Centrum voor Wiskunde en Informatica

REPORTRAPPORT

Changing preferences

J. van Benthem, D.J.N. van Eijck, A. Frolova

Computer Science/Department of Software Technology

Report CS-R9310 February 1993

P.O. Box 4079, 1009 AB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

Changing Preferences

Johan van Benthem, Jan van Eijck and Alla Frolova

Institute for Logic, Language and Computation,
Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
and
Centre for Mathematics and Computer Science,
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Abstract

This is an exploratory document for a new research line in logical semantics which is emerging from several current developments in computer science. Standard logic employs 'flat' unstructured sets of statements for its theories and unstructured classes of models for its semantic universes. Nowadays, however, there is an incipient literature on structured universes of models as well as structured theories, both employing 'preference relations' of some sort. The purpose of this brief report is (1) to propose a more systematic framework for this trend, while also connecting it up with some historical predecessors, (2) to design some new logical systems bringing preferences out explicitly, thereby high-lighting the theoretical properties of 'preferential reasoning' while raising some new kinds of technical question for further research, and (3) to link up with some current computational ideas (emerging also in the field of linguistics), by bringing in the general dynamic logic of procedures as well as logical systems providing suitable fine-structure of information states.

1991 CR Categories: F.4.1, I.2.3, I.2.4
Keywords and Phrases: non-monotonic reasoning, preference logics, knowledge representation languages.

1 Preference Relations in Logic

1.1 Preference in Semantics

In standard logic, valid inference is defined according to Tarski's well-known schema expressing 'transmission of truth':

"Each model of the premises is also a model for the conclusion".

There are two 'flat' quantifications in this definition: one over 'all models' for the premises, and inside that, one over truth for 'all members' of the set of premises. A more flexible perspective arises here when models may be distinguishable as to their relevance, desirability or plausibility in a given style of inference. For instance, in Logic Programming, 'minimal Herbrand models' for a program are most desired, since they make do with (in a sense) the smallest number of objects and facts (cf. the requirements of 'no junk' and 'no confusion' for abstract data types). Likewise, in the method of 'circumscription' in Artificial Intelligence, models for the premises employing fewer objects (or tuples in their predicates) are usually preferred over larger ones. As a common generalization behind these, and other cases, general preference relations between models were introduced in Shoham 1988, allowing us to define various notions of 'plausible inference', on the format:

Report CS-R9310
ISSN 0169-118X
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

"Each *most preferred* model of the premises is a model for the conclusion".

Soon, it was noticed how this approach has its roots far back in the semantics of Conditional Logic (Lewis 1973, van Benthem 1989, Bell 1989), where the truth of a conditional " if A, then B" at a possible world w amounts to truth of its consequent in all worlds verifying in antecedent that are 'most similar' or 'most preferred' from the viewpoint of w. This simple general framework already demonstrates some logical key questions that arise in the 'preferential turn'. Notably, there are clear structural differences between standard reasoning and its preferential variant. For instance, the latter style of inference is neither monotonic nor transitive, because of the changes in minimal model classes for changing sets of premises. Nevertheless, a definite core calculus remains: e.g., it may be shown that Shoham's preferential inference validates precisely the non-iterated Horn-clause fragment of the 'minimal conditional logic' of Pollock (cf. Burgess 1981, Veltman 1985). More concrete preference relations may then generate additional valid principles (although, e.g., the usual 'submodel relation' of circumscription does not: witness van Benthem 1989). Also there turns out to be an interesting non-standard model theory over preferential universes (cf. Morreau 1985).

Even so, the picture of absolute binary preference relations $<$ between models has turned out inadequate for many applications, where the relevant ordering may be determined by various contextual factors (one person can be 'better' than another as a scientist, though not as a human being or a football player). Thus, preference is better viewed as being at least a ternary relation involving some further contextual parameters. In fact, there may be a whole family of relations $<_A$ depending on some class of models A (or even more complex contextual inputs). Indeed, as was noted above, conditional logic itself employs a ternary notion of comparison $C_x yz$, where $\lambda yz \cdot C_x yz$ is a preference relation $<_x$ depending on some 'vantage point' x. (For a concrete example, think of the geometrical relation $|x-y| < |x-z|$.) But the above relations $<_A$, too, have been considered in standard conditional logic (cf. the minimal conditional logics proposed in Chellas 1975).

1.2 Preference in Syntax

Preferential relations also arise in syntax. Again, the history of conditional logic provides examples. In its early syntactic phase, premises were sometimes organized into 'levels of entrenchment', with general 'laws' on top, and 'facts' and 'auxiliary assumptions' in various lower positions. In the evaluation of so-called counterfactual conditionals "if A, then B would have been the case", whose antecedent conflicts with the data, the latter would then have to override the lower levels of the premises in the so-called 'Ramsey Test':

> "Add the antecedent A to the premises and restore consistency 'with minimal effort' (retaining A): then see if the consequent B follows classically from the new revised theory".

Much more elaborate accounts of entrenchment may be found in Gärdenfors and a recent very promising account of structured theories is Ryan 1992, which explores their topology and general model theory, while integrating a number of earlier approaches. The syntactic perspective on preference has not received the attention so far that it deserves, even though standard Proof Theory would support this broader perspective just as well as Model Theory. For instance, in the wide- spread technique of natural deduction, it seems very natural to maintain a structured family of premises, and index the varying assumptions that one makes for insertion at various 'acceptance levels' in the latter.

But even model-theoretic preference relations may depend on the linguistic context of presentation for the relevant data. An example is the reconstruction of Van Fraassen's deontic logic for the purposes of semantic networks in Horty 1992, which (essentially) employs the following preference relation between models:

$$M <_\Sigma M' \quad \text{iff} \quad M \models \phi \Longrightarrow M' \models \phi \quad \text{for all} \quad \phi \in \Sigma.$$

It is easy to see that $<_\Sigma$ may change here along with our premises. Similar ideas occur in the literature on reasoning from inconsistent theories: some models may do better than others in capturing what the theory says in its current syntactic format (which determines, amongst others, what are to be counted as 'large' consistent subsets). Thus, semantic preference relations may be determined by syntactic form.

1.3 Predecessors and Parallels

What the above development suggests is that 'preference' is an interesting candidate for a general logical notion of some unifying power and technical potential. The art in extending the scope of logic is to bring in new primitives that reflect applications, without getting too close to the specific properties of any particular one. Preference seems to be an attractive channel for letting in some practical considerations, while still retaining significant general theory. This feeling is strengthened by the observation that the above development continues a number of themes in the earlier logical literature. For instance, the study of various task-dependent styles of inference dates back to the tradition of logical positivism in the formal philosophy of science, and early forms of circumscription may be discerned in Hempel's account of induction (cf. van Benthem 1989, as well as various references in Doyle 1985). Also, Popper's philosophical research program of 'verisimilitude' (cf. the anthology Kuipers 1987) is based on ternary preference relations of 'better approach to the truth' in both syntactic and semantic senses. Another independent source of 'preferences' as entities in their own right lies in economics and social choice theory (cf. Hahn & Hollis 1979). It may even be worth-while to explore this connection in more detail, given various earlier confluences between mathematical themes in these fields and logical semantics (cf. the articles on social choice and common knowledge in economics and computer science in Halpern 1986, or Gärdenfors and the emergence of semantic primitives).

1.4 Changing Preferences

Even so, it may be thought that 'preference' is an ephemeral phenomenon, merely reflecting diverse and unstable pressures from practice. Does this really deserve a position in the logical pantheon? Perhaps, no particular preference relation would qualify for canonization, but then, at least, the general mechanisms of their introduction and modification may be stable enough to warrant logical attention. ('Connections' may be much more stable than that which they connect.) Indeed, the most 'dynamic' and radical proposal in the field has been to let certain linguistic expressions explicitly trigger changing preferences, witness Spohn 1988, Selman & Kautz 1988, Veltman 1988, 1991, Doherty 1991. In one form or another, these authors all suggest that general information flow involves not just 'updating' but also 'upgrading' of information structures, modifying preferences rather than hard facts. In particular, conditional rules "if A, then B" need no longer be static statements to be tested in a fixed model for truth or falsity. They may also function as instructions that change the current model, telling us to increase our preference for A-worlds that satisfy B over those that are not. Of course, there are many different ways of implementing this idea, and hence there are many different forms of 'preferential dynamics'. These systems also serve as a natural point of entry for modern computational themes into preferential logic, creating some kind of dynamic logic of updating and upgrading.

1.5 Logical Architecture

The current literature in the field has become very diverse. Many different kinds of preference relation have been proposed, and many different styles of updating or inferencing with respect to them. (The monograph Sandewall 1992 gives an extensive analysis, as well as a confrontation with some key problems of temporal reasoning in Artificial Intelligence.) How can one best bring preferential reasoning within the scope of logical systems? There are many options here. For instance, one can take an existing formalism, and just replace its old 'batteries' by a preferential notion of consequence. One example is first-order predicate logic, with Tarskian consequence replaced by McCarthy's circumscription (cf. Lifschitz 1988). But we can also design new logical systems that may be better suited for displaying preferential phenomena. For instance, conditional logic has an explicit operator whose evaluation involves comparison of preferences, and the update logic for default reasoning in Veltman 1991 has operators that explicitly change preferences so as to make the corresponding conditional assertions true in the previous sense. Here are some general desiderata that we see on the design of useful systems of preferential reasoning.

• 'Preference' and 'minimization' (or maximization: the difference is immaterial) with respect to it are two important notions in semantics and reasoning, whatever their precise implementation. Nevertheless, there is no reason to expect that one uniform version will prevail. General cognitive mechanisms may show up in many different special forms, and it is precisely one of our human cognitive abilities that we can handle these simultaneously, and switch between them as the need arises. Thus, what we need are logical systems making various *options for preference* visible, and subject to *systematic manipulation.*

• Moreover, we wish to advocate a move which has been neglected in the literature on 'non-monotonic logic'. Given the richer semantic universe structured by preference relations, we can no longer assume that the standard languages of logic are adequate, and in particular, we need to ask if certain *new logical operations* should emerge in the new setting. For instance, it seems to make obvious sense to introduce modalities reflecting minimal preference, and perhaps even an explicit calculus of preference relations (just as one brings in a calculus of transition relations in Dynamic Logic).

• No deep ideological choices should be enshrined in these systems, but rather the co-existence of various styles of inference, enabling one to make the above smooth switches. In particular, both preferential and classical consequence should be retained, making it a task of the system to describe their *logical interactions.*

To achieve these goals, preference should be studied in a computational perspective straightaway. This requires us to pay attention to dynamics and the nature of transitions, and also to the fine-structure of information states eventually.

2 Dynamic Preference Logic

One can experiment with some simple logical update systems in order to bring out key features of preferential dynamics. These will allow us to use some techniques from dynamic logic that have already been developed for other purposes (cf. van Benthem 1991, de Rijke 1992, van Eijck & de Vries 1991). The first system is one of updates with respect to a fixed absolute preference relation. We analyze the resulting calculus via a translation into standard modal logic (with the usual consequences, such as decidability), first in standard K, then in bimodal $K + S5$. (An alternative analysis may be given using the update calculus of Van Eijck & De Vries 1991.) The next system introduce updates with domain-dependent preference relations, which may be analyzed in a similar fashion, this time using binary modal logic or conditional logic. Finally, we consider several versions of dynamic logic for explicit changing of preferences, one of them based on the 'arrow logic' of van Benthem 1992, Venema 1992 in which preferences are treated as objects in their own right.

2.1 Updating with Absolute Preference

2.1.1 The System

Here is a toy system illustrating the above general ideas on a universe of models or states ordered by one single relation of absolute preference. Take a propositional language as in eliminative update semantics (cf. Veltman 1988):

proposition letters	$p, q, \ldots$
logical operators	$\neg \wedge;\ \vee$

which is interpreted through functional instructions on sets of models (these stand for our 'information states'):

$$\begin{array}{lll} p[X] & := & X \cap [[p]] \quad \text{('atomic update with } p\text{')} \\ \phi \wedge \psi[X] & := & \phi[X] \cap \psi[X] \\ \phi; \psi[X] & := & \psi[\phi[X]] \\ \neg\phi[X] & := & X - \phi[X] \\ \phi \vee \psi[X] & := & \phi[X] \cup \psi[X] \end{array}$$

Now, add an atomic program minimizing the current state with respect to preference:

$$\mu[X] \quad := \quad \min_{<}[X] \qquad \text{(' } < \text{ --minimal part of } X\text{')}$$

2.1.2 Styles of Preferential Reasoning

This system gives us various notions of preferential inference. For instance, we can formulate an analogue of Shoham's notion of preferential consequence as follows:

Preferential Consequence $(P_1, \ldots, P_n \models_p C)$

$\forall X : (P_1; \ldots; P_n; \mu)[X] \subseteq C[X]$ holds in all preferential models.

Given this explication, various structural properties of this notion will be automatic. For instance, we do not have Monotonicity. The sequent $A, \mu \models_p C$ does not imply $A, B; \mu \models_p C$: since $\mu(A \cap B)$ is not necessarily included in $\mu(A)$. Likewise, Transitivity of inference fails. We do retain its 'cautious' version, however, stating that the sequents $X, A \models_p B$ and $X \models_p A$ imply $X \models_p B$. The reason lies again in general properties of the minimizer μ : $\mu(X \cap A) \subseteq B$, $\mu(X) \subseteq A$ imply $\mu(X) \subseteq B$, because $\mu(X) \cap A \subseteq \mu(X \cap A)$. The structural properties of this notion of inference will be a mixture of those for preferential reasoning (encoded in the minimal conditional logic) plus some additional effects of the sequential dynamics in the above interpretation procedure. For instance, the above style of inference no longer admits free permutation of premises if minimizations are included. On the other hand, unlike the original preferential consequence, $\models_p$ does allow monotonic insertion of arbitrary new premises, provided that this happens on the left.

But, there are also other possible options here for dynamic preferential inference. For instance, the above notion refers to arbitrary information states X, whereas Shoham's original notion only refers to the minimal models for the premises, being the result of computing the specific model class $(P_1; \ldots; P_n; \mu)[T]$. Also, the dynamic treatment of the conclusion might be changed: and another attractive style of inference would rather state that, after having processed all premises, followed by minimization, the resulting state $(P_1; \ldots; P_n; \mu)[X]$ should be tested for verification of the conclusion: i.e., further processing of C should have no effect:

Update – Test Inference

$C[(P_1; \ldots; P_n; \mu)[X]] = (P_1; \ldots; P_n; \mu)[X]$.

By way of illustration, we connect up the latter style with standard preferential consequence. Again, we find a mixture of principles from the minimal conditional logic with dynamic motives:

PROPOSITION The complete structural theory of Update-Test Inference may be axiomatized by

Unguarded Left Monotonicity	$X \Longrightarrow A$	$/ \; B, X \Longrightarrow A$
Guarded Right Monotonicity	$X \Longrightarrow A \quad X, Y \Longrightarrow B \; /$	$X, A, Y \Longrightarrow B$
Guarded Right Cut	$X \Longrightarrow A \quad X, A, Y \Longrightarrow B$	$/ \; X, Y \Longrightarrow B$

PROOF By a modification of the representation argument given in van Benthem 1991. Guarded Right Monotonicity expresses the functional character of our updates. □

Note that not even basic structural properties of the original preferential inference will be preserved here. Not just Permutation of arbitrary premises can be harmful, but so can Contraction of identical ones: the function $\mu; A; \mu$ is different from $\mu; A$ or $A; \mu$. Nevertheless, the precise counterparts of the original inferences do retain their old behaviour. The reason is that they inherit some very special semantic properties. In particular, we have for all purely 'descriptive' propositional formulas ϕ, that

$\phi[X] = X \cap [[\phi]]$ where $[[\phi]]$ is the standard denotation of ϕ computed as a set of models.

The key semantic properties guaranteeing this behaviour are the following two (cf. van Benthem 1988 on so-called 'classical dynamic operators'):

$\phi[X] \subseteq X$ Introspection

$\phi(\cup X_i) = \cup \phi(X_i)$ Continuity

All clauses for the propositional connectives given above preserve this combination. There are many interesting technical questions concerning the precise characterization of structural rules for preferential styles of inference (such as the above $\models_p$), as well as the additional effects of special semantic properties (e.g., precisely which logical operations preserve introspection and continuity?), which we must forego here.

Having an explicit minimization instruction gives us even additional flexibility. Here is one new application. Sandewall 1992 argues for 'non-homogeneous' notions of preferential reasoning, where some premises are treated classically and others minimally. In our framework, this can be expressed as a simple variation:

Mixed Classical and Preferential Consequence

$\forall X : (P_1; \ldots; P_i; \mu; P_{i+1}; \ldots; P_n)[X] \subseteq C[X]$ holds in all preferential models.

In this case, structural behaviour becomes still more delicate. There is Monotonicity with respect to the 'classical premises', but not with respect to the 'preferential' ones. Again, the properties of such a notion of inference can be studied via the logic of μ, whose complete axiomatization (assuming no special properties of $<$) may be found in van Benthem 1988, which proves a preferential representation theorem for the following principles:

FACT The complete propositional logic of minimization is given by the three laws

$\mu(A) \subset A$ $\mu(\cup A_i) \subseteq \cup_i \mu(A_i)$ $\cap_i \, \mu(A_i) \subseteq (\cup A_i)$.

Together with some obvious Boolean principles, these imply such useful 'algebraic laws' of minimization as the identity $\mu; \mu = \mu$.

2.1.3 Reduction to Modal Specifications

One can study the inferential properties of our system independently, in a special purpose dynamic logic. But one can also analyze them via their connections with the approprate static logical language over the same model class (this is the line taken in van Eijck & de Vries 1991, van Benthem 1991), often some species of modal logic. In the present case, this is the obvious language over preference universes (viewed as relational possible worlds models) with one unary operator

$[<]\phi$: "ϕ is true in all more $<$ –preferred worlds".

This language allows us to transcribe the above semantics directly. Here is the translation schema for 'next states' of the above update functions:

$$
\begin{array}{lcl}
NS(X,p) & := & X \wedge p \\
NS(X,\phi \wedge \psi) & := & NS(X,\phi) \wedge NS(X,\psi) \\
NS(X,\phi;\psi) & := & NS(NS(X,\phi),\psi) \\
NS(X,\neg\phi) & := & X \wedge \neg NS(X,\phi) \\
NS(X,\phi \vee \psi) & := & NS(X,\phi) \vee NS(X,\psi) \\
NS(X,\mu) & := & X \wedge [<]\neg X
\end{array}
$$

It is easy to show that this has the right behaviour (with 'X' viewed as a new proposition letter, and $[[\alpha]]$ the set of worlds in our total universe where α holds):

FACT. For all state sets X and formulas ϕ, $[[NS(X,\phi)]] = \phi[[X]]$.

2.1.4 Some Applications

Here are some consequences of the existence of this translation. E.g., we have this

FACT. Preferential consequence amounts to modal validity of the implication $NS(X, P_1; \ldots; P_n; \mu) \rightarrow NS(X, C)$.

For instance, non-monotonicity will now show up through the modal invalidity of the principle $NS(X, A; B; \mu) \rightarrow NS(X, A; \mu)$, that is,

$$X \wedge A \wedge B \wedge [<]\neg(X \wedge A \wedge B) \rightarrow X \wedge A \wedge [<]\neg(X \wedge A).$$

This modal logic may also be used to derive basic properties of the minimizer μ directly. Thus, the above law $\mu; \mu = \mu$ corresponds to the valid modal principle

$$X \wedge [<]\neg X \wedge [<]\neg(X \wedge [<]\neg X) + X \wedge [<]\neg X.$$

As the minimal modal logic is decidable, here is a useful

COROLLARY. Preferential consequence in the above dynamic system is decidable.

Similar modal analyses apply to other forms of dynamic consequence. Another route toward the same goal is a Hoare-style calculus as in van Eijck & De Vries 1991.

2.1.5 Epistemic Extensions

We can extend this system with new update constructions using additional modalities. A fair sample is the following 'epistemic test' (Veltman 1988):

$$
\begin{array}{lcll}
\textit{maybe}\ \phi\ [X] & := & X & \text{if } \phi[X] \neq \emptyset \\
 & & \emptyset & \text{if } \phi[X] = \emptyset
\end{array}
$$

This would get transcribed into a suitably extended modal logic as follows:

$$NS(X,\ \textit{maybe}\ \phi) \quad := \quad X \wedge \Diamond NS(X,\phi),$$

where $\Diamond$ is an additional general $S5$-modality stating that something holds at some world in the whole universe. The bimodal logic of $[<]$ and $\Diamond$ supports essentially the same static reduction as above. For further information on the behaviour of such bimodal (or polymodal) combinations, see De Rijke 1990, Spaan 1993.

2.2 *Updates with Domain-Dependent Preference*

2.2.1 *Setting Up the System*

Next, one can move from absolute preference to a domain-dependent version by treating μ no longer as an atomic program, but as a program *operator*. For instance, minimizing over the next state of a procedure, one would get

$$\mu\phi[X] \quad := \quad \min <_{\phi[X]} [\phi[X]]$$

But in applications, one may want to manipulate the next state and the context set independently. In that case, we can use an explicit dynamic 'mode' which takes static propositions (i.e., sets of worlds) to minimization operators:

$$\mu A[X] \quad := \quad \min <_A [X].$$

The absolute preferences of Shoham's original system may then be taken to refer to the total universe of states. That is, the earlier minimizer μ will behave like μT, whereby one obtains an obvious embedding of the earlier calculus.

In the second set-up, our formalism becomes a kind of propositional Dynamic Logic combining static and dynamic propositions, on the pattern of van Benthem 1991:

static	$\longrightarrow$	modes	$\longrightarrow$	dynamic
statements	$\longleftarrow$	projections	$\longleftarrow$	procedures

In particular, this would bring out explicit 'modes' turning static propositions into dynamic operators. Examples are the well-known 'test mode' $(p)?$ or the earlier atomic 'update mode' upd_p intersecting the current state X with some 'truth range' $[[p]]$, while the new operator μ serves as as a 'preferential minimization mode'. Going in the opposite direction, 'projections' will be operators taking dynamic procedures to static statements defining, say, their domains or their fixed points (cf. van Benthem 1991, De Rijke 1992 on this general architecture). This two-level system brings out various interesting themes. For instance, it allows us to display possible interactions between the minimizer μ and the Boolean structure of its argument A. (E.g., should $\mu(A \vee B)$ be the same operation as $\mu A \vee \mu B$?) What would be needed here is a more general account of changing preference relations over different context sets, a subject to which we shall return presently.

2.2.2 *Modal Reduction*

As before, a dynamic preference calculus like this can be reduced to computation of pre- and postconditions in a corresponding static calculus, enriched with suitable new modal operators. For the basic system, this will have to contain a binary modality

$[<_A]\phi$: "ϕ holds in all most $<_A$ -preferred worlds",

which allows us to formulate the appropriate next-state condition as follows:

$$NS(X, \mu A) \quad = \quad X \wedge [<_A]\neg X.$$

Alternatively, this may be seen as the embedding of our update system in a very weak system of 'conditional logic', whose decidability may be inferred from Chellas 1975. On top of this, the above propositional dynamic logic may be regarded as a uniform system performing these reductions 'internally', provided that we add the appropriate modalities to its 'static infra-structure'. (See Van Eijck & de Vries 1991, de Rijke 1992 for technical mechanisms to this effect.)

2.2.3 Context Comparison and Management

The above calculus provides the flexibility to process inferences in various ways, parametrizing preferences to specific context sets which governs the minimization of the set of premise models: whether the latter set itself, or some other contextually given one. This freedom is reminiscent of the way in which contextual dependence functions in natural language. The meaning of many linguistic expressions is context- dependent, and our linguistic ability consists in 'navigating' through their successive changes in reference. For instance, most predicates in natural language have a 'criterial' character, that is, their interpretation depends on comparative standards that depend on context (think of "tall", "good", but also of "blue", "blonde") P where this 'context' may typically change across an utterance or text. Likewise, in the preceding Sections, we have been describing, so to say, criterial *inferences* in varying contexts. Another example is the mechanism of changing 'context sets' for determiners in natural language quantification, as proposed in Westerståhl 1984. Also, the dynamics of changing contexts has recently become a research theme in Artificial intelligence (cf. Shoham 1991). And finally, changing contexts occur naturally in any practical use of preferences. Think of a job application, where one might argue as follows: "If the candidate is male, then prefer an unmarried one, while if she is female, then prefer an married one" P and even further down, toward tertiary preferences ("if a married man, then prefer an elderly one") and so on. In the processing of natural language, where preferential inference may be viewed as a paradigm of 'natural reasoning', two general mechanisms play a pervasive role :

CONTEXT MANAGEMENT. The shifting of relevant domains or contexts (this is only partially encoded in lexicon and syntax, and may depend on phonetic or other clues),

CONTEXT COMPARISON. The connections between preference statements across such contexts (like those expressed in the earlier calculus of indexed preferences $<_A$).

For an example of the latter, see the analysis of comparative relations as generated by contextual judgments in van Benthem 1983. (Comparative relations may be viewed as special preferences satisfying irreflexivity, transitivity, as well as 'almost-linearity': $\forall xyz : (x < y \rightarrow (x < z \vee z < y))$.) In the course of this derivation, various principles are proposed regulating relative preferences across contexts, including

No Reversal	"individual preferences cannot be inverted",
Upward Difference	"preferential differences persist into larger contexts"
	(though not necessarily among the same individuals).

It would be of interest to develop a similar contextual preference calculus in our case. For instance, should preferences from larger contexts persist into all smaller ones? One cue for principles like this may be taken from, again, basic Conditional Logic. On the above schema of interpretation, principles that lie beyond the core logic of Chellas 1975 (though perhaps still inside the minimal Lewis system) will come to express certain conditions of just this kind.

EXAMPLE. Conditional Axioms and Contextual Preference

The connection that we have in mind can be brought out by the usual techniques of 'frame correspondence' for modal logics (cf. van Benthem 1985 for the basic theory). These tell us exactly which formal constraints on indexed preference relations are imposed by various axioms of our conditional logic:

(1) 'Disjunction of Premises' $A \Longrightarrow B,\ C \Longrightarrow B\ \ /\ \ (A \vee C) \Longrightarrow B$
expresses that
" whatever is $<_{A \cup B}$-minimal must be either $<_A$-minimal or $<_B$-minimal"
(2) 'Cautious Cut' $A \Longrightarrow B,\quad A, B \Longrightarrow C\ \ /\ \ A \Longrightarrow C$
expresses that
"the $<_A$-minimal things are also minimal in the ordering $<_{\min<(A)}$"
(3) 'Cautious Monotonicity' $A \Longrightarrow B,\quad A \Longrightarrow C\ \ /\ \ A, B \Longrightarrow C$
expresses essentially the converse inclusion of (2).

2.3 Updating Preferences Themselves

2.3.1 Standard Implementations of Upgrading

'Updating' with a proposition p changes the information state: in our present case, decreasing the set of available worlds. Likewise, there may be dynamic operations changing our preferences, without necessarily changing pure informational content. Such a process may be called 'upgrading' , whose key instruction reads:

Pref(p) 'improve' the status of p-worlds in our ranking.

This may be done in various ways: there are many technical options to this effect. Which notion of 'state' will model these upgrading phenomena, as well as their 'downgrading' analogues (after all, our preferences can be subject to *revision* too)? There are several proposals in the literature, such as the state-preference patterns of the update logic of Veltman 1991, which may be viewed as providing instructions for changing possible worlds models for preferential conditional logic. Another approach would work with abstract state universes as in van Benthem 1991, de Rijke 1992, adding a primitive preference relation at the same level as the primitive *inclusion* relation between information states. And finally, one might deal with preferentially structured universes in the earlier eliminative vein (just like we did for descriptive propositions), by successively removing preference stuctures that do not satisfy our incoming preferential cues. Whatever specific choice is made here, one virtue of these systems is that they allow us to state and pursue such general issues as the logical behaviour of different notions of upgrading or downgrading, as well as possible analogies between the processes of updating and upgrading.

One systematic perspective on upgrading arises in the earlier-mentioned propositional dynamic logic. Atomic updates for a proposition p could be described as actions performing the minimal change of state needed so as to make that proposition true (when tested). Likewise, we can ask which minimal changes in the preference pattern are needed to achieve the truth of some preferential statement, in particular, a conditional " if A, then B". In this perspective, a conditional now functions both as a static 'test' and as a dynamic 'upgrading instruction' telling us to perform the minimal change over a preferential model needed to make the static conditional true. Here, the key question is how to measure 'minimal change' over preference patterns. For instance, there are two natural starting points for an initial state of 'indifference'. If one starts with *universal* preference, then successive further statements may lead to pruning. Note that this process can never remove a counter-example to a preferential conditional (maximal worlds stay maximal). The other option is to start from the empty preference relation, while building up stronger preferences later on. In this case, one can remove counter-examples to the above conditional by putting in suitable preferences over worlds which verify A but falsify B. Either way, it is still an open question whether there exist uniform constructive strategies having this logical effect. The relevant 'distance' between preference relations is measured here by set-theoretic inclusion. One may need more sophisticated measures eventually, allowing for both addition and removal of preferences, such as the comparisons of sets via their relative complements found in the literature on Verisimilitude (cf. van Benthem 1987):

$$A <_B C \quad \text{iff} \quad A \Delta B \subseteq C \Delta B.$$

So far, we have given no concrete definition for the basic operation of upgrading. Here is one, rather modest version:

> *Cautious Upgrading* "When upgrading with respect to p, remove all arrows that encode preferences for$l\neg_p$-states over p-states".

This instruction never adds preference arrows. Here is a more activist one which does:

> *Eager Upgrading* "In case of a tie between a p-state and some $\neg p$-state, add a preference arrow from the latter to the former".

Eager Upgrading has the effect that, if any p-world occurs in the current information state, then no more maximal $\neg_p$-worlds will remain afterwards, whereas Cautious Upgrading does not. In some sense, both of these alternatives still seem too 'uniform': in order to downgrage maximal $\neg_p$-worlds, for instance, one need merely give them *some* preferential pPsuccessor.

REMARK. Unary and Binary Operators
There is still a certain mismatch between this discussion of 'unary' upgrading for propositions p, and the upgrading process for conditionals, which typically involves 'binary' preferences comparing A-worlds and B-worlds. In the latter case, the point is to increase preferences for $(A \wedge B)$-worlds over $(A \wedge \neg B)$-worlds. Nevertheless, the latter task seems decomposable into two orthogonal ones. The first is to locate the proper subdomain for the comparison, and this is provided by the antecedent A (cf. the principle of 'Conservativity' for conditionals discussed in van Benthem 1986): in fact, this is an instance of the earlier-mentioned 'context management'. The second task is to perform the required upgrading in this subdomain, which can then be a unary one with respect to the consequent B. For more strictly 'binary' views of upgrading and preference, one may look at G on 'relative plausibility' initiated by De Finetti 1938 in the foundations of probability.

2.3.2 Arrow Logic

Right now, we want to propose another approach. Let us think of 'preferences' as concrete objects, just like worlds or states, which can be added or removed in our semantic models. A general motivation for this semantic viewpoint is found in the 'Arrow Logic' program of van Benthem 1992, which was originally intended to model concrete transitions in computation. But also, people often think of their 'preferences' as concrete arrows! States in the resulting system will have a two-sorted structure

(set of worlds , set of arrows),

and we need appropriate dynamic and static languages for manipulating these. Here is a sample system, with the following repertoire:

Basic operations

update (p) upgrade (p)

Operational repertoir

$\neg\wedge;\vee$

Test operations

allowing us to 'assess' the current state:

certainly $\phi[X]$	$:=$	X	if $\phi[X] = X$
		$\emptyset$	if $\phi[X] \neq X$
maybe $\phi[X]$	$:=$	X	if $\phi[X] \neq \emptyset$
		$\emptyset$	if $\phi[X] = \emptyset$
likely $\phi[X]$	$:=$	X	if the $(X)_2$-most preferred part of $(X)_1$ 'satisfies' ϕ
		$\emptyset$	otherwise

Here we need to state more precisely what 'satisfaction' is of ϕ in a set of worlds A. One option is to read the argument ϕ dynamically, and take the earlier fixed-point explanation: $\phi[A] = A$. The other, more plausible interpretation is to think of ϕ as a static standard proposition, and demand its truth throughout A. In any case, our preferred design here would be that of the earlier propositional dynamic logic, where such different options can co-exist.

One static description system for such systems uses ordered pairs of formulas

$$A = \langle A_1, A_2\rangle,$$

where the first statement restricts the world-component, and the second the arrow- component of our states. This format gives us a static reduction using the following definition of the next-state function:

$NS(X, \text{update } (p)) \quad := \quad \langle X_1 \wedge P, X_2\rangle$

$NS(X, \text{cautious upgrade } (p)) \quad := \quad \langle X_1, X_2 \wedge (l(P) \rightarrow r(P))\rangle$

$NS(X, \text{eager upgrade } (p)) \quad := \quad \langle X_1, X_2 \vee (l(\neg P) \wedge r(P))\rangle$

where l, r are 'arrow operators' referring to end points,

$NS(X, \text{certainly } \phi) \quad := \quad \langle X_1 \wedge \Box(X_1 \leftrightarrow (NS(X, \phi))1), X_2\rangle$

$NS(X, \text{maybe } \phi) \quad := \quad X_1 \wedge \Diamond(NS(X, \phi))_1, X_2\rangle$

where $\Diamond$, $\Box$ are the general $S5$-modalities of Section 2.1,

$NS(X, \text{likely } \phi) \quad := \quad \langle X_1 \wedge \Box(\neg do(T) \rightarrow \phi), X_2\rangle$

where *do* is the arrow operator expressing 'having some outgoing arrow'.

This binary format will become unary again by encoding states as 'identity arrows'.

3 Towards Further Structure

So far we have brought out various aspects of 'fine-structure' in actual reasoning: possible preferences between information states, and also possible transitions between information states. Now we want to add a more 'internal' computational aspect (the previous two may be considered more 'external'), namely, the internal structuring of preferences and information states. We conclude the present document with a brief discussion of three main lines of possible, and indeed, necessary semantic refinement.

STRUCTURED PREFERENCES

Like facts, preferences need computational maintenance (compare 'truth maintenance' in Artificial Intelligence). Processing information generates reasons for preferences, either because of explicit instructions in the preceding text, or because of more general background assumptions. This requires a calculus of 'indexed preferences' indicating the reasons supporting them. (The earlier indexed preference relations $<_A$ were one step in this direction.) More generally, this would bring in some form of proof theory, because 'reasons' will generally be previous arguments. One suggestive format for such a system would be the labeled deductive systems of Gabbay 1993.

STRUCTURED STATES

The above folklore style of modelling with sets of models has some drawbacks. For some purposes, it is too concrete, and one needs a more abstract framework, such as 'Dynamic Modal Logic' (van Benthem 1991, de Rijke 1992A), where informational inclusion is the key notion for cognitive procedures. Putting in preferences is a natural next move in that enterprise. The result will be a 'dynamic modal preference logic' which serves as a general framework for comparing specific implementations in the literature (cf. De Rijke 1992B on various paradigms for knowledge representation). For other purposes, however, model classes may be too abstract, failing to reveal the 'constructive structure' of worlds, and hence one may want a more 'partial' perspective (cf. Thijsse 1992, van Benthem 1988). For the semantics of conditional logic, this move was first proposed by Turner 1981, who uses a more concrete inclusion order among constructively defined information states, with suitable 'interaction principles' on preference and inclusion. One concrete way of providing the relevant constructive fine-structure reflects earlier historical views of possible worlds, viz. as Carnap-style or Hintikka-style state descriptions. These describe the kinds of individuals inhabiting a world, with their properties and relations, in varying degrees of 'logical detail' (measured by the quantifier depth and variable complexity of the world description). This view has been developed further in Frolova 1986, 1987, 1990, which propose various concrete preference relations based on relative distance between Hintikka's 'distributive normal forms'. These provide different ways of influencing preferences, as one learns more about the current information states. First, the description of the existing individuals may become refined, so that we shift preferences between them based on new predicates, but also, new individuals may enter the scene, changing the 'reference class' for our comparisons. Note the possible importance here of changes in vocabulary when determining preference: the latter may depend on the current 'descriptive format'. (A computational example is the shift in our most preferred 'minimal Herbrand models' when we change the vocabulary of a logic program. Compare also discussions in social choice theory on the effect of describing further options on previous preferences, cf. Hahn & Hollis 1979.) This observation might raise a larger issue, somewhat subversive to our semantics so far. In the final analysis, should not preferences between situations be 'generic', rather than 'individual' ?

WORLDS, PREDICATES AND INDIVIDUALS

The use of the above 'linguistic structuring' of worlds raises another general issue. World descriptions usually refer to individuals and predicates, and many of the examples in the computational literature are indeed about predicate-logical versions of preferential reasoning (cf. Lifshitz 1988, Delgrande 1987). For instance, does a default rule "A's are B's" tell us to grade different situations, or different individuals inside one situation? There are various options here for upgrading individuals, or even adding new individuals (say, a topmost 'typical' A that is B in the above case). What is the appropriate generalization of our proposals to this richer semantic arena?

This concludes our discussion of dynamic frameworks for changing preferences. We have not aimed at technical results here, nor at a complete survey of the literature (cf. e.g., Brown, Mantha & Wakayama 1992A, B for a congenial attempt), but we do hope to have illustrated the benefits of a certain systematic logical style of analysis.

REFERENCES

[1] J. L. BELL (1989). 'The Logic of Non-Monotonicity', *Artificial Intelligence* 41, 365-374.

[2] J. VAN BENTHEM (1983). *The Logic of Time*, Reidel, Dordrecht. (Second revised edition, 1991.)

[3] J. VAN BENTHEM (1985). *Modal Logic and Classical Logic*, Bibliopolis, Naples.

[4] J. VAN BENTHEM (1986). *Essays in Logical Semantics*, Reidel, Dordrecht.

[5] J. VAN BENTHEM (1987). 'Verisimilitude and Conditionals', in T. KUIPERS, (ed.), 1987, 103-128.

[6] J. VAN BENTHEM (1988). *A Manual of Intensional Logic*, The University of Chicago Press, Chicago.

[7] J. VAN BENTHEM (1989). 'Semantic Parallels in Natural Language and Computation', in H-D Ebbinghaus et al., eds., *Logic Colloquium.* Granada 1987, North-Holland, Amsterdam, 331-375.

[8] J. VAN BENTHEM (1991). 'Logic and the Flow of Information', Report LP-91-10, Institute for Logic, Language and Computation, University of Amsterdam. To appear in D. PRAWITZ, B. SKYRMS & D. WESTERST ÅHL, (eds.), *Proceedings 9th International Congress of Logic, Methodology and Philosophy of Science. Uppsala 1991*, North-Holland, Amsterdam.

[9] J. VAN BENTHEM (1992). 'Dynamic Arrow Logic', Report LP-92-11, Institute for Logic, Language and Computation, University of Amsterdam. To appear in J. van Eijck & A. Visser, eds., *Logic and Information Flow*, Studies in Logic, Language and Information, Kluwer, Dordrecht.

[10] J. BURGESS (1981). 'Quick Completeness Proofs for some Logics of Conditionals', *Notre Dame Journal of Formal Logic* 22, 76-84.

[11] A. BROWN, S. MANTHA & T. WAKAYAMA (1992A). 'Exploiting the Normative Aspect of Preference: A Deontic Logic without Actions', Webster Research Center, Xerox Corporation.

[12] A. BROWN, S. MANTHA & T. WAKAYAMA (1992B). 'Preference Logics: Towards a Unified Approach to Nonmomotonicity in Deductive Reasoning', Webster Research Center, Xerox Corporation.

[13] B. CHELLAS (1975). 'Basic Conditional Logic', *Journal of Philosophical Logic* 4, 133-153.

[14] J. DELGRANDE (1987). 'A Semantics for Defaults Using Conditional Logic', *Artificial Intelligence* 36 (1988), 63-90.

[15] P. DOHERTY (1991). *NML3 — A Non-Monotonic Formalism With Explicit Defaults*, dissertation, Department of Computer and Information Science, Linköping University.

[16] J. DOYLE (1985). 'An Essay in Nonmonotonic Logic', Department of Computer Science, Carnegie Mellon University, Pittsburgh.

[17] J. DOYLE (1991). 'Reason Maintenance and Belief Revision: Foundations versus Coherence Theories', Laboratory for Computer Science, Massachusetts Institute of Technology.

[18] J. VAN EIJCK & F-J DE VRIES (1990). 'Dynamic Interpretation and Hoare Deduction', *Journal of Logic, Language and Information* 1, 1-44.

[19] J. VAN EIJCK & F-J DE VRIES (1991). 'A Sound and Complete Calculus for Update Logic', Report CS-R9155, Centre for Mathematics and Computer Science, Amsterdam.

[20] J. VAN EIJCK & F-J DE VRIES (1993). 'Reasoning about Update Logic', Report CS-R9312, Centre for Mathematics and Computer Science, Amsterdam. To appear in the *Journal of Philosophical Logic.*

[21] B. DE FINETTI (1938). 'La Prvision: ses Lois Logiques, ses Sources Subjectives', *Annales de l'Institut Henri Poincaré* 7, 1-68.

[22] A. FROLOVA (1986). 'On Clarifying Several Concepts of Possible Worlds Semantics', *Abstracts IXth International Conference on Logic, Methodology and Philosophy of Science*, Moscow.

[23] A. FROLOVA (1987). 'Distributive Normal Forms as a Means of Clarifying Possible Worlds Semantics', *Modern Logic and Methodology of Science*, Department of Philosophy, Moscow State University.

[24] A. FROLOVA (1990). 'Non-Standard Reasoning and Types of Consequence Relations', *Proceedings Xth Soviet Conference on Logic, Methodology and Philosophy of Science*, Minsk.

[25] A. FROLOVA (1992). 'Structured Information States', manuscript, Centre for Mathematics and Computer Science, Amsterdam.

[26] D. GABBAY (1993). *Labeled Deductive Systems*, Oxford University Press, to appear.

[27] P. GÄRDENFORS (1988). *Knowledge in Flux: on the Dynamics of Epistemic States*, Bradfod Books / MIT Press, Cambridge (Mass.).

[28] P. GÄRDENFORS (1991). 'Social Choice and the Emergence of Semantic Primitives', invited lecture, Eighth Amsterdam Colloquium in Formal Semantics.

[29] F. HAHN & M. HOLLIS, eds. (1979). *Philosophy and Economic Theory*, Oxford University Press, Oxford. J. HALPERN, (ed.)

[30] F. HAHN & M. HOLLIS, eds. (1986). *Proceedings of the First Conference on Theoretical Principles of Reasoning About Knowledge*, Morgan Kaufmann Publishers, Los Altos.

[31] J. F. HORTY (1992). 'Deontic Logic as Founded on Nonmonotonic Logic', to appear in *Annals of Mathematics and Artificial Intelligence.*

[32] TH. KUIPERS, ed., (1987). *What is Closer-to-the-Truth ?*, Rodopi, Amsterdam.

[33] D. LEWIS (1973). *Counterfactuals*, Blackwell, Oxford.

[34] V. LIFSCHITZ (1988). 'Circumscriptive Theories: A Logic-Based Framework for Knowledge Representation'. In R. Thomason, ed., special issue of the *Journal of Philosophical Logic* ('Philosophical Logic and Artificial Intelligence'), 391-441.

[35] M. MORREAU (1985). 'Circumscription, a Sound and Complete Form of Nonmonotonic Reasoning', master's thesis, Philosophical Institute, University of Amsterdam.

[36] M. MORREAU (1991). *Conditionals in Philosophy and Artificial Intelligence*, dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

[37] M. DE RIJKE (1990). 'The Modal Logic of Inequality', Report LP-90-15, dissertation, Institute for Logic, Language and Computation, University of Amsterdam. Also in *The Journal of Symbolic Logic* 57 (1992), 566-584.

[38] M. DE RIJKE (1992A). 'A System of Dynamic Modal Logic', Report LP-92-08, Institute for Logic,Language and Computation, University of Amsterdam.

[39] M. DE RIJKE (1992B). 'Meeting Some Neighbours', Report LP-92-10, Institute for Logic, Language and Computation, University of Amsterdam.

[40] M. RYAN (1992). *Structured Theories*, dissertation, Department of Computing, Imperial College, London.

[41] E. SANDEWALL (1992). *Features and Fluents*, Department of Computer and Information Science, Linköping University.

[42] F. SEBASTIANI (1992). 'A Fully Model-Theoretic Semantics for Model-Preference Default Systems', Istituto di Elaborazione dell'Informazione, Pisa. To appear in *Studia Logica.*

[43] B. SELMAN & H. KAUTZ (1988). 'The Complexity of Model-Preference Default Theories', in *Proceedings of the 1988 Conference of the Canadian Society for Computational Studies of Intelligence*, Edmonton (Alberta) , 102-109. Also in M. REINFRANK, J. DE KLEER, M. GINSBERG & E. SANDEWALL, (eds.), *Nonmonotonic Reasoning*, Springer, Heidelberg, 1989, 115-130.

[44] Y. SHOHAM (1988). *Reasoning about Change: Time and Causality from the Standpoint of Artificial Intelligence*, MIT Press, Cambridge (Mass.).

[45] Y. SHOHAM (1991). 'Varieties of Context'. In V. Lifshitz, ed., *Artificial Intelligence and Mathematical Theories of Computation: Papers in Honor of John McCarthy*, Academic Press, New York.

[46] E. SPAAN (1993). *Complexity of Modal Logics*, dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

[47] W. SPOHN (1988). 'Ordinal Conditional Functions: A Dynamic Theory of Epistemic States', in W. L. HARPER ET AL., (eds.), *Causation in Decision, Belief Change and Statistics* II, Kluwer, Dordrecht, 105-134.

[48] R. STALNAKER (1972). 'Pragmatics', in D. Davidson & G. Harman, eds., *Semantics of Natural Language*, Reidel, Dordrecht, 380-397.

[49] E. THIJSSE (1992). *Partial Logic and Knowledge Representation*, dissertation, Faculty of Arts, University of Tilburg.

[50] R. TURNER (1981). 'Counterfactuals Without Possible Worlds', *Journal of Philosophical Logic* 10, 453-493.

[51] F. VELTMAN (1985). *Logics for Conditionals*, dissertation, Philosophical Institute, University of Amsterdam.

[52] F. VELTMAN (1988). 'Update Semantics' , manuscript, Institute for Logic, Language and Computation, University of Amsterdam.

[53] F. VELTMAN (1991). 'Defaults in Update Semantics', Report LP-91-02, Institute for Logic, Language and Computation, University of Amsterdam. To appear in the *Journal of Philosophical Logic.*

[54] Y. VENEMA (1992). *Many-Valued Modal Logic*, dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

[55] D. WESTERSTÅHL (1984). 'Determiners and Context Sets', in J. van Benthem & A. ter Meulen, eds., *Generalized Quantifiers in Natural Language*, Foris, Dordrecht.45-71.