Institute for Language, Logic and Information

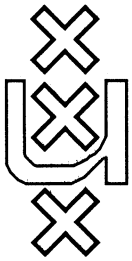
SEMANTIC PARALLELS

IN NATURAL LANGUAGE AND COMPUTATION

Johan van Benthem

ITLI Prepublication Series

for Logic and Philosophy of Language LP-88-06



University of Amsterdam

Table of Contents

Semantic Parallels

1 Convergence

2 Minimality

2.1 Non-Standard Inference

2.2 The Logic of Circumscription

2.2.1 Minimal Models

2.2.2 Conditional Logic

2.2.3 Dynamics of Minimality

2.2.4 First-Order Reduction

2.3 Data Types and Theories

2.3.1 Logic Programs

2.3.2 Abstract Data Types ...

2.3.3 ... and Scientific Theories

3 Dynamics

3.1 Dynamics of Interpretation

3.1.1 Operational Semantics for Programs

3.1.2 Operational Semantics for Assertions

3.2 Dynamics of Information Flow

3.2.1 Reducing Possibilities

3.2.2 Building Up Information

3.2.3 Relational Calculus and Categorial Grammar

3.2.2 Propositional Dynamic Logic

4 Discussion

5 References

1. CONVERGENCE

In recent years, there has been a growing awareness that many central questions of research are quite similar in such diverse disciplines as Logic, Linguistics, Philosophy or Computer Science. One of the major unifying forces here is a shared interest in *reasoning,* whether performed by humans or by machines. Now reasoning, of course, is the prime field of study for *Logic* - so, in a sense, there are logical threads running across the above disciplines. "In a sense", because the newer developments call for a broad view of Logic, encompassing many varieties of reasoning: both as to mechanisms of inference, and as to forms of representing the information that goes into these. There are certainly many more things between earth and heaven here than are dreamt of (or at least, consciously studied) by standard logic as it is.

These recent developments carry two kinds of promise for existing logical theory. On the one hand, they offer many possibilities for wider employment of notions and techniques developed in the narrower setting of the foundations of mathematics, or specific branches of philosophy. But also, there may be genuine opportunities for new research lines inside logic as well, inspired by the newer concerns. To some extent, we are witnessing the latter phenomenon in so-called *logical semantics*, where motivations from philosophy, linguistics and recently also computer science meet and interact. (See e.g. van Benthem 1986a or van Benthem 1988 for several case studies and surveys.)

Perhaps most intriguing are those new research lines which touch our understanding of the basic notion of valid consequence - often held to be codified once and for all in Tarski's semantics for Frege's predicate logic. Especially from a *computational* perspective, broadly conceived, there turns out to be a lot of fine-structure, and indeed variety to valid inference. In this paper, we shall consider two major themes of this kind, viz. techniques for local "strengthening" of logical inference via *minimization* of models (Section 2), and the more general *dynamics* of progressive handling of information in interpretation and argument (Section 3). The purpose is to provide a coherent pattern behind some recent developments in these areas, and also, to discuss their value as affecting Logic in general.

Whatever the outcome of this discussion may be for the short term, the long-term perspective in Logic nowadays seems to be a return to a richer conception of what the study of reasoning should involve. A return, for such richer conceptions were certainly present in the history of Logic. One inspiring example in this respect is Bernard Bolzano (cf. van Benthem 1985c), who still defined the task

of the discipline as studying the *variety* of modes of inference which human beings have available for various intellectual tasks. With Bolzano, this led to the logical study of several notions of valid consequence, differing in their formal properties, without a pious commitment to any particular calculus. And that is certainly also the proper spirit to approach the recent developments to be discussed here.

The paper also makes a number of technical contributions to the topics discussed. The first group of these concerns *minimal models*. We provide a mathematical analysis of the *minimization operator* on classes of models (Section 2.2.3), while also investigating several special systems in which minimal models play a central role. Many such systems are found in the areas of Logic Programming and Abstract Data Types (Section 2.3). Here, we develop an analogy with earlier work in the Philosophy of Science on so-called *Ramsey eliminability* of theoretical terms in scientific theories (Section 2.3.2), showing how the latter can incorporate minimal models (Section 2.3.3). The central example, however, is the notion of *circumscription,* recently developed in Artificial Intelligence: which may be viewed as a semantic generalization of minimal Herbrand modelling for Horn clause formalisms. A technical connection is found between the general inferential properties of circumscription and more traditional *conditional logic* (Section 2.2.2). We also consider possible reductions of circumscriptive inference to standard *first-order logic,* establishing a high *complexity* for the question just when this is possible (Section 2.2.4).

Then, there is also a number of results on dynamical semantics. We provide several *reductions* of proposed dynamic systems to, again, standard ('static') *first-order logic* (Sections 3.1.2, 3.2.1). Also, various proposals are made for connecting up what may be called 'abstract information structures' with already existing types of models, notably in *relational algebra* (Section 3.2.3) and *propositional dynamic logic* (Section 3.2.4). Especially, the latter system seems to provide a promising tool for investigating dynamic modes of handling propositions.

2. MINIMALITY

2.1 Non-Standard Inference

Logically valid inference in the standard sense allows us to draw those conclusions from a set of premises which are true in all models for those premises. This is an important, and safe procedure. But, conclusions drawn in this way may be few in number, when we have to act upon a small amount of explicit information. Accordingly, we may observe the presence of additional mechanisms in cognitive activity, allowing us some further 'contextual' inferences. One pervasive mechanism of this kind is the assumption of a certain *completeness*: the premises as stated give us 'the whole truth' about the matter [and of course, to continue the juridical analogy, hopefully also 'nothing but the truth']. In the Philosophy of Language, this phenomenon was observed by Paul Grice, whose "Maxim of Quantity" expresses that what we say should state our complete information (as far as relevant to the matter at hand). Here is a concrete example of how this convention works. If my partner in a conversation tells me that Basra lies either in Iran or in Iraq, I can take it that she does not know anything more definite about its location than that. So, I can conclude that she does not know that it lies in Iraq - even though this does not logically follow. (She might be just *pretending* ignorance, of course: but, this is not assumed in normal conversation.) A less epistemically oriented example would be this. If I am told that event A happened because of event B, I will tend to assume that, as far as the speaker is concerned, B is the *only* cause of A. Again, there is no logical necessity here: but rather the assumption that I have been supplied with the complete relevant information.

These examples are still rather vague - and indeed, one recurrent complaint about pragmatic principles like Grice's Maxims has been that they can be twisted to suit too many a purpose. Nevertheless, there are already two interesting features of non-standard inference to be observed here. First, we *increase* the number of available inferences, by making some general assumptions about the 'status' of our premises. Next, as a price to be paid for this, we have to be vigilant, since the additional conclusions are *defeasible*, and may have to be given up after all, in the light of further evidence. (I may learn subsequently of another possible reason for event A.) The latter phenomenon is often called *non-monotonicity*, in contrast to the 'monotonicity', or cumulativity, of inferences in standard logic.

One way of making these procedures more precise consists in the following modification of the standard *semantic* picture. A piece of text (story, proof or

program) is often concerned with, not *all* its models in the usual logical sense, but rather with one intended, or at least a severely restricted class of intended models. One reason for this seems to lie in the operation of a general principle allowing us to exploit *absence* as well as *presence* of explicit information:

the world is *the smallest,*

or at least *a minimal* situation verifying our data.

And then, of course, more conclusions will be valid (generally speaking) in just the minimal models of our data than in all of its conceivable models. This perspective turns out to admit of precise mathematical treatment, and indeed, it runs through a number of technical proposals for implementing reasoning in a computational setting. We shall look at a few examples, while pointing at a more general background in the philosophy of language, and also the philosophy of science.

The above idea is prominent in *Logic Programming*. Notably, PROLOG programs are supposed to describe a so-called 'minimal Herbrand model' for their assertions (see Lloyd 1985). In fact, such models already exhibit *two* aspects to minimality. First, they contain no individuals / objects except for those which are explicitly named in the language of the program. But also, they contain no facts about these objects except for those explicitly enforced by the program ('Negation as Failure'). Both aspects are of independent interest. *Individual minimality* gives a minimal delineation of our objects; which can be cashed in, so to speak, as a principle of 'induction':

If a property holds of all objects nameable in our language,

then it is universally true.

In fact, this is close to ordinary mathematical senses of induction. The defeasibility here shows in that addition of further operations forming objects may actually weaken the method of induction (one has to prove more 'inductive steps'). This is also true in ordinary mathematics. Next, *predicate minimality* is actually closer to the original examples given above. In PROLOG terms, its inferential value lies in reading a program as standing for its own 'completion', where a condition P is not only implied by all possible antecedents available for it, but also itself implies their disjunction. (There are no other reasons for P than those explicitly stated.)

Again, it is worth-while to think of similar phenomena in natural language. For instance, many implications function like the above conditions. When the doctor says that you will recover if you do as he says, he does not just utter a promise, but also a threat ('only if'): 'if you don't ..., you won't ...'.

[Similar conventions may be observed, incidentally, in the *juridical* argument called 'a contrario', which looks like the formal fallacy of deriving $\neg A \rightarrow \neg B$ from $A \rightarrow B$: but which in fact resembles the preceding pattern.] But individual minimality occurs too. For instance, answers to questions are often taken 'exhaustively' (see Groenendijk &Stokhof 1985). In the dialogue

"Who are crying?" "John and Mary" ,

the answer will normally be interpreted as a *complete* list, even though the direct logical information is only that "John cries and Mary cries". And this phenomenon persists to more complex answers, such as "a girl", which suggests that exactly one person is crying, and that a girl. In other words, we are invited to form a minimal model in which the answer is true.

Another area of Computer Science where explicit formal models of minimality have emerged is in *Artificial Intelligence*. Here again, action by intelligent agents, whether alive or mechanical, presupposes inference on the basis of incomplete information. In particular, many practical rules which guide us in action involve a 'ceteris paribus' clause: *other things being equal,* doing this or that will produce such and such effect (cf. Shoham 1988). Making certain that no 'other things' have changed will in general require an infinite amount of information about the state of the world: something that we cannot obtain. Instead, we tend to assume the *absence* of disturbing circumstances, as long as they do not appear explicitly. Put differently, until forced to the contrary, we will assume that the case at hand is not an exceptional, but rather a 'normal' one. This strategy, though obviously defeasible, is clearly sound: the odds being in its favour (by the definition of normal versus exceptional cases).

Ceteris paribus rules are quite common in juridical procedures, in ethics (cf. Aqvist 1984), but also in the philosophy of science, in connection with counterfactual statements or dispositional terms ('soluble', 'inflammable'): see Sosa 1975. So, the strategies we are considering are not just little short-cuts for self-deception, but rather necessities, even of scientific life.

Some familiar examples from Artificial Intelligence motivated one basic semantic way of extending the earlier notion of minimality, which is due to McCarthy 1980. All birds can fly, barring some exceptional cases like penguins. Tweety is a bird: now, can Tweety fly? Logically speaking, nothing can be concluded here. But practically, we will want to conclude that Tweety *can* fly. One solution here is to assume the following formalization:

$\forall x((Bx \wedge \neg abx) \rightarrow Fx)$

Bt

That is, 'not abnormal' birds can fly. Now, reading this as describing models with a *minimal* extension for the abnormality predicate, Tweety will not be abnormal, and hence she can fly. In general, however, there may be more abnormality predicates involved, corresponding to different defeasible rules. Then, there may be a genuine ambiguity , as in another well-known example:

Quakers are pacifists	: $\forall x(Qx \wedge \neg ab_1 x) \rightarrow Px)$
Republicans are non-pacifists	: $\forall x(Rx \wedge \neg ab_2 x) \rightarrow \neg Px)$
Nixon is a Quaker	: Qn
Nixon is a Republican	: Rn .

Now, will Nixon be a pacifist, or not? There are two incomparable minimal models here: one with 'ab_1' empty and 'ab_2' consisting of Nixon only, the other with the situation reversed. And this may reflect a genuine intuitive uncertainty here. On the other hand, there may also be cases where we have clear priorities as to which abnormality predicate should be minimalized first, so that we should have to break the tie by *prioritizing* rules.

There is a general technical notion of minimal model for these cases, which makes sense for arbitrary predicate-logical formulas (not just the Horn clauses of the PROLOG formalism), and which allows for the above multiplicity of minimal models. This is the so-called *circumscription* of McCarthy 1980, which will be considered in greater detail below. We can view it as a generalization of the minimal model semantics for logic programs, in which we now consider all those models for a set of premises whose proper *submodels* (in some suitable sense) no longer verify those premises.

Again, it is of interest to observe that there were earlier, similar attempts in the *Philosophy of Science*. Philosophers had already studied such weaker notions of consequence as *confirmation* of a hypothesis by certain evidence. For instance, a universal generalization $\forall xPx$ is, if not implied, at least confirmed by observing instances of it: Pa, Pb, ... Interestingly, Hempel 1965 suggests the following explanation for this. When observing the evidence, we tend to form a *minimal* model in which it holds: and that is one in which the universal generalization is valid. (Again, of course, there is non-monotonicity: a well-confirmed regularity can be refuted after all, by new evidence.) This particular analogy is well in line with the earlier reference to Bolzano, whose logical main work was in fact called 'Wissenschaftslehre'. It was especially in scientific thinking that Bolzano found diverse forms of rational argument, with varying logical properties. Confirmation

itself is just one example of this. Another one would be *explanation* of some fact from certain scientific assumptions. Here, for instance, we would want another kind of minimality, namely the use of only some minimal set of relevant premises deriving the conclusion. In this case too, there will be non-monotonicity: as adding irrelevant premises will disturb the explanatory character of an inference. This observation emphasizes the many possible sources of non-monotonic behaviour: which is a *symptom*, rather than the essence of non-standard inference.

More generally, in this perspective, the choice of a particular mode of reasoning, monotonic or non-monotonic, will be dictated by the specific kind of application intended.What is suitable in one context, need not be suitable in another. (For instance, in faculty politics, many people would be in much better health if they did not jump non-monotonically to ill-informed conclusions all the time.) Thus, human rationality becomes 'parametrized': we have to select some sensible mode of reasoning, prior to performing rationally within that mode. This attitude also fits well with two main aspects to current theories in Artificial Intelligence (of which the above-mentioned circumscription is an example). They are sometimes meant to *describe* intellectual activity, but sometimes also to *design* intelligent systems. And at least the latter purpose requires conscious selection of intellectual tools.

In the following parts of this Section, we shall now consider some technical implementations of minimal modelling in more detail, pointing at various interrelations, open questions, etcetera.

2.2 The Logic of Circumscription

2.2.1 Minimal Models

Individual-minimal models for a set Σ of predicate-logical (or indeed, any kind of) formulas may be defined as those M such that

1) $M \models \Sigma$,
2) for no $M' \subsetneq M$, $M' \models \Sigma$.

Likewise, *predicate-minimal* models are those models M for Σ whose predicate extensions cannot be decreased without losing the truth of Σ. (The latter notion is somewhat like 'Pareto Optimality' in economics.) More generally, we can also consider models only some of whose predicates are minimized in this fashion.

One test of the adequacy of this notion is an application to the earlier topic of exhaustive answers to questions. It is natural to assume that, in such a case, the answer is read as describing a Q-minimal model, where Q is the predicate that the query is about. Thus, the C-minimal models of Cj∧Cm are precisely those in which only John and Mary cry. And likewise, the C-minimal models of ∃x(Gx∧Cx) are those in which exactly one girl forms the whole extension of 'crying'. (This simple analysis obviates the need for the complex 'minimalized generalized quantifier' analysis found in Groenendijk and Stokhof 1985. The point is, so to speak, that the complexity does not reside in the answer, but in the query.) Incidentally, the claim here is *not* that answers are invariably taken in this minimal sense. For instance, on the present account, the two answers 'no girl' and 'no boy' to the previous question would both state that no one is crying. And that is certainly unreasonable: these answers mean no more than they say in standard terms, being ¬∃x(Gx∧Cx), ¬∃x(Bx∧Cx), respectively. More generally, in *applications* of circumscription, whether more linguistic or more computational, there is always a *decision* to be made, not prescribed by the formal theory, as to which predicates are to be minimalized.

Minimal models form an interesting object of mathematical study by themselves. In fact, their behaviour is full of surprises, once *infinite* structures are taken into account. For instance, the integers are a predicate-minimal model for their own first-order theory, being that of unbounded discrete linear orders. (Any removal of a pair in the order relation would destroy linearity.) But, they are not an individual-minimal model for it: the proper substructure of the *even* integers would serve just as well. In fact, this ordering theory has no individual-minimal model at all. On the other hand, what are classically small changes in the *presentation* of a theory may have telling effects now. If one adds two *Skolem functions*, witnessing immediate successors and predecessors, then, viewed as a structure for the expanded similarity type, the integers will become a minimal model for their ordering theory (expanded with the two relevant definitions). This reflects a general phenomenon. With function symbols around, submodels will have to be closed under the corresponding operations: which increases the chances of a model's being minimal. We shall not pursue this mathematical direction here, which would call for comparison with other notions of minimal model in the literature (such as 'prime models': cf. Chang and Keisler 1973).

As for semantic *consequence,* there are several options now, summed up in the following definition:

$\Sigma \models_{ind} \varphi$ if φ is true in all individual-minimal models of Σ

$\Sigma \models_{pred} \varphi$ if φ is true in all predicate-minimal models of Σ

$\Sigma \models_{*} \varphi$ if φ is true in all structures which are both individual- and predicate-minimal models of Σ.

Of the major properties of classical consequence, these new notions preserve some, while losing others. We shall illustrate this for the case of $\models_{*}$.

- $\Sigma \models_{*} \varphi \quad \Rightarrow \quad \Sigma \models_{*} \varphi \vee \psi$

I.e., one may 'weaken consequents'. But, one may not 'strengthen antecedents:

- $\Sigma \models_{*} \varphi \quad \not\Rightarrow \quad \Sigma, \alpha \models_{*} \varphi$.

This is the earlier-mentioned *non-monotonicity*: the minimal models for Σ *plus* α may have shifted from those of Σ by itself.

Likewise we lose the classically valid 'transmission of truth':

- $\Sigma \models_{*} \varphi, \quad \varphi \models_{*} \psi \quad \not\Rightarrow \quad \Sigma \models_{*} \psi$

A counter-example is: $\Sigma = \exists x Px \wedge \exists x \neg Px$, $\varphi = \exists x Px$, $\psi = \forall x Px$.

On the other hand, we keep such classically admissible operations on premises as Permutation, or Contraction of identicals.

Actually, could there be any *gain* in general properties for $\models_{*}$ too, as compared to classical consequence $\models$? As we shall see later on, the answer is negative.

Before continuing with the general logic of circumscription, it may be useful to emphasize a new aspect of reasoning in this situation. Even if a certain pattern of inference is invalid in general, it may still be admissible for certain *types of* statement. For instance, monotonicity will in fact be valid for those additional premises which are *preserved* under the model relations employed in the definitition of minimality. Thus,

for *universal* formulas α,

$\Sigma \models_{ind} \varphi \quad \Rightarrow \quad \Sigma, \alpha \models_{ind} \varphi$.

For , if M is an individual-minimal model for $\Sigma + \alpha$, and M' is a submodel of M verifying Σ, then it also verifies α (by preservation of universal statements under submodels), and so M'=M. This proves that M is also an individual-minimal model for Σ itself, and hence that $M \models \varphi$, by the assumption. ∎

By a similar argument, $\models_{pred}$ is monotone for added *negative* formulas, constructed from negations of atoms (as well as identities) using $\wedge$, $\vee$, $\forall$, $\exists$. Thus, minimal reasoning may require closer attention to syntactic 'fine-structure' of premises.

2.2.2 Conditional Logic

As a notion of predicate-logical consequence, circumscription is highly complex. For instance, as the only minimal model of the usual Peano Axioms (even *minus* Induction) consists of the standard natural numbers, minimal consequence from these premises coincides with actual arithmetical truth: hence, non-axiomatizability (and worse) follows by Tarski's Theorem.

On the other hand, there is also a *general logic* of circumscription, not referring to specific predicate-logical forms, but only to simple (at most) Boolean Structure. Examples of these inferences were such earlier patterns as Transitivity or Weakening Consequents. This raises the issue of what might be called the *propositional logic* of circumscription. The latter system is not trivial: especially, since the break-down of classical principles can often be compensated for by more refined substitutes. For instance, the following form of Transitivity is in fact valid:

$$\Sigma \models_* \varphi, \quad \Sigma, \varphi \models_* \psi \quad \Rightarrow \quad \Sigma \models_* \psi.$$

And, thinking of progressive stages in an argument, this makes sense.

In fact, a further study of valid patterns for notions of minimal consequence brings to light a striking analogy with systems of reasoning developed earlier in *Conditional Logic* (cf. Lewis 1973), which started out as a study of so-called counterfactual assertions 'if A had been the case, then B would have been the case'. [Again, this (non-monotone) notion has strong connections with the philosophy of science: cf. Harper et al., eds., 1981.] For instance, the basic axioms of Conditional Logic are as follows (Burgess 1981, Veltman 1986):

i	$\varphi \Rightarrow \varphi$			
ii	$\varphi \Rightarrow \psi$,	$\varphi \Rightarrow \chi$	imply	$\varphi \Rightarrow \psi \wedge \chi$
iii	$\varphi \Rightarrow \psi$,	$\chi \Rightarrow \psi$	imply	$\varphi \vee \chi \Rightarrow \psi$
iv	$\varphi \Rightarrow \psi$	implies	$\varphi \Rightarrow \psi \vee \chi$	
v	$\varphi \Rightarrow \psi$,	$\varphi \Rightarrow \chi$	imply	$\varphi \wedge \psi \Rightarrow \chi$.

Of these five, all except the last are valid with respect to all three notions of minimal consequence. For the validity of the last, one has to assume a universe of models in

which the submodel relation is *well-founded* for any set of premises. For instance, the universe of all *finite* models would do. (Compare Shoham 1988 on the plausibility of this requirement.) Further principles which are valid turn out to be derivable from these.

Example: Refined Transitivity derived:

$\varphi\wedge\neg\psi \Rightarrow \varphi\wedge\neg\psi$ (i)

$\varphi\wedge\neg\psi \Rightarrow \neg\psi$ (iv) $\qquad$ $\varphi\wedge\psi \Rightarrow \chi$ (assumption)

$\varphi\wedge\neg\psi \Rightarrow \neg\psi\vee\chi$ (iv) $\qquad$ $\varphi\wedge\psi \Rightarrow \neg\psi\vee\chi$ (iv)

$(\varphi\wedge\neg\psi)\vee(\varphi\wedge\psi) \Rightarrow \neg\psi\vee\chi$ (iii)

i.e., $\varphi \Rightarrow \neg\psi\vee\chi$

also, $\varphi \Rightarrow \psi$ (assumption)

so $\varphi \Rightarrow \psi\wedge(\neg\psi\vee\chi)$ (ii)

whence $\varphi \Rightarrow \chi$ (iv).

Note the use of Boolean equivalences throughout.

Digression: Incidentally, here is one other computational way of viewing the minimal conditional logic. The inferences displayed above present a mixture of the usual Boolean rules for *classical* implication $\rightarrow$ and a *non-classical* implication $\Rightarrow$. For instance, axiom iv may be regarded as

$\varphi \Rightarrow \psi$, $\psi \rightarrow \chi$ imply $\varphi \Rightarrow \chi$;

and other rules concern suprema and infima in the $\rightarrow$ ordering combined with $\Rightarrow$. This interaction between simple rules for classical entailments and non-monotone *default* rules is reminiscent of the situation in so-called *semantic networks*. In the latter area, one of the key issues is the creation of sound and computationally tractable inferential algorithms. Perhaps, Conditional Logic can also serve as a useful model for the latter activity.

On the other hand, principles not derivable in the minimal conditional logic turn out to be typically invalid for circumscription too. In a more general perspective, this correspondence is not surprising. There has been a tendency in the literature on circumscription to minimize over other ordering relations between models than just the above notions involving submodels. (Compare the above reference to prioritized circumscription, cf. Lifschitz 1985, or also Shoham 1988). And in fact, only some general properties of such orderings seem relevant to evaluating validity of propositional circumscriptive patterns. But then, we are

precisely back with the original *possible worlds semantics* for conditional statements, due to David Lewis and Robert Stalnaker:

Models are structures (W,R,w), where w is the actual world, from whose perspective we are viewing the other worlds in W, and R orders the latter as to greater or lesser similarity to w. (Technically , R is a strict partial order.) Then, at least on *finite* models (the infinite case is somewhat more subtle), conditionals are evaluated as follows:

$\varphi \Rightarrow \psi$ is true at w if all R-*closest* worlds verifying the antecedent φ
also verify the consequent ψ.

This explains the general analogy between Circumscription and Conditional Logic. What remains is a number of more specialized questions. For instance, can minimality as defined above serve as an adequate concrete model for Lewis Semantics; e.g., in the sense of the following *conjecture:*

A conditional inference $\varphi \Rightarrow \psi$ is derivable in the minimal conditional logic
if and only if it is valid as a circumscriptive consequence $\varphi \models_* \psi$
in a language whose proposition letters stand for formulas of
monadic predicate logic ?

What would be needed here is a *representation* of abstract Lewis models in terms of finite models with the above submodel relation, under a suitable translation of proposition letters into monadic sentences.

Another question in this perspective would concern special logics for special relations to be minimized over. From Conditional Logic, we know that additional conditions on strict partial orders may produce additional validities. For instance, making the relation R *linear* will validate so-called 'conditional excluded middle':

$\varphi \Rightarrow \psi \vee \varphi \Rightarrow \neg\psi$.

So, variants of circumscription may also induce weaker or stronger general logics.

Toward a proof of the above conjecture, at least one suggestive result may be cited [for convenience, attention will be restricted to *individual-minimality*]:

Proposition: The following two assertions are equivalent,
for any propositional inference from $A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n$ to $C \Rightarrow D$:
(1) $C \Rightarrow D$ is derivable from $A_i \Rightarrow B_i$ $(1 \leq i \leq n)$ in the minimal conditional logic,
(2) for every substitution σ replacing proposition letters
by sentences of a monadic predicate logic with identity,
and every sentence φ of such a language,
if $\varphi, \sigma(A_i) \models_{ind} \sigma(B_i)$ for all i $(1 \leq i \leq n)$, *then* $\varphi, \sigma(C) \models_{ind} \sigma(D)$.

Proof: From (1) to (2). This is a straightforward induction on the length of derivations in the minimal conditional logic. The additional premise φ does not affect the earlier soundness argument. (Note here that, for a monadic predicate logic with identity, attention may be restricted to *finite* models, without any changes in semantic satisfiability.)

From (2) to (1). Suppose that $C \Rightarrow D$ is not derivable from the $A_i \Rightarrow B_i$. Then there exists some *finite* possible worlds model W containing a world w from whose perspective all formulas $A_i \Rightarrow B_i$ are true, whereas $C \Rightarrow D$ is false. Passing on to the relevant binary order $<$ on worlds ('greater similarity to w'), we can view this counter-model as some finite ordered set of propositional valuations (sitting on worlds), with conditional statements evaluated as in the Stalnaker-Lewis semantics. Now, take distinct unary predicates $P_v x$ for each world v, and set

$$\alpha_v(x) := P_v x \wedge \bigwedge_{u \in W, u \neq v} \neg P_u x .$$

Let there be exactly $u_1, \ldots, u_n \leq v$. Set

$$\beta_v := \exists x_1 \ldots \exists x_n (\alpha_{u1}(x_1) \wedge \ldots \wedge \alpha_{un}(x_n) \wedge \forall y (y = x_1 \vee \ldots \vee y = x_n)).$$

Moreover, let

$$\varphi := \bigvee_{v \in W} \beta_v .$$

Finally, let σ be the substitution which assigns to each proposition letter p the formula

$$\bigvee_{v \in [[p]]} \beta_v .$$

Claim: Let M be a structure for monadic predicate logic which verifies β_v for some world v. Then, for each purely propositional formula α, $M \models \sigma(\alpha)$ if and only if α is true in v.

Proof: For proposition letters α, 'if' follows by the definition of the substitution σ. Conversely, suppose that $\alpha = p$ is false in v. So, $\sigma(p)$ is a disjunction of formulas β_u with $u \neq v$. It suffices to observe that such β_u exclude β_v, by their definition. The remainder of the argument is a routine induction on Boolean operators. ∎

The desired conclusion then follows from the next assertion:

Claim: (i) $\varphi, \sigma(A_i) \models_{ind} \sigma(B_i)$, for all i ($1 \leq i \leq n$),
(ii) *not* $\varphi, \sigma(C) \models_{ind} \sigma(D)$.

Proof: (i). Let M be a minimal model for $\varphi \wedge \sigma(A_i)$. As φ is true, M verifies some β_v. So, by the first claim, the truth of $\sigma(A_i)$ implies that A_i holds in the world v.

Moreover, v is a <-minimal world where A_i holds. For, if A_i were true at some u<v, then an obvious proper submodel of M could be extracted which verifies β_u and (hence) $\sigma(A_i)$, contradicting the minimality of M itself. So, by the truth of $A_i \Rightarrow B_i$ in the original possible worlds model W, B_i holds in v, whence $\sigma(B_i)$ holds in M.

(ii). Since $C \Rightarrow D$ fails in the possible worlds model W, there must be some <-minimal world v where C holds, while D does not. Now, take the obvious minimal model M for β_v. As before, M verifies $\sigma(C)$, while falsifying $\sigma(D)$. It suffices to show then that M is a minimal model for $\varphi \wedge \sigma(C)$. So, suppose that some proper submodel M' also verified $\varphi \wedge \sigma(C)$: say, β_u holds, as well as $\sigma(C)$. Then, by the definition of the formulas β, u must be a proper <-predecessor of v, where C is true. But, this would contradict the C-minimality of v in W. ∎

Remark: The above characterization essentially describes a more general form of circumscription, where the universe of models being compared can be restricted by some prior condition φ.

The proof presented here derives from an argument by Frank Veltman, produced in response to an earlier version of this paper.

Finally, we return to the earlier question why circumscription should not also validate new inference patterns, not found in classical logic. Here, by 'inference patterns' we mean schemata of the form

'$\varphi_1 \Rightarrow \psi_1$, ... , $\varphi_n \Rightarrow \psi_n$ *imply* $\varphi \Rightarrow \psi$' .

Here, the φ's and ψ's are Boolean forms - as was the case in all earlier examples, and '$\Rightarrow$' stands for the particular notion of consequence being considered. (Thus, in a sense, we are concerned with the *Horn clause logic* of valid consequence.)

Now, if such a schema is not classically valid, then the following single consequence cannot be valid either:

$(\varphi_1 \rightarrow \psi_1) \wedge \ldots \wedge (\varphi_n \rightarrow \psi_n) \models \varphi \rightarrow \psi$.

Otherwise, the schema would be derivable. So, there exists some valuation V making all $\varphi_i \rightarrow \psi_i$ true ($1 \leq i \leq n$), but $\varphi \rightarrow \psi$ false. Now, consider the following substitution σ of formulas for variables in the above schema:

$\sigma(p) =$ T , if V(p)=1

$\bot$, if V(p)=0

It is easy to check that, then, all $\varphi_i \Rightarrow \psi_i$ ($1 \leq i \leq n$) are valid consequences, whereas $\varphi \Rightarrow \psi$ is not. But, for Boolean combinations of T and $\bot$, circumscriptive

consequence and classical consequence coincide. Hence, the above schema is not valid for circumscriptive consequence either.

We have proved the following

Proposition: All valid Horn clause principles of circumscriptive consequence are also valid for classical logic.

Remark: In this connection, note that at the level of *meta-properties* of logical systems, *non-monotonicity* can indeed occur, and has even been well-known for a long time. A notable example is Intuitionistic Logic, which satisfies the so-called Disjunction Property:

if $\vdash_I \varphi \vee \psi$, then $\vdash_I \varphi$ or $\vdash_I \psi$.

The stronger system of Classical Logic, however, has lost this property:

$\vdash_C p \vee \neg p$, but neither $\vdash_C p$ nor $\vdash_C \neg p$.

2.2.3 Dynamics of Minimality

We now turn to another aspect of circumscription. As was remarked before, what we are dealing with in general are various *modes* of taking propositions, as they come in. For instance, answers to questions could be taken 'at face value', but also 'exhaustively'. And likewise, ordinary statements can be taken routinely, so to speak, or 'minimally'. (In fact, the latter mode itself turned out to have at least three varieties.) Here are a few comments on this situation, preparatory to the discussion of 'dynamics' in Section 3 below.

A class of models may be regarded as a state of information, localizing 'the real world' within its range. The ordinary way of taking propositions φ amounts to the following *transformation* of such states:

$X \mapsto X \cap MOD(\varphi)$.

In general, thus, each successive new premise decreases our ignorance, or equivalently, increases our knowledge. Various logical operators then acquire new overtones in this setting. One example is 'conjunction as composition':

$X \cap MOD(\varphi \wedge \psi) = X \cap (MOD(\varphi) \cap (MOD(\psi))$

$= (X \cap MOD(\varphi)) \cap MOD(\psi)$.

Here, we are only interested in the *general properties* of what may be called a *classical transformation* τ, given by

$\lambda X . X \cap F$,for some fixed class of models F.

The following two properties are easily seen to be necessary:

1 $\tau(X) \subseteq X$ (Introversion)

2 $\tau(\cup\{X_i \mid i \in I\}) = \cup\{\tau(X_i) \mid i \in I\}$ (Continuity)

They are also sufficient:

Any introvert continuous operator on sets can be represented by an intersection with some fixed ('information') set .

The argument is this. Define F as $\{x \mid \tau(\{x\}) = \{x\}\}$.

Then $\tau(X) = \tau(\cup\{\{x\} \mid x \in X\}) = \cup\{\tau(\{x\}) \mid x \in X\} = X \cap F$. ∎

Of course, such operators, when defined by formulas of some specific language , may have many additional structural properties. But, the present analysis provides a convenient comparison with our second non-classical mode of transformation. What happens with circumscription may be described as follows. The universe of models now carries some comparative order R. Starting from a certain information state X, one again restricts attention first to $X \cap MOD(\varphi)$, as before, but next only to all R-minimal items in here. Thus, another basic transformation on knowledge states is the following *minimization:*

$\mu(X) = \{x \in X \mid \neg \exists y \in X\ Ryx\}$.

Can we also determine the basic properties of this new process?

Here are again some obvious features:

1 $\mu(X) \subseteq X$ (Introversion)

Of the earlier Continuity, however, only one half remains:

3 $\mu(\cup\{X_i \mid i \in I\}) \subseteq \cup\{\mu(X_i) \mid i \in I\}$

(The side which is lost here is actually equivalent to *monotonicity* .)

To compensate, we do have a new property, namely,

4 $\cap\{\mu(X_i) \mid i \in I\} \subseteq \mu(\cup\{X_i \mid i \in I\})$.

Again, these three conditions are sufficient:

Any operation on sets satisfying 1,3,4 *may be represented as a minimization operator for some suitable model relation* R.

To see this, define the following relation among models:

Rxy iff $y \notin \mu(\{x,y\})$.

Then we have, for all X, $\mu(X) = \{x \in X \mid \neg \exists y \in X\ Ryx\}$.

From left to right. Let $x \in \mu(X)$, $y \in X$. Then $x \in \mu(\{x,y\} \cup (X - \{x,y\})) \subseteq \mu(\{x,y\}) \cup \mu(X - \{x,y\})$ (by 3). By 1 then, $x \in \mu(\{x,y\})$, and hence $\neg Ryx$ (by definition).

From right to left. Let $x \in X$ with $\neg \exists y \in X\ Ryx$. So for all $y \in X$, $x \in \mu(\{x,y\})$, i.e., $x \in \cap\{\mu(\{x,y\}) \mid y \in X\}$. By 4 then, $x \in \mu(\cup\{\{x,y\} \mid y \in X\})$, i.e., $x \in \mu(X)$. ∎

Minimization has several other interesting properties, derivable from the above. One convenient way to proceed here is by organizing the necessary calculations in a *modal logic*, with the following axioms:

1 $\mu p \rightarrow p$

2 $\mu(p \vee q) \rightarrow \mu p \vee \mu q$

3 $\mu p \wedge \mu q \rightarrow \mu(p \vee q)$.

Formal deduction in this modal logic brings out various useful theorems.

Example: i $p \wedge \mu q \rightarrow \mu(p \wedge q)$:

$\mu q \rightarrow \mu((p \wedge q) \vee (\neg p \wedge q)) \rightarrow$(by 3) $\mu(p \wedge q) \vee \mu(\neg p \wedge q)$, so

$(p \wedge \mu q) \rightarrow (p \wedge \mu(p \wedge q)) \vee (p \wedge \mu(\neg p \wedge q))$; but

$\mu(\neg p \wedge q) \rightarrow$(by 1) $\neg p \wedge q \rightarrow \neg p$, so

$p \wedge \mu q \rightarrow p \wedge \mu(p \wedge q) \rightarrow \mu(p \wedge q)$.

ii $\mu p \rightarrow \mu\mu p$:

$\mu p \wedge \mu p \rightarrow \mu(\mu p \wedge p)$ (by part i) , and

$\mu p \wedge p \leftrightarrow \mu p$ (by 1) , so

$\mu p \rightarrow \mu\mu p$. ∎

We can derive an obvious possible worlds semantics for this modal operator μ. But perhaps the easiest observation is just this:

μ is completely described by the following *definition*

inside the *minimal modal logic* K:

$\mu p := p \wedge \neg p$.

We conclude by mentioning a few facts about the *interplay* of a classical transformation τ with minimization μ. Most iterations *collapse,* via the following equations:

$\tau\tau(X) = \tau(X)$

$\mu\mu(X) = \mu(X)$

$\tau\mu(X) = \tau(X) \cap \mu(X)$

$\tau\mu\tau(X) = \mu\tau(X)$

$\mu\tau(X) = \tau\mu(X)$

Proofs can again be given in the above modal logic.

Example: The last identity can be transcribed as the equivalence

$\mu(\mu p \wedge q) \leftrightarrow \mu p \wedge q$.

Here, $\rightarrow$ follows by axiom 1, and $\leftarrow$ from the preceding example:
$\mu p \wedge q \rightarrow \mu\mu p \wedge q \rightarrow \mu(\mu p \wedge q)$. ❦

Thus, the only non-equivalent modes of transforming knowledge states generated by the present perspective are precisely as expected:

τ, μ and $\mu\tau$.

2.2.4 First-Order Reduction

Circumscriptive inference was already quite complex for predicate-logical sentences, as we have seen. The reason for this lies in its *second-order* nature. Using the minimal models of a set of premises Σ amounts to using all models in the *standard* sense for an associated second-order formula. For individual circumscription, that formula is

$$\forall X\ ((\exists y Xy \wedge \Sigma^X) \leftrightarrow \forall y Xy) \quad ,$$

where Σ^X is the *relativization* of Σ to the subdomain X. [In the presence of function symbols in the language, some more care will actually be needed here.]
For predicate circumscription, the corresponding formula is this [written up, for convenience, for some Σ involving only one unary predicate letter P]:

$$\forall X\ (\forall y(Xy \rightarrow Py) \rightarrow (\Sigma(X) \leftrightarrow \forall y(Py \rightarrow Xy))) \quad ,$$

where $\Sigma(X)$ is obtained by *substituting* X for P in Σ.
Both of these second-order formulas are relatively simple, so-called *monadic* Π^1_1-sentences - but still, their standard notion of consequence is already quite complex.

Now, Vladimir Lifschitz has pointed out that this complexity need not worry us, if we can show that all, or most natural *applications* employ premises falling within some narrower syntactic class of formulas, for which these circumscription formulas are in fact equivalent to *first-order* ones. For then, we can employ standard reasoning systems to simulate circumscription after all. And, he does manage to isolate large such classes in Lifschitz 1985a, 1985b, with respect to predicate circumscription. One useful general result is, e.g., that circumscription with respect to P on Σ will define a first-order class of minimal models, in case P occurs only *positively* in Σ. This covers such cases as the following:

$\forall x(Qx \rightarrow Px)$, circumscribed to: $\forall x(Qx \leftrightarrow Px)$,

$\exists x(\neg Qx \wedge Px)$, circumscribed to: $\exists x(\neg Qx \wedge \forall y(Py \leftrightarrow y=x))$.

Similar questions arise for individual-minimality.

Put in general logical terms:

- What is the complexity of 'first-order definability' for circumscriptive monadic Π^1_1-sentences?
- What are large syntactic classes of premises Σ which guarantee the existence of a(n effectively obtainable) first-order equivalent?

As it happens, there is a connection here with earlier work in *Modal Logic*. Modal axioms on possible worlds structures may be viewed in general as monadic Π^1_1-sentences, and their possible first-orderness has been studied extensively (see van Benthem 1984, 1985a). For instance, we know that the general question if an arbitrary monadic Π^1_1-sentence has a first-order equivalent is of hyper-arithmetical complexity. Even so, there are indeed large classes of first-order cases with an effective syntactic description. And, this analogy can be exploited to prove such results as the following.

Proposition: Individual circumscription is first-order for all first-order formulas Σ which are of the syntactic shape
$\exists x_1 ... \exists x_m \forall y_1 ... \forall y_n$ <quantifier-free matrix >.

Proof: By the obvious preservation properties of such a formula, its minimal models can be enumerated as a finite set of finite models whose size does not exceed m. ∎

Simple though it is, this proposition has a useful consequence for a large class of cases where circumscription is actually used:

Corollary: Individual circumscription is first-order for all Σ formulated in *monadic* predicate logic with identity.

Proof: All formulas in this language have a *normal form* of the above syntactic shape. ∎

First-orderness is no longer guaranteed with other quantifier prefixes, such as $\forall\exists$. Then, first-order cases have to be located by more sensitive analysis.

Remark: There is an interesting boundary case of first-order definability for circumscription. For certain first-order formulas Σ, ordinary models and minimal models already *coincide*. For instance, $\forall x(Qx \leftrightarrow Px)$ has only P-predicate-minimal models: and so do all explicit definitions for minimized predicates in terms of the

remaining vocabulary. But, there are further instances of the phenomenon too, witness the following Σ:

' < is a discrete unbounded strict linear order with immediate successors and predecessors, and $\forall x \forall y((x<y \wedge \neg\exists z(x<z \wedge z<y)) \rightarrow (Qx \leftrightarrow \neg Py))$ ' .

Models of Σ are linear orders of copies of the integers, with P interpreted as either the odd or the even integers in each copy. These models are all P-minimal; but even so, P is not definable in terms of just < on the basis of Σ. [E.g., the odd numbers are not explicitly definable in the pure ordering theory of the integers.] Is there a good characterization of those first-order formulas $\Sigma(P)$ which automatically enforce P-predicate-minimality?

Although the general complexity of detecting first-orderness in modal logic may be high, there is a natural specialization. In many cases, the first-order equivalent of a monadic Π^1_1-sentence $\forall X \varphi(X)$ comes in the form of a conjunction of *first-order substitution instances* $\varphi(\psi)$. Now, the class of second-order formulas having such a first-order equivalent is *recursively enumerable*. For, it suffices to enumerate all possible finite conjunctions of first-order instances, checking in each case if

$\varphi(\psi_1) \wedge \ldots \varphi(\psi_n) \models \forall X \varphi(X)$.

(Note that the other direction is automatic.) As the predicate variable X does not occur on the left-hand side of the turn-stile here, the latter consequence is equivalent to the (recursively enumerable) standard first-order consequence

$\varphi(\psi_1) \wedge \ldots \wedge \varphi(\psi_n) \models \varphi(X)$. ∎

Quite probably, however, this property is not *decidable*.

Again, these considerations may be transferred to the case of circumscription.

This time, we shall consider the case with minimal *predicates*.

Proposition: Having a first-order predicate circumscription is not a decidable, in fact not even an arithmetical property of first-order formulas.

Proof: We give an effective reduction of arithmetical truth to first-orderness of predicate circumscriptions. The proposition then follows by Tarski's Theorem.

Consider a finite relational formulation PA^- for Peano Arithmetic minus the Induction Axiom, including the first-order theory of <. Let φ be an arbitrary

arithmetical sentence in this format. Moreover, let N be a new unary predicate letter not occurring in PA^-.

Claim: The next two assertions are equivalent:

1 $IN \models \varphi$ (i.e., φ is arithmetically true)

2 the following sentence Φ has a first-order predicate circumscription:

$< PA^- \wedge \neg\varphi \wedge$ ' N contains 0, and is S-closed and cofinal in $<$ ' $>$.

Proof: $1 \Rightarrow 2$. Let $\mathbf{N} \models \varphi$. Suppose that D is any model for Φ. Because $D \models PA^-$, it consists of some initial copy of **N** followed by a tail of copies of the integers **Z**. (As $D \models \neg\varphi$, this tail must be non-empty.) Now, the extension of N in D must intersect at least one copy of **Z** (by cofinality). Then, removing one N-point together with all its predecessors in that copy of **Z** from the extension of N will leave Φ true. It follows that Φ has no predicate-minimal models at all; whence its predicate circumscription is defined by the first-order sentence FALSE.

$2 \Rightarrow 1$. Let Φ have a first-order predicate circumscription, say α. Suppose that $\mathbf{N} \not\models \varphi$. We derive a contradiction. The model D consisting of **N** itself, with N interpreted as the set of all natural numbers, is a model for Φ. Moreover, it is even a predicate-minimal model: any decrease in some predicate extension would disturb either the first or the last conjunct of Φ. So, $D \models \alpha$. Now take any proper elementary extension D^+ of D. Still, $D^+ \models \alpha$, but D^+ can no longer be a predicate-minimal model for Φ (which is the desired contradiction), for the same reason as above. ∎

Remark: The idea of this argument is that Φ can only have predicate-minimal models on **N** (if it has them at all). Note that the proof works equally well for general predicate-minimality and for minimality only with respect to the special predicate N.

Corollary: There exist formulas having first-order predicate circumscriptions without the latter being obtainable through first-order substitution instances.

Proof: The first class of formulas contains the second, obviously. Moreover, it is not arithmetical, whereas the second is (being recursively enumerable). Hence, the inclusion must be proper. ∎

2.3 Data Types and Theories

2.3.1 Logic Programs

Next, we shall briefly consider another tradition of studying minimal models in Computer Science. First, there is an extensive literature on semantic properties of *logic programming* (see Lloyd 1985, Apt 1987). Usually, this area carries a restriction to a *fragment* of first-order logic, namely universal *Horn clauses*. This restriction leads to some better logical (and computational) behaviour, such as the existence of *unique* minimal (Herbrand) models. The special status of the Horn Clause formalism has been investigated from various angles. For instance, by an early result of Malcev's, as a fragment of first-order predicate logic, Horn clauses are characterized by two model-theoretic preservation properties: they are preserved under the formation of *submodels* and *direct products* of models. There is also a theorem in Mahr & Makowsky 1983, however, which characterizes the Horn clause framework as the maximal one in which every specification is guaranteed to have an *initial model*. We shall return to the latter notion below.

One of the attractions in having unique minimal models for specifications, from an intuitive point of view, is that it corresponds with a natural tendency which many people have. When reading a piece of text, they assume that some unique world is being built up - and it always takes some time to persuade students in a logic class to shift to the standard technical perspective of huge classes of different models for sentences. (See also van Benthem and van Eyck 1982 on such issues.)

Semantic consequence $\models^+$ for logic programs Σ may be defined either in the standard fashion, or via the corresponding minimal Herbrand Model μ_Σ:

$\Sigma \models^+ \varphi$ if $\mu_\Sigma \models \varphi$.

For *atomic formulas* φ, this choice is irrelevant, as $\Sigma \models \varphi$ will be equivalent to $\mu_\Sigma \models \varphi$. This well-known observation implies correctness and completeness for answers to the usual types of question φ computed in logic programming. But, for general conclusions φ, truth in μ_Σ will be a stronger notion than standard consequence from Σ - and characteristically, it is the latter notion being referred to, when we prove certain *correctness statements* about the behaviour of PROLOG programs. Additional principles, over and above classical consequence, which are available for reasoning in the latter setting reflect the earlier two aspects of minimality: individual minimality sanctions certain forms of *induction,* predicate

minimality rather the use of the so-called *completion* of the relevant program (cf. Lloyd 1985).

The connection with our earlier notion of circumscriptive consequence is found in the following simple observation:

$\Sigma \models \varphi$ only if $\Sigma \models *\varphi$ only if $\mu_\Sigma \models \varphi$.

Again, this is because minimal Herbrand models combine individual and predicate minimality. None of these implications can be reversed, however, for arbitrary formulas φ. For instance, $\Sigma = \{ R0S(0) \}$ will imply $\neg RS(0)0$ in its minimal Herbrand model, whereas it does have a trivial one-point minimal model refuting $\neg RS(0)0$. This difference reflects the decision, when working with Herbrand models, not to identify any more terms than is absolutely necessary given the premises. [Similar proposals have been made for circumscription too ('uniqueness of names') - but we will not go into the latter variant here.]

Another point of comparison between the two approaches to semantic consequence concerns the earlier completion *comp*(Σ) of a logic program Σ. It is easy to show that, for all Horn programs Σ, $\Sigma \models *compl(\Sigma)$. But, in a way, circumscription provides a more stable semantic account here, untroubled by some of the more curious syntactic details of forming completions. This may be seen in the following comparative list:

Σ: $\{p \rightarrow q\}$ *compl*(Σ): $\{q \leftrightarrow p\}$ q-minimal models: $q \leftrightarrow p$,

Σ: $\{p \rightarrow q, q \rightarrow q\}$ *compl*(Σ): $\{q \leftrightarrow (p \vee q)\}$ (i.e.,$p \rightarrow q$!) q-minimal models: $q \leftrightarrow p$.

Digression: This example may be viewed as illustrating the superiority of a *semantic* notion like circumscription over a *syntactic* one like completion: in line with current public opinion. Nevertheless, from a different angle, there is also a good deal of interest to the more syntactic reaction of PROLOG to adding 'harmless' rules like $q \rightarrow q$. As is well-known, such rules can trick the inference engine into entering a vicious loop, unsettling earlier proofs - and that is analogous to a phenomenon to be observed in ordinary argumentation. Someone who commits the informal fallacy of *Begging the Question* , by offering q itself as a reason for q, vitiates whatever else she might have advanced already in favour of q. Various logical commentators have found it hard to explain why this should be so, since the new statement is logically true. But again, the explanation lies in the procedural aspects of argumentation and discourse.

We conclude with another observation about circumscription suggested by logic programming. It follows from an earlier remark that, at least for Horn clause

premises and *atomic* conclusions, classical consequence and circumscriptive consequence *coincide*. One interesting question then becomes if this collapse can be *generalized* to larger fragments of predicate logic. For instance, standard consequence and circumscriptive consequence also coincide for all universal Σ, with positive quantifier-free conclusions φ, at least, in languages without function symbols. [For, if $\Sigma \not\models \varphi$, then there exists a *finite* model for Σ where φ fails - and we can obtain a minimal model for Σ with that same failure by stepwise decreasing domain and interpretation. It is crucial here that φ be positive quantifier-free, so that the truth value of its negation cannot change in this process.]

In fact, the universal fragment of predicate logic has many things to recommend it (see also below). It would therefore be of interest to determine the precise complexity of the notion $\Sigma \models_* \varphi$ as restricted to universal statements.

2.3.2 Abstract Data Types ...

The theory of logic programs is hard to distinguish from that of so-called *abstract data types*. Here again, the emphasis is on special models for specifications Σ, namely those which are *initial* in MOD(Σ). That is, such a model must have a *unique homomorphic embedding* into every other model for Σ. Part of the motivation for this notion again amounts to various forms of minimality (be it now with different slogans; such as 'no junk' or 'no confusion': cf. Goguen and Meseguer 1983). But, there is also the interesting idea that 'minimality' should consist in some homomorphic relation to other models of the specification. This might be worth pursuing for the case of circumscription too. In particular, the approach in the theory of abstract data types characterizes its distinguished objects only *up to isomorphism*. But, from that viewpoint, it would be natural to weaken the notion of, say, individual minimality as follows:

> D is a model for Σ and it has no proper *non-isomorphic* submodels verifying Σ.

In the latter case, e.g., the integers as they stand would be a minimal model for their own ordering theory - which is certainly reasonable.

Another interesting aspect to the theory of abstract data types is the greater attention paid to further *syntactic fine-structure* of premises. For instance, in specifications, it often makes sense to distinguish between 'visible variables', standing for features of the system which are accessible to, or at least observable by, the user, and *hidden variables*, helping to structure the specification, without becoming public. Such a division at once introduces a new range of interesting

logical questions. For instance, given the full specification $\Sigma(L_V, L_H)$, what would be a good independent description of its 'observable content'? The literature contains both syntactical proposals here, such as

$$\Sigma \lceil L_V = \{\varphi \in L_V \mid S \vdash \varphi\} \quad ,$$

and semantical ones, such as

$$MOD(\Sigma) \mid L_V = \{ (D \mid L_V) \mid D \models \Sigma \} \quad .$$

The latter class consists of all L_V-*reducts* of models for Σ. (Or equivalently, it consists of all structures for the observable language which can be *expanded* to models of the full specification Σ.) And accordingly, questions concerning minimal models become more complex too. Notably, how are minimal models for the full Σ related to minimal models for its observable part ?

2.3.3 ... and Scientific Theories

Before going into these questions, this is the proper place to point out yet another wider analogy. Within an entirely different tradition, similar questions have come up in the *Philosophy of Science*. There too, it is customary to analyze *scientific theories* as being formulated in a two-tier language $L_e + L_t$. There is an *empirical* , or observational vocabulary corresponding to what we can observe or measure, on top of which one finds a *theoretical vocabulary* corrresponding to postulated theoretical entities. For instance, in mechanics, observables would include predicates of position, velocity, etcetera, whereas, e.g., forces would be theoretical constructs. In general, the statements of a theory will then fall into three kinds:

purely within L_e : empirical facts and regularities ,
purely within L_t : theoretical axioms ,
mixed L_e, L_t : 'bridge principles' .

For empirical theories, this view goes back to Frank Ramsey and Rudolf Carnap: but, it may even be observed for mathematical theories with David Hilbert (where L_e would be the 'finitistically' computable part).

Many logical questions have been studied in this perspective, often concerning the status of the theoretical terms. (See Przelecki 1969, van Benthem 1982). The *practical* value of introducing such terms is not in question, but one would like to know how far this procedure is necessary *in principle*. The latter question brings us back to the issue of adequate independent definitions of the *empirical content* of a theory, which was already mentioned above. Both the above possibilities have been considered in the literature. In particular, $MOD(\Sigma) \mid L_V$ has a

vivid interpretation now. When *applying*, say, mechanics to some empirical system which we have measured, we have to stipulate suitable mass and force functions so as to satisfy Newton's Laws: in order to start mechanical calculations, which can then be used again to explain, or predict, further empirical facts.

Here is one question which has received a good deal of attention. When are the theoretical terms *eliminable* from a theory Σ, in the sense that $MOD(\Sigma) \restriction L_e$ has a first-order definition purely inside L_e? The answer is that this happens if and only if the operators MOD and restriction commute:

$$MOD(\Sigma) \restriction L_e = MOD(\Sigma \restriction L_e).$$

Such commutation does not always occur (in fact, it is an *undecidable* matter when it happens). What we always have is one inclusion:

$$MOD(\Sigma) \restriction L_e \subseteq MOD(\Sigma \restriction L_e).$$

The converse, however, will only be *guaranteed* to hold for certain special syntactic classes of theories Σ. In particular, we always have it for *universal* theories.

The latter restriction even has some further motivation on the above analysis of theories. As soon as we allow quantifier combinations such as $\forall\exists$ inside the 'observational' language , we are really smuggling in theoretical terms through logical complexity. This may be seen explicitly using *Skolem forms*:

$$\forall x \exists y Rxy \leftrightarrow \exists f \forall x Rxf(x).$$

So, it is quite natural to consider only Skolemized theories, keeping the axioms universal, while making our ontological commitments explicit in the function symbols.

This move will make theories more *algebraic* than has been suggested in our discussion of circumscription. Nevertheless, that would be quite in line with research into logic programs and abstract data types, which is usually confined to almost algebraic formalisms.

Now let us return to the issue of *minimal models*. This has not been very prominent in the philosophical literature (with the earlier-mentioned exception of Hempel 1965). In fact, there might be an interesting new research line into elimination of theoretical terms on minimal models. Here, we shall only consider some special cases - while also slightly changing course.

Let Σ be a universal theory in a language L_e+L_t with identity, where each language contains at least one individual constant. What kind of connection can we expect between minimal models for Σ itself and those for its restriction $\Sigma \restriction L_e$ (only with respect to universal formulas)? We define a new, and perhaps more appropriate, notion of model-theoretic restriction:

For any L_e+L_t-structure D, D || L_e is the L_e-structure whose domain consists of all closed L_e-term interpretations in D, with the inherited interpretation for L_e-functions and predicates.

Then, we can define an operation X || L_e on model classes X, with an obvious meaning. Finally, let the operator μ select (say) individual-minimal models.

Proposition: For Σ and $\Sigma\restriction L_e$ as above,

$(\mu MOD(\Sigma))$ || L_e = $\mu MOD(\Sigma\restriction L_e)$.

Proof: '$\subseteq$'. Let D be a minimal model for Σ. A fortiori, D $\models \Sigma\restriction L_e$. Now, the operation || L_e preserves truth of universal L_e-formulas. So, Δ || $L_e \models \Sigma\restriction L_e$. Moreover, individual-minimality of this model is automatic, because the universe has only definable objects.

'$\supseteq$'. Let D be a minimal model for $\Sigma\restriction L_e$. Consider the following set of formulas:

(Δ) $\Sigma \cup <$ 'all true L_e- atoms or true negated L_e-atoms in D' > .

Every finite subset of Δ has a model. [Otherwise, Σ would classically imply a formula $\neg \bigwedge \partial_i$; where the ∂_i are (negated) L_e-atoms occuring in Δ. But then, this formula would belong to $\Sigma\restriction L_e$ and be true in D, whereas by definition, $\bigwedge \partial_i$ is true in D.] By compactness, then, Δ itself has a model, say M. Now, consider the submodel M* of M whose domain is generated by all closed term interpretations from L_e+L_t. By the universal form of Σ, M*$\models \Sigma$. Moreover, M* is a minimal model for Σ, given its term construction. Finally, D is isomorphic to M* || L_e, since M* satisfies a faithful description of its atomic structure. But then, the required extension of D may be constructed by isomorphic copying from M*. ∎

Remark: This type of analysis can be pushed further. For instance, for arbitrary theories $T_1 \subseteq T$, the following statements can be proved equivalent:

- $(\mu MOD(T))$ || L_1 = $\mu MOD(T_1)$
- T is *m-conservative* over T_1 ,
  i.e., for all quantifier-free L_1-sentences ϕ, T $\models \phi$ only if $T_1 \models_{ind} \phi$.

Finally, we note one further similarity between the theory of abstract data types and the philosophy of science. Philosophers have emphasized the existence of a huge network of scientific theories, connected by various *relations*, such as being an 'extension' of another theory, or being 'interpretable' in it. Again, such relations have also been proposed and studied for abstract data types. In fact, a very

interesting *calculus* of operations on and relations between data types has been developed in Bergstra, Heering and Klint 1986. We conclude with one example, which is relevant to the useful property of *modularity*.

The *sum* $\Sigma_1+\Sigma_2$ of two abstract data types is just the union of their axioms, in their combined language. Again the question arises as to the connection between the minimal models for the various components and for the whole. As before, one direction is easy to establish (again, for the purposes of illustration, we consider the case of individual-minimality):

- if D is a minimal model for Σ,
  then $D \| L_i$ is a minimal model for Σ_i $(i = 1,2)$.

For a converse, we would need an *amalgamation* property:

- If D_1 is a minimal model for Σ_1, and D_2 one for Σ_2 ,
  then there exists some minimal model D for Σ such that $D \| L_i = D_i$ $(i = 1,2)$

This will hold only if Σ satisfies a strong *splitting* condition:

- if $\Sigma \models \varphi_1 \vee \varphi_2$ for quantifier-free L_i-sentences φ_i $(i =1,2)$,
  then $\Sigma_1 \models_{ind} \varphi_1$ or $\Sigma_2 \models_{ind} \varphi_2$.

In such a case, Σ_1 and Σ_2 do not really 'interact' in Σ.
As soon as they do, however, the above may be violated.

Example: $\Sigma_1 = \{Pa \vee Qa\}$, $\Sigma_2 = \{\neg Qa \vee Ra\}$. $\Sigma = \Sigma_1 \cup \Sigma_2 \models Pa \vee Ra$,
but neither $\Sigma_1 \models_{ind} Pa$ nor $\Sigma_2 \models_{ind} Ra$.

The interest in the fine-structure of axioms Σ, both as to different kinds of vocabulary involved and as to various syntactic modules, represents a notable tendency in current research. Against the background of the earlier general semantic results, actual performance of logic programs or data specifications will still be crucially affected by their syntactic design. It is important to bring to light useful structures here. (See also Apt and Pugin 1987 on 'stratified' PROLOG programs.) This is not just a technical concern within computer science. Also in general logic, there is a great need for a better theory of the *structure* of premises, if we are to arrive at a deeper understanding of at present intangible phenomena like good *organization* and structuring of arguments.

3. DYNAMICS

The study of minimal models and non-monotonicity almost forces one to acknowledge the more dynamic aspects of knowledge acquisition and revision. But even so, it is only one strand among various motives pointing in the latter direction. In this Section, we shall also consider dynamics at the level of building up interpretations of single statements, and related forms of information flow.

3.1 Dynamics of Interpretation

'Processing mechanisms' play a pervasive role in programming languages, obviously - but also in natural language. For instance, it seems natural to understand various possibilities and impossibilities in *anaphoric linkage* from an algorithmic perspective. And, such considerations are also appropriate in the analysis of conditionals (cf. Stalnaker 1972) or *bare plurals* when viewed as expressing default rules (cf. Pelletier and Schubert 1985). Here, we shall concentrate on the example of anaphora (see Section 3.1.2).

There are various new semantic theories invented especially for their 'dynamic flavour'; but, in fact, ordinary *predicate logic* itself provides an excellent initial model for studying such phenomena. To see this, we take our starting point in a well-known approach to the semantics of programs, due to Floyd and Hoare.

3.1.1 Operational Semantics for Programs

The basic format of interpretation for predicate logic is Tarski's relational schema

$M, I \models \varphi\ [a]$ ('φ is true in M under I, a');

where M is a model structure, I an interpretation of vocabulary into suitable items of M, φ some formula, and a an *assignment* to the variables occurring freely in φ. Here, the assignment is a modest 'auxiliary interpretation function', needed in order to get a recursion going on the structure of φ. Later on, however, this Cinderella met her prince. In Computer Science, assignments may be viewed as *memory states* of a computer (functions from identifiers / addresses to data values) - and as such, they play the leading role in computationally oriented introductions to Logic (such as Gries 1981).

One fundamental generalization of the Tarski schema to the semantics of programming languages has taken place in so-called *operational semantics* . In addition to the set of descriptive formulas, one also defines a set of programs inductively, and then interprets these according to the following schema (in the context of some model M, I):

[[π]] is the set of successful state transitions associated with the program π. Here are some typical steps encountered for simple program constructions:

(1) $[[x:=t]] = \{ (a,a^{x}_{a(t)} \mid a \text{ any assignment} \}$

(2) $[[\pi_1;\pi_2]] = \{ (a,b) \mid \text{for some } c, (a,c) \in [[\pi_1]] \text{ and } (c,b) \in [[\pi_2]] \}$

(3) $[[\text{IF } \varepsilon \text{ THEN } \pi_1, \text{ ELSE } \pi_2]] = \{(a,b) \mid M,I \models \varepsilon[a] \text{ and } (a,b) \in [[\pi_1]],$ or $M, I \not\models \varepsilon[a] \text{ and } (a,b) \in [[\pi_2]] \}$

Predicate-logical formulas function as static assertions in this context. They may be tests as in (3), or statements of *specification* for programs, or statements about program behaviour, such as the well-known *correctness assertions*

$\{\varphi\}\ \pi\ \{\psi\}$:

for all assigments a such that $M,I \models \phi\ [a]$,

and all assignments b such that $(a,b) \in [[\pi]]$,

it holds that $M,I \models \psi\ [b]$.

('pre-condition ϕ for π implies post-condition ψ') .

Although the semantic format is still very much in the usual spirit, even this application yields its own questions. For instance, the logical meta-theory of this framework derives its interest to a large extent from the fact that its predicate-logical counterpart does not transfer smoothly. Notably, in searching for completeness theorems in this area, it turns out that even the correct *formulation* of what should be regarded as 'completeness' is a tricky issue (cf. Cook 1978, Bergstra and Tucker 1982).

Programming languages, at least the traditional imperative ones, exert dynamic control over a series of actions to be performed by the audience, usually the computer. But, this feature is also very conspicuous in natural languages, where speakers direct their listener's representation of information by means of various textual devices. This process occurs explicitly in *reasoning*, when we set up an argumentative structure using imperatives: 'suppose', 'let', 'take', ... But it is also present at the sentence level, witness the illustration given below. Thus, the slogan of *Language as Action* has become a central one in current semantics, and many proposals for 'dynamic formats' of interpretation have appeared (cf. Kamp 1981, Heim 1982, Seuren 1985).

One phenomenon where dynamic aspects emerge is that of *anaphoric connection.* In actual discourse, pronouns are used to refer back to earlier expressions, in such a way that the listener can pick up links intended by the speaker. There are limits to this process, however: not anything goes - and contemporary dynamic theories are often motivated by the desire to explain these constraints by reference to some processing mechanism. Here are three examples of anaphoric facts whose explanation seems to go beyond predicate logic as ordinarily conceived:

'*A speaker* arrived. *He* was late.' (1)

How can the 'existential quantifier' pick up a pronoun in another sentence? Moreover, there is a clear-cut ordering involved, witness the impossibility of an intended link as in the following sentence:

*'*He* was late. *A speaker* arrived.' (2)

And finally, within a single sentence, consider the pattern:

'If *a speaker* arrives, *he* is late.' (3)

This is traditionally transcribed as follows:

$\forall x\ ((Sx \wedge Ax) \rightarrow Lx)$,

interchanging the conditional with the existential quantifier. (It is as if the, invalid, prenex law $(\exists x \phi(x) \rightarrow \psi(x)) \leftrightarrow \forall x(\phi(x) \rightarrow \psi(x))$ had been applied.) Why cannot this sentence be interpreted correctly *as it stands*: its obvious form being

$(\exists x(Sx \wedge Ax) \rightarrow Lx)$?

These problems are solved in the cited work by Kamp and Heim through the intermediary of a level of 'discourse representation', in between syntactic form and eventual truth conditions. But, as was pointed out convincingly in Barwise 1987, they can be handled equally well without such a move, by adopting a more dynamic format of interpreting syntactic forms, inspired by the earlier operational semantics.

3.1.2 Operational Semantics for Assertions

A very clean presentation of this idea arises when the language interpreted is predicate logic *itself*, as is shown in Groenendijk & Stokhof 1987. Thus, the earlier 'static' assertion language for programs itself receives a dynamic interpretation:

$[[Px]] = \{(a,a) \mid M,I \models Px\ [a]\}$:

i.e., atomic formulas are just tests without special side-effects.

$[[\varphi \wedge \psi]] = [[\varphi]] \circ [[\psi]]$:

i.e., conjunction is treated like composition ($\circ$ or ;).

$[[\exists x\varphi]] = \{(a,b) \mid (c,b) \in [[\varphi]] \text{ for some } c \sim_x a\}$:

i.e, existential statements are made true by finding some witness to their matrix.

These stipulations explain the first two anaphoric facts mentioned above. $\exists xAx \wedge Lx$ will have the same interpretation as $\exists x(Ax \wedge Lx)$, but not as $Lx \wedge \exists xAx$. In general, the effect of the above clauses is to widen scopes toward the right.

An explanation of the third phenomenon observed requires a treatment of conditionals. We start with a preliminary notion, however:

$[[\neg \phi]] = \{ (a,a) \mid \text{for no } b, (a,b) \in [[\varphi]] \}$.

Thus, a negation is a strong denial, without side-effects. Then, a clause for implications $\varphi \rightarrow \psi$ may be derived via their traditional equivalence with $\neg(\varphi \wedge \neg\psi)$, or by direct introspection:

$[[\phi \rightarrow \psi]] = \{ (a,a) \mid \text{for all } b \text{ with } (a,b) \in [[\varphi]], \text{ there exists } c$ with $(b,c) \in [[\psi]] \}$.

Finally, we introduce universal quantification, again without side-effects:

$[[\forall x\varphi]] = \{ (a,a) \mid \text{for all } b \text{ with } a \sim_x b, (a,b) \in [[\varphi]] \}$.

Then, it may be checked that the following two formulas indeed have the same associated transition relation:

$\exists xAx \rightarrow Lx$ and $\forall x(Ax \rightarrow Lx)$.

Thus, the anaphoric sentence 3 is vindicated as it stands.

On the other hand, no binding across the conditional occurs with the related syntactic form

$\neg\exists xAx \rightarrow Lx$.

For, the negation in the antecedent 'envelops' the scope of the existential quantifier. But this is as it should be. After all, there is no such link in natural language either, witness the incorrect sequence

*'If *no speaker* arrives, *he* is late.'

It should be admitted at once that this account still has many empirical weaknesses. For instance, it describes anaphoric facts in natural language only indirectly, being dependent on a *translation* into predicate logic. Moreover, those anaphoric facts are much more diverse and subtle than may be apparent from the four examples given here. But, the account does show how standard predicate logic, far from being an obstacle to recognizing the role of dynamical interpretation, can actually be a good testing ground for theorizing about it.

In fact, there is still a close connection between ordinary predicate logic and its interpreted variant. In a sense, the latter amounts to reading 'ordinary ' formulas with a different *scope convention* , extending scopes of existential quantifiers as far

as possible toward the right, until one hits the boundary of an operator like negation, which 'seals off' its subformula. (For a more general discussion of different scoping conventions, as expressing various strategies of interpretation for one single language, cf. van Benthem, 1986b).

More formally, there is a *reduction* to ordinary predicate logic which works as follows. For each formula φ, with free variables $x_1,....,x_n$, the following induction defines a predicate TRANS $(\varphi, x_1,...,x_n, y_1,....,y_n)$
[intuitively, '$(x_1,...,x_n ; y_1,...,y_n)$ is a successful φ-transition between partial assignments']:

$$\begin{aligned}
&\text{TRANS } (Px_i;\ x_i,y_i) &&= && y_i=x_i \wedge Px_i \\
&\text{TRANS } (\varphi\wedge\psi;\ \bar{x},\bar{y}) &&= && \exists\bar{z}\ (\text{TRANS } (\varphi;\ \bar{x},\bar{z}\ ) \wedge \text{TRANS } (\psi;\ \bar{z},\ \bar{y})) \\
& && && \text{(here the } \bar{z} \text{ are new free variables)} \\
&\text{TRANS } (\exists x_j\varphi;\ \bar{x},\bar{y}) &&= && \exists z_j\ \text{TRANS } (\varphi;\ \bar{x}(z_j/x_j),\bar{y}) \\
&\text{TRANS } (\neg\varphi;\ \bar{x},\bar{y}) &&= && \bar{y} = \bar{x} \wedge \forall\bar{z}\neg\text{TRANS } (\varphi;\ \bar{x},\ \bar{z}).
\end{aligned}$$

By an obvious induction, TRANS $(\varphi;\ \bar{x},\bar{y})$ defines the successful transitions for φ. Therefore, any central semantic notion of dynamic predicate logic can be reduced to static assertions in this way.

Nevertheless, the new formalism as it stands does seem to correspond more closely to *practical* uses of predicate logic, or semi-formalisms employed in mathematical prose, where we do tend to use scoping conventions closer to the one mentioned above.

Another attraction of the dynamic framework is that it suggests taking a fresh look at defininining connectives, and logical operators generally. As has often been observed, logical operators do not have meaning only: they also exert control. (For instance, Jennings 1986 points at the use of "and" and "or" as sequencing, or more generally *punctuation* devices.) What we can do in the new setting, for instance, is to define both 'dynamic' order-dependent versions of connectives, and more classical 'parallel' ones, studying their interplay. One instructive example is provided by *conjunction.* The natural 'classical' stipulation, in terms of *intersection* of successful transition sets, will now express a different option from the preceding one: namely, a requirement of parallel execution. And similarly, a classical negation in terms of complement will suddenly acquire a new operational significance.

Finally, the new system also suggest several ways of defining *valid consequence* as arising from successive processing of premises. One natural candidate is the following:

$\varphi_1,\ldots,\varphi_n \models \psi$ if, for all models M,I and assigments $a_1,\ldots a_{n+1}$
such that $(a_1,a_2) \in [[\varphi_1]], \ldots, (a_n, a_{n+1}) \in [[\varphi_n]]$,
there exists an assignment b with $(a_{n+1},b) \in [[\psi]]$.

This notion of consequence, like the ones presented in Section 2, lacks several of the main structural properties of the standard one (be it for different reaons).
It is *not monotone* for instance:

$\exists xAx \models Ax$; but $\exists xAx \wedge \exists x\neg Ax \not\models Ax$.

But this time also, it even lacks such simple 'domestic' properties as insensitivity to *permutation*, or *contraction* of identical premises:

$\exists xAx \wedge Lx \models \exists x(Ax \wedge Lx)$; but $Lx \wedge \exists xAx \not\models \exists x(Ax \wedge Lx)$,

$\exists xAx \wedge \exists x\neg Ax \wedge \exists xAx \models Ax$; but $\exists xAx \wedge \exists x\neg Ax \not\models Ax$.

Still, as in Section 2, certain special cases will remain valid - including the following leftward form of monotonicity:

$\varphi \models \psi$ implies $\chi, \varphi \models \psi$.

Once again, the divergences from classical logic here arise from reasons different from those in Section 2. We have defined 'logical consequence' rather close to one particular interpretation algorithm, and are now feeling its effects. Whether this has been a wise policy, will be discussed further in Section 4 below.

3.2 Dynamics of Information Flow

Changing assignments means no more than changing our links with a certain model. We have been studying the 'dynamics of adjustment', so to speak. But already in Section 2, we also encountered the dynamics of changing *information*. The latter perspective is currently receiving a good deal of attention too. For instance, in the philosophy of science, there has been work on the dynamics of changing *theories*, which has also issued in more general epistemic studies of various operations on knowledge states: in particular, *addition* and *retraction* of information (see Gärdenfors 1988.)

There are at least two ways of thinking about the dynamics of changing information states. One is the classical perspective: common to both standard logic and such less standard frameworks as e.g. possible worlds semantics. Here, incoming information is treated as *reduction* of the space of a priori *possibilities* - as was done already in Section 2, when discussing 'classical' versus 'minimal' transformations on information states. The other perspective takes some more primitive notion of epistemic state, in particular, one in which *partial information* need not be represented by a cloud of all possible total (world) extensions. And

then, 'propositions' can act as more abstract operators on such states. We shall consider both approaches.

3.2.1 Reducing Possibilities

First, the method of 'eliminating uncertainty' is easy to implement, witness current folklore. Here is one particular example, due to Veltman 1987 (but compare also Heim 1982, and others). Consider a *modal propositional language* with the ordinary Boolean connectives as well as modality $\Diamond$ ("might"). Let U be a set of 'possibilities' (say, ordinary valuations), for which it makes sense to call an atom p true or false. Then we can define, for each formula φ, a corresponding transformation $[[\varphi]]$ on subsets of U:

$$[[p]](X) = \{x \in X \mid x \text{ verifies } p\}$$

$$[[\varphi \wedge \psi]](X) = [[\varphi]](X) \cap [[\psi]](X)$$

$$[[\varphi \vee \psi]](X) = [[\psi]](X) \cup [[\psi]](X)$$

$$[[\neg\varphi]](X) = X - [[\varphi]](X)$$

and

$$[[\Diamond \varphi]](X) = \begin{cases} X & \text{if } [[\varphi]](X) \neq \varnothing \\ \varnothing & \text{otherwise} \end{cases}$$

If desired, one might add a sequential conjunction as before:

$$[[\varphi ; \psi]](X) = [[\psi]]([[\varphi]](X))$$

For purely propositional formulas φ, it is easy to see that $[[\varphi]](X)$ merely amounts to intersecting X with the truth range of φ in U computed in the standard fashion. But, already with the modal operator, some interesting phenomena occur, once *consequence* is introduced on the analogy of the previous subsection:

$\varphi_1, ..., \varphi_n \models \psi$ if,

for all X, $[[\varphi_1; ... ; \varphi_n]](X) \subseteq [[\psi]](X)$.

Note, e.g., that $\Diamond\neg p;p$ will be a consistent sequence, implying p, whereas its permutation $p;\Diamond\neg p$ is *inconsistent.* As Veltman argues, this reflects the facts of life for our ordinary use of the epistemic modality "might". And there are various other applications of this formal system too.

Also as before, the new dynamic consequence can be embedded in 'static' standard logic. For instance, in a sense, the above little system is a part of *monadic* predicate logic. To see this, assign unary predicate letters P (uniquely) to each proposition letter p, while also taking one distinguished unary predicate letter X. By

induction, we translate each formula φ into a syntactic operation $\Delta(\varphi)$ on monadic formulas $\alpha = \alpha(x)$ having a free variable x:

$$\begin{aligned}
\Delta(p)\,(\alpha) &= \alpha \wedge Px \\
\Delta(\varphi\wedge\psi)\,(\alpha) &= \Delta(\varphi)\,(\alpha) \wedge \Delta(\psi)\,(\alpha) \\
\Delta(\varphi\vee\psi)\,(\alpha) &= \Delta(\varphi)\,(\alpha) \vee \Delta(\psi)\,(\alpha) \\
\Delta(\neg\varphi)\,(\alpha) &= \alpha \wedge \neg\, \Delta(\varphi)\,(\alpha) \\
\Delta(\Diamond\varphi)\,(\alpha) &= (\exists y\,[y/x]\,\Delta(\varphi)\,(\alpha)) \wedge \alpha \\
\Delta(\varphi;\psi)\,(\alpha) &= \Delta(\psi)\,(\Delta(\varphi)\,(\alpha))
\end{aligned}$$

Evidently, this is a direct transcription of the above 'truth definition' into a simple language quantifying over states in U. Thus, we have the following reduction:

$\varphi_1,\ldots,\varphi_n \models \psi$ if and only if

$\forall x(\Delta(\varphi_1;\ldots;\varphi_n)(Xx) \rightarrow \Delta(\psi)(Xx))$ is valid in monadic predicate logic.

As a bonus, *decidability*, the *finite model property* and other desirable features are immediate for the new dynamic logic.

3.2.2 Building Up Information

Next, implementation of the second approach can actually go in many directions, since we can structure 'knowledge states' in many different ways. (The first approach may in fact be *defended* as being an elegant way of avoiding such decisions: cf. Stalnaker 1986.) In particular, we have to specify the 'grain size', as it were. Are knowledge states like sets of *sentences*, with all their syntactic peculiarities? Or, should we smoothen these somewhat by thinking of deductively closed *theories* in some logic? Could we work on the analogy of *Beth tableaus,* or should we steer away from their particular notational structure, as is done in *Hintikka model sets*? Slight differences in presentation may now become logically significant.

We shall not go into any particular proposal here. Rather, we want to point out a certain *danger*, of merely revamping existing systems. Suppose that we choose the approach via deductively closed theories (as is suggested by Gärdenfors' treatment). Let us work in ordinary classical propositional logic. (Note that this decision itself determines what will be 'deductively closed' theories.) The main idea then becomes just this:

the action of any formula φ on any 'state' T consists in

forming the deductive closure of $T \cup \{\varphi\}$.

Now, these theories come in an obvious relation of inclusion :

$T_1 \sqsubseteq T_2$ if $T_1 \subseteq T_2$.

Moreover, they form a *distributive lattice,* with respect to the available operations of supremum $\bigsqcup$ and infimum $\bigsqcap$ on theories. Then, one can set up a recursive definition of operators $[[\varphi]]$ by merely providing a convenient decomposition:

$$[[\varphi \wedge \psi]](T) = [[\psi]]([[\varphi]](T)) = [[\varphi]](T) \bigsqcup [[\psi]](T) \quad ,$$
$$[[\varphi \vee \psi]](T) = [[\varphi]](T) \bigsqcap [[\psi]](T) \quad .$$

And other connectives could be treated by postulating additional structure on the set of knowledge states (e.g., for implication, one would have to make it into a Heyting algebra). But obviously, nothing much would be gained in this way. So, there is a danger (to be avoided) of trivially achieving a 'dynamic' presentation.

Another approach might be to start from a very abstract notion of *dynamic information structure,* studying specializations as they arise. For instance, the general type might be this:

$$(S, \sqsubseteq, \{\tau_p \mid p \in P\}),$$

where S is a set of 'information states', ordered by increasing strength ($\sqsubseteq$), on which a certain family of transformations acts, indexed by propositions. In other words, we assume a perspective from *Group Theory.*

This perspective makes it easy to ask systematic questions. One is how much structure should be imposed on the transformations. Evidently, they should form a *semi-group* under composition. But, should they also be a *group*: i.e., should there be an *inverse* to every proposition (its 'retraction')? [Gärdenfors 1988 has rather looked at this structure of transformations from the point of view of *Category Theory*, demanding the presence of *equalizers*, modelling equivalence between propositions.] But also, this question interacts with the possible structure to be imposed on information states: should they form a lattice, a Heyting Algebra or even a Boolean Algebra? Then, it would be natural to have corresponding closure conditions on transformations too. E.g., there should be some operation on them such that

$$(\tau_p \bigsqcap \tau_q)(x) = \tau_p(x) \bigsqcap \tau_q(x) \quad , \text{for all } x \in S.$$

Once a class of transformations with certain closure properties has been chosen, we can bring up various additional questions. For instance, what are natural subclasses of transformations satisfying additional mathematical requirements? One natural example are *idempotent* operators, satisfying the condition

$$\tau_p \circ \tau_p = \tau_p \ .$$

This is certainly a very plausible logical requirement too. Another reasonable condition would be *monotonicity* , in the sense of respecting growth of information:

$x \sqsubseteq y$ only if $\tau_p(x) \sqsubseteq \tau_p(y)$, for all $x,y \in S$.

This certainly holds for 'classical' propositions; but also, e.g., for all modal propositions in Veltman 1987 (where $\sqsubseteq$ is identical with $\supseteq$).

Here is one elementary observation in this vein.

Proposition: If S is a Boolean Algebra, and $\{\tau_p \mid p \in P\}$ a set of idempotent operators which are also distributive, i.e.,

$\tau_p(x \sqcap y) = \tau_p(x) \sqcap \tau_p(y)$, for all $x,y \in S$,

then the whole information structure can be represented by a set structure of the 'eliminative' kind described earlier.

Proof: (Compare also the analysis of set transformations given in Section 2.)
For each state $s \in S$, set

$\bar{s} = \{x \mid x \sqsubseteq s\}$.

For each proposition $p \in P$, set

$\bar{p} = \{x \mid \tau_p(x) = x\}$.

(This is the common idea of identifying propositions with their *fixed points*.)
Then it is easy to prove that

(1) $x \sqsubseteq y$ iff $\bar{x} \supseteq \bar{y}$

(2) $\overline{\tau_p(s)} = \bar{s} \cap \bar{p}$. ∎

But, one could also try to apply more sophisticated mathematical results here, on the representation of algebras or lattices with certain types of endomorphisms. (Compare Jónsson and Tarski 1951/2.)

Finally, one basic attraction of the present framework is that it also suggests *new* types of question, beyond the classical case. For instance, where transformations are around, *invariants* cannot be far away (cf. van Benthem 1985b). A relation R between states is invariant with respect to our class of propositions if, say,

$(x,y) \in R$ iff $(\tau_p(x), \tau_p(y)) \in R$, for all $p \in P$.

Is there an interesting invariant structure on S?

3.2.3 Relational Calculus and Categorial Grammar

Nevertheless, there are also other possibilities for abstract information structures. Perhaps, propositions are not really functions on information states, but rather *relations,* which can also take no value, or more than one. For instance,

certain propositions might embody choices, leaving several options. In that case, one natural framework would be to view propositions as forming an algebra in the sense of the *Relational Calculus*, with basic operations such as composition, intersection, union, etcetera. (Recall the earlier discussion of the dynamics of changing assignments.)

One basic question in this perspective too, is what would be the natural operations on propositions. This amounts to asking for some principled account of operations on binary relations. (For an algebraic study of these matters, see Jónsson 1984.) One way of doing this employs the framework of *Type Theory*. For instance, there is a natural notion of *logicality* for this type of operator, in terms of invariance for permutations of states (cf. van Benthem 1987b). Moreover, again, we can introduce additional useful conditions, such as *continuity*, in the sense of respecting arbitrary unions of families of relations. All possibilities can then be classified (cf. van Benthem 1987a), and they form a neat basic set, including the above-mentioned examples.

Proposition: The logical continuous operators on binary relations are exactly those defined by a schema of the form

$\lambda R.\lambda xy.\exists uv.\ Ruv \wedge$ 'some Boolean condition on identities involving x,y,u,v'.

Examples are *converse:*

$\lambda R.\lambda xy.\exists uv.\ Ruv \wedge x= v \wedge y=u,$

or *diagonal*:

$\lambda R.\lambda xy.\exists uv.\ Ruv \wedge x=y=u=v.$

The result is easily specialized to operations from binary to unary relations; with a typical example such as *projection*:

$\lambda R.\lambda x.\exists uv.\ Ruv \wedge x=u.$

It can also be generalized to n-ary operations on binary relations, bringing in (typically) disjunction or *conjunction*:

$\lambda RS.\lambda xy.\exists uv.\exists zw.\ Ruv \wedge Szw \wedge x=u=z \wedge y=v=w\ ,$

or *composition:*

$\lambda RS.\lambda xy.\exists uv.\exists zw.\ Ruv \wedge Szw \wedge x=u \wedge v=z \wedge w=y.$

This kind of systematic classification is important if we are to bring some order into the plethora of possibilities for 'dynamic' logical operators.

Another interesting question is how such a relational structure would translate into a corresponding system of *inference* . Here, we can follow a suggestion arising from Orlowska 1987, and exploit an analogy with *Categorial Grammar*. The logic will be very much like a so-called 'Lambek Calculus', as developed in that area (cf. van Benthem 1986a, 1987a). The main idea can be explained as follows:

- basic propositions p are interpreted as binary relations R_p
- complex propositions p.q are interpreted through the composition $R_p \circ R_q$
- disjunctions may be interpreted by unions

But, what to do about *implications*?

Here, Orlowska notes that two quite plausible 'implicational' operators may be used, introduced recently by Tony Hoare:

$$R\backslash S = \cup\{X \mid R \circ X \subseteq S\}$$
$$S/R = \cup\{X \mid X \circ R \subseteq S\} \quad ,$$

which describe 'weakest pre- and post-specifications'. (This continues the well-known work on pre- and post-conditions in the operational semantics of programming languages. But see also Jónsson 1984 for a purely algebraic introduction of \ and /.) Accordingly, the dynamic perspective suggests introducing *two* implications, one searching for its argument on the left-hand side, and one searching on the right. [A similar idea has been suggested independently by Gordon Plotkin (private communication).] But, this is precisely standard practice in Categorial Grammar, which has developed systems of proof for directed types

$$a\backslash b \quad \text{and} \quad b/a.$$

For instance, the basic *Lambek Calculus* is the following Gentzen-type system:

Axiom: $a \Rightarrow a$

Rules:

$$\frac{X \Rightarrow a \qquad Y,b,Z \Rightarrow c}{Y,X,a\backslash b,Z \Rightarrow c} \qquad \frac{X \Rightarrow a \qquad Y,b,Z \Rightarrow c}{Y,b/a,X,Z \Rightarrow c}$$

$$\frac{X, a \Rightarrow b}{X \Rightarrow b/a} \qquad \frac{a,X \Rightarrow b}{X \Rightarrow a\backslash b}$$

$$\frac{X \Rightarrow a \quad Y \Rightarrow b}{X,Y \Rightarrow a.b} \qquad \frac{X,a,b,Y \Rightarrow c}{X,a.b,Y \Rightarrow c}$$

Derivable sequents in this system L include $a.a\backslash b \Rightarrow b$, but typically exclude $a\backslash b.a \Rightarrow b$. [Here, we have deviated somewhat from the standard version, in allowing *empty* sequences on the left. As a consequence, we can derive, e.g., $(e/e)\backslash t \Rightarrow t$.]

Proposition: The Lambek Calculus L is sound for the above relational calculus.
That is, if the sequent $a_1,...,a_n \Rightarrow b$ is L-derivable, then,
for any assignment of binary relations R_x to primitive types x,
with R_a for complex a computed as above,
$R_{a1} \circ ... \circ R_{an} \subseteq R_b$.

It would be very interesting to have a *converse* too (for the operations . , \ and /, that is.) Thus, we would establish a link between basic logics of categories in natural language and plausible systems of 'dynamic logic'.

3.2.4 Propositional Dynamic Logic

Indeed, there is also a profitable connection to be found with 'Dynamic Logic' in the usual sense of that phrase (see Harel 1984). In that research program, one adopts an enriched modal logic having both *propositions* and *programs*, in which the latter denote transition relations between states, whereas the former stand for functions from states to truth values (their more traditional role). One interesting feature is that there are operators mapping one into the other: *modalities* take programs to operate on propositions, whereas a *test* operator takes propositions to programs. This turns out to be a third convenient abstract framework to explore for present purposes.

Note that here, a possibility comes to the fore which has hitherto been neglected. We have been treating propositions *themselves* as being operations on information states. But in fact, we may want a separation of concerns: into *propositions* expressing a certain informational content, and various *modes of transforming* states (which can *use* certain propositional contents, to be sure). For

instance, the above *test* operator is one such mode, which checks if a state has a certain property, but then leaves it as it is. Another operator might be *addition,* which, given a propositional content, transforms any state into a minimal extension (as measured along some prior inclusion relation among states) having that content. Thus, we are now interested in Propositional Dynamic Logic for its potential as a dynamic logic of propositions, rather than programs.

Again, the general situation here can be analyzed in type theory. Propositional dynamic logic has primitive types t (for truth values) and s (for states). Propositions have the functional type

(s,t) ('from states to truth values') ,

while programs have

(s,(s,t)) ('from pairs of states to truth values').

The above 'switching modes' will then be operators in the type

((s,t), (s,(s,t)).

As before, it makes sense to ask for *logical* items here, being those which are invariant for permutations of states. [Note the formal similarity with the earlier relational calculus case of the converse type ((e,(e,t)), (e,t)).] And in fact, the test operator is logical, while also satisfying the earlier special requirement of *continuity*. Also as before, we can classify all possibilities of the latter kind in the schema

$\lambda P.\lambda xy.\exists u.$ Pu $\wedge$ 'Boolean condition on identities involving x,y,u'.

As a first attempt, one might consider that fragment of dynamic logic which only has basic programs of the form $?\varphi$ (where ? is the test operator) and then the usual program operations

; (sequencing), $\cup$ (choice) and * (iteration) .

But, this will reduce to ordinary propositional logic, because of such equivalences as the following :

$<\pi_1 \cup \pi_2>\varphi \quad \leftrightarrow \quad <\pi_1>\varphi \vee <\pi_2>\varphi$

$<\pi_1;\pi_2>\varphi \quad \leftrightarrow \quad <\pi_1><\pi_2>\varphi$

$< ?\alpha >\varphi \quad \leftrightarrow \quad \alpha \wedge \varphi$

$(?\varphi)^* \quad = \quad I$ (the identity map) .

So, it becomes imperative to add some further structure, as was suggested above. That is, the set D_S of states will now carry a binary inclusion relation. Then,

we add two further modes of handling propositions. First, the mode of *addition* may be defined as follows:

$+\varphi$ is the relation $\{(s,s') \mid s \sqsubseteq s'$ and φ is true at s' and,
either φ is true at s and s'=s, or φ is false at s and φ is false
at all states in between s and s'$\}$.

There is an analogy here with the "until" operator of Tense Logic.

This then suggests having also a dual mode of *subtraction* (compare "since"):

$-\varphi$ is the relation $\{(s,s') \mid s' \sqsubseteq s$ and φ is false at s' and,
either φ is false at s and s'=s, or φ is true at s and φ is true
at all states in between s' and s$\}$.

Of course, the *test* operator ? retains its original definition:

$?\varphi$ is the relation $\{(s,s) \mid \varphi$ is true at s $\}$.

In addition to these three modes + , - , ? , it will also be useful to have a general modality referring to possible extensions along $\sqsubseteq$ in the usual way:

$\Box\varphi$ is true at x if φ is true at all y with $x \sqsubseteq y$,

$\Diamond\varphi$ is true at x if φ is true at some y with $x \sqsubseteq y$.

[Given our dual set-up, it would also make sense to introduce similar 'downward' operators.]

This logic will validate some basic principes for proposition-based transformations.

For instance, all three are *idempotent:*

$?\varphi \,;\, ?\varphi \quad = \quad ?\varphi$

$+\varphi \,;\, +\varphi \quad = \quad +\varphi$

$-\varphi \,;\, -\varphi \quad = \quad -\varphi$

But, there are also other validities, such as

$\varphi \rightarrow (\langle+\varphi\rangle \psi \leftrightarrow \psi)$

$\neg\varphi \rightarrow (\langle-\varphi\rangle \psi \leftrightarrow \psi)$,

or the related

$?\varphi \,;\, +\varphi \quad = \quad ?\varphi$

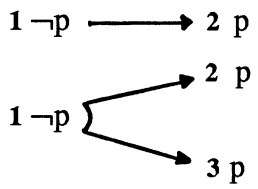
$?\neg\varphi;\, -\varphi \quad = \quad ?\neg\varphi$.

Some non-validities may be worth observing too:

$+\varphi \,;\, -\varphi \quad \neq \quad I$

$-\varphi \,;\, +\varphi \quad \neq \quad I$,

where I is the identity transformation. The reason is shown in the following two pictures of possible information structures:



: (2,2) ∈ [[+p]], (2,1) ∈ [[-p]],

: (2,1) ∈ [[-p]], (1,3) ∈ [[+p]].

Mutatis mutandis, this formalism can also be used to analyse various axioms concerning theory extension and theory revision found in Makinson 1987 or Gardenfors 1988, two studies whose abstract perspective is quite congenial to the one presented here.

The general logic of the dynamic propositional language with only programs formed using $?\varphi$, $+\varphi$, $-\varphi$ and ; , $\cup$, * is probably still *effectively axiomatizable* . [Without the iteration *, this will even be automatic, because there is a *standard translation* then of the semantic truth conditions into a first-order language: compare van Benthem 1984.]

Other questions concerning this logic are of a more general model-theoretic nature. In particular, as before, one can also study special types of transformation here. For instance, tests $?\varphi$ always denote a *(partial) identity function* on states, which does not really 'move forward'. But, so do certain additive cases, such as

$+\Diamond p$,

where $\Diamond$ is the above general modality. If a program $+\varphi$ always denotes a partial identity function, then this must mean (under a reasonable assumption of *well-foundedness* for $\sqsubseteq$) that we never encounter $s \sqsubseteq s'$, φ true in s' but false in s. In other words, φ is *downward-persistent*, in an obvious sense. But, such formulas φ will be equivalent to $\Diamond\varphi$. So, the above example was indeed characteristic. Thus, we get an interplay between semantical properties of transformations and well-known properties of their underlying propositions.

A further specialization would arise if we were to restrict the available propositions beforehand to cases that are *upward- or downward persistent* (or perhaps, merely *convex*).

Another important line of specialization arises, if we impose further restrictions on the structure of states $(S, \sqsubseteq)$. For instance, it might be reasonable (as earlier on) to regard this as a, possibly distributive, *lattice* - a lattice which might even be closed under arbitrary joins and meets. (The earlier case of deductively closed theories is one example.) In the latter case, it becomes possible to make propositions uniformly into *functions*, instead of relations. Namely, one stipulates that

$+\varphi$ maps any state s to the meet of the set

$\{s' \mid s \sqsubseteq s'\} \cup \{ s' \mid \varphi$ is true at $s'\}$ (if that meet exists).

And $-\varphi$ is defined similarly by a join of suitable $\sqsubseteq$-predecessors.

Evidently, the logic of this scheme will be richer than before.

In dynamic logic terms, programs $+\varphi$ and $-\varphi$ have now become *deterministic*.

Digression: A Connection with Minimality

Studying the logic of our dynamic propositional operators involves the earlier issue of *minimal modelling* (see Part II). For, $+\varphi$ may be regarded as a minimized case of a more liberal operator

PLUS φ = $\{ (s,s') \mid s \sqsubseteq s'$ and φ is true at $s' \}$.

Such an operator would reduce to a general modality as follows:

$<\text{PLUS } \varphi>\psi \leftrightarrow \Diamond(\varphi \wedge \psi)$.

But, due to the minimality in the truth clause for $+\varphi$, no similar reduction appears feasible for the latter notion. There is no natural general modal equivalent for the statement $<+\varphi>\psi$. The only plausible reduction would be the special case

$<+\varphi>T \leftrightarrow \Diamond\varphi$;

which will hold provided that we demand *well-foundedness* for the extension relation $\sqsubseteq$.

What further principles are reasonable for the modality $<+\varphi>$?

Of course, there are the standard modal axioms, such as *Distribution:*

$<+\varphi>(\alpha \vee \beta) \leftrightarrow (<+\varphi>\alpha \vee <+\varphi>\beta)$.

For further axioms, again, we can turn toward earlier systems of minimal logic in Part II: notably, those presented in Sections 2.2.2 and 2.2.3. For instance, conditionals $\varphi \Rightarrow \psi$ may now be compared with modal formulas

$[+\varphi]\ \psi$.

Of the principles of the minimal conditional logic, then, rightward Monotonicity and Conjunction follow directly from the Distribution axiom. The remaining ones, however, require various additions, such as

Reflexivity: $[+\varphi]\varphi$.

Disjunction of Antecedents: $([+\varphi]\psi \wedge [+\chi]\psi) \rightarrow [+(\varphi \vee \chi)]\psi$.

Deriving the final principle '$\varphi \Rightarrow \psi, \varphi \Rightarrow \chi$ / $(\varphi \wedge \psi) \Rightarrow \chi$' will require a suitable strengthening of the above Wellfoundedness principle.

Thus, properly viewed, the present dynamic logic encompasses the minimality-based logics of our earlier investigation.

Finally, there are still aspects of dynamic interpretation which go beyond the present framework. For instance, commitment to an *implication* $\varphi \rightarrow \psi$ in ordinary discourse is often explained as obeying a *standing instruction* (compare the rules of a Lorenzen dialogue game):

'as soon as one becomes committed, in further discourse or argument, to the antecedent φ, the obligation to add ψ is incurred'.

In the above terms, this would rather say something about the *internal structure* of a transition process. A successful transition (s,s') gives the two extremal points of a finite sequence of intermediate steps:

$s = s_1 \text{ -- } s_2 \text{ -- } \ldots \text{ -- } s_n = s'$

which may obey, e.g., the constraint that, whenever s_i validated φ, the next step $s_i \text{ -- } s_{i+1}$ was an addition of ψ. [An alternative would be to program standing instructions explicitly, via Kleene *iterations*.] This move towards representing a richer internal structure of processes can also be observed in ordinary Dynamic Logic, namely, when one passes on to some kind of *Process Algebra* (cf. Bergstra and Klop 1984).

Thus, this final perspective too offers us a versatile model for studying the dynamics of information flow.

4. DISCUSSION

The preceding Sections contain enough material to warrant the conclusion that dynamic information structure eminently deserves further logical study. Indeed, the literature already shows some quite promising explorations. Nevertheless, there are also some difficulties to be observed.

One is a kind of general paradox of *complexity*. Although most systems in the area seem inspired by a desire for modelling simple efficient reasoning (of which our own 'domestic reasoning' is taken to be a paradigmatic example), the actual proposals produced seem to go in the wrong direction. This was striking with Circumscription, which produces a notion of consequence whose complexity is vastly higher than that of classical logic. (Compare the conjecture in van Benthem 1985c, that Tarski's classical notion of consequence is distinguished, among all its rivals, as being the *only* one to produce an effectively axiomatizable set of validities). But also, the dynamic systems of Section 3 produce logics whose inferential behaviour, upon closer inspection, is not at all as perspicuous as that of standard predicate logic.

Of course, there may be ways-out here. Perhaps, in actual practice, we shall always avoid the complex cases (compare the Lifschitz analysis of 'elementary' circumscription) - and there may be even a system to that practice which can be brought to light.

But, in some ways, these observations also raise the issue of whether we are going about things in the right fashion. For instance, Franz Guenthner has suggested that we should distinguish between (at least) *two* kinds of inference. One is 'on-line', close to the original linguistic structure of premises, and may be more 'dynamic'. The other operates on more abstract informational content, even after we have forgotten the specific original linguistic formulations. The latter may be closer to classical standard logic.

The latter point is reinforced by the difficulties encountered in analyzing operations on 'information states', where one finds that the properties of particular *representations* chosen tend to get in the way of logical understanding. Perhaps, what is needed is a clear separation of concerns. [The earlier division into propositional contents and modes of transforming information states was already one attempt in this direction.] In fact, many authors have warned that a genuine deep theory of any phenomenon should stay at a reasonable level of abstraction from its particular *implementations*. (Compare recursion-theoretic theories of

computation vis-à-vis actual algorithmics, or the plea for a procedurally *neutral* approach to semantics in Fenstad et al. 1987).

This issue of where to locate what also reminds us of another broad alternative to the *semantic* approach taken throughout in this paper. Many issues treated here really play at the syntactic level of *text and discourse*. For instance, the earlier example of 'retracting' a proposition is difficult to grasp semantically: how to recover a unique source for a given value of a transformation? But, this is a self-inflicted perplexity. At the text level, we *know* what was the previous proposition, and there is no difficulty at all in retracting *it*. And similar considerations apply to such 'modifying speech acts' as 'conditionalizing', which require undoing the effect of some previous transformation:

"P. If Q, that is ..."

Similarly, the earlier *modes* of taking propositions are a simple fact of life at this level. We know in a certain discourse whether we are going to take a certain answer exhaustively (see Section 2) or in the normal sense. For instance, when reading a certain Danish children's book to my sons, called "Auntie Thea", I arrived at a passage where questions are posed and answered. Here is one example:

"What was Auntie Thea wearing?"

"A red hat."

On the theory of Groenendijk and Stokhof 1985, who always predict the exhaustive reading, this would hardly be a children's book ... . But of course, my sons realized the 'language game' we were in , and asked *what else* she was wearing.

So, we may have to look for a theory of text structure and discourse, as indicated by *control expressions*, such as "so", "but", "suppose", "let", etcetera. And the proper model for that may be not so much Semantics as *Proof Theory*, with its account of the structure of proofs: themselves already a nice and rich example of textual phenomena. For instance, there is already a fair amount of anaphoric relations, and dynamic dependency structures in Natural Deduction arguments. (For more technical applications of proof theory, to circumscription, cf. Jaeger 1986.)

But even more generally, we also need a pragmatic theory at the level of logical *language games,* explaining the ease with which we adopt or switch certain modes of reasoning behaviour. (Compare the discussion at the end of Section 2.1.) Logic should not just be about the 'forms' which are the products of reasoning, but also about the 'rules' which guide that activity.

We still have a long way to go.

5. REFERENCES

1. K. Apt, 1987, 'Introduction to Logic Programming',
report TR-87-35, Department of Computer Sciences,
The University of Texas, Austin.
2. K. Apt and J.-M. Pugin, 1987, 'Maintenance of Stratified Databases
Viewed as a Belief Revision System',
Laboratoire d'Informatique, Ecole Normale Superieure, LIENS-87-1, Paris.
3. L. Åqvist, 1984, 'Deontic Logic',
in D. Gabbay and F. Guenthner, eds.,
Handbook of Philosophical Logic, vol.II.,
Reidel, Dordrecht, 605-714.
4. J. Barwise, 1987, 'Noun Phrases, Generalized Quantifiers and Anaphora',
in P. Gärdenfors, ed., 1987, 1-29.
5. D. Batens, 1986, 'Dialectical Dynamics within Formal Logics',
Logique et Analyse 29, 161-173.
6. J. van Benthem, 1982, 'The Logical Study of Science',
Synthese 51, 431-472.
7. J. van Benthem, 1983, *The Logic of Time,*
Reidel, Dordrecht , (Synthese Library, vol. 156).
8. J. van Benthem, 1984, 'Correspondence Theory',
in D. Gabbay and F. Guenthner, eds.,
Handbook of Philosophical Logic, vol. II.,
Reidel, Dordrecht, 167-247.
9. J. van Benthem, 1985a, *Modal Logic and Classical Logic,*
Bibliopolis / Naples and The Humanities Press / Atlantic Heights (N.J.)
10. J. van Benthem, 1985b, 'Situations and Inference',
Linguistics and Philosophy 8, 3-9.
11. J. van Benthem, 1985c, 'The Variety of Consequence,
According to Bolzano',
Studia Logica 44:4, 389-403.
12. J. van Benthem, 1986a, *Essays in Logical Semantics,*
Reidel, Dordrecht, (Studies in Linguistics and Philosophy, vol. 29).
13. J. van Benthem, 1986b, Logical Syntax,
Report 86-06, Institute for Language, Logic and Information,
University of Amsterdam. (To appear in *Theoretical Linguistics.*)

14. J. van Benthem, 1987a, Categorial Grammar and Type Theory, report 87-07, Institute for Language, Logic and Information, University of Amsterdam. (To appear in *Linguistics and Philosophy*.)
15. J. van Benthem, 1987b, 'Logical Constants across Varying Types', to appear in the *Notre Dame Journal of Formal Logic*.
16. J. van Benthem, 1988, *A Manual of Intensional Logic*, Center for the Study of Language and Information, Stanford / Chicago University Press, Chicago. (Second, revised edition of 1985.)
17. J. van Benthem and J. van Eyck, 1982, 'The Dynamics of Interpretation', *Journal of Semantics* 1, 3-20.
18. J. Bergstra, J. Heering and P. Klint, 1986, 'Module Algebra', research report CS-R8617, Centre for Mathematics and Computer Science, Amsterdam.
19. J.A. Bergstra and J.W. Klop, 1984, 'Process Algebra for Synchronous Communication', *Information and Control* 60, 109-137.
20. J. Bergstra and J. Tucker, 1982, 'Expressiveness and the Completeness of Hoare's Logic', *Journal of Computer and System Sciences* 25, 267-284.
21. J. Burgess, 1981,'Quick Completeness Proofs for Some Logics of Conditionals', *Notre Dame Journal of Formal Logic* 22, 76-84.
22. C.C. Chang and H.J. Keisler, 1973, *Model Theory*, North-Holland, Amsterdam.
23. S. Cook, 1978, 'Soundness and Completeness of an Axiom System for Program Verification', *SIAM Journal of Computing 7*, 70-90.
24. J-E Fenstad, P-K Halvorsen, T. Langholm and J. van Benthem, 1987, *Situations, Language and Logic*, Reidel, Dordrecht.
25. P. Gärdenfors, ed., 1987, *Generalized Quantifiers. Linguistic and Logical Approaches*, Reidel, Dordrecht.
26. P. Gärdenfors, 1988, *Knowledge in Flux: Modelling the Dynamics of Epistemic States*, Bradford Books / The MIT Press, Cambridge (Mass.).

27. J. Goguen and J. Meseguer, 1983,
'Initiality, Induction and Computability',
CSL Technical Report 140, SRI International, Menlo Park.

28. D. Gries, 1981, *The Science of Programming*,
Springer, Berlin.

29. J. Groenendijk and M. Stokhof, eds., 1981, *Truth, Interpretation and Information,*
Foris, Dordrecht, (GRASS Series, vol. 2).

30. J. Groenendijk and M. Stokhof, 1985, *On the Semantics of Questions and the Pragmatics of Answers,*
dissertation, Filosofisch Instituut, Universiteit van Amsterdam.
(To appear with Oxford University Press.)

31. J. Groenendijk and M. Stokhof, 1987, 'Dynamic Predicate Logic',
Institute for Language, Logic and Information, University of Amsterdam.

32. D. Harel, 1984, 'Dynamic Logic',
in D. Gabbay and F. Guenthner, eds.,
Handbook of Philosophical Logic, vol. II.,
Reidel, Dordrecht, 497-604.

33. W. Harper, R. Stalnaker and G. Pearce, eds., 1981, *Ifs*,
Reidel, Dordrecht.

34. I. Heim, 1982, *The Semantics of Definite and Indefinite Noun Phrases,*
dissertation, Department of Linguistics,
Massachusetts Institute of Technology.

35. C. Hempel, 1965, *Aspects of Scientific Explanation*,
The Free Press, Glencoe (Ill.).

36. G. Jaeger, 1986, Some Contributions to the Logical Analysis of Circumscription,
report 86-03, Mathematik, Eidgenössische Technische Hochschule, Zürich.

37. R. Jennings, 1986, 'Logic as Punctuation',
in W. Leinfellner and F. Wuketits, eds., 1986.

38. B. Jonsson, 1984, 'The Theory of Binary Relations',
Department of Mathematics, VanderBilt University, Nashville (Tenn.).

39. B. Jonsson and A. Tarski, 1951, 'Boolean Algebra with Operators. I',
American Journal of Mathematics 73, 891-939.

40. B. Jonsson and A. Tarski, 1952, 'Boolean Algebras with Operators. II',
American Journal of Mathematics 74, 127-162.

41. H. Kamp, 1981, 'A Theory of Truth and Semantic Representation',
in J. Groenendijk and M. Stokhof, eds., 1981, 1-41.

42. W. Leinfellner and F. Wuketits, eds.,1986, *The Tasks of Contemporary Philosophy.10th Wittgenstein Symposium, Kirchberg*,
Verlag Hölder-Pichler-Tempsky, Vienna.

43. D. Lewis, 1973, *Counterfactuals*,
Blackwell, Oxford.

44. V. Lifschitz, 1985a, 'Computing Circumscription',
Proceedings IJCAI-85:1, 121-127.

45. V. Lifschitz, 1985b, 'Pointwise Circumscription',
Proceedings AAAI-86:1, 406-410.

46. J.W. Lloyd, 1985, *Foundations of Logic Programming*,
Springer, Berlin.

47. B. Mahr and J. Makovsky, 1983, 'Characterizing Specification Languages which Admit Initial Semantics',
Proceedings 8th CAAP, Springer, Berlin.

48. D. Makinson, 1987, 'On the Status of Recovery',
Journal of Philosophical Logic 16, 383-394.

49. J. McCarthy, 1980, 'Circumscription -
A Form of Non Monotone Reasoning',
Artificial Intelligence 13, 295-323.

50. E. Orlowska, 1987, 'Relational Interpretation of Modal Logic',
Department of Informatics, Polish Academy of Sciences, Warsaw.

51. J. Pelletier and L. Schubert, 1985,
[bare plurals and dynamic interpretation]
Department of Philosophy, University of Edmonton.

52. M. Przelecki, 1969, *The Logic of Empirical Theories*,
Routledge and Kegan Paul, London.

53. P. Seuren, 1985, *Discourse Semantics*,
Blackwell, Oxford.

54. Y. Shoham, 1988, *Reasoning about Change*,
The MIT Press, Cambridge (Mass.).

55. E. Sosa, 1975, *Causation and Conditionals*,
Oxford University Press, Oxford.

56. R. Stalnaker, 1972, 'Pragmatics',
in D. Davidson and G. Harman, eds., *Semantics of Natural Language*,
Reidel, Dordrecht, 380-397.

57. R. Stalnaker, 1986, 'Worlds and Situations',
Journal of Philosophical Logic 15:1, 109-123.

58. F. Veltman, 1986, *Logics for Conditionals*,
dissertation, Filosofisch Instituut, Universiteit van Amsterdam.
(To appear with Cambridge University Press.)

59. F. Veltman, 1987, 'Update Semantics',
Institute for Language, Logic and Information, University of Amsterdam.