraint only says that it must be possible to coherently refine $\pi(d)$ in such a manner that *the result* has this property. It will soon become clear why it would be unreasonable to go further than that.

4.6 DEFINITION

Let W be as before. Then σ is an *(information) state* iff $\sigma = \langle \pi, s \rangle$, and one of the following conditions is fulfilled:

(i) π is a coherent frame on W, and s is a nonempty subset of W;

(ii) π is the frame given by: for every $d \subseteq W$, $\pi(d) = \{\langle i,i \rangle \mid i \in d\}$; $s = \emptyset$.

The minimal information state **1** is given by:

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

The absurd information state is the one described in (ii) above:

$$\mathbf{0} = \langle \iota, \emptyset \rangle, \text{ where } \iota(d) = \{\langle i,i \rangle \mid i \in d\} \text{ for every } d \subseteq W.$$

The differences between these definitions and the corresponding ones in the preceding section 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. Suppose an agent in state $\sigma = \langle \pi, s \rangle$ learns $\phi \leadsto \psi$: within the domain of ϕ-worlds it is normal that $\llbracket\psi\rrbracket$ holds. The relevant domain is given by the proposition $\llbracket\phi\rrbracket$. If the agent decides to accept the new rule, the pattern $\pi(\llbracket\phi\rrbracket)$ on this domain will have to be refined with $\llbracket\psi\rrbracket$. So, the only difference between the updated frame π' and π is that $\pi'(\llbracket\phi\rrbracket) = \pi(\llbracket\phi\rrbracket) \circ \llbracket\psi\rrbracket$. Of course, this is not all there is to it. An agent should not accept $\phi \leadsto \psi$ if the result π' is incoherent.

4.7 DEFINITION

Let π be a frame and $d,e \subseteq W$.

$\pi_{d \circ e}$ is the frame that for every $d' \subseteq W$ is given by:

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

(ii) $\pi_{d \circ e}(d) = \pi(d) \circ e$.

The frame $\pi_{d \circ e}$ is *the result of refining* $\pi(d)$ *in* π *with* e .

4.8 DEFINITION

Let $\sigma = \langle \pi, s \rangle$ be an information state. $\sigma[\phi \leadsto \psi]$ is determined as follows:

- $\sigma[\phi \leadsto \psi] = \mathbf{0}$ if (i) $\pi_{\llbracket\phi\rrbracket \circ \llbracket\psi\rrbracket}$ is incoherent; or (ii) $\llbracket\phi\rrbracket \neq \emptyset$ and $\llbracket\phi\rrbracket \cap \llbracket\psi\rrbracket = \emptyset$;
- otherwise, $\sigma[\phi \leadsto \psi] = \langle \pi_{\llbracket\phi\rrbracket \circ \llbracket\psi\rrbracket}, s \rangle$.

The case that $[\![\phi]\!] \neq \emptyset$ and $[\![\phi]\!] \cap [\![\psi]\!] = \emptyset$ is special: as we saw, $[\![\psi]\!]$ cannot be a rule in the domain $[\![\phi]\!]$ in this case. (Still, by definition, the result of updating $\pi([\![\phi]\!])$ with $[\![\psi]\!]$ is coherent in this case—a technical inconvenience.)

4.9 PROPOSITION

Let π be a coherent frame and $d, e \subseteq W$. Suppose $d \cap e \neq \emptyset$. Then $\pi_{d \circ e}$ is coherent iff there is no $d' \supseteq d$ such that $\mathrm{v}\pi(d') \subseteq d \sim e$.

PROOF

From left to right: Suppose there is some $d' \supseteq d$ such that $\mathrm{v}\pi(d') \subseteq d \sim e$. Given that $d \cap e \neq \emptyset$, $\mathbf{n}\pi_{d \circ e}(d) \subseteq e$. Hence, $\mathrm{v}\pi_{d \circ e}(d') \cap d \subseteq e$. But since $\mathrm{v}\pi_{d \circ e}(d') \subseteq \mathrm{v}\pi(d')$, this means that $\mathrm{v}\pi_{d \circ e}(d') = \emptyset$.

From right to left: Suppose there is no $d' \supseteq d$ such that $\mathrm{v}\pi(d') \subseteq d \sim e$. We have to show that for every d', $\mathrm{v}\pi_{d \circ e}(d') \neq \emptyset$. There are two cases:(i) $d \subseteq d'$. Given that $d \cap e \neq \emptyset$, $\mathrm{v}\pi_{d \circ e}(d') = \mathrm{v}\pi(d') \sim (d \sim e)$. Since $\mathrm{v}\pi(d') \not\subseteq d \sim e$, $\mathrm{v}\pi_{d \circ e}(d') \neq \emptyset$.
(ii) Not $d \subseteq d'$. In this case $\mathrm{v}\pi_{d \circ e}(d') = \mathrm{v}\pi(d')$. So, there is nothing to prove.

4.10 COROLLARY

Let $\sigma = \langle \pi, s \rangle$ be coherent. Then $\sigma[\phi \rightsquigarrow \psi]$ is determined as follows:

- if $[\![\phi]\!] \neq \emptyset$ but $\mathrm{v}\pi(d) \subseteq [\![\phi]\!] \sim [\![\psi]\!]$ for some $d \supseteq [\![\phi]\!]$, $\sigma[\phi \rightsquigarrow \psi] = \mathbf{0}$;
- otherwise, $\sigma[\phi \rightsquigarrow \psi] = \langle \pi_{[\![\phi]\!] \circ [\![\psi]\!]}, s \rangle$.

4.11 EXAMPLES

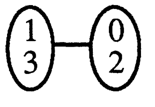
(i) $\mathbf{1}[\textit{normally}\, p]\,[q \rightsquigarrow \neg p] \neq \mathbf{0}$;

(ii) $\mathbf{1}[\textit{normally}\, p]\,[q \rightsquigarrow \neg p]\,[\neg q \rightsquigarrow \neg p] = \mathbf{0}$;

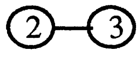
(iii) $\mathbf{1}[\textit{normally}\, p]\,[\textit{normally}\, q]\,[q \rightsquigarrow \neg p] = \mathbf{0}$.

In example (i) we encounter an agent who accepts the general rule *normally* p, but who wants to make an exception for the case that q holds.
The resulting frame π looks like this:

This is $\pi(W)$:



This is $\pi(\{w_2, w_3\})$:

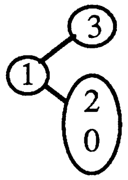


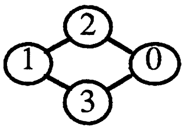
And if $d \neq W$ and $d \neq \{w_2, w_3\}$, $\pi(d) = d \times d$.
In other words, the proposition $[\![p]\!]$ is a rule in the domain of all worlds; the proposition $[\![\neg p]\!]$ is a rule in the domain $[\![q]\!]$; and that's it.

Note that $\nu_\pi(W) = \{w_1\}$. So, despite the fact that it conforms to the general rule *normally* p, the world $w_3 = \{p,q\}$ does not count as a normal world in the domain W. Think of this as follows: By accepting $q \rightsquigarrow \neg p$, the agent decided that the worlds in $[\![q]\!]$ —and w_3 is one of these—are *exempted from* the general rule. So, to say that w_3 conforms to the general rule, as I did above, is a bit misleading as it suggests that w_3 is subjected to the general rule in the first place. But it isn't. It is only subjected to the more specific rule $[\![\neg p]\!]$, to which it happens to be an exception. In other words, w_3 is an exception to an exceptive clause of a more general rule. And we are not going to consider such an 'exception to an exception' as normal.

The example can also be of help in answering the question why the coherence constraint laid down in definition 4.4 is not stronger than it is. Why don't we require that if $d' \subseteq d$, the d'-part of $\pi(d)$ must be a coherent refinement of $\pi(d')$, but just that it must be possible to coherently refine $\pi(d)$ in such a manner that the d'-part of the result will be a refinement of $\pi(d')$?

Note that it is not the case that $\pi(W) \mid \{w_2, w_3\} \subseteq \pi(\{w_2, w_3\})$. But in accordance with the coherence constraint, there is a coherent refinement ε of $\pi(W)$ such that $\varepsilon \mid \{w_2, w_3\} \subseteq \pi(\{w_2, w_3\})$. Here are two such refinements:

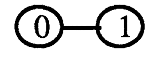




It is easy to check that every refinement ε of $\pi(W)$ with the property that for every d, $\varepsilon \mid d \subseteq \pi(d)$ must be a refinement of one of these patterns. But since neither is a refinement of the other, there is no such thing as *the weakest* coherent refinement ε of $\pi(W)$ such that for every d, $\varepsilon \mid d \subseteq \pi(d)$.

Now, suppose we had taken the stronger coherence requirement as our starting point. Then in the update clause for $\phi \rightsquigarrow \psi$ we should have seen to it that in the updated frame the proposition $[\![\psi]\!]$ would not only be a rule in the domain $[\![\phi]\!]$, but also in the in the $[\![\phi]\!]$-part of all larger domains. However, as the example shows, there are often several ways to refine the pattern of a larger domain so that $[\![\psi]\!]$ becomes a rule in its $[\![\phi]\!]$-part. And in general there is no way to single out one of the alternatives as the right one.

Let us now turn to example (ii). We are dealing with an agent who has accepted the general rule *normally* p and who has made an exception for the case that q holds. On top of this the agent wants to make an exception for the case that q does not hold. Intuitively, this is too much: if you make this exception as well, you make too many exceptions.
Formally: the resulting frame $\pi_{[\![\neg q]\!]\circ[\![\neg p]\!]}$ is the same as the frame π except that $\pi_{[\![\neg q]\!]\circ[\![\neg p]\!]}(\{w_0, w_1\})$ looks like this:



But this means $v_\pi(W) = \emptyset$. The resulting frame $\pi_{[\![\neg q]\!]\circ[\![\neg p]\!]}$ is incoherent. Therefore $\mathbf{1}$[*normally* p] [q $\rightsquigarrow$ ¬p] [¬q $\rightsquigarrow$ ¬p] = $\mathbf{0}$.

The next proposition serves the same purpose as proposition 3.7.

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

4.13 PROPOSITION
(i) $(\pi_{d\circ e})_{d\circ e} = \pi_{d\circ e}$;
(ii) $(\pi_{d\circ e})_{d'\circ e'} = (\pi_{d'\circ e'})_{d\circ e}$;
(iii) π is a refinement of π' iff for every $d, e \subseteq W$, it holds that if $\pi'_{d\circ e} = \pi'$, then $\pi_{d\circ e} = \pi$.

Let $\sigma = \langle \pi, s \rangle$ be an information state. The frame π encodes the rules an agent in state σ is acquainted with, and s his knowledge of the facts. Now, what will an agent in state σ expect? In the previous section, where we dealt with states consisting of just one expectation pattern ε, this question was easy to answer: all we had to do was to take a closer look at the most normal worlds in the s-part of ε. Now things are a lot 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* to which parts of s.
This is the general idea: Suppose e is rule in the domain d, and suppose that there are d-worlds in s. Even if e is not explicitly recorded as a rule in the domain $s \cap d$, e may very well *apply* to $s \cap d$. In this case, the d-worlds in s are

subjected to the rule e, and an agent will expect the real world to be a world in $s \cap (d \cap e)$, or a world in $s \sim d$, rather than a world in $s \cap (d \sim e)$. But the rule e need not apply to the domain $s \cap d$; sometimes it is overridden by other rules. If so, the d-worlds in s are not subjected to e and there is no reason for an agent to expect the real world to be a d-world in which e holds or a world outside the domain d rather than a d-world in which e does not hold.

4.14 DEFINITION

Let π be a coherent frame and assume that e is a rule in $\pi(d)$.
Then e *applies to* d' iff $d' \subseteq d$ and for every $d'' \supseteq d'$ such that $\vee\pi(d'') \subseteq d \sim e$, it holds that $d'' \subseteq d \sim e$.

Here are some trivial consequences of this definition:
If e is a rule in d, then e applies to d.
The d-rule d applies to every $d' \subseteq d$.
If the d-rule e applies to d', then it applies to every d'' such that $d' \subseteq d'' \subseteq d$.
A less trivial, but perhaps more enlightening consequence is the following:

4.15 PROPOSITION

Let π be a frame and assume that e is a rule in $\pi(d)$. Suppose that $d' \subseteq d$ and $d' \cap e \neq \emptyset$. Then e applies to d' iff there is some coherent refinement π' of π such that e is a rule in every d'' such that $d' \subseteq d'' \subseteq d$.

I hope that proposition 4.15 makes definition 4.14 more perspicuous. It says that a d-rule e applies to a subdomain d' of d just in case it is coherent to assume that e is a rule in every domain between d' and d. However, this proposition only covers the case that $d' \cap e \neq \emptyset$; the case that $d' \cap e = \emptyset$ is more difficult to explain: if $d' \cap e = \emptyset$, e cannot be a rule in d'.
Nevertheless, it is very well possible for a d-rule e to apply to a domain d' in which there are no e-worlds. Default rules have much in common with normative rules: worlds can be subjected to a rule and not comply with it. 'The d-rule e applies to the domain d'', means that the d'-worlds have not been exempted from rule in question—they would have been better worlds if they had been e-worlds.
How, then is definition 4.14 to be understood? Recall that a pattern is a kind of preference criterion; every d-rule e provides a criterion for calling some of

the worlds in its domain d more normal—in some respect—than the other worlds in d. Moreover, in a coherent frame a d-rule e can be coherenly used as a preference criterion in every domain larger than d. Now, let d' be a subdomain of d. The rule e may very well *apply* to d'. When will this be? Roughly, the answer is this: *If* the d-rule e can be coherenly used as a preference criterion in d' and in every domain larger than d'.
Consider any domain $d'' \supseteq d'$. If $d'' \subseteq d{\sim}e$, there is hardly a point in using e as a preference criterion in d'', because doing so leaves $\pi(d'')$ as it is: every world in d'' is an exception to the rule, so there are no worlds in d'' that are more normal—in this respect—than the other worlds d''. If, on the other hand, $d'' \not\subseteq d{\sim}e$, the use of the d-rule e as a preference criterion will have consequences. For, consider any world i in $d{\sim}e$, and let j be any world in $d''{\sim}(d{\sim}e)$. If j is an element of $d \cap e$, then j is in at least one respect more normal than i because j satisfies the criterion and i does not. And if j is an element of $d''{\sim}d$, then j is to be prefered to i as well, if only because j *does not have to* to satisfy the criterion, and therefore is no exception to it. So, the effect of using the d-rule e as a preference criterion in d'' is that the d-worlds in d'' in which e does not hold cannot count as normal d''-worlds any more. If $\nu_\pi(d'') \not\subseteq d{\sim}e$, this is alright; but if $\nu_\pi(d'') \subseteq d{\sim}e$, this would mean that there are no normal d''-worlds left. Therefore, if $\nu_\pi(d'') \subseteq d{\sim}e$ for some $d'' \supseteq d'$ such that $d'' \not\subseteq d{\sim}e$, the d-rule e does not apply to d'.

Before proceeding with some examples, I will give a more precise meaning to some of the notions introduced informally above.

4.16 DEFINITION
Let $\langle \pi, s \rangle$ be an information state. Suppose e is a rule in $\pi(d)$,and let $i \in W$.
(i) e *applies within* s iff e applies to $s \cap d$;
(ii) i is subjected to e iff $i \in s \cap d$ and e applies within s;
(iii) i is exempted from e iff $i \in s \cap d$ and e does not apply within s;
(iv) i is an accidental exception to e iff i is subjected to e and $i \notin e$;
(v) i is an nonaccidental exception to e iff i is exempted from e and $i \notin e$.

4.17 EXAMPLES

For each of the states $\sigma_i = \langle \pi_i, s_i \rangle$ we want to know which rules apply in s_i.

(i) $\sigma_1 = \mathbf{1}[\mathit{normally}\, p]\ [q]$;

(ii) $\sigma_2 = \mathbf{1}[\mathit{normally}\, p]\ [q \rightsquigarrow \neg p]\ [q]$;

(iii) $\sigma_3 = \mathbf{1}[\mathit{normally}\, p]\ [q \rightsquigarrow \neg p]\ [q \wedge r]$;

(iv) $\sigma_4 = \mathbf{1}[\mathit{normally}\, p]\ [q \rightsquigarrow \neg p]\ [(q \wedge r) \rightsquigarrow p]\ [q \wedge r]$;

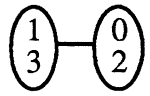
(v) $\sigma_5 = \mathbf{1}[\mathit{normally}\, p][\mathit{normally}\, q]\ [\neg(p \wedge q)]$;

(vi) $\sigma_6 = \mathbf{1}[q \rightsquigarrow \neg p]\ [p]$;

(vii) $\sigma_7 = \mathbf{1}[p \rightsquigarrow q]\ [q \rightsquigarrow \neg p]\ [p]$.

(i).The frame π_1 is given by: If $d \neq W$ then $\pi_1(d) = d \times d$.

And $\pi_1(W)$, looks like this:



(The worlds are indexed in the same manner as before).

The proposotion $\llbracket p \rrbracket$ is a rule in $\pi_1(W)$.It is not difficult to check that it applies to $s_1 = \llbracket q \rrbracket = \{w_2, w_3\}$.This means that both $w_2 = \{q\}$ and w_3 are subjected to the general rule $\llbracket p \rrbracket$; w_2 is an accidental exception to it.

(ii). We already know the frame π_2 from page 30: the proposition $\llbracket p \rrbracket$ is a rule in W and the proposition $\llbracket \neg p \rrbracket$ is a rule in $\llbracket q \rrbracket$. The agent's factual knowledge is given by $s_2 = \llbracket q \rrbracket$. Clearly, $\pi(s_2)$ cannot coherently be refined with $\llbracket p \rrbracket$. So, according to proposition 4.15, $\llbracket p \rrbracket$ does not apply to s_2. It is overridden by the more specific rule $\llbracket \neg p \rrbracket$, which does apply to s_2. The world $w_2 = \{q\}$ is not subjected to the general rule $\llbracket p \rrbracket$; it is a nonaccidental exception to it.

(iii). For this example eight possibilities must be taken into account. But apart from that, the frame π_3 is much like π_2: its only interesting features are that $\llbracket p \rrbracket$ is a rule in W and $\llbracket \neg p \rrbracket$ is a rule in $\llbracket q \rrbracket$. (It is left to the reader to draw the relevant pictures). The agent's factual knowledge is given by $s_3 = \llbracket q \wedge r \rrbracket$. When $\pi_3(\llbracket q \rrbracket)$ is refined with $\llbracket p \rrbracket$, the result is incoherent. Since $s_3 \subseteq \llbracket q \rrbracket \subseteq W$, it follows by proposition 4.15 that the general rule $\llbracket p \rrbracket$ does not apply to s_3. It is overridden by the more specific rule $\llbracket \neg p \rrbracket$, which does apply to s_3.

(iv). In definition 4.14 a three place relation is introduced: The *d*-rule *e* applies to *d'*. Most of the time we suppress the first argument, but sometimes we cannot. This becomes evident when we compare the third example with the fourth. We saw above that in σ_3 the W-rule $[\![p]\!]$ does not apply to $[\![q \wedge r]\!]$. Note however that the result of refining $\pi_3([\![q \wedge r]\!])$ with $[\![p]\!]$ is coherent. There is nothing wrong 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 general W-rule $[\![p]\!]$ does not apply to $[\![q \wedge r]\!]$. It is the $[\![q \wedge r]\!]$-rule $[\![p]\!]$ which does.

(v). We already encountered this example in the preceding section. There we saw that the expectations of an agent in state σ_5 are ambiguous: the real world must be an exception to one of the rules, but it is not clear to which one. This will remain so in the present set up, but now we can ask: what kind of exception is at stake here? By proposition 4.15, both the general rule $[\![p]\!]$ and the general rule $[\![q]\!]$ apply to $[\![\neg(p \wedge q)]\!]$. Hence the worlds in $[\![\neg(p \wedge q)]\!]$—and the real world is known to be like one of these—are all subjected to the rules in question. Therefore, the real world is to be considered as an accidental exception to (at least one of) the rules.

One might have expected a different conclusion here: either the rule $[\![p]\!]$ or the rule $[\![q]\!]$ does not apply to $[\![\neg(p \wedge q)]\!]$—after all the pattern $\pi_5([\![\neg(p \wedge q)]\!])$ cannot coherently be refined with both $[\![p]\!]$ and $[\![q]\!]$. However, given that it is very well possible for one rule to apply to a domain in which it is not satisfied, it shouldn't come as a surprise that two rules can both apply to a domain in which they are not jointly satisfied. (If a more concrete argument is wanted: we want that **1** [*normally* p] [*normally* q] [$\neg$p] $\Vdash$ *presumably* q, and we will get this mainly because in the state **1** [*normally* p] [*normally* q] [$\neg$p] one may safely assume that the rule $[\![q]\!]$ applies within $[\![\neg p]\!]$. But how could it ever be safe to assume that this general rule applies to these specific circumstances, if it is not safe to assume that it applies in the less specific circumstances $[\![\neg(p \wedge q)]\!]$?)

(vi). We will find that $\sigma_6 \Vdash$ *presumably* $\neg$q. In other words, a defeasible form of *Modus Tollens*: is valid$_3$.

$$q \rightsquigarrow \neg p, p \models_3 \textit{presumably}\ \neg q$$

Consider $s_6 = [\![p]\!]$. There we find the world $w_3 = \{p, q\}$. By definition 4.13, the $[\![q]\!]$-rule $[\![\neg p]\!]$ applies to the domain $\{w_3\}$, because the only domain d such that $\nu_\pi(d) \subseteq [\![q]\!] {\sim} [\![\neg p]\!]$ is $\{w_3\}$ itself and $\{w_3\} \subseteq [\![q]\!] {\sim} [\![\neg p]\!]$. Hence, in s_6, w_3 is subjected to the rule in question but does not conform to it. The other world in s_6 is $w_1 = \{p\}$. This world is not subjected to the $[\![q]\!]$-rule $[\![\neg p]\!]$. Actually, for all an agent in state σ_6 knows, this world is utterly normal. Therefore, someone in state σ_6 will expect the real world to be like w_1 rather than like w_3. And in w_1 the proposition $[\![q]\!]$ does not hold.

(vii). In state σ_7 the situation is completely different. Here, the $[\![q]\!]$-rule $[\![\neg p]\!]$ does not apply to $\{w_3\}$: we find that $\nu_\pi(\{w_1, w_3\}) = \{w_3\} \subseteq [\![q]\!] {\sim} [\![\neg p]\!]$, but of course $\{w_1, w_3\} \not\subseteq [\![q]\!] {\sim} [\![\neg p]\!] = \{w_3\}$. In fact, neither w_3 nor w_1 are subjected to the $[\![q]\!]$-rule $[\![\neg p]\!]$. On the other hand, both w_3 and w_1 are subjected to the $[\![p]\!]$-rule $[\![q]\!]$, and w_1 is an (accidental) exception to this rule, whereas w_3 conforms to it. So, in evaluating the situation an agent in state σ_7 must arrive at the conclusion that the real world will be like w_3 rather than like w_1. In other words, $\sigma_6 \Vdash$ *presumably* q, which means that

$$p \rightsquigarrow q, q \rightsquigarrow \neg p, p \models_3 \textit{presumably}\, q$$[1]

(Defeasible Modus Ponens beats Defeasible Modus Tollens).

I am getting ahead of my story. The two definitions I have implicitly been using in the last two examples are these:

4.18 DEFINITION
Let $\sigma = \langle \pi, s \rangle$ be an information state.
The *resulting pattern* ε_σ is the pattern on s given by:
$\langle i, j \rangle \in \varepsilon_\sigma$ iff $i, j \in s$, and j is an accidental exception to every rule to which i is an accidental exception.

It is easy to check that ε_σ is indeed a pattern.

4.19 DEFINITION
Let $\sigma = \langle \pi, s \rangle$ be an information state.
(i) i is *optimal in* σ iff $i \in s$ and $\langle i, j \rangle \in \varepsilon_\sigma$ for every $j \in s$ such that $\langle j, i \rangle \in \varepsilon_\sigma$;
(ii) $\mathbf{m}_\sigma$ is the set of all optimal worlds in σ;

[1] If a concrete example is wanted, take p = 'it snows', and q = 'the temperature is below 0^{o} C'

(iii) σ [*presumably* ϕ] is determined as follows:
- if $\mathbf{m}_\sigma \cap [\![\psi]\!] = \mathbf{m}_\sigma$, $\sigma[\phi] = \sigma$;
- otherwise, $\sigma[\phi] = \mathbf{0}$.

Definition 4.19 is a straightforward adaptation of the corresponding definition in § 3. As for 4.18, perhaps it is easier to understand the contrapositive of this definition. Let $\sigma = \langle \pi, s \rangle$ be an information state. Suppose $i, j \in s$. Then $\langle i, j \rangle \notin \varepsilon_\sigma$ iff there is a rule e in some domain d such that (i) i is subjected to e, but $i \notin e$; whereas (ii) j is not subjected to e, or $j \in e$. If $\langle i, j \rangle \notin \varepsilon_\sigma$, an agent in state σ has some reason to expect the real world to be like j rather than i.

Definition 4.18 can be restated without reference to any rule as follows.

4.20 PROPOSITION

Let $\sigma = \langle \pi, s \rangle$ be an information state. Suppose $i, j \in s$.
Then $\langle i, j \rangle \notin \varepsilon_\sigma$ iff there some $d \subseteq W$ such that
(i) $i \in d$, $i \notin \mathbf{n}\pi(d)$, and $\langle i, j \rangle \notin \pi(d)$;
(ii) for every $d' \supseteq s \cap d$ it holds that if there is some $k \in d'$ such that $\langle i, k \rangle \notin \pi(d)$, then there is some $k \in \nu_\pi(d')$ such that $\langle i, k \rangle \notin \pi(d)$.

PROOF

From left to right: Suppose there is some rule e that applies within s such that i is an accidental exception to e, but j is not an accidental exception to e. Assume that e is a rule in d. In this case we have $i \in d$, $i \notin \mathbf{n}\pi(d)$, and $\langle i, j \rangle \notin \pi(d)$. Furthermore, since e applies to $s \cap d$, we know that if $d' \supseteq s \cap d$, and there is some $k \in d'$ such that $k \notin d{\sim}e$, then there is some $k \in \nu_\pi(d')$ such that $k \notin d{\sim}e$. This can be rephrased as required by noting that $k \notin d{\sim}e$ iff $\langle i, k \rangle \notin \pi(d)$.
From right to left: Let d be such that (i) and (ii) are fulfilled. Set $e = \{l \in d \mid \langle i, l \rangle \notin \pi(d)\}$. The proposition e is a rule in $\pi(d)$. Note that $i \notin e$, whereas if $j \in d$, $j \in e$. So it suffices to show that e applies to $s \cap d$. This means that for any $d' \supseteq s \cap d$ the following must hold: if $\nu_\pi(d') \subseteq d{\sim}e$, then $d' \subseteq d{\sim}e$. That this is indeed the case follows immediately from (ii).

If you are willing to read 4.20 as a definition rather than as a proposition, you can forget everything said between 'The crucial notion here...' half-way down on page 31, and 'It is easy to check that ε_σ is indeed a pattern' on page 36. (But do not throw away the pictures you have drawn—you will need them again below).

4.21 EXAMPLES

(i) $\textit{normally}\,p,\ q \wedge r \models_3 \textit{presumably}\,p$;

(ii) $\textit{normally}\,p,\ q \rightsquigarrow \neg p,\ q \wedge r \models_3 \textit{presumably}\,\neg p$;

(iii) $\textit{normally}\,p,\ q \rightsquigarrow \neg p,\ q \wedge r \rightsquigarrow p,\ q \wedge r \models_3 \textit{presumably}\,p$;

(iv) $\textit{normally}\,p,\ q \rightsquigarrow \neg p,\ q \wedge r \rightsquigarrow p,\ q \models_3 \textit{presumably}(\neg p \wedge \neg r)$;

(v) $\textit{normally}\,p,\ q \rightsquigarrow \neg p,\ q \wedge r \rightsquigarrow p \models_3 \textit{presumably}(p \wedge \neg q)$.

In checking these examples, you will discover that finding the optimal worlds in a state σ is in many cases much less tedious and time consuming than the definitions suggest. There are many shortcuts. Here are some that will always be useful:

Let $\sigma = \langle \pi, s \rangle$. Suppose that $i,j \in s$.

Each of the following conditions implies that $\langle i,j \rangle \in \varepsilon_\sigma$:

- For every $d \subseteq W$ such that $i \in d$, $i \in \mathbf{n}\pi(d)$.
- For every $d \subseteq W$ such that $i \in d$ and $i \notin \mathbf{n}\pi(d)$, it holds that $\langle i,j \rangle \in \pi(d)$.

Each of the following conditions implies that $\langle i,j \rangle \notin \varepsilon_\sigma$:

- There is some $d \subseteq s$ such that $i,j \in d$, $\langle i,j \rangle \notin \pi(d)$;
- There is some $d \subseteq s$ such that $i \in d$, $i \notin \mathbf{n}\pi(d)$, and $j \notin d$;
- There is some $d \subseteq W$ such that $i \in d$, $i \notin \mathbf{n}\pi(d)$ and $\langle i,j \rangle \notin \pi(d)$, while $j \in \mathbf{n}\pi(d')$ for every d' such that $j \in d'$.

The examples in 4.21 illustrate two important characteristics of the system. As the first three examples show, it is very well possible for a frame π' to be a refinement of the frame π without the resulting pattern $\varepsilon_{\langle \pi', s \rangle}$ of the state $\langle \pi', s \rangle$ being a refinement of the resulting pattern $\varepsilon_{\langle \pi, s \rangle}$ of the state $\langle \pi, s \rangle$. This happens in particular when in getting from the state $\langle \pi, s \rangle$ to the state $\langle \pi', s \rangle$ an exceptive clause to one of the rules is learnt—so that this rule, which was applicable in $\langle \pi, s \rangle$, is not applicable any more. Still—and this is illustrated by the fourth and the fifth example—exceptive clauses pertain to *exceptional* circumstances: as long as we do not know that these circumstances obtain, we will expect them not to obtain.

Before proceeding we must not forget to check that the three notions of validity are language independent. I only state the relevant lemma. The relevant proposition is a copy of proposition 2.6(i).

4.22 LEMMA

Suppose $\mathcal{A} \subseteq \mathcal{A}'$; let W be the powerset of $\mathcal{A}$ and W' be the powerset of $\mathcal{A}'$.

g is the function from *P*(W) into *P*(W') that for every $e \subseteq W$ is given by

$$\mathbf{g}(e) = \{ j \subseteq \mathcal{A}' \mid j \cap \mathcal{A} \in e \};$$

h is the function from *P*(W') into *P*(W) that for every $e \subseteq$ W'is given by

$$\mathbf{h}(e) = \{ i \subseteq \mathcal{A} \mid i = j \cap \mathcal{A} \text{ for some } j \in e \}.$$

Let $\sigma = \langle \pi, s \rangle$ be any $\mathcal{A}$-information state. We associate an $\mathcal{A}'$-information state $\sigma^* = \langle \pi^*, s^* \rangle$ with σ as follows:

$s^* = \mathbf{g}(s)$;

π^* is the frame on W' that for every $d \subseteq$ W' is given by:

$$\langle i,j \rangle \in \pi^*(d) \text{ iff } i,j \in d \text{ and } \langle i \cap \mathcal{A}, j \cap \mathcal{A} \rangle \in \pi(\mathbf{h}(d)).$$

Conversely, let $\sigma = \langle \pi, s \rangle$ be any $\mathcal{A}'$-information state. We associate an $\mathcal{A}$-information state $\sigma^\circ = \langle \pi^\circ, s^\circ \rangle$ with σ as follows:

$s^\circ = \mathbf{h}(s)$;

π° is the frame based on W that for every $d \subseteq$ W is given by:

$\langle i,j \rangle \in \pi^\circ(d)$ iff $i,j \in d$, and either (a) $i=j$; or (b) for every $k, l \in \mathbf{g}(d)$ such that $i = k \cap \mathcal{A}$, $j = l \cap \mathcal{A}$, it holds that $\langle k, l \rangle \in \pi(\mathbf{g}(d))$.

Now, for every ϕ of $L_3^{\mathcal{A}}$ the following holds:

ia) if σ is an $\mathcal{A}$-state, then $\sigma[\phi]^* = \sigma^*[\phi]$;

ib) if σ, τ are $\mathcal{A}$-states and $\sigma \neq \tau$, then $\sigma^* \neq \tau^*$;

iia) if σ is an $\mathcal{A}'$-state, then $\sigma[\phi]^\circ = \sigma^\circ[\phi]$;

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

§ 5 INHERITANCE

So far, we have been thinking of the language $L_3^{\mathcal{A}}$ as a propositional language but we can easily give a predicate logical interpretation to it. Think of p, q, and r etc. as atomic predicates rather than atomic sentences. Each such predicate specifies a property and each well-formed expression of $L_0^{\mathcal{A}}$ specifies a boolean combination of properties. Think of W as the set of possible *objects* rather than as the set of possible worlds. A possible object i has the property expressed by 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 object — and think of real objects as tokens: 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 figuring in an information state $\langle \pi, s \rangle$ represents the agent's knowledge, only this time it is not the agent's knowledge about the real world, but about *some* real object. With a formula ϕ of L_0^{λ} it is learnt that this object (which is not explicitly mentioned in ϕ) has a certain property, and all possible objects that do not have this property are removed from s.

A rule e in a pattern $\pi(d)$ now is a property — 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), a formula of the form '$\phi \rightsquigarrow \psi$' can be read as '$\phi$-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 $\nu\pi(d) \neq \emptyset$, this just means that it must be conceivable that there is an object to which all the d-rules apply and which has all the properties it accordingly should 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 $\nu\pi(d)$ 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[1].

[1] For the system to work, it is essential that $n\pi(d) \neq \emptyset$ for every d. But it is not essential that $\nu\pi(d) \neq \emptyset$ for every d. So there is some room for discussion here, at least if by 'Birds normally aren't normal' it is meant that birds normally are so special that at least one of the rules that hold for birds in general does not apply. Within the limits of this paper, I cannot discuss any of the possible alternatives for the coherence constraint in detail. But let me mention the weakest constraint one needs to get things off the ground. It is the following 'noncumulative' version of 4.5 (ii):

The next table should be understood as follows: Let (i, j) be the cell where the i-th column and the j-th row intersect. If i, j > 1, you find in (i, j) some conclusions that can be drawn from the sentences in (1, j) and (i, 1). Read 'x is P' as 'x is adult', 'x is Q' as 'x is student', and 'x is R' as 'x is employed'.

	x is P	x is Q	x is P and Q
P's normally are R	Presumably x is R		Presumably x is R
P's normally are R Q's normally are not R	Presumably x is R Presumably x is not Q	Presumably x is not R Presumably x is not P	
P's normally are R Q's normally are P	Presumably x is R	Presumably x is R Presumably x is P	Presumably x is R
P's normally are R Q's normally are not R Q's normally are P	Presumably x is R Presumably x is not Q	Presumably x is not R Presumably x is P	Presumably x is not R

I will discuss only a few of these examples in detail, leaving the remaining ones as an exercise to the reader.

One of the arguments whose premises meet at (3, 4) is interesting for several reasons:

Q's normally are P, P's normally are R, x is Q $\models_3$ Presumably, x is R (*)

In our propositional formalism this argument looks like this:

$$q \rightsquigarrow p,\ p \rightsquigarrow r,\ q \models_3 \textit{presumably}\, r$$

To see that this is indeed the case, we have to determine the state

$$\mathbf{1}[q \rightsquigarrow p]\,[p \rightsquigarrow r]\,[q] = \sigma = \langle \pi, s \rangle$$

Weak Coherence: If $d' \subseteq d$ and e is a rule in $\pi(d')$, then there exists a coherent refinement ε of $\pi(d)$ such that e is rule in $\varepsilon \mid d'$.

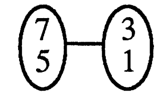
Of course, if the coherence constraint changes, everything else does in its train. In this particular case, however, all you have to do is to substitute '$\mathbf{n}\pi$' for 'v_π' in 4.9, 4.10, 4.14, and 4.20. Having done so, it is easy to see that for all but one of the examples discussed so far, the 'weak' theory and the 'strong' one yield the same results, the one difference being that on the weak account $\mathbf{1}[\textit{normally}\, p]\,[q \rightsquigarrow \neg p]\,[\neg q \rightsquigarrow \neg p] \neq \mathbf{0}$. (I think this suffices to show that this weak coherence constraint is too weak. However, there are some interesting options *between* the weak and the strong one—there is one, for example, according to which $\mathbf{1}[\textit{normally}\, p]\,[q \rightsquigarrow \neg p]\,[\neg q \rightsquigarrow \neg p] = \mathbf{0}$, but $\mathbf{1}[\textit{normally}\, p]\,[\textit{normally}\, q]\,[r \rightsquigarrow \neg p]\,[\neg r \rightsquigarrow \neg q]] \neq \mathbf{0}$).

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

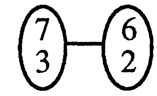
We are dealing with a set $W = \{w_0,..., w_7\}$ of eight possible types of object described in the table on the left.

$s = \{2, 3, 6, 7\}$;
π is the following frame:
If $d \neq \{1, 3, 5, 7\}$ and $d \neq \{2, 3, 6, 7\}$, $\pi(d) = d \times d$;

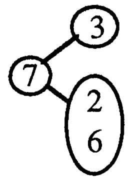
The pattern $\pi(\{1, 3, 5, 7\})$ looks like this:



And the pattern $\pi(\{2, 3, 6, 7\})$ is this:



The resulting pattern ε_σ turns out to be:



Hence, $\mathbf{m}_\sigma = \{7\}$, which means that $\sigma \Vdash$ *presumably* r.

Consider the rules ‘Students are normally adults’ and ‘Adults are normally employed’. Suppose these rules are the only rules you are acquainted with. Given the above, it is correct then 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's normally are R, P's normally are R $\nVdash_3$ Q's normally are R (**)

The *Hypothetical Syllogism* is not valid$_3$; we only have a defeasible version of it, exemplified by (*). The conclusion of (*) is unstable. It will be defeated when for example you learn that students normally are not employed. The conclusion of (**), on te other hand, is not defeasible. Rules are stable. If you were to draw the conclusion that students normally are employed, you cannot later accept that they normally are not.[1]

[1] Within our formalism we cannot handle an argument like this:

Q's normally are P, P's normally are R. So, presumably, Q's normally are R

But I am ready to admit that in an extended version of the theory this should come out valid$_3$.

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

$$\phi_1 \rightsquigarrow \psi_1,\ldots,\phi_n \rightsquigarrow \psi_n, \chi \models_3 \textit{presumably}\, \theta,$$

whereas

$$\phi_1 \rightsquigarrow \psi_1,\ldots,\phi_n \rightsquigarrow \psi_n \not\models_3 \chi \rightsquigarrow \theta.$$

In most of these cases

$$\mathbf{1}[\phi_1 \rightsquigarrow \psi_1]\ldots[\phi_n \rightsquigarrow \psi_n]\,[\chi \rightsquigarrow \neg\theta] \neq \mathbf{0},$$

so that

$$\phi_1 \rightsquigarrow \psi_1,\ldots,\phi_n \rightsquigarrow \psi_n, \chi \rightsquigarrow \neg\theta, \chi \models_3 \textit{presumably}\, \neg\theta.$$

For example, we have a defeasible form of *Modus Tollens*:

$$p \rightsquigarrow q,\ \neg q \models_3 \textit{presumably}\, \neg p,$$

but *Contraposition* fails:

$$p \rightsquigarrow q \not\models_3 \neg q \rightsquigarrow \neg p.$$

We also saw that

$$p \rightsquigarrow q,\ p \wedge r \models_3 \textit{presumably}\, q,$$

but *Strengthening the Antecedent* is not allowed:

$$p \rightsquigarrow q \not\models_3 (p \wedge r) \rightsquigarrow q.$$

Perhaps this is the right place to remind you that the logic generated by the third validity notion is not closed under substitution. The argument under (*) is only valid for predicates that are independent—or at least not known to be dependent. If we substitute 'not Q' for 'R', and determine the optimal worlds in the resulting state, we find

Q's normally are P, P's normally are not Q, x is Q $\not\models_3$ Presumably, x is not Q

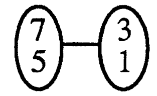
From 'Students normally are adult' and 'Adults normally aren't student' and 'x is a student' it does not follow that x is presumably not a student (but it does follow that John is presumably an adult (Recall example 4.17(vii)).

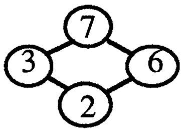
Let's now have a closer look at what happens if you learn on top of the premises of (*) that students normally are not employed. Then we get the argument whose premises meet at the conclusion in (3,5). In our propositional formalism it looks like this:

$$q \rightsquigarrow p,\ p \rightsquigarrow r,\ q \rightsquigarrow \neg r,\ q \models_3 \textit{presumably}\, \neg r.$$

We have to determine the state $\mathbf{1}[q \rightsquigarrow p]\,[p \rightsquigarrow r]\,[q \rightsquigarrow \neg r]\,[q] = \sigma = \langle \pi, s \rangle$.

The set $s = \{2,3,6,7\}$, just like above. As for the frame π, we find: if $d \neq \{1,3,5,7\}$ and $d \neq \{2,3,6,7\}$, then $\pi(d) = d \times d$.

$\pi(\{1,3,5,7\})$ can be depicted as: 

And this is $\pi(\{2,3,6,7\})$: 

The resulting pattern ε_σ is identical to $\pi(\{2,3,6,7\})$; the $[\![p]\!]$-rule $[\![r]\!]$ does not apply within $[\![q]\!]$. So, $\mathbf{m}_\sigma = \{3\}$, which means that $\sigma \Vdash$ *presumably* $\neg r$.

Readers acquainted with the possible worlds semantics of conditionals developed in the early seventies by Stalnaker, Thomason and Lewis, or with the theories of defaults of Delgrande[1988] and Asher&Morreau[1990], will have wondered why I did not choose selection functions[1] 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.
Indeed, there is a simpler version of the present theory in which selection functions instead of expectation frames are used as one of the basic components in an information state. In many cases this simpler version works just as well as the present version. Actually, so long as we restrict ourselves to cases in which for each domain at most one (non trivial) rule has to be taken into account, both versions amount to the same thing. Even *Defeasible Modus Tollens*, an inference principle that neither the theory of Delgrande[1988] nor the theory of Asher&Morreau[1990] account for, is valid$_3$ on this simplified account. But the next two arguments are not:

	Students normally are adult
Students normally are adult	Students normally are not employed
Students normally are not employed	Adults normally are employed
x is a student	Adults normally know how to drive a car
x is employed	x is a student
∴ Presumably, x is an adult	∴ Presumably, x knows how to drive a car

These are instances of a principle that is sometimes called the principle of *Graded Normality*: If an object is exceptional in one respect, this does not

[1] A selection function is a function ν that assigns to each subset d of W, a subset $\nu(d)$ of d. Intuitively, $\nu(d)$ contains the normal elements of d.

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, x is an accidental exception to the rule that students are not employed. So, x is not a normal student—that is, not entirely normal. However, this is no reason to think that the rule that students normally are adults does not apply. You may still presume that x is an adult. As for the example on the right, a formal analysis reveals that the optimal x — the x that conforms to as many (applicable) standards of normality as possible—is an adult who is a nonaccidental exception to the rule that adults are employed, but who knows how to drive a car anyway.

The principle of *Graded Normality* owes its validity$_3$ largely to the fact that for every domain d, $\pi(d)$ can be more than just a bipartition of d in normal and abnormal d-elements. 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. They are too simple.

The expressive power of our propositional formalism is limited. However, it is sufficiently rich to express everything that can be expressed in a (non-monotonic) semantic network. In fact, the theory presented in this paper supplies a semantics for multiple inheritance networks in which cyclic paths and complex predicates are allowed. As such it could perhaps help to narrow the gap between theory and practice in the field of knowledge representation.[1]

Let me try to be more precise. The theory presented in this paper yields a decidable non-monotonic notion of logical consequence, viz. validity$_3$, comparable to the 'support'-relation in inheritance theory. This notion can be used as a basis for answering questions of soundness and completeness: Given an inference algorithm for a suitable[2] class of nets, is it the case that a net Γ

[1] As Thomason & Horty[1988] point out, the market is flooded with inference algorithms for all kinds of semantic networks, but most of these lack a model-theoretic interpretation.

[2] Here 'suitable' means 'everything that can be said in the net language, can be said in the language L_3^{λ}. I am going to be rather sloppy in distinguishing between the two.

belonging to this class supports a conclusion ϕ if and only if it is valid$_3$ to infer *presumably* ϕ from the rules and the facts that make up Γ?

For all inference 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]. Still, from our point of view, this algorithm is not complete. For one thing, *Defeasible Modus Tollens* is valid$_3$, but the net representing the premises 'P's normally are Q', and 'x is not Q' does not support the statement 'x is not P'. Another example can be found in the table on page 41. The argument

P's normally are R, Q's normally are not R, x is P/∴ Presumably, x is not Q

is valid$_3$—and rightly so—but, again, the net that represents the premises of this argument does not support the conclusion.

If we turn to more complicated cases, it appears that the algorithm of Horty *cum suis* is not sound either. If it were, the next argument, which exemplifies the case of cascaded ambiguities would be valid$_3$, but it is not.[1]

Quakers normally are pacifist

Republicans normally are not pacifist

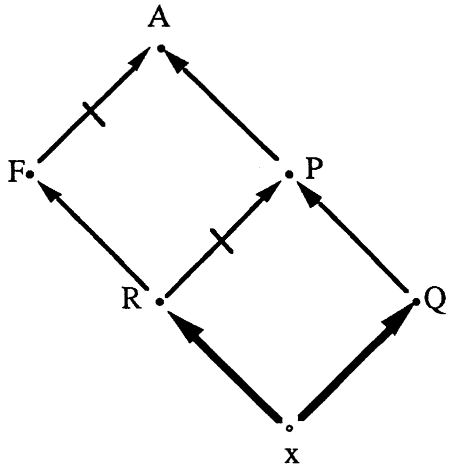
Pacifists normally are anti-military

Republicans normally are footballfan

Foottballfans normally are not antimilitary

x is a quaker and a republican

∴ Presumably, x is not antimilitary



In our propositional formalism this argument has the form

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

[1] From the discussion in Touretzky & Horty & Thomason[1987] it is clear that the authors are not particularly happy with the result that the conclusion 'x is antimilitary' is supported by the net depicted above. However, they present the case as one of the horns in a dilemma. They seem to think that if the next argument is not ambiguous:

$$p \rightsquigarrow q,\ p \rightsquigarrow \neg r,\ q \rightsquigarrow s,\ s \rightsquigarrow \neg t,\ q \rightsquigarrow r,\ r \rightsquigarrow t,\ p \,/\therefore\ \textit{presumably}\ \neg t.$$

the case of the cascaded ambiguities cannot be unambiguous, either. But they are mistaken here. You can have one without the other. The argument above *is* valid$_3$ (mainly because the rule $q \rightsquigarrow r$ does not apply within p).

It is not difficult to see that all the rules given with the premises apply within $[\![q \wedge r]\!]$, so the resulting pattern is highly ambiguous. There turn out to be four optimal objects: $\{q,r,f\}$, $\{q,r,f,p\}$, $\{q,r,p,a\}$ and $\{q,r,f,p,a\}$. So, it is neither valid$_3$ to expect that x is anti-military, nor that x is not anti-military.
A more detailed comparison of the theory of inheritance of Horty *et al.* with the theory offered here will appear in a in a sequel to this paper in which I hope to present an inference algorithm for nonmonotonic nets that is sound and complete with respect to the notion of validity$_3$.

Let $\psi_1,\ldots,\psi_n$, and ϕ be a sentences of $L_3^{\rightsquigarrow}$ in which *presumably* does not occur. From proposition 4.13 it follows almost immediately that

$$\psi_1,\ldots,\psi_n \models_1 \phi \text{ iff } \psi_1,\ldots,\psi_n \models_2 \phi \text{ iff } \psi_1,\ldots,\psi_n \models_3 \phi.$$

In other words, when we are studying the validity and invalidity of arguments in which *presumably* does not occur we can omit the subscripts 1, 2, and 3. It is this kind of arguments that we now turn to.

We saw above that well known principles like the *Hypothetical Syllogism*, *Contraposition* and *Strengthening the Antecedent* fail for the default arrow $\rightsquigarrow$. So, naturally the question arises which principles of implication do hold for $\rightsquigarrow$. If default implication is not strict implication, as the failure of principles like the *Hypothetical Syllogism* shows, is it then perhaps it is a kind of *variable strict implication*?
If it were, the next five principles, which give a complete characterization of the logical interplay of any variable strict implication with the classical connectives, would hold.

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

It turns out, however, that only the first two of these principles, CI and CC, are valid. The remaining three are *almost* valid. We have for instance the following:

Let σ be any information state. If $\llbracket\phi\rrbracket \neq \emptyset$, then

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

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:

$$\Delta,\ \phi \rightsquigarrow \psi,\ \phi \ \models_3 \mathit{presumably}(\psi \vee \chi);$$
$$\Delta,\ \phi \rightsquigarrow \psi,\ \phi \rightsquigarrow \chi,\ \phi \wedge \psi \ \models_3 \mathit{presumably}\,\chi;$$
$$\Delta,\ \phi \rightsquigarrow \chi,\ \psi \rightsquigarrow \chi,\ \phi \vee \psi \ \models_3 \mathit{presumably}\,\chi.$$

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

$$\Delta,\ \phi \rightsquigarrow \psi,\ \psi \rightsquigarrow \chi,\ \phi \ \not\models_3 \mathit{presumably}\,\chi.$$

For ASC and AD we can prove something even stronger: Let π be any frame. Then the following holds:

(a) If $\llbracket\psi\rrbracket$ and $\llbracket\chi\rrbracket$ are rules in $\pi(\llbracket\phi\rrbracket)$, then $\llbracket\chi\rrbracket$ is a rule in $\varepsilon_{<\pi, \llbracket\phi \wedge \psi\rrbracket>}$;

(b) If $\llbracket\chi\rrbracket$ is a rule in $\pi(\llbracket\phi\rrbracket)$ and in $\pi(\llbracket\psi\rrbracket)$, then $\llbracket\chi\rrbracket$ is a rule in $\varepsilon_{<\pi, \llbracket\phi \vee \psi\rrbracket>}$.

Unfortunately, even this cannot compensate for the failure of AD and ASC. It would be nicer if these argument forms were valid, the more so because (a) shows that if $\sigma \Vdash \phi \rightsquigarrow \psi$ and $\sigma \Vdash \phi \rightsquigarrow \chi$, an agent in state σ will at least have to accept $\llbracket\chi\rrbracket$ as an *implicit* rule in the domain $\llbracket\phi \wedge \psi\rrbracket$. And a similar remark can be made with refrence to (b) and AD).

Only for the case of *Weakening the Consequent* I have an argument showing that something would be wrong if this principle were valid. Indeed, it is perfectly alright that

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

Here, I can almost repeat what I wrote near the end of the previous section when we discussed the logical properties of *normally*: 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.

Intuitively, the latter argument is valid$_3$, but the former is not— and the theory confirms this. However, it is difficult to see how this could be so if from 'Tigers normally have four legs' it would follow that tigers normally have four or five legs.

SIDE REMARK

I hope that the ideas underlying this paper can be of help not only to logicians interested in defaults, but also to linguists interested in the semantics of generic sentences. I realize, however, that what I offer here is at best one missing piece in a giant puzzle—nobody knows how many more 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. For one thing, 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 logians, 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

(iii)" Tigers are extinct

And the sentence

(iii)"' Tigers eat people

would imply

(i)"' Tigers normally eat people

And what to think of

(iv)"" A tiger is available

Whatever this means, it's 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 of a long list of problems that surround generic sentences[1]. 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 such a reading amounts to.

ACKNOWLEDGEMENT

Earlier versions of this paper were presented at the workshop on Non-standard Logics for Automated Reasoning (Rocamadour, October 1988), the conference on Semantic Networks and Nonmonotonic Reasoning (Tübingen, December 1988), the Summerschool on Natural Language Processing, Logic & Knowledge Representation (Groningen, June 1989), the Conference on Theories of Partial Information

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

(Austin, January 1990), the Kleene Symposium (Varna, June 1990), and the Third Symposium on Logic and Language (Réfvülöp, August 1990). I want to thank the people present at these occasions for their interest in my ideas at a time that these had hardly (Rocamadour), or not yet fully (Réfvülöp) taken shape.
An earlier draft of this paper was published in Report 2.5.A of the ESPRIT Basic Research Action 3175, DYANA.(Edinburgh, 1990). I have made some substantial changes in sections 4 and 5. Questions and comments by Juan Acero, Marc Moens, Michael Morreau, Lászl6 Pólos, Hans Rott and Mark Ryan have led to many improvements in presentation. I am also much indebted to my colleagues in Amsterdam, in particular to Johan van Benthem, Jeroen Groenendijk en Martin Stokhof, for their support.

REFERENCES

Asher, N. and M. Morreau: 1990, 'Commonsense Entailment: A Modal Theory of Nonmonotonic Reasoning', in J. van Eijck (ed.), *Logics in AI*, Proceedings of JELIA '90, Springer Lecture Notes in Computer Science **478**, 1-30.

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, *A Dynamic Analysis of Paradoxical Sentences*, Master thesis, Department of Philosophy, University of Amsterdam.

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*, technical report CMU-CS-87-175, Carnegie Mellon University, Computer Science Department, 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.

Makinson, D.: 1989, 'General Theory of Cumulative Inference', in *Proceedings of the Second International Workshop on Non-Monotonic Reasoning*, Springer Lecture Notes on Computer Science.

Stalnaker, R.: 1976, 'Possible worlds', in *Nous* **10**, 65-75.

Stalnaker, R.: 1979, 'Assertion', in P. Cole (ed.), *Syntax and Semantics 9 — Pragmatics*, Academic Press, New York, 315-332.

Thomason, R., and J. Horty, 1988, *Logics for Inheritance*, manuscript, University of Pittsburgh, 18 pp.

Touretzky, D., J. Horty, and R. Thomason, 1987, 'A Clash of Intuitions: The Current State of Nonmonotonic Multiple Inheritance Systems', in *Proceedings of IJCAI- 87*, Morgan Kaufman, Los Altos, 476-482.

van Benthem, J.F.A.K.: 1989, 'Semantic Parallels in Natural Language and Computation', in H. Ebbinghaus et al. (eds.) *Logic Colloquium '87*, Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 331-375.

The ITLI Prepublication Series

1990 *Logic, Semantics and Philosophy of Language*

LP-90-01	Jaap van der Does	A Generalized Quantifier Logic for Naked Infinitives
LP-90-02	Jeroen Groenendijk, Martin Stokhof	Dynamic Montague Grammar
LP-90-03	Renate Bartsch	Concept Formation and Concept Composition
LP-90-04	Aarne Ranta	Intuitionistic Categorial Grammar
LP-90-05	Patrick Blackburn	Nominal Tense Logic
LP-90-06	Gennaro Chierchia	The Variablity of Impersonal Subjects
LP-90-07	Gennaro Chierchia	Anaphora and Dynamic Logic
LP-90-08	Herman Hendriks	Flexible Montague Grammar
LP-90-09	Paul Dekker	The Scope of Negation in Discourse, towards a flexible dynamic Montague grammar
LP-90-10	Theo M.V. Janssen	Models for Discourse Markers
LP-90-11	Johan van Benthem	General Dynamics
LP-90-12	Serge Lapierre	A Functional Partial Semantics for Intensional Logic
LP-90-13	Zhisheng Huang	Logics for Belief Dependence
LP-90-14	Jeroen Groenendijk, Martin Stokhof	Two Theories of Dynamic Semantics
LP-90-15	Maarten de Rijke	The Modal Logic of Inequality
LP-90-16	Zhisheng Huang, Karen Kwast	Awareness, Negation and Logical Omniscience
LP-90-17	Paul Dekker	Existential Disclosure, Implicit Arguments in Dynamic Semantics

Mathematical Logic and Foundations

ML-90-01	Harold Schellinx	Isomorphisms and Non-Isomorphisms of Graph Models
ML-90-02	Jaap van Oosten	A Semantical Proof of De Jongh's Theorem
ML-90-03	Yde Venema	Relational Games
ML-90-04	Maarten de Rijke	Unary Interpretability Logic
ML-90-05	Domenico Zambella	Sequences with Simple Initial Segments
ML-90-06	Jaap van Oosten	Extension of Lifschitz' Realizability to Higher Order Arithmetic, and a Solution to a Problem of F. Richman
ML-90-07	Maarten de Rijke	A Note on the Interpretability Logic of Finitely Axiomatized Theories
ML-90-08	Harold Schellinx	Some Syntactical Observations on Linear Logic
ML-90-09	Dick de Jongh, Duccio Pianigiani	Solution of a Problem of David Guaspari
ML-90-10	Michiel van Lambalgen	Randomness in Set Theory
ML-90-11	Paul C. Gilmore	The Consistency of an Extended NaDSet

Computation and Complexity Theory

CT-90-01	John Tromp, Peter van Emde Boas	Associative Storage Modification Machines
CT-90-02	Sieger van Denneheuvel, Gerard R. Renardel de Lavalette	A Normal Form for PCSJ Expressions
CT-90-03	Ricard Gavaldà, Leen Torenvliet Osamu Watanabe, José L. Balcázar	Generalized Kolmogorov Complexity in Relativized Separations
CT-90-04	Harry Buhrman, Edith Spaan, Leen Torenvliet	Bounded Reductions
CT-90-05	Sieger van Denneheuvel, Karen Kwast	Efficient Normalization of Database and Constraint Expressions
CT-90-06	Michiel Smid, Peter van Emde Boas	Dynamic Data Structures on Multiple Storage Media, a Tutorial
CT-90-07	Kees Doets	Greatest Fixed Points of Logic Programs
CT-90-08	Fred de Geus, Ernest Rotterdam, Sieger van Denneheuvel, Peter van Emde Boas	Physiological Modelling using RL
CT-90-09	Roel de Vrijer	Unique Normal Forms for Combinatory Logic with Parallel Conditional, a case study in conditional rewriting

Other Prepublications

X-90-01	A.S. Troelstra	Remarks on Intuitionism and the Philosophy of Mathematics, Revised Version
X-90-02	Maarten de Rijke	Some Chapters on Interpretability Logic
X-90-03	L.D. Beklemishev	On the Complexity of Arithmetical Interpretations of Modal Formulae
X-90-04		Annual Report 1989
X-90-05	Valentin Shehtman	Derived Sets in Euclidean Spaces and Modal Logic
X-90-06	Valentin Goranko, Solomon Passy	Using the Universal Modality: Gains and Questions
X-90-07	V.Yu. Shavrukov	The Lindenbaum Fixed Point Algebra is Undecidable
X-90-08	L.D. Beklemishev	Provability Logics for Natural Turing Progressions of Arithmetical Theories
X-90-09	V.Yu. Shavrukov	On Rosser's Provability Predicate
X-90-10	Sieger van Denneheuvel, Peter van Emde Boas	An Overview of the Rule Language RL/1
X-90-11	Alessandra Carbone	Provable Fixed points in $I\Delta_0+\Omega_1$, revised version
X-90-12	Maarten de Rijke	Bi-Unary Interpretability Logic
X-90-13	K.N. Ignatiev	Dzhaparidze's Polymodal Logic: Arithmetical Completeness, Fixed Point Property, Craig's Property
X-90-14	L.A. Chagrova	Undecidable Problems in Correspondence Theory
X-90-15	A.S. Troelstra	Lectures on Linear Logic

1991 *Logic, Semantics and Philosophy of Langauge*

LP-91-01	Wiebe van der Hoek, Maarten de Rijke	Generalized Quantifiers and Modal Logic
LP-91-02	Frank Veltman	Defaults in Update Semantics

Mathematical Logic and Foundations

ML-91-01	Yde Venema	Cylindric Modal Logic
ML-91-02	Alessandro Berarducci, Rineke Verbrugge	On the Metamathematics of Weak Theories
ML-91-03	Domenico Zambella	On the Proofs of Arithmetical Completeness for Interpretability Logic
ML-91-04	Raymond Hoofman, Harold Schellinx	Collapsing Graph Models by Preorders
ML-91-05	A.S. Troelstra	History of Constructivism in the Twentieth Century
ML-91-06	Inge Bethke	Finite Type Structures within Combinatory Algebras

Computation and Complexity Theory

CT-91-01	Ming Li, Paul M.B. Vitanyi	Kolmogorov Complexity Arguments in Combinatorics
CT-91-02	Ming Li, John Tromp, Paul M.B. Vitanyi	How to Share Concurrent Wait-Free Variables
CT-91-03	Ming Li, Paul M.B. Vitanyi	Average Case Complexity under the Universal Distribution Equals Worst Case Complexity
CT-91-04	Sieger van Denneheuvel, Karen Kwast	Weak Equivalence
CT-91-05	Sieger van Denneheuvel, Karen Kwast	Weak Equivalence for Constraint Sets
CT-91-06	Edith Spaan	Census Techniques on Relativized Space Classes
CT-91-07	Karen L. Kwast	The Incomplete Database
CT-91-08	Kees Doets	Levationis Laus

Other Prepublications

X-91-01	Alexander Chagrov, Michael Zakharyaschev	The Disjunction Property of Intermediate Propositional Logics
X-91-02	Alexander Chagrov Michael Zakharyaschev	On the Undecidability of the Disjunction Property of Intermediate Propositional Logics
X-91-03	V. Yu. Shavrukov	Subalgebras of Diagonizable Algebras of Theories containing Arithmetic
X-91-04	K.N. Ignatiev	Partial Conservativity and Modal Logics
X-91-05	Johan van Benthem	Temporal Logic
X-91-06		Annual Report 1990
X-91-07	A.S. Troelstra	Lectures on Linear Logic, Errata and Supplement
X-91-08	Giorgie Dzhaparidze	Logic of Tolerance
X-91-09	L.D. Beklemishev	On Bimodal Provability Logics for Π_1-axiomatized Extensions of Arithmetical Theories
X-91-10	Michiel van Lambalgen	Independence, Randomness and the Axiom of Choice
X-91-11	Michael Zakharyaschev	Canonical Formulas for K4. Part I: Basic Results
X-91-12	Herman Hendriks	Flexibele Categoriale Syntaxis en Semantiek: de proefschriften van Frans Zwarts en Michael Moortgat