Proceedings of the SNS Logic Colloquium
March 1990

Edited by Peter Schroeder-Heister

Tübingen 2022

Preface

The proceedings of the informal logic colloquium held in March 1990 at the Seminar für natürlich-sprachliche Systeme (SNS) of the University of Tübingen have been available initially as a printed report and later as an internet resource that could be downloaded from my website. As many of their results remain interesting, and as there have been occasional references to these papers, they are republished here with a regular DOI to facilitate access and citation.

Tübingen, June 2022

Peter Schroeder-Heister

Contents

SNS-Bericht 90-58

Proceedings of the SNS Logic Colloquium, March 1990

Peter Schroeder-Heister (ed.)

August 1990

Seminar für natürlich-sprachliche Systeme
der Universität Tübingen
Biesingerstr. 10
D-7400 Tübingen

Informal Logic Colloquium

Universität Tübingen, March 21-22, 1990

Seminar für natürlich-sprachliche Systeme (SNS), Biesingerstr. 10, D-7400 Tübingen

Final Program

Wednesday, 21 March

10.00 - 11.00	Lew Gordeew : Quasi-Ordinals and Proof Theory
11.15 - 12.00	Jörg Hudelmaier: Bounds for Cut Elimination in Intuitionistic Propositional Logic
12.00 - 12.30	Lars Hallnäs: Lambda-Calculus without Alpha-Conversion
14.30 - 15.30	Gerd Döben: Non-Theorems
16.00 - 16.45	Klaus Schulz: Makanin's Algorithm
16.45 - 17.30	Nick Asher: Intentional Paradoxes and Propositional Quantification
17.30 - 18.15	John Derrick: Deductive Databases, Many-Sorted Logic and the Model Theory of Horn Clauses

Thursday, 22 March

10.00 - 11.00	Dov Gabbay/Ruy J.G.B. de Queiroz: Extending the Curry-Howard-Tait Functional Interpretation to Relevance, Linear and Other Logics
11.15 - 12.00	Hans Leiß: Polymorphic Constructs in Natural and Programming Languages
12.00 - 12.45	Urs Egli: Unification Grammar and the Language Theory of AUTOMATH
14.30 - 15.00	Lars Hallnäs: The Structure of Definitions
15.00 - 16.00	Kosta Došen: Rudimentary Kripke Models
16.30 - 17.15	Cesar Mortari: Valuation Semantics for Modal Logics
17.15 - 18.00	Michael Morreau: A Modal Semantics for Genericity: Exorcising the Ghost from the Machine

Preface

In March 1990 an informal logic colloquium was held at the Seminar für natürlich-sprachliche Systeme (SNS) of the University of Tübingen. It was not devoted to a specific topic but covered various logical issues, presented by researchers from Tübingen and elsewhere. The style of these proceedings is informal, as was the colloquium - some authors submitted extended abstracts whereas others submitted drafts of papers or full papers.

The order of papers in this volume corresponds to the order of presentation at the colloquium.

This is a prepublication, so authors retain full copyrights.

Tübingen, June 1990

Contents

BOUNDS FOR CUT ELIMINATION IN INTUITIONISTIC PROPOSITIONAL LOGIC

Jörg Hudelmaier

Synopsis

The central theorem of Gentzen's theory of proofs states that every deduction d (in classical or intuitionistic, propositional or quantifier logic) can be tranformed into a deduction G(d) which does not make use of the *cut rule*. Avoiding the use of a particular proof rule will, obviously, have the effect that G(d) becomes longer than d, and Gentzen's algorithm for cut elimination establishes an upper bound for the length l(G(d)) of G(d) (GENTZEN [35].) In this article, I shall construct a (different) cut free deduction J(d) for the case of intuitionistic propositional logic and derive considerably sharper upper bounds for l(J(d)). Also, I shall use the methods developed for this purpose in order to set up an effective decision method.

Author's address:

Jörg Hudelmaier
Seminar für natürlich-sprachliche Systeme
Universität Tübingen
Biesinger Str. 10
D 7401 Tübingen
West Germany
Phone: (049) 7071-294279
Email: NFHUJA1 @ DTUZDV1 BITNET/EARN

Gentzen's upper bound for $l(G(d))$ depends on both the length $l(d)$ and the *cut degree* $g(d)$ of d, viz. the maximum of the degrees, increased by 1, of cut formulas used in d; it has the form

$$l(G(d)) \leq 2^{2^{\cdot^{\cdot^{2^{g(d)}}}}} \quad (g(d)\ 2\text{'s}) \quad .$$

The reason for the appearance of these hyperexponentials lies in the nature of Gentzen's algorithm which proceeds by a double induction on both $n = l(d)$ and $g = g(d)$ with respect to the lexicographical order of pairs (n,g). It follows from results of WILKIE–PARIS [87] that in the case of quantifier logic the enormity of these bounds cannot be avoided: there is no constant c such that $2_c^{\,n}$ would be an upper bound independent of the cut degree of the particular d.

On the other hand, it is known that in the case of classical propositional logic a deduction d can be transformed into a cut free deduction K(d) such that already

$$l(K(d)) \leq 2^{2^{g(d)} \cdot l(d)} \cdot l(d)$$

will hold (cf. GORDEEV [87]). This leads to the question whether a similar improvement over Gentzen's non elementary bound is also possible in the case of the intuitionistic propositional calculus LJ. I shall answer this question affirmatively by defining an operator J acting on LJ-deductions d with cut and producing cut free LJ-deductions J(d) such that

$$l(J(d)) \leq 2^{2^{2^{2 \cdot g(d)}} \cdot l(d)} \cdot l(d)$$

Having completed this task, I shall apply the methods used for these constructions to attack the problem of giving an efficient, easily implementable decision procedure for intuitionistic propositional logic. This procedure will be based on a special calculus for which, by its internal structure, every sequence of backward applications of its rules must terminate after a number of steps which is in a simple way bounded by the complexity of the starting sequent. This gives a tableau like procedure which may be implemented by a straightforward depth first search program.

The method to be used will obviously be different from Gentzen's lexicographic induction, namely I shall apply the technique of *inversion rules*. Inversion rules for sequent calculi (especially the classical sequent calculus) are well known, and a strategy to use them in order to eliminate cuts from a classical deduction d is described e.g. GORDEEV [87].

Now for the classical calculus we have a complete and well known list of inversion operators for all connectives on both sides of a sequent:

I∧L	$M \Longrightarrow u \wedge v, N$	$M \Longrightarrow u, N$
I∧R	$M \Longrightarrow u \wedge v, N$	$M \Longrightarrow v, N$
I∨	$M \Longrightarrow u \vee v, N$	$M \Longrightarrow u, v, N$
I→	$M \Longrightarrow u \to v, N$	$M, u \Longrightarrow v, N$
E∧	$M, u \wedge v \Longrightarrow N$	$M, u, v \Longrightarrow N$
E∨L	$M, u \vee v \Longrightarrow N$	$M, u \Longrightarrow N$
E∨R	$M, u \vee v \Longrightarrow N$	$M, v \Longrightarrow N$
E→L	$M, u \to v \Longrightarrow N$	$M \Longrightarrow u, N$
E→R	$M, u \to v \Longrightarrow N$	$M, v \Longrightarrow N$

(Here the left column denotes the name of an operator, the middle column shows the sequent it is applied to and the right column gives the resulting sequent.) For intuitionistic logic, however, a calculus is used which has sequents with only one formula on the right hand side. So the **I∨**- and **E→L**-rules are not available because they would produce nonvalid sequents. As for **I∨**, used on the left premiss of a cut with a disjunction, its missing is not serious, and the removal of the cut formula from the subdeduction leading to that premiss can be managed by the familiar methods. The case of **E→L** is a different matter, and for this operator I shall substitute three new ones which perform the transformations made in HUDELMAIER [87] :

if $M,(u\wedge v)\to w \Longrightarrow r$ is derivable then so is $M,u\to(v\to w) \Longrightarrow r$
if $M,(u\vee v)\to w \Longrightarrow r$ is derivable then so is $M,u\to w,v\to w \Longrightarrow r$
if $M,(u\to v)\to w \Longrightarrow r$ is derivable then so is $M,u,v\to w \Longrightarrow r$.

Clearly, these transformations act precisely on the sequents which **E→L** would be applied to, and appropriate applications of the cut rule will show them to be correct. In order to turn them into real inversion rules, I have to construct operators **E→∧**, **E→∨** and **E→→** which transform *cut free* deductions of the left sequents into *cut free* deductions of the respective right sequents. This I do in three lemmas, and it turns out, that these operators may be defined in such a way that **E→→** does not increase the length of the given deduction, while **E→∧** and **E→∨** at most double this length.

These transformations having been constructed, it is easily seen that **E→∧** and **E→∨** may be used for reducing cuts in the same way as the well known classical inversion rules. As for **E→→**, its application is only useful in certain situations. But in this case there is a special transformation which reduces all other cases to this particular one. The only remaining case, therefore, is when the left immediate subformula of our implication is an atomic formula.For this type of formula we do not have any inversion rule at all, but in this case also all the various forms of deductions which may arise, may be reduced to a single possibility, which can be handled in a straightforward way.

Carrying out all these transformations will, obviously, increase the lengths of deductions, and so I will need recursion parameters in order to measure them. I begin by defining a new degree function *deg* for formulas:

$$\begin{aligned}
deg(v) &= 2 && \text{if v is atomic,}\\
deg(u\wedge v) &= deg(u)\cdot(1+deg(v))\\
deg(u\vee v) &= 1+deg(u)+deg(v)\\
deg(u\to v) &= 1+deg(u)\cdot deg(v).
\end{aligned}$$

If d is a deduction and V is a (full) branch of d then I set

$j(V)$:= *Sum* of the degrees of cut formulas on V

and

$j(d)$:= *Maximum* of $j(V)$ for all branches of d .

The reason behind these particular choices is simply the fact that the function j defined in this manner depends only linearly on the length of the deduction and will decrease under application of the operator **RED** to be defined now.

THEOREM 1 There is an operator **RED** converting every deduction d of a sequent s with

$$0 < j(d)$$

into a deduction **RED**(d) of s such that

$$j(\mathbf{RED}(d)) < j(d)$$

which at most doubles the length of d : $l(\mathbf{RED}(d)) \leq 2\cdot l(d)$.

From this result immediately follows

THEOREM 2 There is an operator J converting every deduction d of a sequent s into a deduction J(d) of this same sequent such that

$$J(d) \text{ is cut free and } l(J(d)) \leq 2^{j(d)} \cdot l(d).$$

and since for the traditional cut degree g of a formula v we have $deg(v) \leq 2^{2^{2g}}$ and since the maximal number of cut formulas on a branch of a deduction d is bounded by $l(d)$, this gives the result mentioned above.

Turning now to the problem of giving a decision procedure for the calculus LJ, I introduce the calculus LH which works with the same sequents as LJ and has the same axioms as LJ, namely all sequents of the form $M,v \Longrightarrow v$ and whose rules are

(HI∧) $$\frac{M \Longrightarrow u \qquad M \Longrightarrow v}{M \Longrightarrow u \wedge v}$$

(HE∧) $$\frac{M,u,v \Longrightarrow r}{M,u \wedge v \Longrightarrow r}$$

(HI∨L) $$\frac{M \Longrightarrow u}{M \Longrightarrow u \vee v}$$

(HI∨R) $$\frac{M \Longrightarrow v}{M \Longrightarrow u \vee v}$$

(HE∨) $$\frac{M,u \Longrightarrow r \qquad M,v \Longrightarrow r}{M,u \vee v \Longrightarrow r}$$

(HI→) $$\frac{M,u \Longrightarrow v}{M \Longrightarrow u \to v}$$

(HE→a) $$\frac{M,a,v \Longrightarrow r}{M,a,a \to v \Longrightarrow r}$$ [a atomic]

(HE→∧) $$\frac{M,u \to (v \to w) \Longrightarrow r}{M,(u \wedge v) \to w \Longrightarrow r}$$

(HE→∨) $$\frac{M,u \to w, v \to w \Longrightarrow r}{M,(u \vee v) \to w \Longrightarrow r}$$

(HE→→) $$\frac{M,u,v \to w \Longrightarrow v \qquad M,w \Longrightarrow r}{M,(u \to v) \to w \Longrightarrow r}$$

This calculus has the property that for every one of its rules with premisses s' and s'' and conclusion s we have $deg(s') < deg(s)$ and $deg(s'') < deg(s)$, where $deg(s)$ is the sum of all $deg(v)$ for the formulas v of s. Thus every sequence of backwards applications of rules of LH starting with a particular sequent s breaks off after at most $deg(s)$ steps. Therefore the calculus LH has the properties promised above. All that remains is to prove that it is equivalent with LJ.

While it is easily seen that every LH-deduction may be transformed into an LJ-deduction, the converse part needs two further lemmas dealing with properties of LJ:

LEMMA There is an operator EP acting on cut free LJ-deductions d such that all the impure instances of (E→) have the same principal formula u→v. EP(d) has the same endsequent as d, is cut free, contains only pure instances of (E→) and satisfies $l(EP(d)) \leq 2 \cdot l(d)$.

(Here an instance of (E→) with principal formula u→v is called *pure* if

u is atomic and the left premiss is an axiom, or

u is not atomic and the left premiss is the conclusion of an I–rule with principal formula u.)

LEMMA There is an operator **Q** acting on cut free and pure LJ–deductions. Q(d) is an LH-deduction with the same endsequent s as d satisfying

$$l(Q(d)) \leq 2^{deg(s)} \cdot l(d).$$

Together with lemma 4, theorem 2 and the above remark on transforming LH-deductions into LJ-deductions this shows

THEOREM 3 The calculi LJ and LH are equivalent: every LH-deduction is, essentially, an LJ-deduction, and if P is the iteration of EP then the operator D = P∘Q transforms every cut free LJ-deduction d into an LH-deduction Q(P(d)).

REFERENCES

GENTZEN [35] Gentzen,G.: Untersuchungen über das logische Schliessen. *Math. Zeitschr.* **39** (1935), 176–210, 405–431

GORDEEV [87] Gordeev,L.: On Cut Elimination in the Presence of Peirce Rule. *Arch. math. Logik* **26** (1987), 147–164

HUDELMAIER [87] Hudelmaier,J.: *A Prolog Program for Intuitionistic Logic*. Seminar für natürlich–sprachliche Systeme SNS–Bericht 88–28

WILKIE–PARIS [87] Wilkie A.J., J.B. Paris: On the Scheme of Induction for Bounded Arithmetic Formulas. *Annals of Pure and Applied Logic* **35** (1987), 261–302

RELATED WORK

DOŠEN [87] Došen,K.: A Note on Gentzen's Decision Procedure for Intuitionistic Propositional Logic. *Zeitschr. f. math. Logik und Grundlagen d. Math.* **33** (1987), 453–456

FRANZEN [87] *Algorithmic Aspects of Intuitionistic Propositional Logic*, Swedish Institute for Computer Science SICS Research Report R87010B

Makanin's Algorithm - A Survey and a Reformulation

Klaus U. Schulz
Seminar für natürlich-sprachliche Systeme (SNS)
Universität Tübingen
Biesingerstr. 10, D-7400 Tübingen, W.-Germany
E-mail: nfseja1@dtuzdv1.bitnet

Abstract: Makanin's algorithm [Mak] shows that it is decidable whether a word equation has a solution or not. Makanin's decision procedure is extremely complex and was not designed for a direct implementation. But, since word equations offer a fascinating tool for the general treatment of strings efforts were made to improve the situation (Pécuchet [Péc], Jaffar [Jaf], Kościelski-Pacholski [KoP]) and an implementation is available now (Abdulrab [Abd]). We give a short introduction to the algorithm and present then a pre-algorithm which allows a very simple partial analysis of word equations. In some cases this analysis is even complete, in the remaining cases it influences and simplifies the remaining search which follows then Makanin's strategy. In combination with another result (to be described in a forthcoming paper [Sch]) we find a solution S of a solvable equation E of length d now after at most I transformation steps following d steps of the pre-algorithm, where I is the sum of the lengths of the components X_i of S.

Introduction and Background

A *word equation* is an expression of the form $\sigma_1 \dots \sigma_k == \sigma_{k+1} \dots \sigma_{k+l}$ $(k, l \geq 1)$ with letters σ_i from a two-sorted alphabet $\mathcal{X} \cup \mathcal{C}$ where $\mathcal{X} = \{x_1, \dots, x_n\}$ is a finite set of variables and $\mathcal{C} = \{c_1, \dots, c_m\}$ $(m \geq 1)$ is a finite set of coefficients. A *solution* (unifier) is an assignment of words X_i in the coefficient alphabet (the combined alphabet) to the variables x_i such that both sides of the equation become identical when all occurrences of variables are replaced by these words. Thus, $(X, Y) = (ab, b)$ is a solution of the equation $ayy == xb$ with variables x and y, for example. A forerunner of Makanin's algorithm is Plotkin's semi-decision procedure for a-unification ([Plo], see also Siekmann [Sie]). Since some ideas of this procedure are relevant for Makanin's algorithm and in particular for our pre-algorithm we want to sketch the main idea. Suppose we want to find the solutions of $ayy == xb$. This is an equation with head (a, x). For any solution (X, Y) with nonempty words X and Y, either $X = a$ or $X = aX_1$. Accordingly we may try to solve the two successor equations $ayy == ab$ and $ayy == ax_1b$. But now it amounts to the same to solve $yy == b$ or $yy = xb$. (We use x instead of x_1 as variable. Thus the replacements are $x \to a$ and $x \to ax$.) Equations of the latter type with two variables at the head have three successors corresponding to the possibilities $X = Y$ (replacement $x \to y$), $X = YX_1$

$(x \to yx)$, $Y = XY_1$ $(y \to xy)$. In our case we get the equations $y == b$, $y == xb$ and $yxy == b$. Thus, new variable names are always avoided and all replacement steps are followed by a second step where the resulting two identical symbols at the head are erased. It should be clear on intuitive grounds that the procedure establishes the solvability of any solvable equation. If variables have several occurrences, however, it might happen that the transformation steps lead to equations which are larger and larger and the algorithm may not terminate in the unsolvable case. The equation $xyaxby == yxbyax$, for example, has the successor $xyayxby == yxbyayx$ corresponding to the case $X = YX_1$, and if we continue to apply the same replacement $x \to yx$, then the number of symbols will grow and eventually exceed any given bound. A similar *combinatorial explosion* will occur in every straightforward semi-decision procedure and for this reason it was an open problem for years whether the solvability of word equations is decidable or not. Returning to Plotkin's procedure we may observe, however, that there is an *important subcase* where the length of any successor equation cannot exceed the length of the original equation: if no variable occurs more than twice, then at most two new symbols may result from a replacement step. Then, at the second step, two symbols at the head are erased. Thus the length cannot grow, only a finite number of equations may occur in the search tree. We may now stop every branch as soon as we find an equation which is isomorphic to a predecessor. (The argument is, roughly, the following: whenever we would find a solution, following such a path, then there exists a similar solution in a different path starting at the predecessor.) Following this strategy solvability can be established by means of a finite search tree.

1 Makanin's Algorithm - A Survey

How is it possible to restrict the combinatorical explosion? Makanin's decidability result is strongly based on the following theorem of Bulitko [Bul]:

Theorem 1: *If a solvable equation has length d, then the equation has a minimal solution where exponent of periodicity satisfies* $s \leq (6d)^{2^{d^4}}$.

A solution $S = (X_1, \ldots, X_n)$ is *minimal* if the length of the word $X_1 \ldots X_n$ is minimal with respect to the class of all solutions. The *exponent of periodicity* of S is the maximal number of periodical (consecutive) repetitions of a non-empty subword in a component X_i of the solution. It was Makanin's central idea to use instead of word equations a new type of structures, the so-called position equations, which allow

- to encode constraints which lead to a lower bound for the exponent of periodicity of arbitrary solutions and
- to put the combinatorial explosion (which may be the result of the analysis via iterated transformation steps) into these constraints.

For any number k the number of all (relevant) position equations with associated lower bound $b < k$ is finite. Bulitko's theorem allows to exclude all position equations where b is to large - the search space becomes finite. For the rest of this section we want

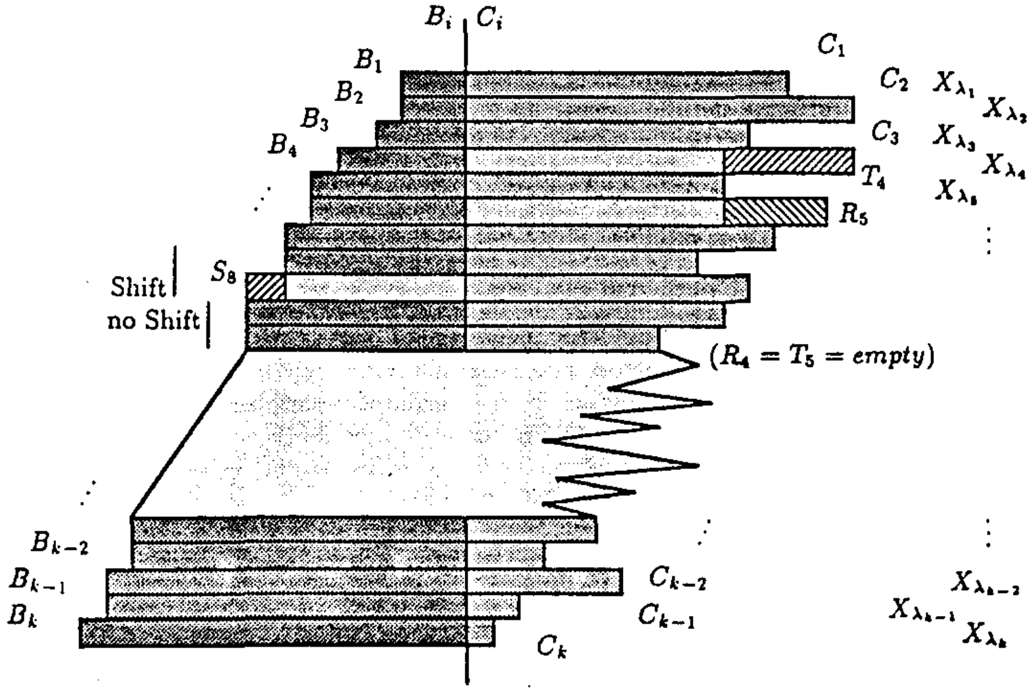
(1.1) to illustrate the concept of a position equation and the type of constraints Makanin uses in order to get a lower bound for the exponent of periodicity of a solution, and
(1.2) to give a short description of the algorithm*.
The second section will then be used to partially describe our own reformulation.

1.1 Domino-Towers, Position Equations and Boundary Connections

Definition 2: The sequence $(X_1, \ldots, X_n)$ of non-empty words ("stone types") may be arranged to a *domino-tower* $< X_{\lambda_i}, B_i, C_i, S_i >_{1 \leq i \leq k}$ of height $k > 0$ if the X_i may be ordered to a sequence

$$(X_{\lambda_1}, X_{\lambda_2}, \ldots\ldots\ldots\ldots\ldots\ldots, X_{\lambda_k})$$

(possibly with many occurrences of the X_i) with decompositions $X_{\lambda_i} = B_i C_i$ (for non-empty words B_i, C_i) $(1 \leq i \leq k)$ such that $B_{i+1} = S_i B_i$ $(1 \leq i \leq k-1)$ for possibly empty words $S_1, \ldots, S_{k-1}$ and $C_i R_i = C_{i+1} T_i$ for possibly empty words R_i and T_i $(1 \leq i \leq k-1)$. The name "domino tower" is motivated by the following figure. Here all parts of consecutive words which have direct contact must be equal. Parts which do not have direct contact are not restricted like that.



* For a complete description it would be necessary to consider an enormous amount of technical details. This is impossible here. Thus we must necessarily stay at a rather informal level. We refer to [Mak] and to our forthcoming paper [Sch].

Let $|X|$ denote the length of the word X.

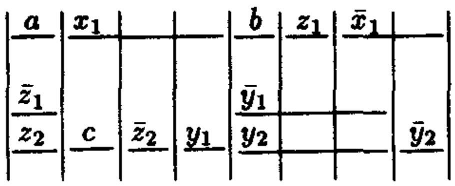
Theorem 3 ([Mak], Lemma 1.4): *Suppose the words $X_1, \ldots, X_n$ may be arranged to a domino tower $< X_{\lambda_i}, B_i, C_i, S_i >_{1 \leq i \leq k}$ of height $k > 0$. If the words $S_1, \ldots, S_{k-1}$ satisfy the "shift condition"*

$$j - i \geq K \Rightarrow |S_i S_{i+1} \ldots S_{j-1} S_j| > 0,$$

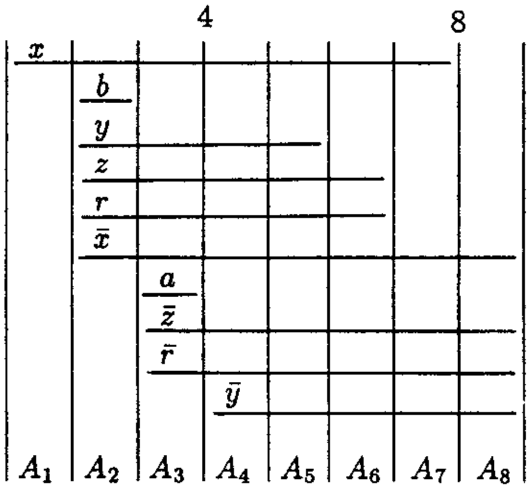
then some word X_t has the form $X_t = P^s Q$, where P is non-empty and has $s \geq \frac{k}{Kn^2} - 1$ consecutive repetitions.

Let us now illustrate the concept of a position equation:

Example 4: Here is the graphical description of one of several position equations which are assigned to the equation $axbzx == zczyyy$:

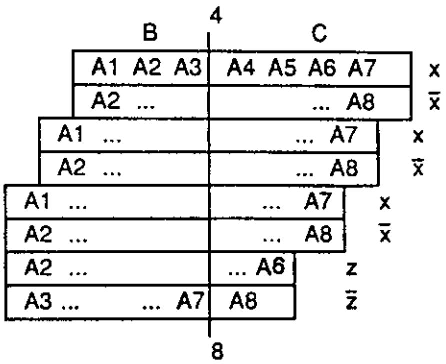


In a position equation, all variables have exactly two occurrences (this may be seen as an artificial imitation of the particular situation where Plotkins algorithm defines a decision procedure). The relative lengths of all occurrences of variables and coefficients are marked by means of boundaries. A solution of such a position equation is an assignment of non-empty words to the indecomposable columns which respects occurrences of coefficients and assigns (via composition) the same word to both occurrences of the same variable. In our example we have eight indecomposable columns. The words $a, c, a, baca, b, a, ca, baca$ represent a solution, the values of the variables of type x, y and z are $(X, Y, Z) = (cabaca, baca, a)$. The exponent of periodicity of a solution of a position equation is, by deinition, the exponent of periodicity of the words which are assigned to the variables (which is trivial in our example). As a result of transformation steps (which are applied to construct successor equations which are "nearer to solutions" in some sense) more complex successor position equations may occur which include some *boundary connections*. The definition of a solution then has to be enriched: a new condition demands that the words which are assigned to the variables may be arranged to a domino tower of a certain length. If the position equation satisfies certain normalization criteria (which will not be discussed here), then these domino towers satisfy the shift condition of theorem 3 and we get in fact a lower bound for the exponent of periodicity of an arbitrary solution, as described above. Let us give an example, to get an impression. The position equation



$(4, x, x, x, z, 8)$.

has a boundary connection $(4, x, x, x, z, 8)$ imposing the following restriction on solutions S: suppose S assigns the nonempty words $A_1, \ldots, A_8$ to the indecomposable columns. Then boundary 4 defines the prefix $B_1^x = A_1A_2A_3$ of the word X assigned to x. Now S assigns the same word X to $\bar{x}$. Thus B_1^x is also a prefix of the word $A_2 \ldots A_8$. The position of x and $\bar{x}$ defines a second prefix $B_2^x = A_1B_1^x$ of X. Again, B_2^x is also a prefix of $A_2 \ldots A_8$ and the relative position of x and $\bar{x}$ shows that $B_3^x = A_1B_2^x$ is a prefix of X. B_3^x is a prefix of $A_2 \ldots A_8$ which now determines the prefix $B_1^z = B_3^x$ of the word Z which is assigned to z (the left boundaries of $\bar{x}$ and of z coincide). Now B_1^z is a prefix of $A_3 \ldots A_8$. S *satisfies* the boundary connection $(4, x, x, x, z, 8)$ if $B_1^z = A_3A_4A_5A_6A_7$ (8 is the right boundary of column 7). To be a solution, S has to satisfy all boundary connections. In our example this implies that the components X and Z of S may be arranged to the following domino tower:



Thus theorem 3 gives a (still trivial) lower bound for the exponent of periodicity of such a solution.

We may now describe the search tree of an equation E in the usual formulation of Makanin's algorithm.

1.2 The Search Tree

In the usual formulation of Makanin's algorithm, the word equation E is immediately translated into a corresponding finite set $SPE(E)$ of position equations. E has a solution if and only if at least one element PE of $SPE(E)$ has a solution. Then a transformation procedure (followed by a normalization step) is used in order to construct for any such position equation PE a finite set of successors $SPE(PE)$ such that PE has a solution if and only if at least one position equation $PE' \in SPE(PE)$ has a solution. Moreover, if E (or PE) has a solution with exponent of periodicity s, then the corresponding solution of the successor PE (or PE') has exponent of periodicity $s' \leq s$. Transformation has the ultimate goal to erase a left part of the position equation, propagating constraints to the remaining part. For this purpose, sometimes boundary connections have to be introduced. Normalization is necessary in order to guarantee that all boundary connections satisfy the shift condition (compare theorem 3). The second crucial property of the transformation and normalization steps is that they never enlarge the number of occurrences of variables and coefficients. Only the length and the number of boundary connections may grow. The complete definition of a position equation (which is technically sophisticated and will not be given here) shows that also a bound for the number of boundary connections may be given and that for given number of occurrences of variables and coefficients and upper bound for the length of connections the number of possible position equations is finite*. Thus, if we follow an arbitrary path of the resulting tree, three possibilities exist: in the first case we find a position equation which is solvable or unsolvable in some trivial sense. Second, we might find a position equation which is isomorphic to a predecessor equation which has occurred earlier in the same path. Then we may stop (the argument is the same as in the special subcase of Plotkin's procedure mentioned in the introduction). In the remaining case, we will eventually find a position equation where the associated lower bound for the exponent of periodicity of an arbitrary solution exceeds the upper bound for the exponent of periodicity which we find for a minimal solution of E, applying theorem 1. Thus, concentrating the search on minimal solutions, we may stop at this point, too. We end up with a finite search tree.

The transformation steps which are given in current formulations of Makanin's algorithm depend from the type of a position equation. A left part is erased only in certain situations. We mention a result of [Sch]:

Lemma 5: *There exists a transformation procedure which applies to arbitrary position equations and has the property that at each transformation step a non-trivial left part of the position equation is erased.*

* More exactly this is true only for the set of all normalized admissible position equations. But we do not want to go too far into technical details.

2 SME-Systems

If E is translated into $SPE(E)$, then the position of the variables in such a position equation reflects exactly the position of the corresponding variables of the word equation (compare example 4). This method has two disadvantages: (1) We have to deal with rather complicated structures quite from the beginning. (2) If the word equation is long, then the corresponding position equations are horizontally very long. Since the analysis via transformation steps proceeds horizontally from left to right a solution is typically found only after a considerable number steps.

We suggest a reformulation of the algorithm. Let us briefly sketch the new picture of the search tree, definitions and details are given immediately. We start translating the word equation E into an equivalent special multi-equation system (sme-system) $SME(E)$. Such an sme-system may be regarded as an equivalence class describing a whole set of position equations. Nevertheless, both the internal representation and the transformation of an sme-system is much simpler than for the corresponding position equations. As long as such a system contains at least one ordinary two-sided equation we continue with transformation steps similar to those of Plotkin's procedure which are followed by simplification steps in some cases. This first part of the search tree, where we use sme-systems only, is called the flat part of the search tree and denoted by $\mathcal{T}_{flat}(E)$. If the word equation E does not have a variable which occurs more than twice, then the whole search tree $\mathcal{T}_{Mak}(E)$ is flat. In the other case, we might eventually reach an sme-system which has only multi-equations with at least three sides (we say that we have reached an ***open leaf*** of the flat tree). These structures are now translated into position equations which are analyzed as usual, using the uniform transformation given in [Sch]. The position equations which we get via translation of sme-systems are horizontally very short and preferable to the structures of the traditional approach.

Definition 6: An l-sided multi-equation ME over $\mathcal{C}, \mathcal{X}$ has the form

$$\sigma_{1,1} \dots \sigma_{1,k_1} == \sigma_{2,1} \dots \sigma_{2,k_2} == \dots == \sigma_{l,1} \dots \sigma_{l,k_l}, \tag{1}$$

where $l \geq 2$, $k_i \geq 1$ $(1 \leq i \leq l)$ and $\sigma_{i,j} \in \mathcal{C} \cup \mathcal{X}$ ($1 \leq j \leq k_i, 1 \leq i \leq l$). A ***solution*** of ME is a sequence

$$S = (X_1, \dots, X_n) \tag{2}$$

of non-empty words over $\mathcal{C}$ such that all sides of (1) become graphically identical when we replace every occurrence of x_i by X_i $(1 \leq i \leq n)$. For $k \leq k_j$, the word $S(\sigma_{j,1} \dots \sigma_{j,k})$ is defined in the obvious way, regarding S as a morphism fixing the coefficients $(1 \leq j \leq l)$. A ***special*** multi-equation system (sme-system) is a system of multi-equations where no variable occurs more than twice. A sequence (2) is a solution of the system $SME = \{ME_1, \dots, ME_r\}$ if S is a simultaneous solution of all multi-equations ME_i $(1 \leq i \leq r)$. An sme-system SME may be empty, in this case every sequence of the form (2) is a solution.

Definition 7: Let E be a word equation. The sme-system $SME(E)$ *canonically associated with* E is defined as follows: for every variable x_i, the n_i occurrences of x_i in E are replaced by distinct new variables $x_{i,1}, \ldots, x_{i,n_i}$. To the resulting equation we add a multi-equation $x_{i,1} == \ldots == x_{i,n_i}$ in case $n_i > 1$.

Example 8: The equation $axbzx == zczyyy$ with variables x, y and z is translated into

$$\begin{vmatrix} ax_1bz_1x_2 \\ z_2cz_3y_1y_2y_3 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} \begin{vmatrix} z_1 \\ z_2 \\ z_3 \end{vmatrix} \begin{vmatrix} y_1 \\ y_2 \\ y_3 \end{vmatrix}$$

with *principal column* $(ax_1bz_1x_2, z_2cz_3y_1y_2y_3)^t$ and the *columns associated with x, y and z.*

Lemma 9: *If E has a solution $S = (X_1, \ldots, X_n)$, then $SME(E)$ has a corresponding solution S' which assigns X_i to every variable $x_{i,j}$. If $SME(E)$ has a solution S', then the words $S'(x_{i,j})$ coincide ($1 \leq j \leq n_i$). The assignment $S(x_i) = S'(x_{i,1})$ ($1 \leq i \leq n$) defines a solution of E.*

We are now ready to start the description of $\mathcal{T}_{flat}(E)$.*

Definition 10: For every word equation E, the *flat search tree* $\mathcal{T}_{flat}(E)$ is defined as follows:

- The top node of $\mathcal{T}_{flat}(E)$ is labelled with $SME(E)$.

Suppose η is any node of $\mathcal{T}_{flat}(E)$, labelled with the sme-system SME. In the following cases, η is a leaf of $\mathcal{T}_{flat}(E)$:

- If SME is empty, then η is a *successful leaf.*
- If SME is non-empty and all two-sided equations been completely resolved, i.e. if all multi-equations of SME are at least three-sided, then η is an *open leaf* of $\mathcal{T}_{flat}(E)$ (the term "open" indicates that η is not a leaf of $\mathcal{T}_{Mak}(E)$).
- If SME is isomorphic to the label of a predecessor node, then η is a *blind leaf.*
- If SME contains an equation ME with head (a_i, a_j) with two distinct coefficients a_i and a_j, then η is a blind leaf.

In the other case, if $SME \neq \emptyset$ has a two-sided equation and is not an isomorphic copy of a predecessor, then the successors of η are defined by means of a transformation and a simplification procedure. For every sme-system $SME_i \in Simpl(Trans(SME))$ the node η has one successor η_i labelled with SME_i. The transformation steps follow exactly the corresponding transformation steps of Plotkin's procedure as described above.

- **Transformation** (of the sme-system SME with two-sided equation ME):

* To simplify discussion we will not mention the various possibilities to recognize sme-systems which are unsolvable in a more or less trivial sense. Thus, a more complete description would include various methods of pruning blind branches.

(T_1) Suppose that ME has head (σ,σ) with two identical entries. Then delete the head symbols of ME and leave the other multi-equations unmodified. The resulting system is the label of the unique successor of η.

(T_2) Suppose that ME has head $(x_{i_1,j_1}, x_{i_2,j_2})$ with two distinct variables. $Trans(SME)$ has three elements SME_i $(1 \leq i \leq 3)$:

(1) To get SME_1, replace all occurrences of x_{i_2,j_2} in SME by x_{i_1,j_1}. Then delete the head symbols of the distinguished equation and leave the other multi-equations unmodified.

(2) To get SME_2, replace all occurrences of x_{i_2,j_2} in SME by $x_{i_1,j_1}x_{i_2,j_2}$. Then delete the head symbols of the distinguished equation and leave the other multi-equations unmodified.

(3) To get SME_3, replace all occurrences of x_{i_1,j_1} in SME by $x_{i_2,j_2}x_{i_1,j_1}$. Then delete the head symbols of the distinguished equation and leave the other multi-equations unmodified.

(T_3) Suppose that ME has head $(x_{i,j}, a_h)$ or $(a_h, x_{i,j})$, where a_h is a coefficient symbol. $Trans(SME)$ has two elements SME_1, SME_2:

(1) To get SME_1, replace all occurrences of $x_{i,j}$ in SME by a_h. Then delete the head symbols of the distinguished equation and leave the other multi-equations unmodified.

(2) To get SME_2, replace all occurrences of $x_{i,j}$ in SME by $a_h x_{i,j}$. Then delete the head symbols of the distinguished equation and leave the other multi-equations unmodified.

After the transformation it might happen that one side of the distinguished two-sided equation of a structure SME_i is empty while the other is not. Then SME_i is not an sme-system in the sense of definition 1. We erase it. If both sides of the two-sided equation are empty after the transformation, then this equation is erased (resolved). If there is another two-sided equation left, then we continue. Otherwise, if the system is non-empty we have reached an open leaf.

- **Simplification** (of the sme-system SME):

The following *simplification rules* are applied until the system is erased or a system SME' is reached which cannot be further simplified by the rules.

(S_1) If the multi-equation ME in SME has two identical sides of the form $x_{i,j}$ (where $x_{i,j}$ is a variable), then erase both sides. If now ME has only one side, then erase ME.

(S_2) If SME contains a multi-equation ME which has a side of the form a_i (where a_i is a coefficient symbol) and if all other sides of ME have length 1 and are variables, then replace all occurrences of these variables in SME by a_i. Erase ME.

Theorem 11: (a) *The maximal length of a path in $\mathcal{T}_{flat}(E)$ does not exceed the number $(d!)^3$, where $d = 2nl(E)$.*
(b) *E has a solution if and only if $\mathcal{T}_{flat}(E)$ has a successful leaf or $\mathcal{T}_{flat}(E)$ has an open leaf which is labelled with a solvable sme-system.*
(c) *If no variable occurs more than twice in E, then $\mathcal{T}_{flat}(E)$ does not have an open leaf.*

E is solvable if and only if $\mathcal{T}_{flat}(E)$ has a successful leaf.

Remark 12: It is simple to keep track of the solutions. We augment $SME(E)$ by the additional substitution list

$$(x_{1,1}, x_{2,1}, \ldots, x_{n,1}).$$

All replacements which occur during transformation or simplification are also applied to this sequence. If the rule (S_1) is applied with the variable $x_{i,1}$ and ME is not empty, then we replace $x_{i,1}$ in the substitution list by one of the remaining sides, etc. If we reach a leaf labelled with an empty sme-system, then the substitution list defines a unifier of the equation E. The technique may even be optimized since the structure of the columns associated with the variables reflects the actual substitution list.

Example 13: Consider the equation $axbzx == zczyyy$. $SME(axbzx == zczyyy)$ has the following representation:

$$\begin{vmatrix} ax_1bz_1x_2 \\ z_2cz_3y_1y_2y_3 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} \begin{vmatrix} y_1 \\ y_2 \\ y_3 \end{vmatrix} \begin{vmatrix} z_1 \\ z_2 \\ z_3 \end{vmatrix}.$$

Replacing z_2 by a and using (S_2) we get

$$\begin{vmatrix} x_1bax_2 \\ cay_1y_2y_3 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} \begin{vmatrix} y_1 \\ y_2 \\ y_3 \end{vmatrix} \qquad (\_,\_,a).$$

Here a is the z-entry of the substitution list. It is not necessary to store x- and y-entries since we might use any line of the corresponding columns as substitution value. Combining now two steps, we replace x_1 by cax_1. After this step we may replace y_1 by x_1:

$$\begin{vmatrix} x_1bax_2 \\ y_1y_2y_3 \end{vmatrix} \begin{vmatrix} cax_1 \\ x_2 \end{vmatrix} \begin{vmatrix} y_1 \\ y_2 \\ y_3 \end{vmatrix} \quad (\_,\_,a) \quad \longrightarrow \quad \begin{vmatrix} bax_2 \\ y_2y_3 \end{vmatrix} \begin{vmatrix} cax_1 \\ x_2 \end{vmatrix} \begin{vmatrix} x_1 \\ y_2 \\ y_3 \end{vmatrix} \quad (\_,\_,a).$$

Combining several steps, we may replace y_2 by bay_2, x_2 by y_2x_2 and y_3 by x_2:

$$\begin{vmatrix} cax_1 \\ y_2x_2 \end{vmatrix} \begin{vmatrix} x_1 \\ bay_2 \\ x_2 \end{vmatrix} \qquad (\_,\_,a).$$

Now the principal column is completely resolved. We continue with the column associated with x and add the value cax_1 to the substitution list (as x-entry). Combining two steps, we may replace y_2 by ca, then x_2 by x_1:

$$\begin{vmatrix} x_1 \\ x_2 \end{vmatrix} \begin{vmatrix} x_1 \\ baca \\ x_2 \end{vmatrix} \quad (cax_1,\_,a) \quad \longrightarrow \quad \begin{vmatrix} x_1 \\ baca \\ x_1 \end{vmatrix} \quad (cax_1,\_,a).$$

By (S_1), the matched occurrences of x_1 may be erased, substituting *baca* for x_1 in the substitution list. The resulting column has only one line *baca* which is used as substitution entry for y and erased. The final sme-system is empty and the substitution list shows that $(X, Y, Z) = (cabaca, baca, a)$ is a solution of $axbzx == zczyyy$.

Remark 14: When we reach an open leaf of $\mathcal{T}_{flat}$, then the naive transformation strategy leads again to an explosion of the number of symbols: if, for example, a multi-equation has the three-valued head (x, y, z) for variables x, y and z with two occurrences and we consider the case where X is assumed to be a proper prefix of Y and of Z, then *four* new occurrences of x are introduced, by the replacement. Then, when we take the tail, only three occurrences of x are erased and the number of symbols grows. Thus we have reached the point where we have to return to the concept of a position equation. The translation of an sme-system into a set of position equations is obvious. We only have to consider all possibilities how the entries of distinct lines of the same multi-equation may be positioned with respect to each other. Then, after introducing the corresponding boundaries, variables which have only one occurrence are erased.

Remark 15: Since the structure of the variable columns of $SME(E)$ reflects the structure of the components of a solution $S = (X_1, \ldots, X_n)$ of E lemma 5 shows that any solution S of an equation E of length d is found after at most $I = |X_1 \ldots X_n|$ transformation steps following the first d transformation steps of the flat tree.

References

[Abd] H.Abdulrab, *Résolution d'équations sur les mots: Etude et implémentation LISP de l'algorithme de MAKANIN*, Thesis, Laboratoire d'informatique, Rouen 1987.

[Bul] V.K.Bulitko, *Equations and Inequalities in a Free Group and a Free Semigroup*, Tul. Gos.Ped.Inst.Učen.Zap.Mat.Kafedr Vyp.2, Geom. i Alg. (1970), 242-252 (Russian).

[Jaf] J.Jaffar, *Minimal and Complete Word Unification*, Research Note, IBM Thomas J. Watson Research Center, New York, March 1989, to appear in the JACM.

[KoP] A.Kościelski,L.Pacholski, *On the Exponent of Periodicity of a Minimal Solution of a Word Equation (Part II)*, Research Note, Math.Inst., Polish Academy of Sciences, Wroclaw 1989.

[Mak] G.S.Makanin, *The problem of solvability of equations in a free semigroup*, Math. USSR Sbornik **32**, 2 (1977), 129-198.

[Péc] J.P.Pécuchet, *Equations avec constantes et algorithme de Makanin*, Thèse de doctorat, Laboratoire d' informatique, Rouen 1981.

[Plo] G.Plotkin, *Building-in Equational Theories*, Machine Intelligence **7** (1972), 73-90.

[Sch] K.U.Schulz, *Makanin's Algorithm - Two Improvements and a Generalization*, CIS-Report, Centrum für Informations- und Sprachforschung, Ludwig-Maximilians-Universität München, to appear.

[Sie] J.Siekmannn, *Unification and Matching Problems*, Ph.D.Thesis, Essex University, Memo CSA-4-78, 1978.

Two Theories of Propositional Quantification[1]

Nicholas Asher
The University of Texas at Austin & IMSV Stuttgart

1.Introduction

Natural language is replete with reference to abstract entities like propositions, facts and their companion (and less abstract) events. Sentential nominalizations like those underlined in (1) apparently refer to such abstract entities. We also appear to quantify over such entities in (2):

(1.a) That Mary is wise is true.

(1.b) Mary's having won the Mathematics prize surprised the professors.

(1.c) The fact that Stan was Director made the Institute a lively place.

(1.d) The idea of Stan's flying to Austin was intriguing.

(2.a) Everything Mary believes is true.

(2.b) Every fact you discover may be relevant.

(2.c) Nothing you have said convinces me.

The question I would like to pose in this essay is a familiar one from analytic philosophy since the turn of the century:[2] what is the logical form underlying this apparent reference and quantification over abstract entities? Two general theories emerge, one a first order theory, the other a higher order theory of quantification.[3] The difficult task for such theories is to develop a coherent theory of quantification over abstract objects that are suitably descriminated to be objects of attitudes. The task is

---

[1]I would like to thank Dan Bonevac, Anil Gupta, Herbert Hochberg, Rob Koons, Geoffrey Laforte, and Per Lindstrom for helpful discussions on these matters. I would also like to thank Rich Thomason for comments on an earlier draft of this paper read at the Second Conference on Logic and Linguistics in Tuscon CA (1989). A drastically shortened version of some of the ideas in this paper appeared in the proceedings of TARK III. Unfortunately also some errors occur therein that I have tried to correct here. More specifically, I did not in the TARK paper make clear in my construction that I was claiming a completeness proof for the axiomatization presented here below as R_1 relative to the class of all fixed points that could be defined by consistent additions to base partial models. I hope to have corrected this here.

[2]See for instance Russell's (1901) arguments in *The Principles of Mathematics.* The concern with abstract entities and their logic remained a concern throughout Russell's life.

[3]Many people have been suggesting a first order theory of abstract entities in the past few years-- for instance Bealer (1983), Turner (1987), (1989), Aczel (1989). Higher order theories have found advocates like Russell (1901) (1911), Ramsey (1926), Prior (1960) e.g., and others like Fine, Cocchiarella, and Thomason (1980), and Menzel (1986). I will use Turner and Thomason as my main sources here, but that is not because I have made a detailed survey of all the proposals.

difficult because many attempts to do so have led to paradoxes concerning abstract entities. These paradoxes have bedeviled philosophers and logicians since ancient times.

There are two generally recognized families of paradoxes. One contains paradoxes having to do with sentences and direct quotation contexts like the Liar. Then there are paradoxes of application like the property version of Russell's paradox and the family of associated set theoretic paradoxes (Burali-Forti, Russell, etc). Arthur Prior (1961) and more recently Rich Thomason (1982) have argued that there is a third family of paradoxes, the so called "paradoxes of indirect discourse," which have to do with the nature of propositions. Here is an example of such a paradox originally due to Jean Buridan, embellished by Prior and Thomason: Suppose Prior is thinking to himself only the following thought:

(3) Either everything that I am thinking at the present moment or everything that Tarski will think in the next instant, but not both, is false.

Suppose that at the next moment Tarski thinks that snow is white. By reasoning that is valid in the simple theory of types, we conclude that Tarski was not able to think only that snow is white, a bizarre and unwanted consequence of a logic for belief. Note that this is a paradox about "entertaining," "thinking about" or "explicit" belief, not about implicit belief. Nevertheless, since explicit belief is a seemingly plausible and useful notion, the paradox has a bite to it.

In this paper I examine the intentional paradoxes from two points of view: (i) a first order perspective in which the intentional paradoxes are merely a special case of paradoxes direct discourse, and (ii) a higher order perspective like the one Prior and Thomason advocate. Beginning with a representationalist's view of attitudes and abstract entities, one can arrive at a natural formulation in a first order language of what Prior is thinking to himself. This is the most congenial perspective perhaps to a semanticist committed to a representational theory of attitudes and to a conceptualist understanding of abstract entities in general. But the Prior-Thomason reconstruction of the intentional paradoxes in higher order logic-- most notably in the simple theory of types-- has merits of its own with regard to natural language semantics. I will give an inductive definition for the propositional quantifiers, which appear to solve the intentional paradox discussed by Thomason and Prior and one other difficulty for the simple theory of types which Russell noticed. I end by drawing some comparisons between the two theories.

The category of paradoxes of indirect discourse is potentially very varied. The defining characteristic of a paradox of indirect discourse is that it does not directly involve a quotational context. Clearly these paradoxes are not restricted simply to attitude contexts. But many of these paradoxes of indirect discourse also have ties to the paradoxes of direct discourse, as the following example about facts makes out. Those who believe in facts and take sentences to be bearers of truth would espouse something like the doctrine that a sentence φ denotes (expresses, corresponds to or whatever relation you like) a fact iff ⌜φ⌝ is true. But then the statement

(4.a) This sentence does not denote a fact

exhibits the same pathological behavior as the Liar sentence for obvious reasons. (4.a) might be taken to be yet another example of the paradox of direct discourse, except that it leads to a paradox of indirect quotation

(4.b) It is a fact that (4.a) does not denote a fact.

The paradox of indirect discourse appears when we take the sort of logic that ought to govern contexts like *it is a fact that.* Surely, one might think (naively), if it is a fact that p, then p, and if p, then it is a fact that p. In higher order logic we can easily express this thought as a sentence; in first order logic it could be taken as an axiom schema.

Other ways of constructing paradoxes of indirect discourse do not depend on direct discourse at all. There are paradoxes of intention (similar to Newcomb's Problem and explored recently by Gaifman) that resemble at least semantic paradoxes. Gaifman's puzzle gives a *prima facie* plausible example of a very odd, but desirable goal. By having the intention to reach the goal, you in effect have the intention of not getting it, because you know that if you have the intention to reach the goal you won't reach it. Conversely, by having the intention not to reach the goal, you have the intention of reaching it. This supposition results in a diagonal intention of achieving φ iff you don't intend to achieve φ. This diagonalized intention appears to yield similar difficulties for the logic of intention. Yet it has nothing to do with direct quotation at least on the face of it; they appear to be properly classified as paradoxes of indirect discourse.[4]

2. Representationalism, Conceptualism, Indirect Discourse and First Order Logic

One thesis about abstract entities is the conceptualist's. The analysis of propositional attitudes as attitudes towards representations (sentences in the language of thought) with a certain determinate content[5] suggests that abstract entities like propositions are constructs from representations. If one adopts a representationalist view of attitudes, then the role of propositions and representations coincide. The conceptualist makes propositions and other abstract entities equivalence classes of

[4]Koons (1987) contains a detailed exposition of some game-theoretic paradoxes and argues for their similarity to the semantic paradoxes.
[5]For details on the particular representational theory I espouse, see Asher (1986), (1987), (in press). But many other philosophers and workers in CS espouse such theories.

representations. For instance, propositions, the objects that are referred to by the *that* clause in (1.a) and quantified over in (2.a) and (2.c), are equivalence classes of representations under some suitable equivalence relation like sameness of functional role. An equivalence relation like this one is needed to give propositions enough structure to handle problems about the semantics of propositional attitudes.

The conceptualist should also say something about the structure of these representations and their relations to natural language. Typically he does so by giving a construction procedure from natural language discourse. For my brand of conceptualist, the class of propositions is a countable, recursive set (assuming at most a countable number of possible human languages). A similar story goes for facts, the type of abstract object denoted in (1.b) and quantified over in (2.b), but I won't go into details here. For my purposes here I can simplify the conceptualist's connection between representations and sentences. I will model the conceptualist's propositions with sentences (or equivalence classes of alphabetic variant sentences) of a language with enough expressive power to express the content and structure of our beliefs as in Asher and Kamp (1986) (1989). Such an approach construes quantification over propositions as quantification over sentences, and from this philosophical perspective, indirect discourse is much like direct discourse.

A cautionary note is in order, however. The use of sentences to model representations allows us to make use of well-established techniques of recursion theory. But we should not necessarily identify a system of representations with a particular language, in particular a particular formal langauge. A system of representation yields instead of sentences well-formed information structures.[6] These are not linguistic objects, though like sentences they are complexes constructed by means of recursive rules from meaningful constituents. A more accurate way of thinking of an information structure is to consider it a partial model. This analogy also cautions one from making too close a connection between a language (like that of first order logic) and a system of representations.

These observations yield a first order framework for quantifying over propositions that exploits the sort of quantification found in the arithmetization of syntax. On such an approach, we take various predicates like 'true', 'believe', etc... to take sentences as their objects. In writing down a translation of a sentential nominal like (1.a), however, we must not put the sentence itself but a name of the sentence. Consider a first order language L with identity, a denumerable infinity of individual constants, and one distinguished predicate S (to be read as *is a sentence*). L(B, T) is the language L expanded with a 2-place predicate B (to be read as *believes that*) and a 1-

---

[6] We should think of a system of representations at least as an interpreted language-- not merely a syntactic engine.

place truth predicate T. We include within the domain the set S of all sentences of L(B, T), and we now relativize our quantifiers over propositions to quantifiers restricted to S. Thus, (2.a) becomes on this framework:

(5) $\forall x\ (S(x) \rightarrow (B(m, x) \rightarrow T(x)))$

We would like to extend L to express beliefs about arithmetic. Thus, we should countenance the possibility that some suitable extension of L be able to talk about the syntax of its sentences. This makes it possible to prove instances of Goedel's diagonalization lemma $\psi \leftrightarrow \xi(\ulcorner\psi\urcorner)$, where the instance of the 1-place predicate variable ξ are B and T. As Tarski (1931), Montague (1963) and Thomason (1980) showed, these instances of the diagonalization lemma are incompatible with very weak assumptions about the logic of belief or truth. Let's call any theory containing instances of the diagonalization lemma a *self-referential theory*. Techniques of Kripke (1975) and Herzberger (1982) and Gupta (1982) permit the construction of models for self-referential theories of truth and belief. A semantics for L can exploit these techniques in several ways.[7] A *model for* L is a quintuple $M = \langle W, A, R, D, [\![\,]\!] \rangle$ such that:

(i) W is a set (of possible worlds);

(ii) A is a subset of D (a set of agents)

(iii) R is a function from agents in A to binary relations on W (wR_aw' means that w' is a doxastic alternative for K in w for agent a; $[wR_a]$ is the set of alternatives to w for $a \in A$);

(iv) D is a non-empty set (the domain of individuals);(iv) $[\![\,]\!]$ is a function which assigns to each non-logical constant of L at each world a suitable extension: if c is an individual constant of L, $[\![c]\!]_w \in D$; and if Q is an n-ary predicate of L, $[\![Q]\!]_w \subseteq D^n$;

(v) for each $w \in W$, $[\![S]\!]_w$ is the set of sentences of L;

(vi) each individual constant c is a *rigid designator*, i.e., for all $w, w' \in W$, $[\![c]\!]_w = [\![c]\!]_{w'}$.

(vii) for each $d \in D$ and $w \in W$, there is a constant c of L such that $[\![c]\!]_{w,M} = d$.

A *model* for L(B, T) is a triple $\langle M, [\![B]\!], [\![T]\!] \rangle$ where M is a model for L, $[\![B]\!]$ is an intension for B relative to M (i.e., a function from W_M into $\wp(D_M)$) and $[\![T]\!]$ an intension for T such that $\forall w \in W_M\ [\![B]\!]_w \subseteq [\![S]\!]_w$. I will call models for L(B, T) simply *models* and models for L *model-structures*. A model structure M is *extensional* just in case W_M is a singleton and $\langle w,w \rangle \in R_M$.

[7]I here rehearse the work of Asher & Kamp (1986) and (1989).

An important notion for this conception of a model is the idea of *model coherence.* A model $\mathcal{M}$ is (*doxastically and alethically*) *coherent* iff the following statement is satisfied for each sentence ψ and each world $w \in W_{\mathcal{M}}$:

(i) $<a, \psi> \in [\![B]\!]_{\mathcal{M},w}$ iff $[\![\psi]\!]_{\mathcal{M},w'} = 1$ for all $w' \in [wR_a]$

(ii) $\psi \in [\![T]\!]_{\mathcal{M},w}$ iff $[\![\psi]\!]_{\mathcal{M},w} = 1$

A model structure M is *essentially incoherent* iff every model that expands M is incoherent. The notion of coherence brings together two, independent features of the models that are essential to the semantics of the attitudes, the alternativeness relation and the extension of the B predicate. The alternativeness relation in the model structure encodes plausible doxastic principles of reasoning and the basic doxastic facts that the agent may uncover through reflection; the predicate B's initial extension represents what an agent might in fact consciously believe.[8] Coherent models are those models in which the agent believes (or could come to believe through reasoning) all that is doxastically possible for him to come to believe. Coherent models are those in which the agent can use all the principles of reasoning encoded in the alternativeness relation to their full effect. A similar story goes for truth: coherence connects the sentences in the extension of the truth predicate with their extensions (truth values) in the model.

Models that are incoherent may become coherent through the process of *model revision.* To define this notion, however, I need some auxillary notions. Define an *interpolation function on* a set A to be any function f from $\wp(A)^2$ into $\wp(A)$ such that whenever $A_1, A_2 \subseteq A$ and $A_1 \cap A_2 = \emptyset$ then $f(A_1, A_2) \supseteq A_1$ and $f(A_1, A_2) \cap A2 = \emptyset$. A *revision scheme* is a function $\mathcal{R}$ defined on the class of all limit ordinals such that for each λ $\mathcal{R}(\lambda)$ is an interpolation function on the set S_L of sentences of L. Given a model $\mathcal{M}$ and a revision scheme $\mathcal{R}$, the *revision sequence starting from* $\mathcal{M}$ *according to* $\mathcal{R}$ is the sequence $\{\mathcal{M}^{\alpha,\mathcal{R}}\}_{\alpha \in On}$, such that: $\mathcal{M}^{\alpha,\mathcal{R}} = <W_{\mathcal{M}}, D_{\mathcal{M}}, R_{\mathcal{M}}, [\![]\!]^{\alpha,\mathcal{R}}>$, where $[\![\theta]\!]^{\alpha,\mathcal{R}} = [\![\theta]\!]_{\mathcal{M}}$ for all nonlogical constants θ other than B and T, and $[\![B]\!]^{\alpha,\mathcal{R}}$ and $[\![B]\!]^{\alpha,\mathcal{R}}$ are defined as follows:

i) $[\![B]\!]^{o,\mathcal{R}}{}_w = [\![B]\!]_w$; $[\![T]\!]^{o,\mathcal{R}}{}_w = A^{\mathcal{R}}{}_w$

ii) $[\![B]\!]^{\alpha+1,\mathcal{R}}{}_w = \{<a, \varphi>: (\forall w' \in R_{a,\mathcal{M}}) [\![\varphi]\!]_{\mathcal{M}^\alpha, w'} = 1\}$

iii) $[\![T]\!]^{\alpha+1,\mathcal{R}}{}_w = \{\varphi : [\![\varphi]\!]_{\mathcal{M}^\alpha, w} = 1\}$

iv) $[\![B]\!]^{\lambda,\mathcal{R}}{}_{\mathcal{M},w} = \mathcal{R}(\lambda)(B^+{}_w, B^-{}_w)$, where $B^+{}_w = \{<a, \varphi>: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow \varphi \in [\![B]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w})\}$ and $B^-{}_w = \{<a, \varphi>: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow \varphi \notin [\![B]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w})\}$.

[8]The fact that in some models the extension of the belief predicate at a world w may be inconsistent with Th([wR]) (see below) might be taken to be a drawback. But this feature serves a purpose; it models those situations in which agents hold conscious beliefs that upon reflection they would discard as being false.

v) $[\![T]\!]^{\lambda,\mathcal{R}}{}_{\mathcal{M},w} = \mathcal{R}(\lambda)(T^+{}_w, T^-{}_w)$, where $T^+{}_w = \{\varphi: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow \varphi \in [\![T]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w})\}$ and $B^-{}_w = \{\varphi: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow \varphi \notin [\![T]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w})\}$.

There are many different choices for revision schemes $\mathcal{R}$ obeying the local stability principle, on which those sentences that become members of B+ (T+) at some world w ought to be counted in the extension of B (T) at w at a limit ordinal stage λ and those sentences that become members of $B\text{-}_w$ $(T+_w)$ ought not to be counted in the extension of B (T) at w at a limit ordinal stage. $\mathcal{R}$ could be defined to be one of the Kripke proposals for an inductive definition of the intensions of B and T or it could be one of a wide variety of "semi-inductive" definitional schemes. In this section I shall most often use an intensional version of Herzberger's semi-inductive definitions of the problematic concepts.[9] By preserving classical logic, this approach appears to be slightly more conservative in spirit than Kripke's inductive method (though often more difficult to use). The *Herzberger revision scheme* $\hbar$ decrees that $A^{\hbar}{}_w = \emptyset$ and $\hbar(\lambda)(B^+{}_w, B^-{}_w) = B^+{}_w$. I will call *Herzberger revision sequences* those revision sequences that employ the Herzberger revision scheme $\hbar$.

There are certain conditions under which coherence cannot be achieved no matter how many revisions are undertaken; in general, models in which paradoxical forms of self-reference are present will not be coherent. The presence of incoherent models leads to the following distinctions. φ is *doxastically positively (negatively) stable in* a model $\mathcal{M}$ *with respect to* a revision scheme $\mathcal{R}$ *at* a world w iff $\varphi \in [\![B]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w}$ for all β ($\varphi \notin [\![B]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w}$ for all β). φ *doxastically stabilizes at* an ordinal α *in* a model $\mathcal{M}$ *with respect to* a revision scheme $\mathcal{R}$ (*at* a world w) iff α is the first ordinal β such that φ is doxastically positively or negatively stable (at w) in $\mathcal{M}^{\beta}$ with respect to $\mathcal{R}$. Similarly, φ is *alethically positively (negatively) stable in* a model $\mathcal{M}$ *with respect to* a revision scheme $\mathcal{R}$ *at* a world w iff $\varphi \in [\![T]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w}$ for all β ($\varphi \notin [\![T]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w}$ for all β), and φ *alethically stabilizes at* an ordinal α *in* a model $\mathcal{M}$ *with respect to* a revision scheme $\mathcal{R}$ (*at* a world w) iff α is the first ordinal β such that φ is alethically positively or negatively stable (at w) in $\mathcal{M}^{\beta}$ with respect to $\mathcal{R}$. α is a *doxastic (alethic) stabilization ordinal for* $\mathcal{M}$ (*at* w) *with respect to* $\mathcal{R}$ iff every φ that doxastically (alethically) stabilizes in $\mathcal{M}$ (at w) with respect to $\mathcal{R}$ stabilizes at some ordinal $\leq \alpha$ in $\mathcal{M}$ (at w) with respect to $\mathcal{R}$. There is also the more general notion of a *stabilzation ordinal: or* $\mathcal{M}$ (*at* w) *with respect to* $\mathcal{R}$: α is a *stabilization ordinal for* $\mathcal{M}$ (*at* w) *with respect to* $\mathcal{R}$ iff α is a doxastic and alethic stabilization ordinal for $\mathcal{M}$ at w with respect to $\mathcal{R}$. If β is any ordinal greater or equal to the first (doxastic, alethic) stabilization ordinal for $\mathcal{M}$ with respect to $\mathcal{R}$, the model $\mathcal{M}^{\beta,\mathcal{R}}$ is called a (*doxastically*, *alethically) metastable* model. Call γ a *doxastic (alethic) perfect stabilization ordinal for*

[9]The Herzberger revisions scheme is the easiest of the semi-inductive schemes to manipulate.

$\mathcal{M}$ with respect to $\mathcal{R}$ just in case γ is a doxastic (alethic) stabilization ordinal for $\mathcal{M}$ with respect to $\mathcal{R}$, and $\varphi \in [\![B]\!]^{\gamma}_{\mathcal{M}, w}$ iff φ doxastically stabilizes at some ordinal $\leq \gamma$ in $\mathcal{M}$ at w with respect to $\mathcal{R}$ ($\varphi \in [\![T]\!]^{\gamma}_{\mathcal{M}, w}$ iff φ alethically stabilizes at some ordinal $\leq \gamma$ in $\mathcal{M}$ at w with respect to $\mathcal{R}$). Finally call γ a *perfect stabilization ordinal for* $\mathcal{M}$ with respect to $\mathcal{R}$ just in case γ is a doxastic & alethic perfect stabilization ordinal for $\mathcal{M}$ with respect to $\mathcal{R}$, and call $\mathcal{M}^{\gamma}$ a *semi-stable* model.

Just as when one predicate is defined semi-inductively, one can show that a model revision sequence with the two predicates B and T defined as above eventually yields a semi-stable model.

Proposition 1: Any Herzberger model revision sequence for L(B, T) yields a semi-stable model $\mathcal{M}^{\lambda}$ and a periodic sequence of metastable models thereafter.

The proof of proposition 1 follows a well-known path. A standard cardinality argument immediately implies that any Herzberger model revision sequence must yield a model $\mathcal{M}^{\gamma}$ with γ a stabilization ordinal. The observation that the revision sequence must yield a stabilization ordinal implies that every sentence that does not stabilize must after a certain number ζ of revisions begin to show a cyclic pattern of evaluations after ζ with period say δ. Suppose that the longest such period is θ. We may assume without loss of generality that $\theta < \gamma$. Now consider the limit ordinal λ of the sequence $\gamma + \theta$, $\gamma + 2\theta, \ldots$ λ is a perfect stabilization ordinal, and so $\mathcal{M}^{\lambda}$ is semi-stable. Thereafter, perfect stabilization ordinals must occur with a regular pattern, with all models $\mathcal{M}^{\alpha}$ being semistable, for $\alpha \geq \lambda$.

If we confine our attention to the class of coherent models and chose the appropriate alternativeness relation, we get a well-behaved theory of belief and truth--for instance a sentential version of S4 + the Tarski biconditional for every sentence of L.[10] But in the general case, matters are much less satisfying. Essentially incoherent model structures yield counterexamples to the axiom schema $B(\ulcorner B\ulcorner\varphi\urcorner \rightarrow \varphi\urcorner)$ and to the Tarski biconditional schema at arbitarily large successor, limit and perfect stabilization ordinal stages in the revision process. Moreover, given essentially incoherent model structures such that the revision procedure yields only incoherent models, the other axiom schemata and axioms for the truth predicate like those in Turner (1987) are not closed under the rule: if φ is an L-instance of one of the axioms then $B\ulcorner\varphi\urcorner$ is a theorem. Nevertheless, the quantificational theory of first order logic is left intact. In particular the axioms $\forall x\, \varphi \rightarrow \varphi(t/x)$, $\forall x(\varphi \rightarrow \psi) \rightarrow (\forall x\varphi \rightarrow \forall x\psi)$ and $\psi \rightarrow \forall x\, \psi$ where x does not occur free in ψ are valid in every $\mathcal{M}^{\alpha}$ for $\alpha \geq 1$. That the variables of quantification may range over sentences does not alter this fact.

[10]For details see Gupta (1982), Asher & Kamp (1989) thm 14. Care must be taken to state this theory given the limited resources of L needed to insure that models for L be coherent. But I won't bother to give the details here. They are discussed at some length in Asher & Kamp (1989).

The first order framework here entails that variables of quantification only occur in argument positions to relational symbols; there are no variables occuring in predicate positions. In particular variables do not occur in 0-place predicate positions-- i.e., in the positions of sentences or formulas. So propositions are quantified over only insofar as they are arguments to properties. This has to some extent a natural language analogue.[11] There is no quantification over the denotations of sentences in natural language, it would seem, only over the denotation of sentential nominalizations. So, for instance, there is no way to express in natural language the higher order formula $\forall p\ (p \vee \neg p)$ except by quantifying over nominalizations and using a predicate like 'true' as follows.

(6) Every proposition is such that either it or its negation is true.

If we attempt to quantify in a natural language over sentences rather than sentential nominalizations, we lose the verbs and predicates necessary to make grammatical sentences in natural language. The first order approach to propositional quantification appears to reflect the logical form of natural language, at least at first glance.

The first order quantification over propositions in this framework makes it possible to translate Prior's problematic belief, which I expressed as (3). Before doing so, however, we must include within our notion of an L(B, T) model a set of times I as well as worlds and objects. Our predicates for belief and truth will also now contain argument places for terms referring to times. The revision procedures for predicates will also be relativized to times, and there are a variety of ways the new revision procedure might go. The most obvious is to require the following revisions to the recursion clauses:

i) $[\![B]\!]^{0,\mathcal{R}}{}_{w,t} = [\![B]\!]_{w,t};\ [\![T]\!]^{0,\mathcal{R}}{}_{w,t} = A^{\mathcal{R}}{}_{w,t}$

ii) $[\![B]\!]^{\alpha+1,\mathcal{R}}{}_{w,t} = \{<a, \varphi, t'>: (\forall w' \in R_{a,\mathcal{M}})\ [\![\varphi]\!]_{\mathcal{M}^{\alpha}, w', t'} = 1)\}$

iii) $[\![T]\!]^{\alpha+1,\mathcal{R}}{}_{w,t} = \{<\varphi, t'> : [\![\varphi]\!]_{\mathcal{M}^{\alpha}, w, t'} = 1)\}$

iv) $[\![B]\!]^{\lambda,\mathcal{R}}{}_{\mathcal{M},w,t} = \mathcal{R}(\lambda)(B^{+}{}_{w}, B^{-}{}_{w})$, where

$B^{+}{}_{w,t} = \{<a, \varphi, t'>: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow <a, \varphi, t'> \in [\![B]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w,t}$
$)\}$

and

$B^{-}{}_{w,t} = \{<a, \varphi, t'>: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow <a, \varphi, t'> \notin [\![B]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w,t}$
$)\}$.

v) $[\![T]\!]^{\lambda,\mathcal{R}}{}_{\mathcal{M},w,t} = \mathcal{R}(\lambda)(T^{+}{}_{w}, T^{-}{}_{w})$, where

$T^{+}{}_{w} = \{<\varphi, t'>: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow <\varphi, t'> \in [\![T]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w,t})\}$

and

$T^{-}{}_{w} = \{<\varphi, t'>: (\exists\gamma<\lambda)(\forall\beta)(\gamma<\beta<\lambda \rightarrow <\varphi, t'> \notin [\![T]\!]^{\beta,\mathcal{R}}{}_{\mathcal{M},w,t})\}$.

[11]Noted for instance by Bealer (1982) Turner .(1987) As will be evident I do not think of this evidence as that conclusive.

Let's suppose that we also introduce designated constants t_0, t_1, etc. to designate elements of I and that predicates like T and B acquire an additional argument place for the temporal parameter. Call this language L_1. The reformulation of the revision procedure for L_1 is quite obvious. Finally, now we can formalize and evaluate Prior's thought within L_1.

$$(7)\ [\forall x\ (S(x) \rightarrow (B(a, x, t_0) \rightarrow \neg true(x, t_0))) \vee \forall x\ (S(x) \rightarrow (think(b, x, t_1) \rightarrow \neg true(x, t_1)))]\ \&\ [\exists x\ (S(x)\ \&\ (B(b, x, t_1)\ \&\ true(x, t_1)) \vee \exists x\ (S(x)\ \&\ think(a, x, t_0)\ \&\ (true(x, t_0))]$$

This somewhat longwinded formula can be treated without paradox in the semi-inductive theories of truth and belief with the Herzberger revision scheme.[12] Consider those models in which the model structures verify the snow is white for all times t. The Herzberger revision scheme predicts that in the situation in which Tarski only believes at t_1 that snow is white and Prior only believes (7) at t_0, (7) will be unstable, alternatively true at one model revision and false at the next model revision. Kripke's theory predicts that in such a circumstance (7) will get the value gap. In general the first order theories of belief and truth using either inductive or semi-inductive schemes of model revision treat (7) as pathological in a way similar to the Liar.

This theory of propositional quantification also does not lack for expressive power. One can talk about common or mutual belief,[13] and one can make headway on the semantics of attitude reports. Nevertheless, one can object to this treatment of indirect discourse for two reasons. The first is that the rules for simple modalities (like necessity) have to be rather drastically altered. Montague (1963) and Montague and Kaplan (1960) showed that the axioms of T formulated for any 1-place predicate of sentences was together with Robinson arithmetic inconsistent.[14] The representationalist theory of attitudes and propositions-- in particular the parallel that I have been stressing between representations and sentences-- make these conclusions of Montague inescapable. But on the other hand, models for syntactic treatments of modalities are now known, and a unified treatment of truth along with other complementizers appears to be an advance for semantics. The standard modal logics might be seen as approximations (or idealizations of) the logic of modal and attitudinal complementizers. The axioms for these logics hold for the unproblematic parts of our language-- those sentences that don't employ any self-reference. One is tempted to say that the axioms are "usually true," and in this they resemble rules of default reasoning.

---

[12] I should note that Kripke's inductive definition of truth may also be used with (6) to solve the Prior-Thomason paradoxes of indirect discourse. or a number of other semi-inductive shcmes .

[13] See my ' Common Knowledge and Model Revision ', talk delivered at the American Philosophical Association, Central Division Meetings in St. Louis, MO, 1986. The manuscript is available from the author.

[14] Later Thomason (1980.a) showed that weak S4 (S4 - the axiom schema $B(\ulcorner\varphi\urcorner) \rightarrow \varphi$) formulated for any 1-place predicate was, together with Robinson arithmetic and the assumption that one believed something, inconsistent. For Montague the predicate might represent necessity or 'it is known that'. In Thomason's result, we might suppose that the one place predicate stands for 'it is believed that'

It is only when we come across unusual propositions in which paradoxical self-reference occurs that the standard axioms for truth and belief fail. we assume stability and reason as if all statements are stable. Sometimes we have to revise our assumptions.

A second objection to the syntactic approach to modalities relies on an old argument of Church's. Church argued that sentences were an inappropriate object of attitudes because that entailed that two monolingual speakers of different languages could not share beliefs. But this objection obviously doesn't hold of representational theories of attitudes. Representationalism is compatible and even encourages the idea that there is a common representational system across humans who speak different languages. Stalnaker (1984) argues that humans and animals may share beliefs too, and indeed here the postulation of a common representational system across species is less plausible. But I think the representationlist has many replies to this challenge too. The most attractive, I think, is to think of a relation between the representation and the mental state it is supposed to characterize as somewhat flexible-- certainly more flexible when reporting beliefs of other types of agents than adult humans. This proposal does not entail a common representational system across species.

If one is persuaded by anti-representationalist arguments, however, one can abstract away from the syntactic approach to modalities and still remain within a first order theory of propositional quantification. I don't applaud this move, since it is the conceptualist's thesis about propositions as sentence-like entities that yields useful formal tools for dealing with propositional quantification and theories of truth and belief. A more general approach would introduce a set of propositions P instead of sentences of L as a subset of the domain of an LBT model. With each sentence of L, we would associate an extension and an intension; the intensions of sentences on this view would naturally in such a theory be an element of P, the extension a truth value.[15] One would then add as in Turner (1989) an operator $\triangledown$ of the language that makes the embedded sentence denote its intension rather than its extension. By supposing that every sentence yields a unique element of P -- i.e., by taking as a valid principle $\exists! p\ p = \triangledown\varphi$ for all φ-- we may use the self-referential properties of L sentences to get at the properties of self-referential propositions. On the other hand, we may exploit $\triangledown$, the truth predicate of propositions and identity to define the liar directly as a sentence of our language. Let 'p' be a propositional constant of L and let L contain identity, a truth predicate of propositions and the operator $\triangledown$. Then the Liar is expressed by the following sentence of L:

$$p = \triangledown\neg \mathrm{true}(p)$$

---

15 This use of extension/intension is due to Turner (1987) (1989).

Though perhaps more abstract, it does not appear as though the generalization of the sentential approach to a theory of first order propositions yields substantially different results. Extensions to this fragment would include propositional functions and propositions as one way of modelling properties and relations as well as propositions (Aczel (1989), Turner (1989)).[16]

2. Propositional Quantification and Higher Order Logic

2.1 Motivations

The first order theory of propositional quantification I have just sketched is essentially limited in that it quantifies only over "argument place" positions. We could quantify also over relations and properties, considering propositions to be 0-place properties. Quantification over predicate positions is the syntactic criterion for a higher order logic. The expressive power of higher order logic is quite attractive when thinking about mathematical theories. When we think of a theory like standard set theory or arithmetic we think of a certain canonical structure. We find the Lowenheim Skolem Tarski theorems surprising, even paradoxical when applied to theories of these structures (as we think of them naively) Higher order logic can describe these structures up to isomorphism, and the Lowenheim Skolem Tarski theorems don't hold for higher order theories.[17] This is one sign that our mathematical views might be couched in higher order, not first order terms.

There is also evidence in natural language of at least an indirect sort that we do directly quantify over higher order objects, and not just their first order correlates that some have assumed to be the denotions of sentential and verbal nominals. The evidence has to do both with anaphoric reference to abstract entities and quantification. Consider the following counterpart for verb phrases to the argument given by Bealer (1982) for *that* clauses or sentential nominals.

John does everything that Mary does.

Mary solves math problems.

Therefore, John solves math problems.

I take this to be a valid inference, of the same general form as,

Everything that Mary believes is true.

---

[16] Aczel (1989) proves completeness and soundness of a generalized predicate logic with quantification over propositions-- something which we cannot do. But this is because Aczel's language is quite restricted in expressive power; there are operators on propositions like □ and the truth functional operators, but the language contains no predicates of propositions and no machinery with which to construct instances of the diagonalization lemma that are the mark of self-referential theories.

[17] Per Lindstrom suggested in conversation that one might explain our categorial mathematical thinking by resorting to some direct grasp of certain structures. But if we think of the set of all subsets of a given set as being a well-defined notion, then for all intents and purposes we have espoused extensional higher order logic. For a very good defense of the view that second order logic underlies mathematical practice see Shapiro (1985).

Mary believes that the collection of $[\![T]\!]_{\mathcal{M}}$ for all metastable $\mathcal{M}$ is non-recursive.
Therefore it is true that the collection of $[\![T]\!]_{\mathcal{M}}$ for all metastable $\mathcal{M}$ is non-recursive.

But while in the second argument it indeed looks like we are quantifying over the values of *that* clauses or sentential nominalizations -- and one can make a good case that that is first order quantification-- the first argument suggests a logical form in which quantification is directly over a verb phrase (VP) denotation-- a function from individuals to sentence extensions (or intensions). It is much less plausible linguistically to suggest that all finite VPs are in fact singular terms of a first order theory. For we must give some account of how the singular term denoting a property and a noun phrase combine to give a sentence denotation, and at the same time we must give a uniform account of the semanics of finite and infinite VPs.[18] The much simpler and more plausible hypothesis is that this inference appears to involve a quantification over higher type objects-- second order quantification or higher.

Another bit of evidence for direct quantification over propositions occurs with anaphoric reference to what a sentence expresses. Consider for instance.

Fred was an alcoholic. But none of us **believed it** until he announced he was taking a leave of absence to go to a clinic for treatment.

The boldfaced pronoun is of interest, because it is linked anaphorically with the previous sentence in the discourse above. The semantics of anaphora involves some relation between the denotation of the pronoun (some sort of variable) and the denotation or semantic value of the antecedent. In abstract entity anaphora it appears, however, that here we are asked to identify the variable introduced by *it* with the proposition expressed by the sentence (its intension). One could construct a theory of abstract entity anaphora in which the variable is bound to a sentential nominalization-- the anaphoric process transforms the sentence into its nominal correlate. But a simpler hypothesis is possible if one quantifies over sentence denotations or intensions; one could give the logical form of the discourse above as $\exists p$ (p = Fred is an alcoholic & none of us believed p...). Sentential quantification like this is not in general first order definable, as Fine (1977) showed. So facts about anaphoric reference to propositions in natural language indicates that quantification over sentential argument places leads to a theory of higher order quantification.

A similar argument arises from VP ellipsis, an example of which is *Fred likes a drink after work and Sue does ø too.* One plausible semantic analysis of VP ellipsis is that the null VP (indicated by ø in the example) introduces a variable bound to value of

[18]Chierchia's (1985) argues for distinguishing finite VPs as denoting unsaturated properties and infinite VPs are nominalizations of finite VPs. The latter denote hypostizations of unsaturated properties, "nominal correlates" Chierchia calls them. A crucial feature of his account is that these nominal correlates are of the wrong type to combine with NP denotations.

the antecedent VP. For analogous reasons to those I have just sketched, VP ellipsis also seems to offer *prima facie* evidence of higher order quantification in natural language. One could make do with a first order theory of quantification, but the higher order approach is more natural and in keeping with general assumptions about anaphora and the semantics of VPs.

2.2 Standard Theory of Types

This train of thought leads to a different theory of propositional quantification, the one that Thomason and Prior had in mind.[19] Syntactically, propositional variables and constants are 0-place property variables and constants. The language of propositional quantification, L_2, is thus a second order language. However, I shall consider a natural extension, L_ω, the language of the theory of simple types.[20] Formulas are constructed in the usual manner from the truth functional connectives and quantifiers. L_ω is a language containing individual and temporal constants and variables for all finite types formed from the basic primitive types-- P (the set of propositions), E (the set of individuals) and T (the set of truth values {0, 1}). Formulas are defined for each type using λ-abstraction and functional application. So for instance, if ζ is a formula of type τ and x is a variable of type τ', then $\lambda x\zeta$ is a formula of type $\tau \rightarrow \tau'$, and if ψ is of type $\tau \rightarrow \tau'$ and β is of type τ, then $\psi(\beta)$ is of type τ'.

$L\omega$ has extensional and intentional versions of the connectives and quantifiers. $\forall$, $\exists$, &, $\vee$, $\rightarrow$, $\neg$ will be the truth functional operators and quantifiers, while Π, Σ, $\cap$, $\cup$, $\Rightarrow$ and $\sim$ will be the intensional correlates. Extensional identity, =, also has an intentional correlate, $\approx$. I shall also assume that in the language there is also a function constant $\vee$ from propositions to their truth values as in Thomason (1980) (manuscript). Note that ${}^\vee p$ is not considered to be a proposition!

We insure a homomorphism between extensional and intensional correlates if we take the following as axioms:[21]

(HOM)

for all p, q: ${}^\vee[p \cap q] = {}^\vee p \,\&\, {}^\vee q$ ${}^\vee[p \cup q] = {}^\vee p \vee {}^\vee q$ ${}^\vee[p \Rightarrow q] = {}^\vee p \rightarrow {}^\vee q$

for all ζ: ${}^\vee[\,\Pi x\tau\,\zeta] = \forall x\tau{}^\vee\zeta$ ${}^\vee[\Sigma x\tau\,\zeta] = \exists x\tau{}^\vee\zeta$

for all p: ${}^\vee[\sim p] = \neg{}^\vee p$

for all t, t': ${}^\vee[t \approx t'] = \vee\,[t = t']$

---

[19]There are arguments for getting rid of types in doing natural language semantics. But I want to sidestep those here, as they usually revolve around a treatment of properties (with one or more argument places!) and this would lead us too far afield here.

[20]It is interesting to note that some difficulties such as those in the last section of the paper arise in full type theory but not simple quantification over propositions and properties in intentional logic. This seems to cast doubt on the equivalence in intentional logic between second order and full type theory. This equivalence is a fact of extensional, higher order logic.

[21]A weaker theory of propositions without (ABS) call it P-ABS would need to require of HOM in addition that

$\vee\, \lambda x^\tau A(\beta) = {}^\vee A(\beta/x^\tau)$ if β is of type τ

P-ABS is already alluded to in footnote 21 above.

To get complete freedom in choosing one's intentional logic for the attitudes, it is better to give for each usual extensional quantifier and connective an intentional operator and quantifier. But for the statement of various truth definitions, it is very tiresome to read recursive clauses for each quantifier and connective; so in what follows I shall illustrate the various definitions by just exploiting the connectives, quantifiers and operators in the first column of the above table. The rest of the cases are always entirely obvious, and the interested reader may easily fill them in.

Once variables range over sentence denotations, it no longer make sense to take these to be truth values a la Frege, Carnap and Montague, if we wish to justice to propositional attitudes and other intensional contexts. Rather, we must take the denotations of sentences to be propositions. A sentence will be true iff the proposition it denotes is true. Thus, (2.a) expresses the proposition,

(8) $\Pi p\,(\text{believe}(\text{mary}, p) \Rightarrow p)$.

(8) is a formula of L_ω; in L_ω 'believe' is a second order predicate of individuals and propositions. By the correspondence rules in (HOM) (7) and hence (2.a) are true just in case,

$\forall p\,({}^{\vee}\text{believe}(\text{mary}, p) \rightarrow {}^{\vee}p)$,

where p ranges over the domain of propositions.

A *standard intentional model* with times of L_ω consists of a quadruple $<\underline{\underline{E}}, [\![\,]\!], f, \mathcal{F}>$. $\underline{\underline{E}}$ is an inductively defined set of domains of various types, with non-empty sets E_0, E_P, E_I and E_T (of individuals, propositions, times and truth values respectively) as the basic types of objects. Other types are constructed from basic types as functions from types to types. In a standard model, if $\tau_1, \ldots \tau_n$ are types, then the set of all objects of type $<\tau_1, \ldots, \tau_n>$, $E(<\tau_1, \ldots, \tau_n>)$, $= \wp(E(\tau_1) \times E(\tau_2) \times \ldots \times E(\tau_n))$. The interpretation of expressions of the other types are the functions constructible from these basic types. I shall also assume that types are closed under functional application.[22] So

(FA) if ν is of type $\tau \rightarrow \tau'$ and ζ of type τ, then $\nu(\zeta) \in E_{\tau'}$

$[\![\,]\!]$ assigns an (intentional) interpretation to each expression of type τ; the interpretation is some element of $E(\tau)$. The interpretation function of an intentional model respects λ abstraction and application in its assignments. That is, we have for any term α of type τ and any term $\lambda x\,\beta$ of type $\tau \rightarrow \tau'$,

(ABS)

$$[\![\lambda x \beta]\!]([\![\alpha]\!]) = [\![\beta(x/\alpha]\!].$$

Our theory is intentional so the objects assigned to predicates of a language by $[\![\,]\!]$ are properties and relations, not sets. Since sets are useful in the truth definition,

22(ABS) is an optional constraint. One might require simply that $\lambda x^\tau A(\beta)$ and $A(x^\tau/\beta))$ coincide in truth value in every model, which could be imposed by HOM below however, intentional models have a function f that assigns to each object in a type a certain extension. Let [] be the extension of ⟦⟧ and f to include the assignment of denotations to complex terms of the form $^{\vee}\varphi$. Then $\mathcal{F}$ is a function from P X I into T = {0, 1} such that:

i. $\mathcal{F}_t(G(a_1, .. a_n)) = 1$ iff $\langle⟦a_1⟧, .., ⟦a_n⟧\rangle \in f(⟦G⟧)$

ii. $\mathcal{F}_t(p \cap q) = 1$ iff $\mathcal{F}_t(p) = \mathcal{F}_t(q) = 1$

iii. $\mathcal{F}_t(\sim q) = 1 - \mathcal{F}_t(q)$

iv. $\mathcal{F}_t(\Pi x^{\tau} \zeta) = 1$ iff $\mathcal{F}_t(\zeta(a)) = 1$ for all objects a of type τ

v. $\mathcal{F}_t(\alpha \approx \beta) = 1$ iff $[\alpha]_t = [\beta]_t$

(similarly for the other operators)

If ⟦φ⟧ is a proposition, $^{\vee}\varphi$ is a singular term denoting in $\mathcal{M}$ the truth value of ⟦φ⟧ in $\mathcal{M}$. It requires a special interpretation. Further, these singular terms may combine with truth functional operators and quantifiers, which will have the usual recursive, semantic clauses. Let us write $[A]_{t, M} = 1$ if A denotes in M truth at t; $[A]_{t, M} = 0$ otherwise.

a. If A is of the form $^{\vee}\varphi$ where $⟦\varphi⟧_{\mathcal{M}}$ is a proposition, then $[A]_{t, M} = \mathcal{F}(\varphi)$

b. If A is of the form B & C, then $[A]_{t, M} = 1$ iff $[A]_{t, M} = 1$ and $[A]_{t, M} = 1$.

c. If A is of the form $\neg B$, then $[A]_{t, M} = 1$ iff $[A]_{t, M} = 0$.

d. If A is of the form $\forall x^{\tau} \zeta$, $[A]_{t, M} = 1$ iff $[\zeta(a/x)]_{t, M} = 1$ for all a of type τ.

e. If A is of the form $^{\vee}[\alpha = \beta]$, $[A]_{t, M} = \mathcal{F}_t(\alpha \approx \beta)$.

f. If A is of the form at($^{\vee}\varphi$, t), $[A]_{t, M} = \mathcal{F}_t(\varphi)$.

(Similarly for the other operators)

Let T_0 be the theory given by the axioms below and closed under the rule modus ponens.

(i) $^{\vee}\varphi$, where φ is a tautologous proposition.

(ii) $\forall x^{\tau} (^{\vee}\varphi \rightarrow {}^{\vee}\psi) \rightarrow (\forall x^{\tau\vee}\varphi \rightarrow \forall x^{\tau\vee}\psi)$, where x^{τ} is a variable of any type τ.

(iii) $\forall x^{\tau}\, {}^{\vee}\varphi \rightarrow {}^{\vee}\varphi(t/x)$ where t is substitutable in φ for x^{τ}, a variable of any type τ.

(iv) $^{\vee}\varphi \rightarrow \forall x^{\tau}\, {}^{\vee}\varphi$, where x^{τ} does not occur free in φ and is a variable of any type τ.

(v) The usual axiom and rule for identity: $\mathcal{F}(\alpha \approx \alpha)$, and if $\mathcal{F}(\beta \approx \beta') = 1$, then $\mathcal{F}(\psi(\beta)) = \mathcal{F}(\psi(\beta'))$

(vi) $\lambda x^{\tau} \varphi[\alpha] = \varphi(\alpha/x^{\tau})$, where α is a term of type τ.[23]

T_0 contains desirable axioms for identity, quantification and the truth functional connectives. Given this definition of intentional models, every intentional model for L_{ω} $\mathcal{M}$, verifies (HOM) as well as the usual rules of predicate logic and β-conversion.

[23]The appropriate axiom for P-ABS instead of the identity for functional application is the schema, $\lambda x^{\tau} \varphi[\alpha] \leftrightarrow \varphi(\alpha/x^{\tau})$

Proposition 2: Let $\mathcal{M}$ be any intentional model. Then $\mathcal{M}$ is a model for for (HOM) and for T_0.

The models for L_ω impose a structure on P.[24] P is closed under the operations $\cap$, $\sim$; Π must be a function from PF $\rightarrow$ P, where PF is the set of propositional functions $\{f \mid f: E \cup P \rightarrow P\}$. I will take P to be an algebra whose atoms are given by the atomic sentences of L_ω.

Let's now formulate the intentional paradoxes or paradoxes of indirect discourse within this theory. I'll assume some standard addition of constants for times and set of times in the models for $L\omega$. The proposition denoted by (3) is easily expressed in L_ω, and it is true just in case (9) holds.

$$(9)\ (\forall p\ ({}^{\vee}B(prior, p, t_0) \rightarrow \neg at({}^{\vee}p\ t_0) \vee \forall p\ ({}^{\vee}B(tarski, p, t_1) \rightarrow \neg at({}^{\vee}p, t_0))) \ \&$$
$$(\exists p\ ({}^{\vee}B(tarski, p, t_0\ ) \ \&\ at({}^{\vee}p, t_0)) \vee \exists p\ (\ {}^{\vee}B(prior, p, t_1) \ \&\ at({}^{\vee}p, t_0)))$$

We can easily show:

Proposition 3: There is no intentional model for L_ω $\mathcal{M}$ such that Prior thinks (9) at t_0 in $\mathcal{M}$, Tarski thinks that snow is white at $t_1$in $\mathcal{M}$ and 'snow is white' is true at t_1 in $\mathcal{M}$.

One should note that from the perspective of the simple theory of types, Prior's "paradox" differs from the semantic paradoxes like the Liar and paradoxes of application and comprehension like Russell's predicative paradox. There is no question of inconsistency in the theory T_0 or in HOM, and the simple intentional theory of types is after all a highly restricted framework (in comparison, for instance, to ZF). Nevertheless, Prior's thought experiment yields entirely unsatisfactory results.

2.3 The Partial Theory of Types

The reason why this theory of propositional quantification gets into difficulties is not hard to discover, if we contrast the higher order theory of propositions with the first order theory of the previous section. As the translations for (2.a) and (3) in higher order logic make evident, the truth predicate has disappeared into the theory of propositional quantification. The higher order theory of quantification (as Ramsey and Prior might naturally have suggested) yields a "pro-sentential theory of truth," on which the truth predicate in English is just an anaphor, or perhaps even more simply a dummy or redundant predicate needed because of the limitations of natural language syntax. The theory of quantification has in effect swallowed up the truth predicate. To fix the sort of difficulties that Priorean thought experiments like (3) give rise to, then, the

[24]A couple of facts about ${}^{\vee}$ are immediate once we realize it is a function constant in fact denoting $\mathcal{F}$. First of all, ${}^{\vee}$ does not iterate; so ${}^{\vee}\,{}^{\vee}\,\varphi$ isn't well-defined. Thus any identity statement like $p = {}^{\vee}\,[\sim p]$ is false in every model! Further, we might symbolize the Liar as ${}^{\vee}\,p \leftrightarrow \neg{}^{\vee}\,p$. But this sentence too is false in every model; it is a simple contradiction. Thus, the stipulative version of the Liar does not pose any problems in this higher order logic. Higher order logic says that the liar is false in every model. Note also that the strong liar, whichsays that the Liar is false is logically true!

natural suggestion is to do for quantification what Kripke and Gupta-Herzberger have done for predicates like truth. Just as truth is defined inductively or semi-inductively mirroring the restrictions of the Tarskian hierarchy, so too is quantification to be similarly bounded by types until the construction is finished. My proposal complicates the connection between propositions and their truth values in intentional models by using either semi-inductive or inductive definitions for the domains of quantification.

2.3.1 A Semi-Inductive Theory of Propositional Quantification

Let me make the suggestion a bit more precise by looking at the semi-inductive case first. Let $\mathcal{M}_0$ be a standard intentional model for type-theory. I distinguish a subset of E_P, P_0, which contains just those propositions not containing propositional variables or constants. We now define a revision sequence of models $\mathcal{M}_{QH}{}^{\alpha}$ as follows. Let $\mathcal{M}_{QH}{}^{\alpha} = <\underline{E}, [\![]\!], f, \mathcal{F}^{\alpha}>$. We now define a recursion for $\mathcal{F}^{\alpha}$ on the ordinals. $\mathcal{F}^0 = \mathcal{F}\!\restriction\! P_0 \cup (P\text{-}P_0 \times \{0\})$. All the definitions for $\mathcal{F}$ and the assignment of truth values to terms of the form $^{\vee}\varphi$ largely the same as before with the exception of the quantified clauses:

i. $\mathcal{F}_t{}^{\alpha}(G(a_1, .. a_n)) = 1$ iff $<[\![a_1]\!], .., [\![a_n]\!]> \in f([\![G]\!])$

ii. $\mathcal{F}_t{}^{\alpha}(p \cap q) = 1$ iff $\mathcal{F}_t{}^{\alpha}(p) = \mathcal{F}_t{}^{\alpha}(q) = 1$

iii. $\mathcal{F}_t{}^{\alpha}(\sim q) = 1 - \mathcal{F}_t{}^{\alpha}(q)$

iv. $\mathcal{F}_t{}^{\alpha+1}(\Pi x^{\tau} \zeta) = 1$ iff $\mathcal{F}_t{}^{\alpha}(\zeta(a)) = 1$ for all a of type $\tau \neq P$.

v. $\mathcal{F}_t{}^{\alpha}(\alpha \approx \beta) = 1$ iff $[\alpha]_{\mathcal{M}^{\alpha}} = [\beta]_{\mathcal{M}^{\alpha}}$

vi. If A is of the form $^{\vee}\varphi$ where $[\![\varphi]\!]_{\mathcal{M}}$ is a proposition, then $[A]_{\mathcal{M}^{\alpha}} = \mathcal{F}_t{}^{\alpha}(\varphi)$

vii. If A is of the form B & C, then $[A]_{t, \mathcal{M}^{\alpha}} = 1$ iff $[B]_{t, \mathcal{M}^{\alpha}} = 1$ and $[C]_{t, \mathcal{M}^{\alpha}} = 1$.

viii. If A is of the form $\neg B$, then $[A]_{t, \mathcal{M}^{\alpha}} = 1$ iff $[B]_{t, \mathcal{M}^{\alpha}} = 0$.

ix. If A is of the form $\forall x^{\tau} \zeta$, $[A]_{t, \mathcal{M}^{\alpha}} = 1$ iff $[\zeta(a^{\tau}/x)]_{t, \mathcal{M}^{\alpha}} = 1$ for all a^{τ} $\tau \neq P$.

x. If A is of the form $^{\vee}[\alpha = \beta]$, $[A]_{t, \mathcal{M}^{\alpha}} = \mathcal{F}_t{}^{\alpha}(\alpha \approx \beta)$.
(Similarly for the other operators and non-propositional quantifiers)

The clauses for the propositional quantifiers must be defined relative to previous models in the sequence. We need a pair of clauses for successor and limit ordinal cases.

xi.a. $\mathcal{F}_t{}^{\alpha+1}(\Pi x^P \zeta) = 1$ iff $\mathcal{F}_t{}^{\alpha}(\zeta(t^P)) = 1$ for all t^P.

xii.a. If A is of the form $\forall x^P \zeta$, $[A]_{t, \mathcal{M}^{\alpha+1}} = 1$ iff $[\zeta(t^P/x)]_{t, \mathcal{M}^{\alpha}} = 1$ for all t^P.

xi.b. $\mathcal{F}_t{}^{\lambda}(\Pi x^P \zeta) = 1$ iff $\exists\beta\, \forall\alpha\, (\beta \leq \alpha < \lambda \rightarrow \mathcal{F}_t{}^{\alpha}(\Pi x^P \varphi) = 1)$.

xii.b. If A is of the form $\forall x^P \zeta$, $[A]_{t,\, \mathcal{M}\lambda} = 1$ iff $\exists\beta\, \forall\alpha\, (\beta \leq \alpha < \lambda \rightarrow [A]_{t,\, \mathcal{M}\alpha} = 1$.

(Similarly for Σx^P)

The first stage of our model revision procedure now may have a quantificational incoherence in there. For instance, a quantificational proposition of the form $\pi x^P \varphi$ will be false in $\mathcal{M}^0$ even though all it's instances may be true. But this incoherence is erased once the revision procedure gets started. We can still show that every model $\mathcal{M}^{\alpha}{}_{QH}$ in the revision sequence defined verifies (HOM).

Proposition 4: Let $\mathcal{M}^{\alpha}{}_{QH}$ be a revision model for $\alpha > 1$. Then $\mathcal{M}^{\alpha}{}_{QH}$ is a model for HOM.

The proof of proposition 4 proceeds by induction as before. The only interesting case comes with the propositional quantifiers, and that case is easily proved from clauses vi., xi., and xii. But notice that this gives us a different correlation between a quantificational statement and its instances:

$$\mathcal{M}_{QH}{}^{\alpha+1} \models \forall p\, {}^{\vee}\varphi \text{ iff } \mathcal{F}^{\alpha}(\varphi(t/p)) = 1 \text{ for any } t \in P.$$

$$\mathcal{M}_{QH}{}^{\lambda} \models \forall p\, {}^{\vee}\varphi \text{ iff } \exists\beta\, \forall\alpha\, (\beta \leq \alpha < \lambda \rightarrow \mathcal{F}^{\alpha}(\Pi p\, \varphi) = 1)$$

Our model revision procedure now yields eventually a *higher order semistable* model, as all sentences with a string of propositional quantifiers of a given depth that will stabilize eventually do so. $\mathcal{M}^{\delta}$ is a *higher order semistable* model just in case δ is a perfect stabilization ordinal for $\mathcal{M}$ with respect to the revision sequence above and $\mathcal{F}$. Let $\mathcal{M}^{\gamma}$ be such a model. Prior's belief, (9), is false at $\mathcal{M}^{\gamma}$, if Tarski's belief is true. Moreover, the truth of Tarski's belief, if it is a simple proposition, does not depend upon Prior's thinking (9) or not thinking (9). So far so good. But a rather surprising result is in store for us:

Proposition 5: There is no semi-inductive model such that such that (i) Prior thinks (9) at t_0 in $\mathcal{M}$ and nothing else, (ii) Tarski thinks that snow is white at t_1 in $\mathcal{M}$ and nothing else, (iii) 'snow is white' is true at t_1 in $\mathcal{M}$, and (iv) $\mathcal{M}$ is a model of T_0.

The proof proceeds by an examination of cases. We observe that on such a theory (9) also has a 2 cycle interpretation. Any $\mathcal{M}^0$ cannot be a model of T_0, because the T_0 theorem $\varphi(c^P) \rightarrow \exists x^P \varphi(x^P)$ is false at $\mathcal{M}^0$, where c^P is a propositional term. Successor states $\mathcal{M}^{\gamma+1}$ either fail to verify $\forall x^P\, \psi(x^P) \rightarrow \psi(c^P/\, x^P)$, where ψ is either the subformula

$${}^{\vee}B(\text{prior}, p, t_0) \rightarrow \neg at({}^{\vee}p\ t_0)$$

or

$${}^{\vee}B(\text{tarski}, p, t_1) \rightarrow \neg at({}^{\vee}p, t_0)$$

of (9); or they share the following difficulty with limit stages $\mathcal{M}^{\lambda}$. $\mathcal{M}^{\lambda} \models$ (9) iff $\mathcal{M}^{\lambda} \models$ $\forall x^p$ $({}^{\vee}B(a, x^p, t_0) \rightarrow \neg^{\vee} x^p)$ iff $\exists\beta\ \forall\gamma\ (\beta \leq \gamma < \lambda \rightarrow \mathcal{M}^{\gamma} \models \forall x^p\ ({}^{\vee}B(a, x^p, t_0) \rightarrow \neg^{\vee} x^p)$. So $\mathcal{M}^{\lambda} \models \neg(9)$. But then by ordinary quantificational logic, $\mathcal{M}^{\lambda} \models \neg(9)$ iff $\mathcal{M}^{\lambda} \models \exists x^p$ $({}^{\vee}B(a, x^p, t_0)\ \&\ {}^{\vee} x^p)$. But by the constraint (i) of the proposition, $\mathcal{M}^{\lambda} \models \neg(9)$ iff $\mathcal{M}^{\lambda} \models$ (9).

Proposition 5 uses the Herzberger approach. But the same result holds for all of the semi-inductive schemes that have been proposed in the literature: Herzberger's Gupta's and Belnap's. Further, it appears that no supervaluation over successor stages will give us T_0, as long as the successor stage revision models are defined in the way above. Actually, the result isn't all that surprising. The semi-inductive approach "preserves" classical logic in the first order case, because there is a truth predicate to revise. Classical logic is preserved there at the cost of falsifying the Tarski truth scheme. In the higher order case, the semi-inductive scheme falsifies the quantificational axioms for propositional quantification. But these should, I think, be considered part of the backbone of higher order logic. Any approach which does not preserve these rules in some format is not a satisfactory approach to the version of the intentional paradoxes cast in higher order logic.

2.3.2 An Inductive Definition of Propositional Quantification

A more satisfactory construction is available with an inductive definition like the one used by Kripke (1975). I will first consider a simple case in which $\mathcal{M}$ is a *partial* standard intentional model for L_{ω} satisfying (FA) and (ABS). There are many reasaons for looking at partial models besides those involving the paradoxes.[25] A partial intentional model is just like a standard intentional model, *except* that the assignment function f is partial and assigns extensions and anti-extensions to the basic properties. I will refer to such functions with the symbol f^*. Recall that the distinguished subset of $P_{\mathcal{M}}$, P_0, contains just those propositions not containing propositional variables. An inductive revision sequence is defined by setting $\mathcal{F}^0 = \mathcal{F}P_0$ and the *base partial model* $\mathcal{M}_{QK}{}^0 = \langle \underline{E}, [\![ ]\!], f^*, \mathcal{F}^0\rangle$, $\mathcal{M}_{QK}{}^{\alpha} = \langle \underline{E}, [\![ ]\!], f^*, \mathcal{F}^{\alpha}\rangle$, and then requiring the following constraint on $\mathcal{F}$ (which I call the *partial model constraint* PMC):

(PMC)

1. $\mathcal{F}\alpha$ and $[\,]_{\mathcal{M}^{\alpha}}$ are closed under the usual semantical rules for a strong Kleene interpretation of the truth functional connectives and non-propositional quantifiers.

[25]See for instance discussions in Langholm (1988), Muskens (1989). The types I have given are still total (though not on the set of truth values due to the partiality of ${}^{\vee}$). Muskens notes that it would perhaps be better to have partial types in an extensional theory, and I might agree with his reasoning. If so one could then modify the theory in the way he sketches.

2. All $\mathcal{M}_{QK}{}^{\alpha}$ verify identity statements of the form $\beta = \beta$, where β is any term. Otherwise,

   $\mathcal{M}_{QK}{}^{\alpha} \models \beta = \beta'$ iff $[\beta] = [\beta']$ and both $[\beta]$ and $[\beta']$ are defined in $\mathcal{M}_{QK}{}^{\alpha}$

   $\mathcal{M}_{QK}{}^{\alpha}$ $\beta = \beta'$ iff $[\beta] \neq [\beta']$ and both $[\beta]$ and $[\beta']$ are defined in $\mathcal{M}_{QK}{}^{\alpha}$

3. For propositional quantifiers (again I illustrate only for Π; the case for Σ is completely analogous),

   A. With regard to the successor case:

   i.a. $\mathcal{F}^{\alpha+1}(\Pi x^P\, \zeta) = 1$ if $\mathcal{F}^{\alpha}(\zeta(t^P)) = 1$ for all t^P.

   i.b. $\mathcal{F}^{\alpha+1}(\Pi x^P\, \zeta) = 0$ if $\mathcal{F}^{\alpha}(\zeta(t^P)) = 0$ for some t^P.

   i.c. $\mathcal{F}^{\alpha+1}(\Pi x^P\, \zeta)$ undefined otherwise.

   ii. If A is of the form $\forall x^P\, \zeta$,

   a.$[A]_{\mathcal{M}^{\alpha+1}} = 1$ if $[\zeta(t^P/x)]_{\mathcal{M}^{\alpha}} = 1$ for all t^P.

   b. $[A]_{\mathcal{M}^{\alpha+1}} = 0$ if $[\zeta(t^P/x)]_{\mathcal{M}^{\alpha}} = 1$ for some t^P.

   c. $[A]_{\mathcal{M}^{\alpha+1}}$ = undefined otherwise.

   (Similarly as in i. and ii. for the existential quantifier)

   B. The limit case may defined quite simply.

   a. $\mathcal{F}^{\lambda} = \bigcup_{\beta < \lambda} \mathcal{F}^{\beta}$

   b. $[]_{\mathcal{M}^{\lambda}} = \bigcup_{\beta < \lambda} []_{\mathcal{M}^{\beta}}$

The QK sequence of models builds up inductively the values of the partial function $\mathcal{F}_{\alpha}$ and the extensional definition [] for each α. In $\mathcal{M}_{QK}{}^{0}$ no propositionally quantified statements are given truth values. $\mathcal{F}P_0$, however, does assign every atom in the propositional algebra a truth value. After the first application of the inductive definition $\mathcal{M}_{QK}{}^{1}$ now verifies many propositions that quantify over propositions-- e.g. ${}^{\vee}\Sigma pp$. But notice that (9) will not get a value in $\mathcal{M}_{QK}{}^{1}$. In fact (9) will not get a value throughout the QK sequence. I will call the models in the QK sequence *standard partial models* for L_{ω}. The class Δ_0 of fixed points of the QK sequence as defined by PST gives us the *minimal fixed point* models definable on the class of all base partial models for L_{ω}.

Standard partial models are not models of (HOM). But they are models for a closely related theory. We must make two changes to (HOM). First, we must define correspondences for each pair of intensional and extensional connectives. Second we must replace the identities in (HOM) with rule equivalences. Call the resulting theory (HOM'):

(HOM')

$$\frac{{}^{\vee}[p \cap q]}{{}^{\vee}p \,\&\, {}^{\vee}q} \qquad \frac{{}^{\vee}[p \cup q]}{{}^{\vee}p \vee {}^{\vee}q} \qquad \frac{{}^{\vee}[p \Rightarrow q]}{{}^{\vee}p \rightarrow {}^{\vee}q} \qquad \frac{{}^{\vee}[\sim p]}{\neg{}^{\vee}p}$$

$$\underline{\underline{\forall\zeta\ {}^{\vee}[\Pi x^{?}\zeta]}} \qquad \underline{{}^{\vee}[t \approx t']} \qquad \underline{\underline{\exists\zeta\ {}^{\vee}[\Pi x^{?}\zeta]}}$$

$\forall x^\tau \vee \zeta$ $\qquad\qquad$ $\vee [t = t']$ $\qquad\qquad$ $\exists x^\tau \vee \zeta$

Notice that as with the QH revision models, the correspondence between $\mathcal{F}$ and [] is not complete:

(i) $\mathcal{M}_{QK}{}^{\alpha+1} \vDash \forall p\, {}^\vee\varphi$ iff $\mathcal{F}^\alpha(\varphi(t/p)) = 1$ for any $t \in P$.

$\mathcal{M}_{QK}{}^{\alpha+1}$ $\forall p\, {}^\vee\varphi$ iff $\mathcal{F}^\alpha(\varphi(t/p)) = 0$ for some $t \in P$.

(ii) $\mathcal{M}_{QK}{}^{\alpha+1} \vDash \exists p\, {}^\vee\varphi$ iff $\mathcal{F}^\alpha(\varphi(t/p)) = 1$ for some $t \in P$.

$\mathcal{M}_{QK}{}^{\alpha+1}$ $\exists p\, {}^\vee\varphi$ iff $\mathcal{F}^\alpha(\varphi(t/p)) = 0$ for every $t \in P$.

Nevertheless, just as for standard intentional models for type theory, we may show that the axioms in (HOM) are verified in the following sense.

Proposition 6: Any partial intentional model $\mathcal{M}^\alpha$ is a model for (HOM')

The proof of proposition 6 follows immediately from the constraints on partial models.

The rule equivalences in (HOM') form a weaker theory than (HOM) to be sure. We only have a partial homomorphism from propositions to truth values respecting the propositional and truth functional connectives and quantifiers. But we can still prove the following with it. Define a L_ω formula φ' in ${}^\vee$ *normal form* such that ${}^\vee$ occurs only in front of atomic formulas. The rules in (HOM') allow us to prove

Proposition 7: Let φ be a formula of L_ω. Then given (HOM'), there is a formula φ' in ${}^\vee$ normal form such that $\varphi \vdash \varphi'$.

Proposition 7 allows us to ignore the carrots once again.

Because the QK sequence of models is inductively defined and there is a fixed set of propositions, one can show by the standard argument that the sequence reaches a fixed point. I'll call any $\mathcal{M}_{QK}{}^\gamma$ model that is a fixed point of the definition a *standard fixed point* model for L_ω. Let R_1 be the following set of rules (corresponding to the strong Kleene interpretation of the connectives and quantifiers):

R_1

1. The usual introduction and elimination rules for $\exists$ $\forall$ & and $\vee$ generalized to all types

2. The equivalences

$$\frac{\neg\neg A}{A} \qquad \frac{\neg(A \,\&\, B)}{\neg A \vee \neg B} \qquad \frac{\neg(A \vee B)}{\neg A \,\&\, \neg B}$$

3. The rule $\psi \,\&\, \neg\psi \vdash \varphi$

4. Suppose $\varphi(\psi)$ is a positive context (ψ is a constituent that is not under the scope of any negations or relation symbols in prenex disjunctive form). Then if $\psi_1 \vdash \psi_2$, $\varphi(\psi_1) \vdash \varphi(\psi_2)$.

5. The axioms

a. $\beta = \beta$

b. $\lambda x^\tau A(\beta) = A(\beta/x)$

6. If α and α' are of type $\tau \to \tau'$ and β, β' of type τ', then $\alpha = \alpha'$ & $\beta = \beta' \vdash \alpha(\beta) = \alpha'(\beta')$

7. $\forall u\ \lambda x\ \varphi(u) = \lambda x\ \varphi'(u) \vdash \lambda x\ \varphi = \lambda x\ \varphi'$.

8. The rules in (HOM')

Say that φ is a logical consequence of ψ relative to the class of models Δ_0 just in case if $\mathcal{M} \in \Delta_0$ and $\mathcal{M} \models \psi$, then $\mathcal{M} \models \varphi$.

Proposition 8: If $\varphi \vdash \psi$ is a rule of R_1, then $\varphi \models_{\Delta_0} \psi$.

To illustrate, let us take one of the quantifier rules, the universal exploitation rule $\forall x\ \varphi \vdash \varphi\ (t/x)$. Suppose for some $\mathcal{M} \in \Delta_0$, $\mathcal{M} \models \forall x\ \varphi$. If x is of other than propositional type, then by the constraints on $\mathcal{F}$ given by the strong Kleene interpretation of the truth functional connectives and non-propositional quantifiers, $\mathcal{M} \models \varphi(t/x)$ for any suitable term t. Now suppose that x is of propositional type. By the construction of the sequence QK $\forall x\varphi$ will be true only if all its instances are verified at some previous stage, if α is a successor or limit ordinal. In either case, since the construction is inductive, this assures that $\mathcal{M} \models \varphi(t/x)$. The only other rule that may not be obvious is the rule 4. Assume that $\psi_1 \models_{\Delta_0} \psi_2$ and that for some $\mathcal{M} \in \Delta_0$, $\mathcal{M} \models \varphi(\psi_1)$ where $\varphi(\psi_1)$ is positive. We must now show $\mathcal{M} \models \varphi(\psi_2)$. We do this by induction on the complexity of φ. If φ is empty then the result is obvious, so now assume for all positive ζ of complexity less φ, the result holds. Since φ is positive, there are several cases to consider. First, $\varphi = \zeta_1 \vee \zeta_2$. Suppose that ψ_1 occurs in ζ_2. Then $\mathcal{M} \models \zeta_1 \vee \zeta_2(\psi_1)$ iff $\mathcal{M} \models \zeta_1$ or $\mathcal{M} \models \zeta_2(\psi_1)$ iff, by the inductive hypothesis, $\mathcal{M} \models \zeta_1$ or $\mathcal{M} \models \zeta_2(\psi_2)$ iff $\mathcal{M} \models \varphi(\psi_2)$. The other truth functional cases are similar. Now suppose that $\varphi = \exists x\ \zeta$. Since $\mathcal{M}$ is a fixed point model, it must contain the relevant instances to the quantifier. So by the inductive hypothesis again, we see easily that the desired result follows.

Let $\vdash_R$ be the derivation relation defined by the rules in R_1. A standard argument will now prove the soundness of R_1 relative to Δ_0. In fact we may consider a wider class Δ of models, the class of all fixed points of partial base models that employ the inductive revision procedure defined in (PMC) for successor and limit ordinals. To define these models, we begin with a base partial model $\mathcal{M} = \langle \underline{E}, [\![]\!], f^*, \mathcal{F}^0 \rangle$ and then consider a *partial base model expansion*, $\mathcal{M}^* = \langle \underline{E}, [\![]\!], f^*, \mathcal{F}^{*0} \rangle$ where $\mathcal{F}^{*0}$ is some extension of $\mathcal{F}^0$ consistent with the strong Kleene valuation rules. We now exploit the inductive revision procedure relative to all such $\mathcal{M}^*$. Let $\models_\Delta$ be the consequence relation defined over Δ. Then,

Proposition 9: For a set of sentences Γ, if $\Gamma \vdash_R \varphi$, then $\Gamma \models_\Delta \varphi$.

It appears that if we loosen the notion of a standard partial model for L_ω to get *general partial models* L_ω, we may also be able to prove a completness result for R_1.

A *general model* for L_ω from Henkin (1950) is a model in which the domains of propositions, truth values, and individuals are as before and where if $\tau_1, \ldots \tau_n$ are types, then the interpretation of a type $< \tau_1, \ldots, \tau_n>$ is a *subset* of $\wp([\![\tau_1]\!] \times [\![\tau_2]\!] \times \ldots \times [\![\tau_n]\!])$; in a standard model $[\![< \tau_1, \ldots, \tau_n>]\!] = \wp([\![\tau_1]\!] \times [\![\tau_2]\!] \times \ldots \times [\![\tau_n]\!])$. Just as there are standard partial models, there are also general partial models for L_ω. Let $\models_G$ be the consequence relation defined over all *general fixed point models*-- all those fixed point models definable relative to *partial base general model expansions*.

Proposition 10: For a set of sentences Γ, if $\Gamma \models_G \varphi$, then $\Gamma \models_R \varphi$.

The outline of the proof relies on an adaptation of the Henkin method to partial models proposed by Kamp (1984). What I shall do is show that if not $\Gamma \models_R \varphi$ then there is a partial model that verifies Γ but does not verify φ. So suppose not $\Gamma \vdash \varphi$. Define φ to be a positive formula just in case all negation signs in φ occur only on atomic formulae. We may show that for every φ there is a positive φ' that is R-equivalent to it (i.e. $\varphi \vdash \varphi'$).,

Using an enumeration of all positive formulae of L_ω, we build up two maximal sets Ω and Σ from Γ and $\{\varphi\}$ respectively as follows. I assume that infinitely many constants of each type do not occur in the enumeration of the positive formulae.

1. $\Omega_0 = \Gamma$; $\Sigma_0 = \{\varphi\}$

2.a. if not $(\Omega_n \cup \{\psi_{n+1}\} \vdash \Sigma_n)$ and ψ_{n+1} is not existential, then $\Omega_{n+1} = \Omega_n \cup \{\psi_{n+1}\}$; $\Sigma_{n+1} = \Sigma_n$

b. if not $(\Omega_n \cup \{\psi_{n+1}\} \vdash \Sigma_n)$ and $\psi_{n+1} = \exists v\zeta$, then $\Omega_{n+1} = \Omega_n \cup \{\psi_{n+1}, \zeta(c_j/v)\}$ where c_j is the first individual constant not appearing in $\Omega_n \cup \Sigma_n \cup \{\psi_{n+1}\}$; $\Sigma_{n+1} = \Sigma_n$

c. if $\Omega_n \cup \{\psi_{n+1}\} \vdash \Sigma_n$ and ψ_{n+1} is not universal, then $\Omega_{n+1} = \Omega_n$; $\Sigma_{n+1} = \Sigma_n \cup \{\psi_{n+1}\}$

d. if $\Omega_n \cup \{\psi_{n+1}\} \vdash \Sigma_n$ and $\psi_{n+1} = \forall v\zeta$, then $\Omega_{n+1} = \Omega_n$; $\Sigma_{n+1} = \Sigma_n \cup \{\psi_{n+1}, \zeta(c_j/v)\}$ where c_j is the first individual constant not appearing in $\Omega_n \cup \Sigma_n \cup \{\psi_{n+1}\}$

3. $\Omega = \bigcup_{n \in \omega} \Omega_n$; $\Sigma = \bigcup_{n \in \omega} \Sigma_n$

The next step is to show

Lemma 11: not $\Omega \vdash \Sigma$

This is proved by an induction on Ω_n and Σ_n.

I now construct a base partial intentional model $\mathcal{M} = <\underline{E}, [\![]\!], f^*, \mathcal{F}^0>$ from these sets. First I inductively define the type structure $\underline{E}$. Let $E_{0\mathcal{M}} = \{[c^0]_\Omega$: c is an individual constant occurring in $\Omega \cup \Sigma\}$, where$[c]_\Omega = \{d : \Omega \vdash d = c\}$, and let $E_{p\,\mathcal{M}} = \{[\psi]_\Omega$: ψ is a sentence occuring in $\Omega \cup \Sigma\}$. I define E_T as the set of truth values using the sentences and their negates in Ω. $\top = \{\psi: \psi \in \Omega\}$; $\bot = \{\psi: \neg\psi \in \Omega\}$. Now assume that E_τ and $E_{\tau'}$ are already defined as equivalence classes $[\alpha]_\Omega$ and $[\beta]_\Omega$ respectively. We define $E_{\tau \to \tau'}$ to be $\{[\zeta]_\Omega$: ζ is of the form $\lambda x^\tau \gamma$ and γ occurs in $\Omega \cup \Sigma\}$. We further define for each $[\zeta]_\Omega$ in $E_{\tau \to \tau'}$ to be a function such that $[\zeta]_\Omega([\alpha]_\Omega) = [\gamma(x^\tau/\alpha)]_\Omega$ for any element $[\alpha]_\Omega$ of E_τ. We can easily check this definition by exploiting (6) and (7) of R_1 and noting that for any $\alpha_1, \alpha_2 \in [\alpha]_\Omega$ and $\lambda x\gamma_1, \lambda x\gamma_2$ in $[\zeta]_\Omega$, $\lambda x\gamma_1(\alpha_1) = \lambda x\gamma_2(\alpha_2)$. Further, if $[\zeta_1]_\Omega = [\zeta_2]_\Omega$, it follows that ζ_1 and ζ_2 agree on all arguments. So $\forall u\ \lambda x\gamma_1(x)(u) = \lambda x\gamma_2(x)(u)$. By (7) we may conclude $\zeta_1 = \zeta_2$. The set of all types $\underline{E}$ for $\mathcal{M}$ are those constructed from the basic types by this procedure, and it obeys (FA).

The second step in defining the model is to specify the interpretation function. Define $[\![\,]\!]$ as follows. If φ is a term of type τ, then $[\![\varphi]\!] = [\varphi]_\Omega \in E_\tau$. Because of my definition of the type structure and because of (5.b), $[\![\,]\!]$ obeys (ABS).

The third step is to specify extensions for intentional objects. Define f^* such that for $[\beta]_\Omega \in E_{<\tau_1, \ldots, \tau_n>}$ $f^*(\beta) = <\{<[\alpha_1]_\Omega, \ldots, [\alpha_n]_\Omega>: \beta(\alpha_1, \ldots \alpha_n) \in \Omega\}$, $\{<[\alpha_1]_\Omega, \ldots, [\alpha_n]_\Omega>: \beta(\alpha_1, \ldots \alpha_n) \in \Sigma\}$. The first member of this set is the extension of β, the second its antiextension.

The final step is to assign truth values to propositions. Define $\mathcal{F}^0{}_{\mathcal{M}}$ to be a function from E_{P_0} to E_T such that: if φ is of the form $R(\beta_1, \ldots \beta_n)$, then $\mathcal{F}_{\mathcal{M}}(R(\beta_1, \ldots \beta_n)) = 1$ iff $R(\beta_1, \ldots \beta_n) \in \Omega$ and $\mathcal{F}_{\mathcal{M}}(R(\beta_1, \ldots \beta_n)) = 0$ iff $R(\beta_1, \ldots \beta_n) \in \Sigma$.. Given my definition of f^*, $\mathcal{F}^0$ is correctly defined.

Now we must extend $\mathcal{F}^0{}_{\mathcal{M}}$ to a function that verifies all of Ω. We do this by considering $\mathcal{F}^0{}_{\mathcal{M}} \cup \{\Omega_p \times \{1\}\} = \mathcal{F}^{*0}{}_{\mathcal{M}}$, where Ω_p is the part of Ω that contains propositional variables or constants. $\mathcal{F}^{*0}{}_{\mathcal{M}}$ is obviously a fuction, and we can extend $\mathcal{F}^{*0}{}_{\mathcal{M}}$ to a partial function $\mathcal{F}^\alpha$ from E_P to E_T using the inductive revision procedure defined in (PMC) for successor and limit ordinals. Let $\mathcal{M}^\alpha = <E, P, [\![\,]\!], f, \mathcal{F}^\alpha>$ be the fixed point of that revision process. $\mathcal{F}^\alpha$ is easily shown to be consistent, since if not then $\mathcal{F}^{*0}{}_{\mathcal{M}}$ must assign a formula and its negation both 1 or the same formula belongs both to Ω and to Σ, which is impossible by the construction. So $\mathcal{M}^\alpha$ is a fixed point model. Now we can show the following:

Lemma 12: $\mathcal{M}^\alpha$ is a partial model that verifies all of Ω and fails to verify Σ. Hence $\mathcal{M}^\alpha$ verifies Γ and fails to verify φ.

We prove this by induction on the complexity of $\vartheta \in \Omega \cup \Sigma$. Suppose that ϑ is atomic of the form $R(\beta_1, \ldots \beta_n)$ The construction of $\mathcal{F}^0$ insures that $\mathcal{M} \models \vartheta$ if $\vartheta \in \Omega$ and not $\mathcal{M} \models \vartheta$ if $\vartheta \in \Sigma$. Suppose $\vartheta = \neg\psi$ and that $\vartheta \in \Omega$. By the construction of Ω, $\psi \notin \Omega$ and ψ must be atomic. But then $\psi \in \Sigma$ and so again by the definition of $\mathcal{F}^0$, $\mathcal{M}$ ψ and so $\mathcal{M} \models \vartheta$. An entirely parallel argument holds if $\vartheta \in \Sigma$. The truth functional cases and ordinary quantificational cases are straightforward. Suppose $\vartheta = \lambda x\alpha(\beta) \in \Omega$. By (5.b) in R_1, $\alpha' = \alpha(\beta/x) \in \Omega$, and by the inductive hypothesis $\mathcal{M} \models \alpha'$ if $\alpha' \in \Omega$. $\mathcal{M} \models$

α' if $\alpha' \in \Omega$. Since $\mathcal{M}$ obeys (FA) and (ABS) as seen above, $\mathcal{M} \models \vartheta$. A similar argument holds for the case $\vartheta \in \Sigma$. The only non-straightforward step involves quantified statements of the form $\exists p\psi$ and $\forall p\ \psi$ where p is a propositional quantifier. Let $\vartheta = \exists p\psi$ and suppose $\vartheta \in \Omega$. By the construction of Ω, if $\exists p\psi \in \Omega$, then $\psi(cP_j/p) \in \Omega$. By the inductive hypothesis $\mathcal{M}^{\alpha} \models \psi(cP_j/p)$ and so $\mathcal{M}^{\alpha} \models \exists p\psi$, since $\mathcal{M}^{\alpha}$ is a fixed point. Now suppose that $\vartheta \in \Sigma$. $\vartheta \in \Sigma$ only if it implies φ or is itself φ. So by the construction procedure of Σ and Ω, every instance $\psi(cP_j/p)$ of ψ must be in Σ, since $\psi(cP_j/p) \vdash \exists p\psi$. But $\mathcal{M}^{\alpha} \models \exists p\psi$ iff for some proposition cP_1, $\mathcal{M}^{\alpha} \models \psi(cP_j/p)$, since $\mathcal{M}^{\alpha}$ is a fixed point. Then by the inductive hypothesis it is not the case that $\mathcal{M}^{\alpha} \models \psi(cP_j/p)$ for any instance $\psi(cP_j/p)$ of ψ, and so not $\mathcal{M}^{\alpha} \models \exists p\psi$. The arguments where $\vartheta = \forall p\psi$ are analogous to those for the existential case. Suppose $\vartheta \in \Omega$. We must show $\mathcal{M}^{\alpha} \models \forall p\ \psi$. By the construction procedure and the fact that $\forall p\psi \vdash \psi(cP_j/p)$, every instance $\psi(cP_j/p) \in \Omega$, and by the inductive hypothesis $\mathcal{M}^{\alpha} \models \psi(cP_j/p)$. So $\mathcal{M}^{\alpha} \models \forall p\ \psi$. Now assume that $\varphi \in \Sigma$. By the construction of Σ, an instance $\psi(cP_j/p) \in \Sigma$. By the inductive hypothesis then, not $\mathcal{M}^{\alpha} \models \psi(cP_j/p)$. But this suffices to show that t is not the case that $\mathcal{M}^{\alpha} \models \vartheta$.

With this lemma the end of the proof of proposition 9, the completeness proof, is at hand. This completeness proof establishes a logic for partial fixed point models of propositional quantification, a logic which I'll call *partial, simple theory of types* (PST).[26] But it does so by using general models. If we define first-order logic by means of the model theoretic properties of their *standard* models rather than by their syntax, the use of general models for PST essentially assigns a higher order logic syntax a first order logic semantics. But in this PST is no different from the standard simple theory of types (ST); with respect to standard partial intentional models PST is sound just as with respect to standard intentional models, (ST) is sound. By the completeness proof, we are also able to show that logical consequence for PST relative to the class of general partial models is Σ_1 definable, just as (ST) is Σ_1 definable relative to general models; with respect to standard models, consequence in ST, and, I conjecture, in PST, is only $\Sigma^1{}_1$ definable.[27] The major accomplishment of course is that in PST, the intentional paradoxes are rendered harmless.

Of course PST isn't conservative in one respect. Classical logic is not valid in these models. We can do better at the price of some elegance. Let us divide our language into those sentences in which propositional constants or variables occur and those sentences in which they do not. Call the former L_p and the latter $L - L_p$. R_1 be the following set of rules and axioms.

[26]Of course there is a weaker logic than PST the one corresponding to the theory of propositions P-(FA). Such a theory would replace the axiom of PST $\lambda x^{\tau}\ \varphi[\alpha] = \varphi(\alpha/x^{\tau})$ with $\lambda x^{\tau}\ \varphi[\alpha] \leftrightarrow \varphi(\alpha/x^{\tau})$. This system is highly intentional and non-standard. I'll call it PST - FA

[27]The proof of this claim would follow the lines of that given by Van Bentham and Doets (1984).

R_2

1. All instances of T_0 restricted to L - L_p

2. All instances of the rules of R_1 in L.

Now consider a restricted class of partial base models, those models $\mathcal{M}$ = <$\underline{\mathbf{E}}$, [], f, $\mathcal{F}^0$>in which the assignment functions f are total. The base models then are classical then and classical logic is valid in them when restricted to L-L_p. I'll call such base models *classical base models.* PST relative to the class of classical base models yields a class of minimal fixed point models Φ_0. Classical base models also may have model expansions. Call Φ_1 the class of all fixed point general models definable relative to classical base model expansions using the inductive revision procedure given in PST for successor and limit ordinals. $\Phi_0 \subset \Phi_1$. Just as before, we have

Proposition 13: $\Gamma \vdash_{R_2} \varphi \Rightarrow \Gamma \models_{\Phi_1} \varphi$

By constructing first an L-L_p maximal saturated set in the ordinary way and then doing the partial construction for L_p like the one given for proposition 9, one can then get the desired converse to proposition 13:

Proposition 14: $\Gamma \models_{\Phi_1} \varphi \Rightarrow \Gamma \vdash_{R_2} \varphi$

The logic for the class of models Φ_1 I'll call PST+.

This still doesn't yield classical logic for propositional quantification. This is not as bad as it seems, however; for all of mathematics might be done in L_ω at the level of individuals (we would have sets as individuals and axioms for them). For many purposes outside natural language semantics and modelling cognitive attitudes, it would seem as though we could dispense with the propositional part of this logic. If one really wanted to have classical logic for propositional logic too, however, I conjecture that one could resort to supervaluations as defining the interpretation of the connectives and quantifiers rather than the strong Kleene rules and use the same construction procedure as here.

2.4 Russell's Problem with the Theory of Types

Thomason's paper discusses another problem for the simple theory of types, mentioned in an appendix to Russell's *Principles of Mathematics.* It motivates Thomason's proposal for dealing with the intentional paradoxes, which uses a free logic for the propositional quantifiers. My proposal solves this difficulty too though in a manner different from what Thomason suggests.

The difficulty, due originally to Russell (1903), is that the simple theory of types is too liberal in what it countenances as propositions and propositional functions. For example in L_ω the term

(10) $\lambda x^p \exists f^{<p,p>} ({}^\vee Ff = x \ \& \ \neg {}^\vee fx)$

denotes a property of propositions,[28] for any given F. Let's call the property of propositions in (10) w. Then assuming $^{\vee}w(Fw) \vee \neg {}^{\vee}w(Fw)$, we get the following disturbing result.[29]

$$(11)\ \exists f^{<p,p>} \exists g^{<p,p>} ({}^{\vee}[Ff = Fg]\ \&\ \neg \forall x^{p} ({}^{\vee}fx \leftrightarrow {}^{\vee}gx))$$

Since (11) holds for arbitrary F (underlying it is a simple cardinality argument), it holds for the particular definition of F in (12)

$$(12)\ F = \lambda g^{<p,p>} \forall x^{p} ({}^{\vee}gx \rightarrow {}^{\vee}x)$$

By the principles of identity (11) and (12) have worrisome consequences for the theory of attitudes formulated within the simple theory of types. One such consequence is (13):

$$(13)\ \exists f^{<p,p>} \exists g^{<p,p>} (\Box \forall x^{o} \Box ({}^{\vee}Bel(x, Ff) \leftrightarrow {}^{\vee}Bel(x, Fg))\ \&\ \neg \forall x^{p} ({}^{\vee}fx \leftrightarrow {}^{\vee}gx))$$

Such consequences may have led Russell to develop the Ramified Theory of Types. Thomason points out correctly that by limiting what expressions denote higher order objects in the models and by employing a free logic, one can avoid this consequence. So one doesn't need the Ramified Theory to solve this difficulty.

Thomason's proposal won't work in the partial logic for propositional quantifiers as I have defined it. It is a valid principle of the partial logic PST that

$$(14)\ \exists p\ {}^{\vee}[p = \varphi]$$

This partial logic for the theory of types is no different from the classical theory of types in this respect. But Thomason's proposal leads as he points out to unintuitive consequences when dealing with the Intentional Paradoxes: it implies among other things that the existence of propositions is a context dependent, speaker relative matter. This collides with our intuitions about propositions. (14) also appears to be a needed principle in the analysis of propositional anaphora in natural language. Even though

---

[28]The superscripts in the formulas (10)-(13) are there to make clear the types of variables involved.

[29]The proof is as follows:
Dropping carrots we have

1) $w(Fw) \vee \neg w(Fw)$

Now suppose that

2) $w(Fw)$

and that

3) $\forall g \forall h (F(g) = F(h) \rightarrow \forall y(g(y) \leftrightarrow h(y)))$

Then by 2) and the definition of w,

4) $\exists f(F(f) = F(w)\ \&\ \neg f(F(w)))$

So for some f_0

4) $F(f_0) = F(w)\ \&\ \neg f_0(F(w))$

By (3) and (5),

6) $\forall y(f_0(y) \leftrightarrow w(y))$

By (2) and (6),

7) $f_0(F(w))$

which is a contradiction. So now suppose

8) $\neg w(Fw)$

By the definition of w again, and 8)

9) $\forall f(F(f) = F(w) \rightarrow f(F(w)))$

So by the laws of identity,

10) $w(F(w))$

Again this is a contradiction. Note that this proof is valid in T_0.

propositions are paradoxical or non-sensical, we may refer to them anaphorically. Imagine the following dialogue:

(15) Cretan: Everything I say is false_i

Socrates: I don't believe that_i

According to the free logic proposal, the Cretan did not manage to express a proposition in the circumstance in which the first sentence of (15) is the only sentence he manages to utter. But then it appears that Socrates doesn't manage to have a belief-- or express a belief-- about the Cretan. The analysis of anaphora in (15) is a semantic mystery, unless we assume there is some proposition the Cretan expresses. This seems to cast doubt on the free logic approach, at least if we are interested in applying our theory of propositional quantification to natural language semantics.

Russell's argument culminating with (11) is not valid in PST for the simple reason that it relies on the excluded middle. So that motivation for introducing type-free logic for higher order quantification dissolves. It's also not clear, however, that (13) is such a bizarre consequence for a theory of simple types to countenance. The real import of the difficulty hinges on what one takes to be the criterion of identity for types. Our models say little about what identity of types should amount to. If one thinks of how propositional functions might operate compositionally with propositions in a standard model, cardinality arguments would dictate that the function from a tuple consisting of a propositional function and its arguments to propositions could not be 1-1; (11) then is simply a special case of a much more general argument. But this need not be troublesome; one could have criteria of type identity such that $\psi(\beta) = \psi(\beta')$ but $\beta \not\approx \beta'$. This actually makes much more sense than (SIT) if one thinks that predicate terms denote propositional functions.[30] What this goes against, however, is a certain natural criterion of identity for intentional objects that one might call a *structural* criterion of identity for types (SIT):

(SIT) Let β and β' be of type τ and let ψ, ψ' be of type $\tau \rightarrow \tau'$. Then $\psi(\beta) = \psi(\beta')$ implies $\psi = \psi' \;\&\; \beta = \beta'$.

(SIT) together with the principle of indiscernibility of identicals contradicts (11). Thus (SIT) + the principle of indiscernibility of identicals is inconsistent with the simple theory of types (ST). There are at least trivial models of (PST), in which (SIT) + the principle of indiscernibility of identicals are never refuted and are verified in the trivial cases of where $\alpha(\beta) = \alpha(\beta)$ (which must be true according to the constraints on $\mathcal{F}$ in models for PST). This may be small consolation to the lover of (SIT), but it seems that one could fill out such models with more interesting examples of the application of (SIT).

[30]Aczel (1989) warns that the application relation should not be taken to be structure creating for such reasons. That is, he wants to deny that $\alpha(\beta) = \alpha'(\beta') \rightarrow \alpha = \alpha' \;\&\; \beta = \beta'$, our principle (SIT).

My solution to Russell's problem actually gives us an interesting comparison with Russell's own solution-- the Ramified Theory of Types (RT). We have in effect constructed models for certain versions of RT. But the orders of our theory are entirely semantic and in the models not in the syntax or the proof theory-- as they should be. In all versions of (RT), there is a function, Ord, from propositions to ω that recursively assigns orders. In some versions,[31] it is defined as follows. Let Dom(Q) in ψ be the set of objects satisfying ζ, where the structure of φ is $Q\delta\ (\zeta, \theta)$.

Ord: $E_P \rightarrow \omega$ such that:

If $\varphi \in P_0$, then $Ord(\varphi) = 1$.

If $\varphi = \alpha = \beta$ then $Ord(\varphi) = Max\{Ord(\alpha), Ord(\beta)\} + 1$.

If $\varphi = \neg\psi$, then $Ord(\varphi) \leq Ord(\psi)$.

If * is a boolean two place connective and $\varphi = \alpha*\beta$, then $Ord(\varphi) \leq Max\{Ord(\alpha), Ord(\beta)\}$.

If Q is a quantifier and $\varphi = Q\delta\psi$, then $Ord(\varphi) = Max\{Ord(\delta)$ for $\delta \in Dom(Q)$ in $\psi\} + 1$

Within PST Ord must be a partial function, because there are many propositions in our setup that cannot be assigned an order-- Prior's proposition for instance. This definition of order suggests a correlation between order and stages of revision in our model theoretic framework. All intentional identities are verified at every stage of our revision procedure, whereas in an RT model this is not the case. So let us define a translation function *, such that:

If φ is atomic of the form $R(a_1, \ldots, a_n)$, then $\varphi^* = \varphi$

If φ is atomic of the form $\alpha = \alpha'$, then $\varphi^* = \exists p \exists q\ (p = \alpha\ \&\ q = \alpha'\ \&\ \alpha = \alpha'\ \&\ {}^{\vee}p = {}^{\vee}q)$.

If φ is of the form $\zeta\ \&\ \psi$, $\zeta \vee \psi$, $\zeta \rightarrow \psi$, $\exists v\zeta$, etc. then $\varphi^* = \varphi$.

Recall the class of models Φ used in proposition 14 and the subclass of minimal fixed points Φ_0.

Proposition 15: Suppose φ is a proposition for which Ord is defined. Then for $\mathcal{M}^0$ a classical base model, $\mathcal{M}^n \models \varphi^*$ iff $Ord(\varphi) \leq n$, where φ^* is defined above. ${}^{\vee}\alpha = {}^{\vee}\beta$, if φ is $\alpha = \beta$ and $\varphi^* = \varphi$ otherwise.

The proof of proposition 14 is by induction on n. The upshot of this proposition is that if $\mathcal{M} \in \Phi_0$ also yields a model of RT. If $\mathcal{M} \in \Phi_0$, then *its RT reduct* is just like $\mathcal{M}$ except that the domain of propositions in the reduct is just those set of propositions φ such that φ^* gets a truth value in $\mathcal{M}$. The domain of propositions of the RT reduct also has an order imposed on it by ORD.

To sum up then, there appear to be two solutions to the paradoxes of indirect discourse. One familiar route uses a first order theory of quantification and a truth

[31] I folow Thomason (1989) and Church (1976) here.

predicate. The other uses higher order logic, in particular the intentional version presupposed by Russellians and spelled out in Thomason (1980.b). By giving an inductive definition of propositional quantification, we avoid the difficulties associated with other solutions to the paradoxes concerning truth in higher order logic. The partiality of PST and its cousins is located within what truth values propositions take on, not, as in Thomason's proposal, the existence of propositions.

Let us now return briefly to the picture with the first order theory. The set of valid sentences in all metastable models or semi-stable models, for instance, is clearly not r.e., regardless of ones approach to propositions. If we consider a conceptualist approach to intentional objects coupled with partial logic, we still do not get very far for reasons adumbrated in Visser and Burgess-- at least if we want to talk of truth of propositions in models defined over standard models of arithmetic.[32] The analogue in partial logic to the proposal by Turner for classical, first order theories, however, appears to have a straightforward axiomatization if we follow the techniques given here. The drawback is that one cannot have first order logic. The pro-sentential theory of truth incorporated into propositional quantification appears to mitigate Liar-like paradoxes, in that one can get a natural logic that includes first order logic. Somewhat surprisingly, the system with the higher order syntax-- partial type theory-- (or at least with the typing of variables) turns out to have a more tractable notion of validity than that of the classical, first order theory of propositions with a truth predicate for propositions.

3. Models for Belief

This last section concludes with an extension of the theory of propositional quantification to attitude contexts. One must be wary in concluding that higher order logic is "safe" from other paradoxes related to the Liar. Once we have a belief or knowledge predicate, we could, for a given propositional constant c, stipulate $[\![c]\!] = \neg B(a, c)$. Alternatively, it seems as though we could stipulate:

(16) $p = \sim B(a, p)$

By our constraints on $\mathcal{F}$ it follows that in every model in which (16) is true,

(17) ${}^{\vee}p \vdash \neg{}^{\vee}B(a, p)$

Now suppose our semantics for attitude predicates is such that for every agent we assign a *belief state*, a collection of propositions which is subject to certain closure conditions. Then we may encode by means of these closure conditions the usual doxastic reasoning principles and validate rules which correspond, say, to the logic presented in Thomason (1980).[33] We can still have such identities between

[32]See Burgess (1986).

[33]Here would be the relevant closure principles for the S4 logic of Thomason (1980):

$p \rightarrow q, p \in S \rightarrow q \in S$

propositions as in (16). But p will never get assigned a truth value, and so it will be undetermined whether a believes p. We must be careful not introduce any "essentially ungrounded" propositions with such predicates as knowledge and belief into our domain; if we do so completeness will vanish (we can no longer construct the models) and the higher order theory of propositions becomes an uninteresting variant of the first order theory.[34] But PST can admit "contingent self-reference" without harmful effects.

PST gives us more than just a logic with a moderate amount of self-reference. It also permits a variety of logics for the attitudes which go far beyond what a possible worlds framework yields. The reason for this is simple. If our semantics for attitudes ascribes to an agent a set of propositions, then we may choose from a variety of closure conditions. In particular we may assume very weak closure conditions-- such as those detailed in Asher (1986). Nothing forces us in PST to require a closure condition on S that exploits logical equivalence. In PST it is consistent to assume that two propositions may be necessarily even logically equivalent without being identical. So PST does not validate $B(p)$ and $\vdash p \leftrightarrow q \rightarrow B(q)$, a rule which is provable in most possible worlds semantics of attitudes. This rule leads to well-known, unintuitive results in the semantics of attitudes. Thus, PST offers a semantics for attitudes beyond that provided by possible worlds semantics in at least two ways. The PST semantics for attitudes allows at least as much self-reference as any possible worlds semantics but also does not succumb to problems of logical equivalence.

Let us get more concrete and define a variety of explicit logics of belief as well as a logic for implicit belief within PST.[35] I will take a belief model in PST to be a pair $<\mathcal{M}, \$>$, consisting PST model $\mathcal{M}$ and a collection of states \$. For each agent a in the domain of individuals of $\mathcal{M}$, one assigns a subset S_a of \$. One element of S_a, s_0, designates the beliefs of a, while the other elements designate beliefs of other agents. $\forall s \in S_a\ s \subseteq E_P$. We may now constrain the elements of S_a to various closure conditions. Postulates of "minimal rationality" might be the following, with quantification over all s in all $S_a \in \$$.

$\varphi \,\&\, \psi \in s \rightarrow \varphi, \psi \in s$

$\varphi(b) \in s \rightarrow \exists x\varphi(x) \in s$

One could add many other closure conditions. Of particular interest might be the fragment of relevance logic, advocated as an appropriate logic for belief by Levesque (1984), the Kleene logic encoded in R_1, and the rules of positive and negative

---

$p \in S \rightarrow Bp \in S$

$p \in S$ and $p \vdash q \rightarrow q \in S$.

$Bp \in S \rightarrow p \in S$.

One could introduce any sort of constraints on S and get a vareity of logics for belief and other attitudes in this way. One could do better than the ordinary possible worlds semantics here because the logic of propositions is not constrainted by possible worlds. ***

[34]Thus, we cannot introduce an expression relation between sentences and propositions. See Asher & Kamp (1986), Parsons (1974) for a discussion.

[35]One could choose the slightly stronger logic PST+ if one wanted to also.

introspection. A formulation of the latter is slightly problematic, so I give one possible way here. Suppose s represents the state of agent a. Then:

$\varphi \in s \Rightarrow B(a, \varphi) \in s$ (positive introspection)

$\varphi \notin s \Rightarrow \neg B(a, \varphi) \in s$ (negative introspection)

These logics, with the exception of the minimally rational closure conditions (I would argue),[36] all idealize the cognitive capacities of agents to some degree. So one might wonder what sort of logic of implicit belief PST models furnish. To get such models we may replace each collection of propositions in \$ with a partial world, a world with a Kleene valuation defined on it. At each world, w there will be designated one a-belief world w_a, a partial world depicting what a believes from the perspective of w. We can think of this as a selection function taking worlds and agents as arguments; I shall write the agent a world from the perspective of w as f(w, a). I evaluate formulas as before with the following exception for belief formulas. I will write $\models_w$ to designate that a formula is satisfied at a given world relative to a model structure.

$<\mathcal{M}, \$> \models_w B(a, \varphi)$ iff $<\mathcal{M}, \$> \models_{f(w,a)} \varphi$.

$<\mathcal{M}, \$> \pmb{\models}_w B(a, \varphi)$ iff $<\mathcal{M}, \$> \pmb{\models}_{f(w,a)} \varphi$.

These are natural definitions for a semantics for belief within a partial setting. One other natural assumption concerns what are agent worlds at a given world. Suppose one assumes that if f(w', a) = w, then f(w, a) = w. Of interest is the axiomatization that follows from them. In particular, the axiom corresponding to the closure principle of positive introspection, $B(a, \varphi) \vdash B(a, B(a, \varphi))$, as well as its converse $B(a, B(a, \varphi)) \vdash B(a, \varphi)$ is valid. The corresponding axioms for negative introspection is not valid. Also valid are the rules of R_1 and the following rule of belief closure:

$\varphi_1, \ldots, \varphi_n \vdash \psi$, then $B(a, \varphi), \ldots, B(a, \varphi_n) \vdash B(a, \psi)$

Let us call this system R_3 and the set of all general fix point models based on the semantics for belief systems Θ. I then close with a theorem, which falls out from the same technique used for proposition 9 generalized to the usual sort of construction for completeness proofs in modal logic.

Proposition 16: $\varphi \vdash_{R_3} \psi$ iff $\varphi \models_\Theta \psi$.

Bibliography

Aczel, P. (1989): 'Algebraic Semantics for Intensional Logics, I' in in *Properties, Types and Meaning, Volume I: Foundational Issues*, eds. Chierchia G., B. Partee and R. Turner, Dordrecht: Kluwer Academic Publishers, pp. 17-46.

Asher, N. (1986): 'Belief in Discourse Representation Theory', *Journal of Philosophical Logic* 15, pp. 127-189.

---

[36]See for instance Asher (1986).

Asher, N. (1988): "Reasoning about Knowledge and Belief with Self-reference and Time', *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, ed. M. Vardi.

Asher, N. & H. Kamp (1986): 'The Knower's Paradox and Representational Theories of Attitudes', in *Theoretical Aspects of Reasoning about Knowledge*, ed. J. Halpern. Los Angeles: Morgan Kaufmann, pp, 131-148.

Asher, N. & H. Kamp (1989): 'Self-Reference, Attitudes and Paradox', in *Properties, Types and Meaning, Volume I: Foundational Issues*, eds. Chierchia G., B. Partee and R. Turner, Dordrecht: Kluwer Academic Publishers, pp. 85-158.

Bealer, G. (1982): *Quality and Concept*, Oxford: Clarendon Press.

Gupta, A. (1982): 'Truth and Paradox,' *Journal of Philosophical Logic* 12, pp. 1-60.

Henkin, L. (1950): 'Completeness in the Theory of Types', *Journal of Symbolic Logic*, 15, pp. 81-91.

Herzberger, H. (1982): 'Notes on Naive Semantics,' *Journal of Philosophical Logic* 12, pp. 61-102.

Herzberger, H. (1982): 'Naive Semantics and the Liar Paradox,' *Journal of Philosophy* 79, pp. 479-497.

Kaplan, D. & R. Montague (1960): 'A Paradox Regained,' *Notre Dame Journal of Formal Logic* 1, pp. 79-90.

Kamp, H. (1983): 'A Scenic Tour Through The Land of Situations', manuscript.

Kripke, S. (1975): 'Outline of a New Theory of Truth,' *Journal of Philosophy* 72, pp. 690-715.

Langholm, T. (1988): *Partiality, Truth and Persistence*, CSLI Lecture Notes Series.

Levesque, H. (1984): 'A Logic for Explicit and Implicit Belief', *AAAI* Proceedings.

Montague, R. (1963): 'Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability,' *Acta Philosophica Fennica* 16, pp. 153-167.

Muskens, R. (1989): *Meaning and Partiality*, Ph.D. dissertation, University of Amsterdam.

Parsons, C. (1974): The Liar Paradox, *Journal of Philosophical Logic* 3, pp.381-412.

Prior, A. (1961): 'On a Family of Paradoxes', *Notre Dame Journal of Formal Logic*, 2, pp. 16-32.

Russell, B. (1903): *The Principles of Mathematics*, London.

Ramsey, F. (1927), 'Facts and Propositions', *Proceedings of the Aristotelian Society supplementary volume* 7, pp. 153-170.

Shapiro, S. (1985), 'Second-order Languages and Mathematical Practice', *Journal of Symbolic Logic*, 50 (1985), pp. 714-742.

Thomason, R. (1982): 'Paradoxes of Intentionality?', manuscript of a paper read at the 1989 Conference on Attitudes and Logic at The University of Minnesota, Minneapolis-Saint Paul, Minnesota.

Thomason, R. (1980.b): 'A Note on Syntactic Treatments of Modality', *Synthese* 44, pp.391-395.

Thomason, R. (1980.b): 'A Model Theory for Propositional Attitudes', *Linguistics and Philosophy* 4, pp.47-70.

Turner, R. (1989): 'Two Issues in the Foundations of Semantic Theory' in *Properties, Types and Meaning, Volume I: Foundational Issues*, eds. Chierchia G., B. Partee and R. Turner, Dordrecht: Kluwer Academic Publishers, pp. 63-84.

Turner, R. (1987): 'Intentional Semantics', manuscript.

Van Bentham, J. and Doets, K. (1984): 'Higher Order Logic' in *Handbook of Philosophical Logic* , vol. I, eds. Gabbay, D. & Guenthner G., Dordrecht: Reidel Publishing Company, 274-329.

EXTENDING THE CURRY-HOWARD-TAIT INTERPRETATION TO LINEAR, RELEVANT AND OTHER RESOURCE LOGICS[1]

by DOV M. GABBAY[2] and RUY J. G. B. DE QUEIROZ[3]

(Working Draft May 2, 1990.)

1. Motivation

The so-called Curry-Howard-Tait interpretation (Curry 1934, Curry & Feys 1958, Howard 1980, Tait 1965, 1967) is known to provide a rather neat term-functional account of intuitionistic implication. Could one refine the interpretation for other neighbouring logics to obtain an almost as good account of the so-called 'resource' implications (e.g. linear, relevant) ?

We answer this question positively by demonstrating that just by working with side conditions on the rule of *assertability conditions* for the connective representing implication ('$\rightarrow$') one can capture those 'resource' logics. In the Curry-Howard-Tait interpretation such a rule involves a λ-*abstraction*, which usually has a number of hidden assumptions (e.g., when abstracting 'x' from 'T' to make '$\lambda x.T$', 'T' could have one, many, or even no free occurrences of 'x'). Type-theoretic presentation systems are particularly useful in handling a not-quite-declarative feature of resource logics such as linear logic and relevant logic, namely the special requirement saying that in order to allow '$\mathbf{A} \rightarrow \mathbf{B}$' to be derived as a theorem the assumption '$\mathbf{A}$' must be used in order to obtain '$\mathbf{B}$'. This is because they are based on the identification of propositions with types and of proofs/constructions with elements, thus allowing the manipulation of proofs/constructions in the object language.

Based on such an extension of the propositions-are-types identification (Curry-Howard) combined with the convertibility-based intensional interpretation (Tait), we attempt at a classification of different systems of propositional implication (**W**, Ticket Entailment, linear, relevant, entailment, strict, minimal[4], intuitionistic, classical, linear classical, relevant classical, deductive relevant[5]) based on which axioms of the implicational calculus are allowed to be derived from the presentation of '$\rightarrow$' subject to side conditions on the rule of $\rightarrow$-*introduction*. As each axiom corresponds to the type-scheme[6] of a stratified pure term of combinatory logic (Curry & Feys 1958, Hindley & Seldin 1986), we can classify combinators through systems of implication and vice-versa.[7] E.g., I, B, B′ and C are linear, whilst S is not linear but is

[1]A preliminary version of this paper was presented at the *Informal Logic Colloquium*, held at the Seminar für natürlich-sprachliche Systeme (SNS), Universität Tübingen, on March 21–22, 1990, and the current version shall appear in one volume of the series SNS-Berichte edited by P. Schroeder-Heister. (We are grateful to Prof Schroeder-Heister for his kind invitation to participate in the event.) While this is still a working draft, a more developed version shall be presented at the *Logic Colloquium '90*, European Summer Meeting of the Association for Symbolic Logic, Helsinki, Finland, July 15–22, 1990.

[2]E-mail: (janet) dg@doc.ic.ac.uk.

[3]Supported by MEDLAR, ESPRIT Basic Research Action 3125, CEC. E-mail: (janet) rjq@doc.ic.ac.uk.

[4]In the sense of Johansson 1936.

[5]Developed in Gabbay 1989 where it is also called '**H**-relevant'.

[6]For an elegant presentation of the notion of type-schemes (and 'principal type-schemes'), including its relevance to the formulae-as-types interpretation, see chapter 14 of Hindley & Seldin 1986.

[7]We thank Dr Kosta Došen for pointing out that similar work on classifying subsystems of implication was done by Y. Komori (1983, 1989) and H. Ono (1988, and 1985 with Komori), although their framework was

relevant (indeed, the derivation of the axiom corresponding to the type-scheme of S involves a multiple and branching assumption discharge); K is not relevant but is (minimal) intuitionistic (the derivation of the axiom for K involves a non-relevant/vacuous assumption discharge); etc. By treating the proposition '**A** $\rightarrow$ **B**' as an $\rightarrow$*-type* of λ-terms, we shall demonstrate how to formalise resource logics with *Type Theory* by working on side conditions on the λ-calculus *abstraction* rule ($\rightarrow$*-introduction*). We shall give type-theoretic derivations of desired axioms identifying the combinators (from combinatory logic) to which they correspond, as well as conditions to invalidate derivations of undesired axioms for each particular system of implication.[8]

Additionally, we develop a type-theoretic counterpart to *Peirce's rule*, which allows us to add as an extra condition to the presentation of the type/proposition '**A** $\rightarrow$ **B**' giving us classical implication. The idea is to introduce an extra condition to the λ-abstraction rule which discharges an assumption in the form of '$y \in$ **A** $\rightarrow$ **B**' introducing a λ-abstraction term as a member of an atomic type '**A**', given that the latter is obtained from the former. This extra condition gives us a combinator we here call 'P''.[9]

The use of the term 'resource' has its origins in Gabbay's investigations of systems of implication through a technique which combines features of the object language and the meta-language: the Metabox technique. In order to illustrate the use of the Metabox technique, an algorithmic proof system methodology based on 'Labelled Deductive Systems (LDS)' de-

---

not the Curry-Howard-Tait interpretation with natural deduction, but Gentzen's sequent calculi. We thank Drs Komori and Ono for having sent us the still unpublished typescript. Dr Došen has also told us about the 'Lambek calculus' (Lambek 1958), and van Benthem's interpretations (e.g., van Benthem 1989), as well as about his own recent work Došen 1988, 1989.

[8] In his treatise on λ-calculus (Barendregt 1981), Barendregt refers to a dissertation by G. Helman as an application of restricted λ-abstraction to relevant logic:

> "The formulae-as-types idea gave rise to several investigations connecting typed λ-calculus, proof theory and some category theory, (...). Another direction is the connection between subsystems of logic and restricted versions of the typed λ-calculus (e.g. relevance logic and the typed λI-calculus), see Helman [1977]."
>
> (Barendregt 1981, p. 572.)

At the present moment we have not got hold of Helman's work (Helman 1977), but it looks as though there might be strong connections with part of what we are doing here.

[9] This is an attempt at further extending the Curry-Howard-Tait interpretation to a sound semantical instrument which can be capable of handling logics as rule-based calculi (as opposed to truth-value-based calculi). In other words, we claim that the interpretation should not be restricted to the intuitionistic case, therefore we want the so-called *Peirce's axiom* to be provable. But, of course, we want to do it on the conditions that an assumption can be discharged where it would not be possible to discharge it just by using the rules of the calculus for the intuitionistic case. We are obviously moving away from the strictly intuitionistic principles underlying the framework of, e.g., Howard 1980:

> "Results following from cut elimination in P($\supset$) (e.g.) the nonderivability of Peirce's Law ($\alpha \supset \beta . \supset \alpha) \supset \alpha$) seem to be obtainable at least as easily from the normalizability of constructions."
>
> (Howard 1980, p. 483.)

We are trying to follow the trend initiated by Curry which is to devise systems of implication including classical (NB.: classical implication, not classical logic, and we shall end up with something like an *implicational* logic (to use a term of van Benthem 1989); e.g., we do not want '**A** $\vee \neg$**A**' to be provable regardless of **A**, but we want '$\neg\neg$**A** $\rightarrow$ **A**' to be a theorem under the condition of negation being defined as '$\neg$**A** $\equiv$ **A** $\rightarrow \mathcal{F}$'.) For a consistency proof of our modified framework, it is sufficient to show that one cannot prove a proposition which does not have implication as its major connective. The notion of *provable* here is similar to Martin-Löf's:

> "A formula is *provable* if there is a deduction of it all of whose assumptions have been discharged."
>
> (Martin-Löf 1972, p. 96.)

Moreover, as we shall see below, such a consistency proof would find a parallel in Martin-Löf's *consistency theorem*. "*No atomic formula is provable.*" (Ibid., p. 102.)

scribed in Gabbay 1989, a class of logics called *resource logics* is defined as a generalisation of linear logic (Girard 1987). In a proof system with resource characteristic as $\rho(m,n)$, where $m < n$, it is required that each assumption be used at least m times but not more than n times. Linear logic would require a proof system with resource characteristic $\rho(1,1)$, and for relevant logic a proof system with resource characteristic $\rho(1,\infty)$ would be required. Here, instead of using the Metabox technique, which deals with proofs via a 'labelling' discipline where each assumption/step is given a new atomic label ('*label* : *formula*'), we deal with those *resource logics* via a type-theoretic presentation system based on the so-called 'Curry-Howard-Tait' interpretation where the form of judgement '*proof* $\in$ *proposition*' finds an immediate parallel with the one used in LDS.[10]

In Gabbay 1989 the different logical implications are presented in a Hilbert system as:

Linear
$\mathbf{A} \to \mathbf{A}$ (*reflexivity*)
$(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B}))$ (*left transitivity*)
$(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{B} \to \mathbf{C}) \to (\mathbf{A} \to \mathbf{C}))$ (*right transitivity*)
$(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{B} \to (\mathbf{A} \to \mathbf{C}))$ (*permutation*)

Modal T-strict
Add the schema below to linear implication:
$(\mathbf{A} \to (\mathbf{A} \to \mathbf{B})) \to (\mathbf{A} \to \mathbf{B})$ (*contraction*)

Relevant
Add the schema below to linear implication:
$(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to ((\mathbf{A} \to \mathbf{B}) \to (\mathbf{A} \to \mathbf{C}))$ (*distribution*)

(Minimal) Intuitionistic
Add the schema below to relevant implication:
$\mathbf{A} \to (\mathbf{B} \to \mathbf{A})$ (*truth*)

(Full) Intuitionistic
Add the schema below to minimal implication:
$\mathcal{F} \to \mathbf{A}$ (*absurdity*)

Classical
Add the schema below to intuitionistic implication:
$((\mathbf{A} \to \mathbf{B}) \to \mathbf{A}) \to \mathbf{A}$ (*Peirce's rule*)

[10] With respect to the rôle of labels in deductive systems, we have found an interesting remark by Lambek & Scott in their book on *An Introduction to Higher Order Categorical Logic*:

"Logicians should note that a deductive system is concerned not just with unlabelled entailments or sequents $A \to B$ (as in Gentzen's proof theory), but with deductions or proofs of such entailments. In writing $f : A \to B$ we think of f as the 'reason' why A entails B."

(Lambek & Scott 1986, p. 47.)

In the framework we discuss here, the 'reason' is represented by the witnessing of a closed λ-term (such as, e.g., '$\lambda x.x \in \mathbf{A} \to \mathbf{A}$'), whereas in Gabbay 1989, where the main data consist of axioms in a Hilbert-style presentation, it is an auxiliary tool which plays a crucial rôle in the description of the proof methodology.

Now, according to the Curry-Howard-Tait interpretation one can treat a proposition of the form '**A** $\rightarrow$ **B**' as a $\rightarrow$-*type* of λ-terms, and to say that the proposition is true is the same as to say that one can find a pure closed λ-term which is contained in it. That is the main principle underlying the so-called constructive notion of validity, which supports the 'propositions are types' identification, and whose seminal ideas stem from Curry's theory of functionality, Howard's formulae-as-types notion of construction, and Tait's notion of convertibility which establish the connections between cut-elimination and normalisation.[11]

Notational remark. In what follows, the sign '□' denotes the end of a definition, and '■' indicates the end of a proof.

2. Preamble

The idea of reading a formula as a type originates with Curry (1934) and is used to give a λ-calculus interpretation of an intuitionistic theorem. A formula of intuitionistic implicational logic is a theorem if and only if, when read as a type, it can be shown to be non-empty using the rules of term-construction, namely *abstraction* and *application*. By varying the natural abstraction principles available in the λ-calculus, we are able to extend the point of view of formulae-as-types to some weak systems of implication (relevance, linear, etc.) as well as to a system which is stronger than intuitionistic, namely classical implicational logic. The weaker logics are called resource logics in the framework of *Labelled Deductive Systems (LDS)* of Gabbay 1989. The research reported here can also be understood in the spirit of *LDS*, where the labels are not words of a certain grammar (as in Gabbay 1989) but λ-terms. As pointed out in Gabbay 1989, the framework of *LDS* generalises the usual consequence relation '$A_1, ..., A_n \vdash A$' between formulas to the more general notion '$t_1 : A_1, ..., t_n : A_n \vdash t : A$' where '$t_i$' are labels. The logical 'unit' in *LDS* is not a well-formed formula A but a labelled well-formed formula $t : A$, t being a label which conveys some 'meta-level' information about A. In modal logic t can be a possible world index and in the resource logics (which include intuitionistic, linear and relevance logics), the label t indicates what assumptions and rules we used to prove A. The Curry-Howard-Tait interpretation can be viewed as a labelling scheme for intuitionistic well-formed formulae and this paper generalises this scheme for other resource logics. We take Church's λ-calculus and Curry's combinatory logic as the building blocks supporting our framework.

Let us then take a standard definition of the terms and operators needed to obtain a λ-calculus, and let us examine the abstraction rule more closely. In his treatise *The Lambda Calculus* Barendregt defines:

2.1.1. DEFINITION. (i) *Lambda terms* are words over the following alphabet:

$v_0, v_1, ...$	variables,
λ	abstractor,
(,)	parentheses.

[11] "H. Curry (1958) has observed that there is a close correspondence between *axioms* of positive implicational propositional logic, on the one hand, and *basic combinators* on the other hand. (...) The following notion of construction, for positive implicational propositional logic, was motivated by Curry's observation. More precisely, Curry's observation provided *half* the motivation. The other half was provided by W. Tait's discovery of the close correspondence between cut elimination and reduction of λ-terms (W. W. Tait, 1965)."
(Howard 1980, p. 480.)

(ii) The set of λ-terms is defined inductively as follows:
(1) $x \in \Lambda$;
(2) $M \in \Lambda \Longrightarrow (\lambda x.M) \in \Lambda$;
(3) $M, N \in \Lambda \Longrightarrow (MN) \in \Lambda$;
where x in (1) or (2) is an arbitrary variable.

(Barendregt 1981, p. 22.)

Note that there are many hidden assumptions in the case (2) of the definition of λ-terms, e.g.:

(a) M may have *no* free occurrence of x:
  (a.1) M is an open term, but contains no free occurrence of x;
  (a.2) M is a closed term, thus contains no free variable at all;
(b) M may have *one* free occurrence of x:
  (b.1) M may be of the form '(Tx)' (or 'APPLY(T, x)');
  (b.2) M may be of the form '(xT)' (or 'APPLY(x, T)');
(c) M may have *more than one* free occurrence of x:
  (c.1) the *λ-abstraction* may cancel *exactly one* of the free occurrences of x;
  (c.2) the *λ-abstraction* may cancel *all* free occurrences of x;

Moreover, in (3), where *application* is being defined (which can be done by juxtaposition as in '(MN)', or by an explicit non-canonical operator 'APPLY(M, N)' in the terminology used here in this paper), 'M' is assumed to be of 'higher' level than 'N': 'M' is supposed to be the 'course-of-value' of a function, while 'N' is assumed to be the argument.

Now, by working with some of these hidden assumptions one can use the simple typed λ-calculus together with the Curry-Howard-Tait interpretation to formalise a number of systems of implication, as we shall demonstrate below.[12]

3. Types and propositions

As pointed out above, the identification of propositions with types of their proofs/constructions (the latter indicates the distinction from proof-trees), usually referred to as the 'formulae-as-types' notion of construction, goes back at least to Curry's results on the isomorphism between the principal type-schemes of combinators and the axioms of the positive implicational fragment of intuitionistic implication (Curry 1934). It has been given a more precise presentation in Howard's investigations on the isomorphism between natural deduction proofs

[12]The classification of a number of systems of implication has been made by various people in the context of the 'Lambek calculus' (Lambek 1958), giving rise to what is sometimes referred to as the 'Categorial Hierarchy', and has been used by many of those interested in the connections between the language of category theory, λ-calculus and proof theory, such as, e.g.:

"The general linguistic framework which arises here is that of a *Categorial Hierarchy* of different logical calculi ('categorial engines'). At the lower end lies the standard calculus of Ajdukiewicz [Ajdukiewicz 1937], at the upper end lies the full constructive, or intuitionistic conditional logic, whose derivations correspond to arbitrary lambda/application terms. In between lies a whole spectrum, not necessarily linearly ordered, of calculi with stronger or weaker intermediate rules of inference. For instance, one important principle of classification concerns the number of occurrences of premises which may be withdrawn in one application of conditionalization. Only *one* occurrence at a time was withdrawn in Examples 1 and 3. This particular restriction gives a very natural intermediate logic, which was already studied by Lambek as early as 1958, and is often called after him."

(van Benthem 1990, p. 10.)

and terms of the λ-calculus (Howard 1980) within the framework of Heyting arithmetic. Such an identification, which also finds support in Gödel functional interpretation of intuitionistic logic via a system **T** of finite types (1958), has played an important rôle in most developments of some key notions of modern logic such as 'constructive validity' (see, e.g., Läuchli 1965,[13] 1970, Scott 1970) and 'theory of constructions' (see, e.g., Goodman 1970), as well as in some attempts at reconciling category theory with constructive logics (such as, e.g., Lambek & Scott 1986).

In order to make logical sense of the identification between propositions and types, we can recall that it was in Frege's *Begriffschrift* that the symbol '⊢' (then meaning 'is true') first appeared. It was motivated by the need for characterising a proposition (Frege's horizontal bar, i.e. '—**A**': '**A** is a proposition') and distinguishing it from a judgement (Frege's vertical bar, i.e. '⊢ **A**': '**A** is a true proposition'). In other words:

$$\boxed{\boxed{\mathbf{A}}\text{ is true}}$$

'**A**' is a *proposition*, whereas '**A** is true' is a *judgement*. A *judgement* of the form '**A** is true' can only be made on the basis of the existence of a proof of the proposition **A**. Contrary to the classical view, a proposition is not the same as a truth value. And in contrast to the traditional proof-theoretic account of propositions and inference rules, a logical inference is to be made from *judgement(s)* to *judgement*, and not from *proposition(s)* to *proposition*. Both premisses and conclusions of inference rules are not *propositions* as in the usual case even in traditional natural deduction presentations of logics, but *judgements*. This seems to be a highly relevant refinement of the usual formalisation of mathematical procedures into rules of inference, such as e.g., natural deduction style *à la* Gentzen. The difference between usual natural deduction presentation rules which have *propositions* as premises and conclusions, and *Intuitionistic Type Theory* where *judgements* are the objects on which the rules of inference operate, is explained briefly in Martin-Löf's illuminating account of the often neglected distinction between the two logical concepts of *proposition* and *judgement*, namely the written account of a series of lectures entitled 'On the Meanings of the Logical Constants and the Justifications of the Logical Laws' given in Siena, Italy, in April 1983.[14]

Now, the identification of propositions with types gives us instruments to deal with *judgements* which include its justification: in '$a \in \mathbf{A}$' we are basically saying that '**A** is true because

[13] "*Theorem: A is a tautology of intuitionistic propositional calculus if and only if t(A) is definably non-empty.*"

[14] There he says:

"We must remember that, even if a logical inference, for instance, a conjunction introduction, is written

$$\frac{A \quad B}{A \,\&\, B}$$

which is the way in which we could normally write it, it does not take us from the propositions A and B to the proposition A & B. Rather, it takes us from the affirmation of A and the affirmation of B to the affirmation of A & B, which we may make explicit, using Frege's notation, by writing it

$$\frac{\vdash A \qquad \vdash B}{\vdash A \,\&\, B}$$

instead. It is always made explicit in this way by Frege in his writings, and in Principia, for instance. Thus we have two kinds of entities here: we have the entities that the logical operations operate on, which we call propositions, and we have those that we prove and that appear as premises and conclusion of a logical inference, which we call assertions."

(Martin-Löf 1985, pp. 204–5.)

of a'. (E.g. in '$\lambda x.x \in \mathbf{A} \rightarrow \mathbf{A}$' we say that '$\mathbf{A} \rightarrow \mathbf{A}$' is true because we have a closed term '$\lambda x.x$' which inhabits it)

We shall be using here what has been named *'Meaning-As-USE' Type Theory* (de Queiroz & Maibaum 1990), a recent reformulation of *Intuitionistic Type Theory* (Martin-Löf 1975, 1984), where instead of Martin-Löf's rules for the definition of types/propositions:

formation
introduction
elimination
equality

we have the following rules, with corresponding purpose:

formation: to show how to form the type-expression as well as when two type-expressions are equal.
introduction: to show how to form the canonical value-expressions via the `constructor(s)`, as well as when two canonical value-expressions are equal.
reduction: to show how to normalise non-canonical value-expressions, by demonstrating the effect of DESTRUCTOR(S) on the terms built up by `constructor(s)`.
induction: minimality rule.

The $\rightarrow$-type of λ-terms is presented as:[15]

$\rightarrow$-formation

$$\frac{\mathbf{A}\ type \qquad \mathbf{B}\ type}{\mathbf{A} \rightarrow \mathbf{B}\ type} \qquad \frac{\mathbf{A} = \mathbf{C} \qquad \mathbf{B} = \mathbf{D}}{\mathbf{A} \rightarrow \mathbf{B} = \mathbf{C} \rightarrow \mathbf{D}}$$

$\rightarrow$-introduction

$$\frac{\begin{matrix}[x \in \mathbf{A}] \\ b(x) \in \mathbf{B}\end{matrix}}{\lambda x.b(x) \in \mathbf{A} \rightarrow \mathbf{B}} \qquad \frac{\begin{matrix}[x \in \mathbf{A}] \\ b(x) = d(x) \in \mathbf{B}\end{matrix}}{\lambda x.b(x) = \lambda x.d(x) \in \mathbf{A} \rightarrow \mathbf{B}}$$

$\rightarrow$-reduction

$$\frac{a \in \mathbf{A} \qquad \begin{matrix}[x \in \mathbf{A}] \\ b(x) \in \mathbf{B}\end{matrix}}{\mathtt{APPLY}(\lambda x.b(x), a) = b(a/x) \in \mathbf{B}}$$

$\rightarrow$-induction

$$\frac{c \in \mathbf{A} \rightarrow \mathbf{B}}{\lambda x.\mathtt{APPLY}(c, x) = c \in \mathbf{A} \rightarrow \mathbf{B}}$$

□

We shall be concerned here mainly with the first *$\rightarrow$-introduction* and *$\rightarrow$-reduction*. In the actual proof-trees, we shall be making use of *$\rightarrow$-elimination*:

[15] An attempt at finding useful connections between the presentation of an $\rightarrow$-type as a theory of λ-terms, and the axiomatic presentation of a λ-theory (such as in Barendregt 1981) is given in de Queiroz & Maibaum 1991.

$$\frac{a \in \mathbf{A} \quad c \in \mathbf{A} \to \mathbf{B}}{\text{APPLY}(c, a) \in \mathbf{B}}$$

which follows from the presentation of the →-type above (as shown in de Queiroz & Maibaum 1990).

Observe that unlike Barendregt's set of *Lambda terms*, which included variables and *application*-terms, our →-type only contains λ-abstraction terms (the rule of →-*induction*, which is the counterpart to λ-calculus η-rule, is a kind of formal counterpart to that). Here those Barendregt's terms which are not λ-*abstractions* can only be subterms. Thus, in order to work with the hidden assumptions of the *abstraction* rule discussed in the previous section, we need to look at the '$b(x)$' of our first →-*introduction*, which can have the form of any of Barendregt's *Lambda terms*: a variable, an *abstraction*-term, or an *application*-term.

Now, we have to show that one can construct derivations of the Hilbert style axioms given for linear, relevant and intuitionistic implication in Gabbay 1989, and how one can draw the appropriate distinctions for each implication. At least since Curry's theory of functionality (1934) it is well known that there is a correspondence between the type-schemes of combinators and the axioms of intuitionistic implication. Within the propositions-are-types paradigm there is a correspondence between axioms of implication and →-types which contain λ-terms as elements or proofs/constructions of the corresponding axioms. So, combinators are mathematical objects which correspond to λ-terms, which in their turn are elements/proofs/constructions which belong to an →-type.[16]

Moreover, just to make clear our own proof methodology, we should say that we read the first rule of →-*introduction*, namely:

$$\frac{\begin{array}{c}[x \in \mathbf{A}]\\ b(x) \in \mathbf{B}\end{array}}{\lambda x.b(x) \in \mathbf{A} \to \mathbf{B}}$$

as follows: having made the assumption '$x \in \mathbf{A}$', and arriving at the conclusion '$b(x) \in \mathbf{B}$' by means of one (or none) of the rules available, then we can discharge the assumption by making a λ-abstraction of the assumption-term ('x') over the conclusion-term ('$b(x)$'). In other words, when constructing a proof-tree one can discharge an assumption if there is at least one proof step between the assumption and the conclusion where the assumption is discharged. So, in the construction of a proof of '$\lambda x.x \in \mathbf{A} \to \mathbf{A}$', as we shall see below, we need at least *reflexivity* in order to arrive at a conclusion of the form '$x \in \mathbf{A}$' from the assumption '$[x \in \mathbf{A}]$'.

[16] When approaching a presentation of the Curry-Howard-Tait interpretation one has to be warned to specific terminological diversions from conventional logical frameworks. In the framework of the interpretation 'proofs' refer to constructions, and not to the actual proof-trees. Without such a terminological warning misconceptions may arise. E.g., in Lambek's:

"The association of entities with proofs becomes even more striking when we compare the free typed Schönfinkel algebra (generated by a set of letters) with pure intuitionistic implicational logic. Then
combinators = proofs."

(Lambek 1980, p. 385.)

'proofs' should be understood as constructions (terms).

Such a methodological requirement makes our framework slightly different from the one used in chapter 15 of Hindley & Seldin 1986, where a proof of the latter is constructed as follows:

"EXAMPLE 15.4. In any system containing (→e) and (→i),

$$\vdash \mathsf{I} \,\varepsilon\, \alpha \to \alpha.$$

Proof.

$$\frac{\overset{1}{[x \,\varepsilon\, \alpha]}}{\lambda x.x \,\varepsilon\, \alpha \to \alpha}(\to \text{i-1}).$$

□"

(Hindley & Seldin 1986, p. 208.)

(Here '(→ i-1)' indicates that at that particular step the assumption numbered '1' was being discharged by the introduction of the '→'.)[17]

There is another relevant remark to be pointed out with respect to our proof methodology and notation, and that is the following: our abstractor 'λ' is used to abstract an element-variable from a term, and not a type-variable. In this respect we agree with Hindley & Seldin, who assign a type to λ-terms as '$(\lambda x^{\alpha}.M^{\beta})^{\alpha\to\beta}$' (see p. 205), but we diverge from Howard's (1980) type-abstraction as in '$(\lambda X^{\alpha}.F^{\beta})^{\alpha\supset\beta}$'.[18]

[17] It must be noted, however, that Hindley & Seldin seem to adopt the same reading, in spite of the divergence in the example just mentioned. Cf.:

"It [the rule of →-introduction] is usually written thus:

(→i)

$$\frac{\begin{matrix}[x \,\varepsilon\, \alpha]\\ M \,\varepsilon\, \beta\end{matrix}}{\lambda x.M \,\varepsilon\, \alpha \to \beta}.$$

(...)

In such a system, rule (→i) is read as "If $x \notin FV(L_1...L_n)$, and $M \,\varepsilon\, \beta$ is the conclusion of a deduction whose not-yet-discharged assumptions are $x \,\varepsilon\, \alpha, L_1 \,\varepsilon\, \delta_1, ..., L_n \,\varepsilon\, \delta_n$, then you may deduce

$$(\lambda x.M) \,\varepsilon\, (\alpha \to \beta),$$

and whenever the assumption $x \,\varepsilon\, \alpha$ occurs undischarged at a branch-top above $M \,\varepsilon\, \beta$, you must enclose it in brackets to show that it has now been discharged." "

(Hindley & Seldin 1986, p. 206.)

(Our emphasis. Note that in the 'I' example, as well as in the 'K' example below, there is no conclusion of the form '$M \,\varepsilon\, \beta$'. The introduction of the '→', and corresponding assumption-discharge, is made straight from the assumption.)

[18] "*2. Type symbols, terms and constructions*

Byt a type symbol is meant a formula of P(⊃). We will consider a λ-formalism in which each term has a type symbol α as a superscript (which we may not always write); the term is said to be of type α. The rules of term formation are as follows.

(2.1) Variables X^{α}, Y^{β}, ... are terms

(2.2) λ-abstraction: from F^{β} get $(\lambda X^{\alpha}.F^{\beta})^{\alpha\supset\beta}$.

(2.3) Application: from $G^{\alpha\supset\beta}$ and H^{α} get $(G^{\alpha\supset\beta}H^{\alpha})^{\beta}$."

(Howard 1980, pp. 480–1.)

Note that Howard's *constructions* belonging to the formula/type of the form '$\alpha \supset \beta$' are built with a λ-abstraction which operates on type-variables rather than element-variables. That is not the case for the framework we present here.

Proofs and corresponding conditions for invalidation:

1. $\mathbf{A} \to \mathbf{A}$ (*reflexivity*)

$$\cfrac{\cfrac{\cfrac{[x \in \mathbf{A}]}{x = x \in \mathbf{A}}}{x \in \mathbf{A}}}{\lambda x.x \in \mathbf{A} \to \mathbf{A}}$$

and we have the 'identity' construction, which corresponds to combinator 'I' $\equiv \lambda x.x$ (Curry & Feys 1958, p. 152; Hindley & Seldin 1986, p. 191).[19] ■

2. $(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B}))$ (*left transitivity*)

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{[z \in \mathbf{C}] \quad [y \in \mathbf{C} \to \mathbf{A}]}{\mathtt{APPLY}(y,z) \in \mathbf{A}} \quad [x \in \mathbf{A} \to \mathbf{B}]}{\mathtt{APPLY}(x, \mathtt{APPLY}(y,z)) \in \mathbf{B}}}{\lambda z.\mathtt{APPLY}(x, \mathtt{APPLY}(y,z)) \in \mathbf{C} \to \mathbf{B}}}{\lambda y.\lambda z.\mathtt{APPLY}(x, \mathtt{APPLY}(y,z)) \in (\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B})}}{\lambda x.\lambda y.\lambda z.\mathtt{APPLY}(x, \mathtt{APPLY}(y,z)) \in (\mathbf{A} \to \mathbf{B}) \to ((\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B}))}$$

which corresponds to combinator 'B' $\equiv \lambda x.\lambda y.\lambda z.\mathtt{APPLY}(x, \mathtt{APPLY}(y,z))$ (Curry & Feys 1958, p. 152; Hindley & Seldin 1986, p. 191). In terms of a calculus of functions, 'B' would correspond to the functor for the *(left) composition* of two functions.[20] ■

3. $(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{B} \to \mathbf{C}) \to (\mathbf{A} \to \mathbf{C}))$ (*right transitivity*)

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{[z \in \mathbf{A}] \quad [x \in \mathbf{A} \to \mathbf{B}]}{\mathtt{APPLY}(x,z) \in \mathbf{B}} \quad [y \in \mathbf{B} \to \mathbf{C}]}{\mathtt{APPLY}(y, \mathtt{APPLY}(x,z)) \in \mathbf{C}}}{\lambda z.\mathtt{APPLY}(y, \mathtt{APPLY}(x,z)) \in \mathbf{A} \to \mathbf{C}}}{\lambda y.\lambda z.\mathtt{APPLY}(y, \mathtt{APPLY}(x,z)) \in (\mathbf{B} \to \mathbf{C}) \to (\mathbf{A} \to \mathbf{C})}}{\lambda x.\lambda y.\lambda z.\mathtt{APPLY}(y, \mathtt{APPLY}(x,z)) \in (\mathbf{A} \to \mathbf{B}) \to ((\mathbf{B} \to \mathbf{C}) \to (\mathbf{A} \to \mathbf{C}))}$$

---

[19] Here we have used one of the general rules of equality available in our type-theoretic framework, namely the *reflexivity* rule:

$$\frac{x \in \mathbf{A}}{x = x \in \mathbf{A}}$$

followed by either one of the *equality left* or *right*:

$$\frac{a = b \in \mathbf{A}}{a \in \mathbf{A}} \qquad \frac{a = b \in \mathbf{A}}{b \in \mathbf{A}}$$

[20] It also guarantees, together with the previous combinator 'I', that there is a left identity function such that, for all $f : A \to B$, '$f1_A = f$', as in the definition of a *category* as a deductive system in Lambek & Scott 1986, p. 52.

which corresponds to a combinator which results from applying combinator 'C' to combinator 'B', or what Curry has called combinator 'B′' in Curry & Feys 1958, p. 379, and in Curry 1963, p. 118: 'B′' $\equiv \lambda x.\lambda y.\lambda z.\mathsf{APPLY}(y, \mathsf{APPLY}(x, z))$ ('CB' in Hindley & Seldin 1986, p. 191). In terms of a calculus of functions, 'B′' would correspond to the *(right) composition* of two functions.[21]

In order to invalidate the derivation above one would have to impose the restriction on the λ-*abstraction* rule such that the abstractions have to occur in the order 'from higher to lower subterms'. ■

4. $(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{B} \to (\mathbf{A} \to \mathbf{C}))$ (*permutation*)

$$\dfrac{\dfrac{\dfrac{\dfrac{[y \in \mathbf{B}] \qquad \dfrac{[z \in \mathbf{A}] \qquad [x \in \mathbf{A} \to (\mathbf{B} \to \mathbf{C})]}{\mathsf{APPLY}(x, z) \in \mathbf{B} \to \mathbf{C}}}{\mathsf{APPLY}(\mathsf{APPLY}(x, z), y) \in \mathbf{C}}}{\lambda z.\mathsf{APPLY}(\mathsf{APPLY}(x, z), y) \in \mathbf{A} \to \mathbf{C}}}{\lambda y.\lambda z.\mathsf{APPLY}(\mathsf{APPLY}(x, z), y) \in \mathbf{B} \to (\mathbf{A} \to \mathbf{C})}}{\lambda x.\lambda y.\lambda z.\mathsf{APPLY}(\mathsf{APPLY}(x, z), y) \in (\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{B} \to (\mathbf{A} \to \mathbf{C}))}$$

which corresponds to combinator 'C' $\equiv \lambda x.\lambda y.\lambda z.\mathsf{APPLY}(\mathsf{APPLY}(x, z), y)$ (Curry & Feys 1958, p. 152; Hindley & Seldin 1986, p. 191). It is the counterpart to the rule of exchange of Gentzen's sequent calculi. In a calculus of functions it would correspond to the associativity of *composition*.

In order to invalidate the derivation above one would have to impose the restriction on the λ-*abstraction* rule such that the abstractions have to occur in the order 'from higher to lower subterms' within the order 'from inner to outer subterms'. ■

5. $(\mathbf{A} \to (\mathbf{A} \to \mathbf{B})) \to (\mathbf{A} \to \mathbf{B})$ (*contraction*)

$$\dfrac{\dfrac{\dfrac{\boxed{[y \in \mathbf{A}]} \qquad \dfrac{\boxed{[y \in \mathbf{A}]} \qquad [x \in \mathbf{A} \to (\mathbf{A} \to \mathbf{B})]}{\mathsf{APPLY}(x, y) \in \mathbf{A} \to \mathbf{B}}}{\mathsf{APPLY}(\mathsf{APPLY}(x, y), y) \in \mathbf{B}}}{\boxed{\lambda y.}\mathsf{APPLY}(\mathsf{APPLY}(x, y), y) \in \mathbf{A} \to \mathbf{B}}}{\lambda x.\lambda y.\mathsf{APPLY}(\mathsf{APPLY}(x, y), y) \in (\mathbf{A} \to (\mathbf{A} \to \mathbf{B})) \to (\mathbf{A} \to \mathbf{B})}$$

which corresponds to combinator 'W' $\equiv \lambda x.\lambda y.\mathsf{APPLY}(\mathsf{APPLY}(x, y), y)$ (Curry & Feys 1958, p. 152; Hindley & Seldin 1986, p. 191). It is also the counterpart to the rule of contraction of Gentzen's sequent calculi.

The assumption '$\boxed{[y \in \mathbf{A}]}$' is used twice and in a nested way. So, the restriction one has to impose here is rather obvious: a λ-abstraction will cancel *one* free occurrence of the variable at a time. ■

6. $(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to ((\mathbf{A} \to \mathbf{B}) \to (\mathbf{A} \to \mathbf{C}))$ (*distribution*)

[21] Similarly to the previous case, this also guarantees the identity to the right '$1_B f = f$', Ibid.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\boxed{[z \in \mathbf{A}]} \quad [x \in \mathbf{A} \to (\mathbf{B} \to \mathbf{C})]}{\texttt{APPLY}(x,z) \in \mathbf{B} \to \mathbf{C}} \qquad \dfrac{\boxed{[z \in \mathbf{A}]} \quad [y \in \mathbf{A} \to \mathbf{B}]}{\texttt{APPLY}(y,z) \in \mathbf{B}}}{\texttt{APPLY}(\texttt{APPLY}(x,z),\texttt{APPLY}(y,z)) \in \mathbf{C}}}{\boxed{\lambda z.}\texttt{APPLY}(\texttt{APPLY}(x,z),\texttt{APPLY}(y,z)) \in \mathbf{A} \to \mathbf{C}}}{\lambda y.\lambda z.\texttt{APPLY}(\texttt{APPLY}(x,z),\texttt{APPLY}(y,z)) \in (\mathbf{A} \to \mathbf{B}) \to (\mathbf{A} \to \mathbf{C})}}{\lambda x.\lambda y.\lambda z.\texttt{APPLY}(\texttt{APPLY}(x,z),\texttt{APPLY}(y,z)) \in (\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to ((\mathbf{A} \to \mathbf{B}) \to (\mathbf{A} \to \mathbf{C}))}$$

which corresponds to combinator 'S' $\equiv \lambda x.\lambda y.\lambda z.\texttt{APPLY}(\texttt{APPLY}(x,z),\texttt{APPLY}(y,z))$ (Curry & Feys 1958, p. 153; Hindley & Seldin 1986, p. 191). ■

Note that the assumption '$\boxed{z \in \mathbf{A}}$' is used twice, and both occurrences are discharged in one single abstraction '$\boxed{\lambda z.}$'. To obtain a linear implication one has to restrict the discharging abstraction to one occurrence of the assumption only. In other words, each discharge affects only one (linear) path in the proof-tree, instead of affecting all branching occurrences like in the proof above.

6a. $(\mathbf{A} \to \mathbf{B}) \to (((\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{A} \to \mathbf{C}))$ (*variant of distribution*)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\boxed{[z \in \mathbf{A}]} \quad [x \in \mathbf{A} \to \mathbf{B}]}{\texttt{APPLY}(x,z) \in \mathbf{B}} \qquad \dfrac{\boxed{[z \in \mathbf{A}]} \quad [y \in \mathbf{A} \to (\mathbf{B} \to \mathbf{C})]}{\texttt{APPLY}(y,z) \in \mathbf{B} \to \mathbf{C}}}{\texttt{APPLY}(\texttt{APPLY}(y,z),\texttt{APPLY}(x,z)) \in \mathbf{C}}}{\boxed{\lambda z.}\texttt{APPLY}(\texttt{APPLY}(y,z),\texttt{APPLY}(x,z)) \in \mathbf{A} \to \mathbf{C}}}{\lambda y.\lambda z.\texttt{APPLY}(\texttt{APPLY}(y,z),\texttt{APPLY}(x,z)) \in (\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{A} \to \mathbf{C})}}{\lambda x.\lambda y.\lambda z.\texttt{APPLY}(\texttt{APPLY}(y,z),\texttt{APPLY}(x,z)) \in (\mathbf{A} \to \mathbf{B}) \to (((\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{A} \to \mathbf{C}))}$$

which corresponds to a variation of the 'S', precisely:
'SC' $\equiv \lambda x.\lambda y.\lambda z.\texttt{APPLY}(\texttt{APPLY}(y,z),\texttt{APPLY}(x,z))$. ■

Observe that the same remarks as to the 'non-linearity' of the discharge/abstraction made for the previous case also applies for the case here.

7. $\mathbf{A} \to (\mathbf{B} \to \mathbf{A})$ (*truth*[22])

[22] This axiom essentially represents that 'a true proposition is implied by anything', so we have accordingly called it *truth*. As we shall see, a derivation of this axiom from the presentation of the $\to$-type involves a non-relevant abstraction under the condition (a.1) of section 2 above, which is when abstraction is made over an open term which contains no free occurrence of the variable. Additionally, by allowing a non-relevant abstraction over closed terms (condition (a.2) above), such as, e.g., in '$\lambda x.\lambda y.y \in \mathbf{B} \to (\mathbf{A} \to \mathbf{A})$', one can see how to relate this axiom to the following axiom of a deductive system defined in Lambek & Scott 1986, p. 48, 'R2. $A \xrightarrow{O_A} T$', which comes from the categorical notion of *terminal* object and its existence linked to the existence of a unique arrow $O_A : A \to T$ for all objects A. If one has *permutation* (recall the type-scheme of combinator 'C' above), it is easy to see that:
$(\mathbf{B} \to (\mathbf{A} \to \mathbf{A})) \to (\mathbf{A} \to (\mathbf{B} \to \mathbf{A}))$
and vice-versa.

Now we want to build a derivation of the above axiom and show where it can be invalidated by the appropriate side condition. By imposing the condition that the *abstraction* can only be made when there is indeed at least one free occurrence of the variable being abstracted from the expression ('$b(x)$' in the case below), one can obtain a 'relevant' *abstraction*:

→-introduction

$$\frac{\begin{array}{c}[x \in \mathbf{A}] \\ b(x) \in \mathbf{B}\end{array}}{\lambda x.b(x) \in \mathbf{A} \to \mathbf{B}}$$

The proof-tree is:

$$\cfrac{\cfrac{\begin{array}{c}\boxed{[y \in \mathbf{B}]} \\ [x \in \mathbf{A}]\end{array}}{\boxed{\lambda y.x} \in \mathbf{B} \to \mathbf{A}}}{\lambda x.\lambda y.x \in \mathbf{A} \to (\mathbf{B} \to \mathbf{A})}$$

which corresponds to combinator 'K' ≡ $\lambda x.\lambda y.x$ (Curry & Feys 1958, p. 153; Hindley & Seldin 1986, p. 191). It is also the counterpart to the structural rule of thining of Gentzen's sequent calculi. ■

Note that the discharge/abstraction of the assumption '$\boxed{[y \in \mathbf{B}]}$' is made over the expression 'x' in '$\boxed{\lambda y.x}$', which prevents it from being considered 'relevant', given that the expression 'x' does not contain any free occurrence of 'y'. (Such a 'non-relevant' discharge/abstraction is called 'vacuous discharge' in Hindley & Seldin 1986.[23]) So, the restricted λ-abstraction to be adopted in order to invalidate the derivation above is exactly the relevant abstraction, i.e., there must be at least one free occurrence of the variable in the term on which the abstraction is operating.

---

[23] "EXAMPLE 15.3. In any system containing (→e) and (→i).

$$\vdash \mathsf{K} \; \varepsilon \; \alpha \to \beta \to \alpha.$$

Proof. Here is a deduction of the required formula. In it, the first application of (→-i) discharges all assumptions y ε β that occur. But none in fact occur, so nothing is discharged. This is perfectly legitimate; it is called 'vacuous discharge', and is shown by '(→i-v)'.

$$\cfrac{\cfrac{\overset{1}{[\mathrm{x} \; \varepsilon \; \alpha]}}{\lambda \mathrm{y.x} \; \varepsilon \; \beta \to \alpha}(\to\text{i-v})}{\lambda \mathrm{xy.x} \; \varepsilon \; \alpha \to \beta \to \alpha}(\to\text{i-1}).$$

□"

(Hindley & Seldin 1986, p. 208.)

As pointed out in Lambek 1989 (p. 234), in his *The Calculi of Lambda-Conversion* Church already distinguished the relevant from the non-relevant λ-*abstraction*.[24] Most current textbooks, however, still omit such a restriction in the *abstraction* rule (see, e.g. Barendregt 1981, and Hindley & Seldin 1986).

8. $\mathcal{F} \to \mathbf{A}$ (*absurdity*)

We assume that there is a distinguished proposition '$\mathcal{F}$' which is taken to be empty (i.e. no term, whether open or closed, is a member of it), and a distinguished closed term '$\lambda\bot.$' such that

$$\lambda\bot. \in \mathcal{F} \to \mathbf{A}$$

for any '$\mathbf{A}$'. We say that '$\mathcal{F} \to \mathbf{A}$', for any '$\mathbf{A}$', is the type-scheme for a combinator we call '$\mathsf{I}_{\mathcal{F}}$'.

With this axiom, and taking '$\neg\mathbf{A} \equiv \mathbf{A} \to \mathcal{F}$', we can prove Heyting's axioms involving (intuitionistic) negation,[25] namely:

(i) $\neg\mathbf{A} \to (\mathbf{A} \to \mathbf{B})$ and
(ii) $((\mathbf{A} \to \mathbf{B}) \wedge (\mathbf{A} \to \neg\mathbf{B})) \to \neg\mathbf{A}$.

(i) $(\mathbf{A} \to \mathcal{F}) \to (\mathbf{A} \to \mathbf{B})$:

$$\dfrac{\dfrac{\dfrac{\dfrac{[y \in \mathbf{A}] \quad [x \in \mathbf{A} \to \mathcal{F}]}{\mathtt{APPLY}(x,y) \in \mathcal{F}} \qquad \lambda\bot. \in \mathcal{F} \to \mathbf{B}}{\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x,y)) \in \mathbf{B}}}{\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x,y)) \in \mathbf{A} \to \mathbf{B}}}{\lambda x.\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x,y)) \in (\mathbf{A} \to \mathcal{F}) \to (\mathbf{A} \to \mathbf{B})}$$

(ii) $((\mathbf{A} \to \mathbf{B}) \wedge (\mathbf{A} \to (\mathbf{B} \to \mathcal{F}))) \to (\mathbf{A} \to \mathcal{F})$:

$$\dfrac{\dfrac{\dfrac{\dfrac{[y \in \mathbf{A}] \quad \dfrac{[x \in (\mathbf{A} \to \mathbf{B}) \wedge (\mathbf{A} \to (\mathbf{B} \to \mathcal{F}))]}{\mathtt{FST}(x) \in \mathbf{A} \to \mathbf{B}}}{\mathtt{APPLY}(\mathtt{FST}(x), y) \in \mathbf{B}} \qquad \dfrac{[y \in \mathbf{A}] \quad \dfrac{[x \in (\mathbf{A} \to \mathbf{B}) \wedge (\mathbf{A} \to (\mathbf{B} \to \mathcal{F}))]}{\mathtt{SND}(x) \in \mathbf{A} \to (\mathbf{B} \to \mathcal{F})}}{\mathtt{APPLY}(\mathtt{SND}(x), y) \in \mathbf{B} \to \mathcal{F}}}{\mathtt{APPLY}(\mathtt{APPLY}(\mathtt{SND}(x), y), \mathtt{APPLY}(\mathtt{FST}(x), y)) \in \mathcal{F}}}{\lambda y.\mathtt{APPLY}(\mathtt{APPLY}(\mathtt{SND}(x), y), \mathtt{APPLY}(\mathtt{FST}(x), y)) \in \mathbf{A} \to \mathcal{F}}}{\lambda x.\lambda y.\mathtt{APPLY}(\mathtt{APPLY}(\mathtt{SND}(x), y), \mathtt{APPLY}(\mathtt{FST}(x), y)) \in ((\mathbf{A} \to \mathbf{B}) \wedge (\mathbf{A} \to (\mathbf{B} \to \mathcal{F}))) \to (\mathbf{A} \to \mathcal{F})}$$

[24] "If M does not contain the variable x (as a free variable), then $(\lambda x M)$ might be used to denote a function whose value is constant and equal to (the thing denoted by) M, and whose range of arguments consists of all things. This usage is contemplated below in connection with the calculi of λ-K-conversion, but is excluded from the calculi of λ-conversion and λ-δ-conversion – for technical reasons which will appear."

(Church 1941, pp. 6–7.)

[25] See the two axioms below:

X. $\vdash \neg p \to (p \to q)$.
XI. $\vdash ((p \to q) \wedge (p \to \neg q)) \to \neg p$.

(Heyting 1956, p. 101.)

Moreover, to prove two of Ackermann's axioms for negation (quoted in Gabbay 1988, p. 106), namely

$(\mathbf{A} \to \neg\mathbf{B}) \to (\mathbf{B} \to \neg\mathbf{A})$

$(\mathbf{A} \to \neg\mathbf{A}) \to \neg\mathbf{A}$,

we put '$\neg\mathbf{A} \equiv \mathbf{A} \to \mathcal{F}$' and they become, respectively:

$(\mathbf{A} \to (\mathbf{B} \to \mathcal{F})) \to (\mathbf{B} \to (\mathbf{A} \to \mathcal{F}))$

$(\mathbf{A} \to (\mathbf{A} \to \mathcal{F})) \to (\mathbf{A} \to \mathcal{F})$

which are instances of the type-schemes of combinators 'C' (*permutation*), and 'W' (*contraction*), respectively.

We can also prove intuitionistically that '$\neg\neg(\mathbf{A} \vee \neg\mathbf{A})$':

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[y \in \mathbf{A}]}{\texttt{inl}(y) \in \mathbf{A} \vee (\mathbf{A} \to \mathcal{F})} \quad [x \in (\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}]}{\texttt{APPLY}(x, \texttt{inl}(y)) \in \mathcal{F}}}{\lambda y.\texttt{APPLY}(x, \texttt{inl}(y)) \in \mathbf{A} \to \mathcal{F}}}{\texttt{inr}(\lambda y.\texttt{APPLY}(x, \texttt{inl}(y))) \in \mathbf{A} \vee (\mathbf{A} \vee \mathcal{F})} \quad [x \in (\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}]}{\texttt{APPLY}(x, \texttt{inr}(\lambda y.\texttt{APPLY}(x, \texttt{inl}(y)))) \in \mathcal{F}}}{\lambda x.\texttt{APPLY}(x, \texttt{inr}(\lambda y.\texttt{APPLY}(x, \texttt{inl}(y)))) \in ((\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}) \to \mathcal{F}}$$

Note that we started with '$y \in \mathbf{A}$', literally 'y is a proof of $\mathbf{A}$', as an assumption. One can also prove the same theorem by starting with '$y \in \mathbf{A} \to \mathcal{F}$', literally '$y$ is a proof of $\mathbf{A} \to \mathcal{F}$ (i.e. $\neg\mathbf{A}$)', but in this case one would need classical implication as we shall see below.

Similarly to the proof of '$((\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}) \to \mathcal{F}$' above, we can prove intuitionistically that '$(\neg\mathbf{A} \vee \mathbf{B}) \to (\mathbf{A} \to \mathbf{B})$':

$((\mathbf{A} \to \mathcal{F}) \vee \mathbf{B}) \to (\mathbf{A} \to \mathbf{B})$:

$$\dfrac{\dfrac{\dfrac{[x \in (\mathbf{A} \to \mathcal{F}) \vee \mathbf{B}] \quad \dfrac{\dfrac{[z \in \mathbf{A}] \quad [y \in \mathbf{A} \to \mathcal{F}]}{\texttt{APPLY}(y, z) \in \mathcal{F}} \quad \lambda\bot. \in \mathcal{F} \to \mathbf{B}}{\texttt{APPLY}(\lambda\bot., \texttt{APPLY}(y, z)) \in \mathbf{B}} \quad \dfrac{\dfrac{[t \in \mathbf{B}]}{t = t \in \mathbf{B}}}{t \in \mathbf{B}}}{\texttt{WHEN}(x, \iota y.\texttt{APPLY}(\lambda\bot., \texttt{APPLY}(y, z)), \iota t.t) \in \mathbf{B}}}{\lambda z.\texttt{WHEN}(x, \iota y.\texttt{APPLY}(\lambda\bot., \texttt{APPLY}(y, z)), \iota t.t) \in \mathbf{A} \to \mathbf{B}}}{\lambda x.\lambda z.\texttt{WHEN}(x, \iota y.\texttt{APPLY}(\lambda\bot., \texttt{APPLY}(y, z)), \iota t.t) \in ((\mathbf{A} \to \mathcal{F}) \vee \mathbf{B}) \to (\mathbf{A} \to \mathbf{B})}$$

It is proper to remark here that in such an interpretation of *absurdity* we have just given, it is not the case that there are open terms of type '$\mathcal{F}$'. What we are saying here is that there are closed terms of the form '$\lambda\bot.$' of type '$\mathcal{F} \to \mathbf{A}$' for any '$\mathbf{A}$'.[26]

9. $((\mathbf{A} \to \mathbf{B}) \to \mathbf{A}) \to \mathbf{A}$ (*Peirce's rule*)[27]

In his *Foundations of Mathematical Logic* (1963) Curry presents an 'inferential' counterpart to the axiomatic form of the *Peirce's rule* as:

$$\frac{\begin{array}{c}[\mathbf{A} \to \mathbf{B}]\\ \mathbf{A}\end{array}}{\mathbf{A}} \quad \text{(in type-theoretic presentation:} \quad \frac{\begin{array}{c}[x \in \mathbf{A} \to \mathbf{B}]\\ b(x) \in \mathbf{A}\end{array}}{\lambda x.b(x) \in \mathbf{A}})$$

in p. 182 of Curry 1963.

One of the usual presentations of the rule for *reductio ad absurdum* such as:

[26] Howard's observation that by introducing an *absurdity* type such as '$\mathcal{F}$' one introduces open terms, does not seem to be applicable to our case:

"(i) For ¬: add a new prime formula f to P(⊃). Then, for each formula α, introduce a term $A^{f\supset\alpha}$. (...) There *are* open terms of type f; for example, the variable X^f–which is a construction of f→f."

(Howard 1980, p. 483.)

In the present framework, a construction of '$\mathcal{F} \to \mathcal{F}$' is also of the form '$\lambda\bot.$', therefore a closed term. After all, in our interpretation *absurdity* implies anything, including *absurdity* itself. But that does not imply that our framework is inconsistent in the sense that it produces an open term (such as Howard's 'X^f') as a member of a closed type (such as Howards 'f→f'), as we shall see from our consistency result. (Briefly: similarly to Howard's case, any judgement of the form '$a \in \mathcal{F}$' will not be a closed judgement, i.e., it will contain at least one free variable. But, unlike Howard's 'X^f', 'a' could not be a construction of '$\mathcal{F} \to \mathcal{F}$'.) This leaves us to justify the equivalence of a term like '$\lambda x.x$' to the term '$\lambda\bot.$', given that both are terms of type '$\mathcal{F} \to \mathcal{F}$', but we need not worry too much about it.

[27] As it is well known, this axiom does not find a straight counterpart in the type-schemes of Curry's combinators. Nonetheless, it seems unlikely that Curry intended his theory of functionality to be applicable only to intuitionistic implication. Rather, he appeared to be more interested in defining families of *calculi* of implication, which would also include a calculus of classical implication, such as his LC- (HC-, TC-) systems as classical counterparts to intuitionistic LA- (HA-, TA-) systems, 'A' standing for 'absolute':

"**4. The classical positive propositional algebra.** In Sec. 4C5 we saw that the scheme

$$(A \supset B) \supset A \leq A \qquad (15)$$

was not an elementary theorem scheme of an absolute implicative lattice, and in Sec. 4D1 a classical implicative lattice was defined, in effect, as an implicative lattice for which (15) holds. This classical implicative lattice is here called the *system EC*.

Acting by analogy with the absolute system, we can define classical positive propositional systems HC and TC by adjoining to HA and TA, respectively, postulates in agreement with (15). The postulate for HC is the scheme

Pc $\qquad \vdash A \supset B. \supset A :\supset A\ddagger$

which is commonly known as "Peirce's law"; that for TC is the rule

Pk $\qquad \dfrac{\begin{array}{c}[A \supset B]\\ A\end{array}}{A}$"

(Curry 1963, p. 182.)

So, pursuing what we believe to have been Curry's 'methodology', which was to direct the chief concern at the establishment of calculi (intuitionistic, classical, etc.) rather than at the interpretations, we want to obtain a calculus of classical implication within the 'propositions-are-types' interpretation by extending the conditions for closing a term (therefore binding a free variable and discharging an assumption), so that one can obtain a closed pure term for '$((\mathbf{A} \to \mathbf{B}) \to \mathbf{A}) \to \mathbf{A}$'.

$$\frac{\begin{matrix}[\neg \mathbf{A}]\\ \mathbf{A}\end{matrix}}{\mathbf{A}}$$

can be seen as a particular case of *Peirce's law* in its deductive (non-axiomatic) presentation, when negation is introduced (taking $\neg\mathbf{A} \equiv \mathbf{A} \rightarrow \mathcal{F}$):

$$\frac{\begin{matrix}[\mathbf{A} \rightarrow \mathcal{F}]\\ \mathbf{A}\end{matrix}}{\mathbf{A}}$$

where '$\mathbf{B}$' is instantiated with '$\mathcal{F}$'.[28]

Now, if we use the formulation given by Curry to construct the proof-tree for the axiom scheme '$((\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A}) \rightarrow \mathbf{A}$' similarly to the one given above, we get:

$$\dfrac{\dfrac{\dfrac{\boxed{[y \in \mathbf{A} \rightarrow \mathbf{B}]} \qquad [x \in (\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A}]}{\mathtt{APPLY}(x, y) \in \mathbf{A}}}{\boxed{\lambda y.}\mathtt{APPLY}(x, y) \in \mathbf{A}}(*)}{\lambda x.\lambda y.\mathtt{APPLY}(x, y) \in ((\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A}) \rightarrow \mathbf{A}}$$

where the step '$(*)$' is justified by Curry's inferential presentation of *Peirce's law* above. It allows the rewriting of '$(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A}$' to '$\mathbf{A}$', in this direction.[29] In the opposite direction the rewriting can be made with a weaker implication such as strict implication, as shown by the following proof-tree:

[28] Note that the axiom corresponding to *Peirce's rule* has nothing to do with the axiom which introduces negation (or better, absurdity). It only requires that the rules for assumption discharge with *λ-abstraction* be changed to cover the full power of classical implication. There are some slightly different views on this particular point, such as, e.g. Lambek's:

"The negationless formula $A \Leftarrow (A \Leftarrow (B \Leftarrow A))$ is a theorem classically but not in the system without negation."
(Lambek 1980, p. 384.)

The system without negation referred to by Lambek corresponds to the system we had before introducing the *absurdity* judgement. *Peirce's axiom* cannot be expected to be a theorem of that system, given that the extra conditions of assumption-discharge were not present in that system. But those extra conditions are not introduced specifically to allow the handling of negation.

In the presentation of propositional calculus as a deductive system, Lambek & Scott also insist on the fact that it is by introducing (double) negation that one obtains classical implication:

"If we want *classical* propositional logic, we must also require
R7. $\perp \Leftarrow (\perp \Leftarrow A) \rightarrow A$."
(Lambek & Scott 1986, p. 50.)

The point here is that the double negation above (R7) follows from the more general characteristic of classical implication which is captured by *Peirce's law* and which is not intrinsically bound to the introduction of (double) negation. In other words, one can have a negationless classical implication by dropping the *absurdity* axiom (or *λ⊥.-abstraction*) from full intuitionistic implication and adding *Peirce's law*.

[29] It should come as no surprise that the λ-term inhabiting the type '$((\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A}) \rightarrow \mathbf{A}$' is the same as '$((\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A}) \rightarrow ((\mathbf{A} \rightarrow \mathbf{B}) \rightarrow \mathbf{A})$' as well as '$(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow (\mathbf{A} \rightarrow \mathbf{B})$', which is the term corresponding to Curry's 'I_η' $\equiv \lambda x.\lambda y.\mathtt{APPLY}(x, y)$ (Curry & Feys 1958, p. 379):

$$\dfrac{\dfrac{\begin{array}{c}\boxed{[y \in \mathbf{A} \to \mathbf{B}]}\\ [x \in \mathbf{A}]\end{array}}{\boxed{\lambda y.}\,x \in (\mathbf{A} \to \mathbf{B}) \to \mathbf{A}}}{\lambda x.\lambda y.x \in \mathbf{A} \to ((\mathbf{A} \to \mathbf{B}) \to \mathbf{A})}$$

(Note that the abstraction '$\boxed{\lambda y.}$' is not a relevant abstraction, but it is an intuitionistic one.)
■

To prove the classical double negation '$((\mathbf{A} \to \mathcal{F}) \to \mathcal{F}) \to \mathbf{A}$' we can use the inferential presentation of the *Peirce's law*:

$$\dfrac{\dfrac{\dfrac{\dfrac{[y \in \mathbf{A} \to \mathcal{F}] \quad [x \in (\mathbf{A} \to \mathcal{F}) \to \mathcal{F}]}{\mathtt{APPLY}(x,y) \in \mathcal{F}} \qquad \lambda\bot. \in \mathcal{F} \to \mathbf{A}}{\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x,y)) \in \mathbf{A}}}{\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x,y)) \in \mathbf{A}}}{\lambda x.\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x,y)) \in ((\mathbf{A} \to \mathcal{F}) \to \mathcal{F}) \to \mathbf{A}}$$

As we have mentioned above, we can prove classically '$\neg\neg(\mathbf{A} \vee \neg\mathbf{A})$' starting from the assumption that '$y \in \neg\mathbf{A}$' (i.e., '$y \in \mathbf{A} \to \mathcal{F}$').

Indeed, using Curry's $\dfrac{\begin{array}{c}[\mathbf{A} \to \mathbf{B}]\\ \mathbf{A}\end{array}}{\mathbf{A}}$, we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[y \in \mathbf{A} \to \mathcal{F}]}{\mathtt{inr}(y) \in \mathbf{A} \vee (\mathbf{A} \to \mathcal{F})} \quad [x \in (\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}]}{\mathtt{APPLY}(x, \mathtt{inr}(y)) \in \mathcal{F}} \qquad \lambda\bot. \in \mathcal{F} \to \mathbf{A}}{\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, \mathtt{inr}(y))) \in \mathbf{A}}}{\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, \mathtt{inr}(y))) \in \mathbf{A}}}{\mathtt{inl}(\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, \mathtt{inr}(y))) \in \mathbf{A} \vee (\mathbf{A} \to \mathcal{F})} \qquad [x \in (\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}]}{\mathtt{APPLY}(x, \mathtt{inl}(\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, \mathtt{inr}(y))) \in \mathcal{F}}}{\lambda x.\mathtt{APPLY}(x, \mathtt{inl}(\lambda y.\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, \mathtt{inr}(y))) \in ((\mathbf{A} \vee (\mathbf{A} \to \mathcal{F})) \to \mathcal{F}) \to \mathcal{F}}$$

Despite working well in most cases, Curry's inferential counterpart to Peirce's law does not seem to be sufficient to prove the following theorem of classical implication (taking $\neg\mathbf{A} \equiv \mathbf{A} \to \mathcal{F}$):

$$((\mathbf{A} \to \neg\mathbf{A}) \to \mathbf{B}) \to ((\mathbf{A} \to \mathbf{B}) \to \mathbf{B})$$

$$\dfrac{\dfrac{[x \in (\mathbf{A} \to \mathbf{B}) \to \mathbf{A}]}{\lambda y.\mathtt{APPLY}(x,y) \in (\mathbf{A} \to \mathbf{B}) \to \mathbf{A}}(*)}{\lambda x.\lambda y.\mathtt{APPLY}(x,y) \in ((\mathbf{A} \to \mathbf{B}) \to \mathbf{A}) \to ((\mathbf{A} \to \mathbf{B}) \to \mathbf{A})}$$

where the step '$(*)$' is justified by the rule of *$\to$-induction*, the counterpart to λ-calculus η-rule. Of course, the additional condition in Curry's rule is that the consequent '$(\mathbf{A} \to \mathbf{B}) \to \mathbf{A}$' is identified with '$\mathbf{A}$'.

whose proof is left as an exercise by Curry in p. 279 of his *Foundations of Mathematical Logic* (Curry 1963). We have been able to prove it using a reformulation of the inferential counterpart to Peirce's axiom which is framed as follows:

$$\frac{\begin{array}{c}[\mathbf{A} \to \mathbf{B}] \\ \mathbf{B}\end{array}}{\mathbf{A}} \qquad \text{(provided `}\mathbf{A} \to \mathbf{B}\text{' is used as both minor and ticket)}$$

In the framework of the propositions-are-types interpretation it would be framed as:

$$\frac{\begin{array}{c}[x \in \mathbf{A} \to \mathbf{B}] \\ b(x, ..., x) \in \mathbf{B}\end{array}}{\lambda x.b(x, ..., x) \in \mathbf{A}}$$

meaning that if from the assumption that a term 'x' belongs to a type of the form '$\mathbf{A} \to \mathbf{B}$' one obtains a term '$b(x)$' belonging to the consequent '$\mathbf{B}$' where 'x' appears both as a 'higher' and a 'lower' subterm of '$b(x)$', then we can apply a $\lambda x.$-abstraction over the '$b(x, ..., x)$' term and obtain a term of the form '$\lambda x.b(x, ..., x)$' belonging to the antecedent '$\mathbf{A}$', discharging the assumption '$x \in \mathbf{A} \to \mathbf{B}$'.

Such an alternative presentation of the inferential counterpart to *Peirce's axiom* finds a special case in another one of the standard presentations of the proof-theoretic *reductio ad absurdum*, namely:

$$\frac{\begin{array}{c}[\neg \mathbf{A}] \\ \mathcal{F}\end{array}}{\mathbf{A}} \qquad \text{which can also be presented as} \qquad \frac{\begin{array}{c}[\mathbf{A} \to \mathcal{F}] \\ \mathcal{F}\end{array}}{\mathbf{A}}$$

By using such an alternative to Curry's formulation we can also prove the classical double negation:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{[y \in \mathbf{A} \to \mathcal{F}] \quad [x \in (\mathbf{A} \to \mathcal{F}) \to \mathcal{F}]}{\mathtt{APPLY}(x, y) \in \mathcal{F}} \qquad \lambda\bot. \in \mathcal{F} \to \mathbf{A}}{\mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, y)) \in \mathbf{A}} \qquad [y \in \mathbf{A} \to \mathcal{F}]}{\mathtt{APPLY}(y, \mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, y))) \in \mathcal{F}}}{\lambda y.\mathtt{APPLY}(y, \mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, y))) \in \mathbf{A}}}{\lambda x.\lambda y.\mathtt{APPLY}(y, \mathtt{APPLY}(\lambda\bot., \mathtt{APPLY}(x, y))) \in ((\mathbf{A} \to \mathcal{F}) \to \mathcal{F}) \to \mathbf{A}}$$

For the present framework, however, we shall restrict ourselves to Curry's inferential counterpart to *Peirce's law*.[30]

6. Consistency Proof

[30] In fact, we are also investigating the possibility of having the axiom for Lukasiewicz' many-valued logics (implication), namely:

$$((\mathbf{A} \to \mathbf{B}) \to \mathbf{B}) \to ((\mathbf{B} \to \mathbf{A}) \to \mathbf{A})$$

as a more general schema which would entail *Peirce's law*. We are looking for a proof discipline for Lukasiewicz' logics which would perhaps gives us an insight as to which discipline to adopt for classical implication.

Following a technique presented in Martin-Löf 1972, we shall prove that our systems of implication are consistent by arguing that no judgement of the form '$m \in \mathbf{M}$' can be obtained from the presentation of the $\rightarrow$-type such that 'm' is not a closed term (i.e., contains a free variable). Furthermore, because all final judgements (at the bottom of proof-trees) must involve closed terms only, one cannot prove '$\mathbf{A} \rightarrow \mathbf{B}$' for any $\mathbf{A}$, $\mathbf{B}$.

First of all, we recall that the only assumptions allowed in any step of a proof is of the form '$x \in \mathbf{A}$' where 'x' is simply a (new) variable. If '$y \in \mathbf{A}$' already appears in the proof, then 'y' has to be used instead of 'x', which means that once a proposition/formula is given a label, it cannot be given a new name.

Exceptions to the general rule that assumptions are placed in the top of the proof-tree are introduced by:

(i) the rule of ***truth***, where an assumption is allowed to be placed anywhere in the proof-tree (except in the bottom, of course) to play the rôle of the '$b(x)$' in the $\rightarrow$-*introduction*:

$$\cfrac{\cfrac{\cfrac{[y \in \mathbf{B}]}{\boxed{[x \in \mathbf{A}]}}}{\lambda y.x \in \mathbf{B} \rightarrow \mathbf{A}}}{\lambda x.\lambda y.x \in \mathbf{A} \rightarrow (\mathbf{B} \rightarrow \mathbf{A})}$$

(Note that the assumption '$\boxed{[x \in \mathbf{A}]}$' is introduced in the middle, not at the beginning, of a path in the proof-tree, to play the rôle of the '$b(x)$' for the λy.-abstraction made in the following step.)

(ii) the rule of ***absurdity***, where a distinguished λ-abstraction term, namely '$\lambda\bot.$', is taken for granted and does not involve an assumption discharge:

$$\lambda\bot. \in \mathcal{F} \rightarrow \mathbf{A}$$

Now, looking at the general form of the rules of proof that we have available:

$$\rightarrow\textit{-introduction}\quad \cfrac{\begin{matrix}[x \in \mathbf{A}]\\ b(x) \in \mathbf{B}\end{matrix}}{\lambda x.b(x) \in \mathbf{A} \rightarrow \mathbf{B}} \qquad \rightarrow\textit{-elimination}\quad \cfrac{a \in \mathbf{A} \qquad c \in \mathbf{A} \rightarrow \mathbf{B}}{\texttt{APPLY}(c,a) \in \mathbf{B}}$$

Note that with respect to the discharging of assumptions, only the $\rightarrow$-*introduction* can discharge.

If we want to extend the framework with conjunction and disjunction:

$$\wedge\textit{-i}\quad \cfrac{a \in \mathbf{A} \qquad b \in \mathbf{B}}{\texttt{pair}(a,b) \in \mathbf{A} \wedge \mathbf{B}} \qquad \wedge\textit{-e}\quad \cfrac{c \in \mathbf{A} \wedge \mathbf{B}}{\texttt{FST}(c) \in \mathbf{A}} \qquad \cfrac{c \in \mathbf{A} \wedge \mathbf{B}}{\texttt{SND}(c) \in \mathbf{B}}$$

$$\vee\textit{-i}\quad \cfrac{a \in \mathbf{A}}{\texttt{inl}(a) \in \mathbf{A} \vee \mathbf{B}} \qquad \cfrac{b \in \mathbf{B}}{\texttt{inr}(b) \in \mathbf{A} \vee \mathbf{B}} \qquad \vee\textit{-e}\quad \cfrac{c \in \mathbf{A} \vee \mathbf{B} \qquad \begin{matrix}[x \in \mathbf{A}]\\ d(x) \in \mathbf{C}\end{matrix} \qquad \begin{matrix}[y \in \mathbf{B}]\\ e(y) \in \mathbf{C}\end{matrix}}{\texttt{WHEN}(c, \iota x.d(x), \iota y.e(y)) \in \mathbf{C}}$$

we note that with respect to the capacity of discharging assumptions is:

$\wedge$-*i*: none $\wedge$-*e*: none

$\vee$-*i*: none $\vee$-*e*: some

Back to the $\rightarrow$-case, which is the only case where all assumptions are discharged, the only 'odd' case is that of the counterpart of *Peirce's law*, where a λ-*abstraction* term is said to belong to a 'lower' type and the assumption on the higher type is discharged:

$$\frac{\begin{array}{c}[y \in \mathbf{A} \rightarrow \mathbf{B}]\\ b(y) \in \mathbf{A}\end{array}}{\lambda y.b(y) \in \mathbf{A}}$$

Here the higher type '$\mathbf{A} \rightarrow \mathbf{B}$' is discharged and the lower type '$\mathbf{A}$' is said to contain a λ-abstraction term '$\lambda y.b(y)$'.

7. Systems of implication and combinators

In Gabbay 1989 a classification of different systems of implication is given in terms of the axioms chosen from a certain stock of basic axioms (pp. 31–2):

1. *identity*
$\mathbf{A} \rightarrow \mathbf{A}$, which corresponds to the type-scheme of combinator 'I'.

2. *right transitivity*
$(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow ((\mathbf{B} \rightarrow \mathbf{C}) \rightarrow (\mathbf{A} \rightarrow \mathbf{C}))$, which corresponds to the type-scheme of combinator 'B′'.

3. *left transitivity*
$(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow ((\mathbf{C} \rightarrow \mathbf{A}) \rightarrow (\mathbf{C} \rightarrow \mathbf{B}))$, which corresponds to the type-scheme of combinator 'B'.

4. *distribution*
$(\mathbf{A} \rightarrow \mathbf{B}) \rightarrow (((\mathbf{A} \rightarrow (\mathbf{B} \rightarrow \mathbf{C})) \rightarrow (\mathbf{A} \rightarrow \mathbf{C}))$, which is a variation of the type-scheme of combinator 'S'.

$5m,n$ (m,n) *contraction*
$(\mathbf{A}^m \rightarrow \mathbf{B}) \rightarrow (\mathbf{A}^n \rightarrow \mathbf{B})$, which is a generalisation of the type-scheme of combinator 'W', namely $(\mathbf{A} \rightarrow (\mathbf{A} \rightarrow \mathbf{B})) \rightarrow (\mathbf{A} \rightarrow \mathbf{B})$.

6α. α-*deduction*
$\alpha \rightarrow (\mathbf{B} \rightarrow \alpha)$, which comprises variations of the type-scheme of combinator 'K'. For example, by saying that α has to be in implicational form we get a $\mathsf{K}_{\rightarrow}$ for strict implication.

7γ. γ-*permutation*

$(\mathbf{A} \to (\gamma \to \mathbf{C})) \to (\gamma \to (\mathbf{A} \to \mathbf{C}))$, which corresponds to variations of the type-scheme of combinator 'C'. For entailment implication, one has to impose the condition that γ must be in implicational form (e.g., '$\mathbf{P} \to \mathbf{Q}$'), and we here call the corresponding combinator '$\mathsf{C}_{\to}$'. If we impose that '$\mathbf{A} \nvdash \gamma$' in the particular system, we have the variation '$\mathsf{C}_{\nvdash}$'.

8. *restart (Peirce's rule)*
   $((\mathbf{A} \to \mathbf{B}) \to \mathbf{A}) \to \mathbf{A}$, which corresponds to the type-scheme of our combinator 'P′'.

Based on the correspondence between the axioms and the type-schemes of combinators, we can do the same classification done in Gabbay 1989, p. 33, but now in terms of the combinators which would be obtained according to the side conditions on the rule of *assertability conditions* for the logical connective of implication, namely *→-introduction*.

Systems of implication:

system name :	combinators :
W	I, B, B′
Linear	I, B, B′, C
Modal T-Strict	I, B B′, C, W
Relevant	I, B, B′, C, S
Entailment	I, B, B′, S, $\mathsf{C}_{\to}$
Ticket Entailment	I, B, B′, S, SC
Strict	I, B, B′, C, $\mathsf{K}_{\to}$
Minimal	I, B, B′, C, S, K
Intuitionistic	I, B, B′, C, S, K, $\mathsf{I}_{\mathcal{F}}$
Classical	I, B, B′, C, S, K, P′
Linear Classical	I, B, B′, C, P′
Relevant Classical	I, B, B′, C, S, P′
Deductive Relevant	I, B, B′, $\mathsf{C}_{\nvdash}$, S
LE-Linear Entailment	I, B, B′, $\mathsf{C}_{\to}$
Linear Intuitionistic	I, B, B′, C, K, $\mathsf{I}_{\mathcal{F}}$

WC I, B, B′, P′

The systems are roughly ordered by proof-theoretic strength, and the system **W** is considered to be the weakest implicational system for which a reasonable deduction theorem exists.[31] Note that the most primitive combinators are I, B, B′(≡ CB), C and W, instead of I, K and S. These were, in fact the primitive combinators in Curry's earliest results on combinatory logic, unlike Schönfinkel's independent pioneering results using B, C, I, K, and S.[32]

7. Finale

With the help of the distinction between a *proposition* ('**A**') and a *judgement* ('$a \in \mathbf{A}$'), together with the identification of *propositions* with *types*, one can have proof-objects 'coded' into the object language, so to speak. Such an 'improvement' on the syntactical tools of a proof calculus seems to be particularly helpful in dealing with the so-called resource logics. By making the type-theoretic equivalent of the λ-calculus' *abstraction* rule into a 'resource' *abstraction* where an extra condition is included requiring the existence of at least a free occurrence of the variable being abstracted, one can provide a workable framework to present resource logics via the so-called 'Curry-Howard-Tait'-interpretation in a reasonably simple way. We are currently working on the extension of the classification presented here for the case of implication to first-order quantification, given that in the Curry-Howard-Tait interpretation the universal quantifier is dealt with in a similar manner to implication.[33] And indeed, by presenting the universal quantifier in a similar way to implication as:

[31] Here we should mention a recent attempt by Y. Komori (in a handwritten memo – Komori 1990(?) – which was kindly sent to us) to answer the question 'What is the weakest meaningful logic?' by saying that: "The weakest meaningful logic is **B** Logic.", where **B** is a logic with only one axiom (the one corresponding to the type-scheme of combinator B) and a rule of modus ponens.

[32] "The earliest work of Curry (till the fall of 1927), which was done without knowledge of the work of Schönfinkel [Schönfinkel 1924], used B, C, W, and I as primitive combinators.", p. 184 of Curry & Feys 1958.

[33] In Howard's account of the formulae-as-types notion of construction he defines constructions as terms built up from prime terms by means of term formation as indicated by *Prime terms*, *λ-abstraction* and *Application*:
"(i) *Type symbols* The prime type symbols are: 0 and every equation of $H(\supset, \wedge, \forall)$. From these we generate all type symbols by the following two rules.
(a) From α and β get $\alpha \supset \beta$ and $\alpha \wedge \beta$.
(b) From α and a number variable x get $\forall x\alpha$.
(ii) *Prime terms* These are:
(a) number variables x, y,...; constants 0 and 1; function symbols for plus and times,
(b) variables X^{α}, Y^{β},...,
(c) certain special terms, mentioned in §8, below, corresponding to axioms and rules of inference of $H(\supset, \wedge, \forall)$.
(iii) *λ-abstraction*:
(a) From F^{β} get $(\lambda X^{\alpha}.F^{\beta})^{\alpha \supset \beta}$ as in §2.
(b) If x does not occur free in the type symbol of any free variable of F, form $(\lambda x F^{\beta})^{\forall x \beta}$.
(iv) *Application*:
(a) From F^{α} and $G^{\alpha \supset \beta}$ form $(GF)^{\beta}$ as in §2.
(b) From $G^{\forall x \alpha(x)}$ and t of type 0 form $G(t)^{\alpha(t)}$."

(Howard 1980, p. 485.)

Observe that for both implication ('(i.a)', '(iii.a)' and '(iv.a)') and universal quantification ('(i.b)', '(iii.b)',

∀-formation

'(iv.b)') a λ-system is used (with *abstraction* and *application*). Later he gives the following axioms for the reducibility of terms:

"*11. Normalisation of terms*
For the theory of reducibility of terms we postulate the following contraction schemes
(i) $(\lambda X.F(X))^{\alpha\supset\beta}G$ contr $F(G)^{\beta}$
$(\lambda x.F(x))^{\forall x\alpha(x)}t$ contr $F(t)^{\alpha(t)}$
(...)"

(Ibid., p. 487.)

The equivalent of β-normalisation is postulated to the contraction of terms characterising implication, as well as terms characterising universal quantification.

In the description of his type system F Girard also makes use of the notions of *abstraction* and *application* in the definition of both implication and universal quantification, although in a way which is different from Howard's:

"*Types* are defined starting from *type variables* $X, Y, Z, \ldots$ by means of two operations:
1. if U and V are types, then $U \to V$ is a type.
2. if V is a type, and X a type variable, then $\Pi X.V$ is a type.

There are five schemes for forming *terms:*
1. *variables:* $x^T, y^T, z^T, \ldots$ of type T,
2. *application:* tu of type V, where t is of type $U \to V$ and u is of type U,
3. λ-*abstraction:* $\lambda x^U.v$ of type $U \to V$, where x^U is a variable of type U and v is of type V,
4. *universal abstraction:* if v is a term of type V, then we can form $\Lambda X.v$ of type $\Pi X.V$, so long as the variable X is not free in the type of a free variable of v.
5. *universal application* (sometimes called *extraction*): if t is a term of type $\Pi X.V$ and U is a type, then tU is a term of type $V[U/X]$.

As well as the usual *conversions* for application/λ-abstraction, there is one for the other pair of schemes:
$$(\Lambda X.v)U \rightsquigarrow v[U/X]"$$

(Girard 1989, pp. 81–2.)

Note that Howard's way is to abstract on *type*-variables ('$(\lambda X^{\alpha}.F^{\beta})^{\alpha\supset\beta}$') for implication and on *element*-variables ('$(\lambda x F^{\beta})^{\forall x\beta}$') for universal quantification, whereas Girard's way is to abstract on *element*-variables ('$\lambda x^U.v$') for implication (clause 3 above) and on *type*-variables ('$\Lambda X.v$') for universal quantification (clause 4 above). Similarly to Martin-Löf (1984) we abstract on *element*-variables in both cases, the difference being that for implication we can say that '$\lambda x.b(x) \in \mathbf{A} \to \mathbf{B}$' provided '$b(x) \in \mathbf{B}$' on the assumption that '$x \in \mathbf{A}$', whereas for universal quantification '$\Lambda x.b(x) \in \forall x \in \mathbf{A}.\mathbf{B}(x)$' provided '$b(x) \in \mathbf{B}(x)$' (where '$\mathbf{B}(x)$' is a type indexed by '$x$') on the assumption that '$x \in \mathbf{A}$'. But unlike Martin-Löf's unified treatment with a Π-type (and associated definitional equalities distinguishing '$\to$' from '$\forall$', namely, '$A \to B \equiv (\Pi x \in A)B$' and '$(\forall x \in A)B(x) \equiv (\Pi x \in A)B(x)$', Martin-Löf 1984, p. 32.), we follow Howard's and Girard's approach of dealing with implication and universal quantification by using separate type definitions.

Curry's original insight concerning the treatment of universal quantification in a similar way to implication is clear from his early 'The Universal Quantifier in Combinatory Logic' (Curry 1931):

"The combinator here defined I have called the formalizing combinator, because by means of it it is possible to define the relation of formal implication for functions of one or more variables in terms of ordinary implication. Thus if P is ordinary implication, it follows from Theorem 1 (below) that $(\phi_n P)$ is that function of two functions of n variables, whose value for the given functions $f(x_1, x_2, \ldots, x_n)$ and $g(x_1, x_2, \ldots, x_n)$ is the function $f(x_1, \ldots, x_n) \to g(x_1, \ldots, x_n)$. Thus formal implication for functions of n variables is $(B\Pi_n(\phi_n P))$."

(Curry 1931, pp. 165–6.)

and later comes the axioms:

"(ΠB). (f) $[(x)fx \to (g,x)f(gx)]$.
(ΠC). (f) $[(x,y)f(x,y) \to (x,y)f(y,x)]$.
(ΠW). (f) $[(x,y)f(x,y) \to (x)f(x,x)]$.

$$\frac{\mathbf{A}\ type \quad \begin{matrix}[x \in \mathbf{A}] \\ \mathbf{B}(x)\ type\end{matrix}}{\forall x \in \mathbf{A}.\mathbf{B}(x)\ type} \qquad \frac{\mathbf{A} = \mathbf{C} \quad \begin{matrix}[x \in \mathbf{A}] \\ \mathbf{B}(x) = \mathbf{D}(x)\end{matrix}}{\forall x \in \mathbf{A}.\mathbf{B}(x) = \forall x \in \mathbf{C}.\mathbf{D}(x)}$$

∀-introduction

$$\frac{\begin{matrix}[x \in \mathbf{A}] \\ b(x) \in \mathbf{B}(x)\end{matrix}}{\Lambda x.b(x) \in \forall x \in \mathbf{A}.\mathbf{B}(x)} \qquad \frac{\begin{matrix}[x \in \mathbf{A}] \\ b(x) = d(x) \in \mathbf{B}(x)\end{matrix}}{\Lambda x.b(x) = \Lambda x.d(x) \in \forall x \in \mathbf{A}.\mathbf{B}(x)}$$

∀-reduction

$$\frac{a \in \mathbf{A} \quad \begin{matrix}[x \in \mathbf{A}] \\ b(x) \in \mathbf{B}(x)\end{matrix}}{\mathtt{EXTR}(\Lambda x.b(x), a) = b(a/x) \in \mathbf{B}(a)}$$

∀-induction

$$\frac{c \in \forall x \in \mathbf{A}.\mathbf{B}(x)}{\Lambda x.\mathtt{EXTR}(c, x) = c \in \forall x \in \mathbf{A}.\mathbf{B}(x)}$$

□

one can see the correspondence of type-schemes of combinators, axioms of implication, and axioms of universal quantification:[34]

(ΠK).	(p)	$[p \to (x)Kpx]$.
(ΠP).	(f, g)	$[(x)(fx \to gx) \to ((x)fx \to (x)gx)]$."

(Ibid., p. 170.)

The combined treatment is also made in the second volume of *Combinatory Logic* (with R. Hindley and J. Seldin). In both cases the fundamental rule is a sort of modus ponens (universal instantiation):

"Rule Π. $\Pi X, EU \vdash XU$,
Rule P. $PXY, X \vdash Y$."

(Curry, Hindley & Seldin 1972, p. 427.)

[34]In Curry, Hindley & Seldin 1972 one already finds some of the parallels listed here, such as, e.g.:

"(PK)	$\vdash \alpha \supset .\beta \supset \alpha$,
(PS)	$\vdash \alpha \supset .\beta \supset \gamma :\supset: \alpha \supset \beta. \supset .\alpha \supset \gamma$,
(Π_0)	$\vdash (\forall x)\alpha x. \supset \alpha U$,
(Π_2)	$\vdash \alpha \supset (\forall x)\alpha$,
(ΠP)	$\vdash (\forall x)(\alpha x \supset \beta x). \supset .(\forall x)\alpha x \supset (\forall x)\beta x$,"

(Ibid., p. 433.)

Note the parallel between PK and Π_2, as well as between PS and ΠP. Furthermore, following the same line of reasoning another Π-rule is soon defined mirroring PC – type-scheme for combinator 'C', namely '$\gamma \supset \alpha. \supset \beta :\supset: \alpha \supset \gamma. \supset \beta$' –, which is called Π_1:

"(Π_1) $\vdash (\forall x)(\alpha \supset \beta x). \supset .\alpha \supset (\forall x)\beta x$" (Ibid., p. 439.)

(In fact, Π_1, Π_2 and ΠP, though not the parallel with the propositional PC, PK and PS, were already presented in Curry's own earlier work Curry 1963, p. 344.)

That would seem to justify why the structural similarity between implication and universal quantification, which goes back at least as far as Heyting's intuitionistic predicate calculus (Heyting 1946), is so naturally reflected in the Curry-Howard-Tait interpretation. (And indeed, a form of both ΠP and Π_1 without the

I:
$\mathbf{A} \to \mathbf{A}$
$\forall x \in \mathbf{A}.\mathbf{A}(x)$

B:
$(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{C} \to \mathbf{A}) \to (\mathbf{C} \to \mathbf{B}))$
$\forall x \in \mathbf{A}.\mathbf{B}(x) \to (\forall x \in \mathbf{C}.\mathbf{A}(x) \to \forall x \in \mathbf{C}.\mathbf{B}(x))$

B′:
$(\mathbf{A} \to \mathbf{B}) \to ((\mathbf{B} \to \mathbf{C}) \to (\mathbf{A} \to \mathbf{C}))$
$\forall x \in \mathbf{A}.\mathbf{B}(x) \to (\forall x \in \mathbf{B}.\mathbf{C}(x) \to \forall x \in \mathbf{A}.\mathbf{C}(x))$

C:
$(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to (\mathbf{B} \to (\mathbf{A} \to \mathbf{C}))$
$\forall x \in \mathbf{A}.(\mathbf{B} \to \mathbf{C}(x)) \to (\mathbf{B} \to \forall x \in \mathbf{A}.\mathbf{C}(x))$

W:
$(\mathbf{A} \to (\mathbf{A} \to \mathbf{B})) \to (\mathbf{A} \to \mathbf{B})$
$(\mathbf{A} \to \forall x \in \mathbf{A}.\mathbf{B}(x)) \to \forall x \in \mathbf{A}.\mathbf{B}(x)$

S:
$(\mathbf{A} \to (\mathbf{B} \to \mathbf{C})) \to ((\mathbf{A} \to \mathbf{B}) \to (\mathbf{A} \to \mathbf{C}))$
$\forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x)) \to (\forall x \in \mathbf{A}.\mathbf{B}(x) \to \forall x \in \mathbf{A}.\mathbf{C}(x))$

K:
$\mathbf{A} \to (\mathbf{B} \to \mathbf{A})$
$\mathbf{A} \to \forall x \in \mathbf{B}.\mathbf{A}$

For the counterpart of *Peirce's axiom*
$((\mathbf{A} \to \mathbf{B}) \to \mathbf{A}) \to \mathbf{A}$,
one would have
$(\forall x \in \mathbf{A}.\mathbf{B}(x) \to \mathbf{A}) \to \mathbf{A}$.

As an example, we can see that the following axiom would be valid only if the universal quantifier is not linear:

$\forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x)) \to (\forall x \in \mathbf{A}.\mathbf{B}(x) \to \forall x \in \mathbf{A}.\mathbf{C}(x))$ (*distribution over individuals*) (parallel to 'S')

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\boxed{[t \in \mathbf{A}]} \quad [z \in \forall x \in \mathbf{A}.\mathbf{B}(x)]}{\mathtt{EXTR}(z,t) \in \mathbf{B}(t)} \qquad \cfrac{\boxed{[t \in \mathbf{A}]} \quad [y \in \forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x))]}{\mathtt{EXTR}(y,t) \in \mathbf{B}(t) \to \mathbf{C}(t)}}{\mathtt{APPLY}(\mathtt{EXTR}(y,t), \mathtt{EXTR}(z,t)) \in \mathbf{C}(t)}}{\boxed{\Lambda t.}\mathtt{APPLY}(\mathtt{EXTR}(y,t), \mathtt{EXTR}(z,t)) \in \forall x \in \mathbf{A}.\mathbf{C}(x)}}{\lambda z.\Lambda t.\mathtt{APPLY}(\mathtt{EXTR}(y,t), \mathtt{EXTR}(z,t)) \in \forall x \in \mathbf{A}.\mathbf{B}(x) \to \forall x \in \mathbf{A}.\mathbf{C}(x)}}{\lambda y.\lambda z.\Lambda t.\mathtt{APPLY}(\mathtt{EXTR}(y,t), \mathtt{EXTR}(z,t)) \in \forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x)) \to (\forall x \in \mathbf{A}.\mathbf{B}(x) \to \forall x \in \mathbf{A}.\mathbf{C}(x))}$$

leftmost universal quantification already appear in Heyting 1946 as 'Rule (γ1)' and 'formula (7)', respectively, where it is said they both come from Hilbert & Ackermann's *Grundzüge der theoretischen Logik*, 2nd edition, Berlin, 1938.)

(Note that the $\boxed{\Lambda t.}$-abstraction is discharging assumptions non-linearly and cancelling more than one free occurrence of the variable 't' in the expression '$\mathtt{APPLY}(\mathtt{EXTR}(y,t),\mathtt{EXTR}(z,t))$') and similarly the following variant would not be valid if the quantifier is linear, given that the universal abstraction is cancelling more than one free occurrence of the variable and in a branching way:

$\forall x \in \mathbf{A}.\mathbf{B}(x) \to (\forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x)) \to \forall x \in \mathbf{A}.\mathbf{C}(x))$ (*variant of distribution over individuals*) (parallel to 'SC')

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\boxed{[t \in \mathbf{A}]} \quad [y \in \forall x \in \mathbf{A}.\mathbf{B}(x)]}{\mathtt{EXTR}(y,t) \in \mathbf{B}(t)} \quad \dfrac{\boxed{[t \in \mathbf{A}]} \quad [z \in \forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x))]}{\mathtt{EXTR}(z,t) \in \mathbf{B}(t) \to \mathbf{C}(t)}}{\mathtt{APPLY}(\mathtt{EXTR}(z,t),\mathtt{EXTR}(y,t)) \in \mathbf{C}(t)}}{\boxed{\Lambda t.}\mathtt{APPLY}(\mathtt{EXTR}(z,t),\mathtt{EXTR}(y,t)) \in \forall x \in \mathbf{A}.\mathbf{C}(x)}}{\lambda z.\Lambda t.\mathtt{APPLY}(\mathtt{EXTR}(z,t),\mathtt{EXTR}(y,t)) \in \forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x)) \to \forall x \in \mathbf{A}.\mathbf{C}(x)}}{\lambda y.\lambda z.\Lambda t.\mathtt{APPLY}(\mathtt{EXTR}(z,t),\mathtt{EXTR}(y,t)) \in \forall x \in \mathbf{A}.\mathbf{B}(x) \to (\forall x \in \mathbf{A}.(\mathbf{B}(x) \to \mathbf{C}(x)) \to \forall x \in \mathbf{A}.\mathbf{C}(x))}$$

There is room for further extending the Curry-Howard-Tait interpretation to deal with modal logics, if one makes a special (and useful) reading of the modal connective '$\Box$'. Looking at implication and universal quantification as being fundamentally characterised by modus ponens (or universal extraction), which in the Curry-Howard-Tait interpretation it is captured by β-normalisation, one can see that each implication/universal quantifier changes only its *assertability conditions* rule according to the logic, the rules corresponding to the explanation of the *consequences* – the 'ultimate' semantical rules, according to a particular semantical standpoint explored in de Queiroz 1989 – remaining fixed. Now, looking at the modal '$\Box$' as a sort of second-order universal quantification, in a way such that:

$$\Box\mathbf{A} \equiv \forall \mathbf{X} \in \mathcal{W}.(\mathbf{X} \to \mathbf{A})$$

(where '$\mathcal{W}$' would be a collection of types, or a *type of types* – 'worlds' –) we can see that the same reasoning made for implication and first-order quantification can be carried through for the case of modal logics. For example the axiom for the modal logic K finds a parallel in the axioms for implication and first-order universal quantification which correspond to the S (*distribution*) combinator:

Modal Logic K:
$\Box(\mathbf{A} \to \mathbf{B}) \to (\Box\mathbf{A} \to \Box\mathbf{B})$
which could be rewritten as:
$\forall \mathbf{X} \in \mathcal{W}.(\mathbf{X} \to (\mathbf{A} \to \mathbf{B})) \to (\forall \mathbf{X} \in \mathcal{W}.(\mathbf{X} \to \mathbf{A}) \to \forall \mathbf{X} \in \mathcal{W}.(\mathbf{X} \to \mathbf{B}))$

and the parallel with the S combinator is quite clear: $\mathbf{X}$ is being distributed over the implication. Similarly to the case of implication and first-order universal quantification one can find a parallel between axioms characterising different modal logics and the type schemes of combinators. For example, the weakest rule characterising the so-called 'standard normative logics according to Chellas (1980), namely:

$$\text{RM} \quad \frac{\mathbf{A} \rightarrow \mathbf{B}}{\Box\mathbf{A} \rightarrow \Box\mathbf{B}}$$

mirrors the type-scheme of the weakest combinator, namely B (i.e. *left transitivity*, in terms of implication), if one thinks of '□' as a combination of second-order universal quantification and implication.

Concerning the 'jump' to second-order quantification observe that unlike the previous cases (implication and first-order universal quantification) where we had only element variables ($x, y, \ldots$, etc.), now we have type variables (e.g. '**X**'). And indeed, in modal logics one is dealing with higher-order objects, and therefore some kind of 'higher-order modus ponens (and universal abstraction)' is needed.

Now, in order to characterise the 'second-order' normalisation one can make use of the seminal results independently obtained by Girard (1971) and Reynolds (1974) on second-order typed λ-calculus and polymorphism. In fact, an attempt at a formulation of the second-order normalisation in a type-theoretic framework has been made with the development of a '*type of types*'-Fix operator described in de Queiroz & Maibaum 1990.

References

Ajdukiewicz, K.: 1937, 'Die syntaktische Konnexität', in *Studia Philosophica*, vol. 1, pp. 1–27. (English translation in *Polish Logic 1920–1939*, Storrs McCall (ed.), Clarendon Press, Oxford, 1967, viii+406pp, pp. ???–???.)

Anderson, Alan R. & Nuel D. Belnap: 1975, *Entailment. The Logic of Relevance and Necessity*, Volume I, Princeton University Press, Princeton, xxxii+541pp.

Barendregt, Henk P.: 1981, *The Lambda Calculus. Its Syntax and Semantics*, (revised edition 1984), series Studies in Logic and The Foundations of Mathematics, vol. 103, North-Holland, Amsterdam, xv+621pp.

van Benthem, Johan: 1989, 'Categorial Grammar and Type Theory', in *Journal of Philosophical Logic*, vol. 18, pp. ???–???. (54pp.)[35]

— : 1990, 'Categorial grammar meets unification', (to appear in *Unification Formalisms: Syntax, Semantics and Implementation*, J. Wedekind et. al. (eds.); (obtainable from Institute for Language, Logic and Information, Dept. of Mathematics and Computer Science, Plantage Muidergracht 24, NL-1018 TV Amsterdam, The Netherlands), (final version, February 1990), 41pp.

Chellas, Brian F.: 1980, *Modal Logic. An Introduction*, Cambridge University Press, Cambridge, xii+295pp.

Church, Alonzo: 1941, *The Calculi of Lambda-Conversion*, (second printing 1951), series Annals of Mathematical Studies, N. 6, Princeton University Press, Princeton, ii+82pp.

Curry, Haskell B.: 1931, 'The Universal Quantifier in Combinatory Logic', in *Annals of Mathematics*, vol. 32, pp. 154–180.

— : 1934, 'Functionality in Combinatory Logic', in *Proceedings of the National Academy of Sciences of USA*, vol. 20, pp. 584–590.

— : 1963, *Foundations of Mathematical Logic*, (reprinted with corrections 1977), Dover Publications, New York, viii+408pp.

Curry, Haskell B. & Robert Feys: 1958, *Combinatory Logic*, Vol. I, series Studies in Logic and The Foundations of Mathematics, North-Holland, Amsterdam, XVI+417pp.

Curry, Haskell B., J. Roger Hindley & Jonathan P. Seldin: 1972, *Combinatory Logic*, Vol. II, series Studies in Logic and The Foundations of Mathematics, North-Holland, Amsterdam, XIV+520pp.

Došen, Kosta: 1988, 'Sequent systems and grupoid models', in *Studia Logica*, vol. 47, pp. ??–??.

— : 1989, 'Sequent systems and grupoid models II', in *Studia Logica*, vol. 48, pp. 41–65.

Gabbay, Dov M.: 1988, 'What is Negation in a System?', in Proceedings of *Logic Colloquium '86*, held in Kingston-upon-Hull, UK, F. R. Drake & J. K. Truss (eds.), series Studies in Logic and The Foundations of Mathematics, North-Holland, Amsterdam, pp. 95–112.

[35] Prof van Benthem has kindly sent us a copy of the first proof, which unfortunately does not contain the page numbers of the volume in which it will be published.

— : 1989, 'LDS – Labelled Deductive Systems', preprint, Department of Computing, Imperial College, London, SW7 2BZ, UK, First Draft September 1989, Current Draft January 1990, 95pp.

Girard, Jean-Yves: 1971, 'Une Extension de l'Interpretation de Gödel à l'Analyse, et son Application à l'Elimination des Coupures dans l'Analyse et la Theorie des Types', in *Proceedings of the Second Scandinavian Logic Symposium*, held in Oslo, June 18–20 1970, J. Fenstad (ed.), series Studies in Logic and The Foundations of Mathematics, vol. 63, North-Holland, Amsterdam, viii+405pp, pp. 63–92.

— : 1987, 'Linear Logic', in *Theoretical Computer Science*, vol. 50, pp. 1–102.

— : 1989, *Proofs and Types*, (Translated and with appendices by Paul Taylor and Yves Lafont), (Reprinted with minor corrections 1990), series Cambridge Tracts in Theoretical Computer Science (Managing Editor: Professor C. J. van Rijsbergen), Cambridge University Press, Cambridge, xii+175pp.

Gödel, Kurt: 1958, 'Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes', in *Dialectica*, vol. 12, pp. 280–287.(English translation in *Journal of Philosophical Logic*, vol. 9, 1980, pp. 133–142.)

Goodman, Nicolas D.: 1970, 'A theory of constructions equivalent to arithmetic', in *Intuitionism and Proof Theory*, Proceedings of the *Summer Conference* at Buffalo, New York, 1968, A. Kino, J. Myhill & R. E. Vesley (eds.), series Studies in Logic and The Foundations of Mathematics, North-Holland, vii+516pp, pp. 101–120.

Helman, Glen H.: 1977, *Restricted Lambda-abstraction and the Interpretation of Some Non-classical Logics*, Dissertation, University of Pittsburgh.

Heyting, Arend: 1946, 'On weakened quantification', in *Journal of Symbolic Logic*, vol. 11, pp. 119–121.

— : 1956, *Intuitionism. An Introduction*, series Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, viii+133pp.

Hindley, J. Roger & Jonathan P. Seldin: 1986, *Introduction to Combinators and λ-Calculus*, series London Mathematical Society Student Texts, (Managing Editor: Prof. E. B. Davies), vol. 1, Cambridge University Press, Cambridge, vi+360pp.

Howard, William A.: 1980, 'The formulae-as-types notion of construction', in *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, J. P. Seldin & J. R. Hindley (eds.), Academic Press, London, xxv+606pp, pp. 479–490.

Johansson, I.: 1936, 'Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus', in *Compositio Mathematica*, vol. 4, pp. 119–136.

Komori, Yuichi: 1983, 'The Variety Generated by BCC-Algebras Is Finitely Based', in *Reports of Faculty of Science, Shizuoka University*, Vol. 17, pp. 13–16.

— : 1989, 'Illative Combinatory Logic based on *BCK*-Logic', in *Mathematica Japonica*, n. 4, pp. 585–596.

— : 1990(?), 'BI Logic and its related logics', handwritten memo, 8pp.

Lambek, Joachim: 1958, 'The Mathematics of Sentence Structure', in *American Mathematical Monthly*, vol. 65, pp. 154–170.

— : 1980, 'From λ-calculus to cartesian closed categories', in *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, J. P. Seldin & J. R. Hindley (eds.), Academic Press, London, xxv+606pp, pp. 375–402.

— : 1989, 'Multicategories Revisited', in *Contemporary Mathematics*, Volume 92, pp. 217–239.

Lambek, Joachim & P. J. Scott: 1986, *Introduction to Higher Order Categorical Logic*, series Cambridge Studies in Advanced Mathematics, (Editorial Board: D. J. H. Garling, D. Gorenstein, T. tom Dieck, P. Walters), vol. 7, Cambridge University Press, Cambridge, x+293pp.

Läuchli, H.: 1965, 'Intuitionistic propositional calculus and definably non-empty terms', (abstract), in *Journal of Symbolic Logic*, vol. 30, p. 263.

— : 1970, 'An abstract notion of realizability for which intuitionistic predicate calculus is complete', in *Intuitionism and Proof Theory*, Proceedings of the *Summer Conference* at Buffalo, New York, 1968, A. Kino, J. Myhill & R. E. Vesley (eds.), series Studies in Logic and The Foundations of Mathematics, North-Holland, vii+516pp, pp. 227–234.

Martin-Löf, Per: 1972, 'Infinite terms and a system of natural deduction', in *Compositio Mathematica*, vol. 24, pp. 93–103.

— : 1975, 'An intuitionistic theory of types: predicative part', in Proceedings of the *Logic Colloquium '73*, held at Bristol, U. K., H. E. Rose & J. C. Shepherdson (eds.), series Studies in Logic and The Foundations of Mathematics, vol. 28, North-Holland, Amsterdam, pp. 73–118.

— : 1984, *Intuitionistic Type Theory*, (Notes by Giovanni Sambin of a series of lectures given in Padova, June 1980), series Studies in Proof Theory, Bibliopolis, Naples, iv+91pp.

— : 1985, 'On the meanings of the logical constants and the justifications of the logical laws', in *Atti degli incontri di logica matematica*, vol. 2, C. Bernardi & P. Pagli (eds.), series Scuola di Specializzazione in

Logica Matematica, Dipartimento di Matematica, Università di Siena, pp. 203–281.

Ono, Hiroakira: 1988, *Structural Rules and a Logical Hierarchy*, preprint, Faculty of Integrated Arts and Sciences, Hiroshima University, Higashisendamachi, Hiroshima 730, Japan. (To appear in the Proceedings of the *Heyting '88* Conference, held at Varna, Bulgaria.)

Ono, Hiroakira & Yuichi Komori: 1985, 'Logics without the contraction rule', in *Journal of Symbolic Logic*, vol. 50, pp. 169–201.

de Queiroz, Ruy J. G. B.: 1989, 'The Mathematical Language and its Semantics: to Show the Consequences of a Proposition is to Give its Meaning', in *Reports of the Thirteenth International Wittgenstein Symposium 1988*, (Edited by Paul Weingartner & Gerhard Schurz), series Schriftenreihe der Wittgenstein-Gesellschaft, Vol. 18, (Edited by Elisabeth Leinfellner, Rudolf Haller, Adolf Hübner, Werner Leinfellner & Paul Weingartner), Hölder-Pichler-Tempsky, Vienna, 304pp, pp. 259–266.

de Queiroz, Ruy J. G. B. & Thomas S. E. Maibaum: 1990, 'Proof Theory and Computer Programming' (to appear in vol. 36 of the *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*.)

— : 1991, 'Abstract Data Types and Type Theory: Theories as Types' (to appear in vol. 37 of the *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*.)

Reynolds, John C.: 1974, 'Towards a theory of type structure', in *Paris Colloque sur la Programmation*, series Lecture Notes in Computer Science, vol. 19, pp. 408–425.

Schönfinkel, Moses: 1924, 'Über die Bausteine der mathematischen Logik', in *Mathematische Annalen*, vol. 92, pp. 305–316. (English translation 'On the building blocks of mathematical logic' in J. van Heijenoort's *From Frege to Gödel*, pp. 355–366.)

Scott, Dana S.: 1970, 'Constructive Validity', in the Proceedings of the *Symposium on Automated Deduction*, held in Versailles, December 1968, (Edited by M. Laudet, D. Lacombe, L. Nolin & M. Schützenberger), series Lecture Notes in Mathematics, vol. 125, Springer-Verlag, v+310pp, pp. 237–275.

Tait, William W.: 1965, 'Infinitely long terms of transfinite type', in *Formal Systems and Recursive Functions*, J. N. Crossley & M. A. E. Dummett (eds.), series Studies in Logic and The Foundations of Mathematics, North-Holland, Amsterdam, pp. 176–185.

— : 1967, 'Intensional interpretations of functionals of finite type I', in *Journal of Symbolic Logic*, vol. 32, pp. 198–212.

Dov M. Gabbay and Ruy J. G. B. de Queiroz
Department of Computing
Imperial College of Science, Technology and Medicine
University of London
180 Queen's Gate, London SW7 2BZ
UK

First Draft February 2, 1990.
Second Draft February 14, 1990.
Third Draft April 25, 1990.
Fourth Draft May 2, 1990.

Polymorphic Constructs in Natural and Programming Languages (Preliminary Version)

Hans Leiß
Siemens AG, ZFE IS INF 2
Otto-Hahn-Ring 6, D-8000 Munich 83
email: leiss%ztivax@unido.uucp

April 18, 1990

Abstract

Problems in parsing conjoined phrases with conjuncts of different syntactic type can apparently be overcome with techniques developed for type synthesis in λ-calculus and programming languages. We briefly review the notions of parameter-, subtype- and let-polymorphism developed in functional programming, and then exhibit an analogy to parsing of conjoined phrases in natural languages.

In particular, we show that a full exploitation of let-polymorphism allows to treat the parsing of conjoined phrases by means of subsumption constraints, while sticking to unification based parsing techniques. This refines S.Shieber's [12] proposal to use subsumption constraints in describing 'polymorphic conjunction'. However, our refinement shows that there is a way of parsing polymorphic conjunctions that avoids to raise instances of the semi-unification problem, which has recently been proved undecidable.

Thus, while type synthesis for 'polymorphic recursive definitions' in programming is reducible to semi-unification and hence undecidable, parsing of polymorphic conjunctions is less complex (at least for the forms studied here). In contrast to suggestions of B.Rounds and J.Dörre[1], polymorphic conjunctions do not lead to semi-unification problems, and I am not aware of other linguistic phenomena that do.

1 Introduction

Simple examples - mostly taken from S.Shieber's Thesis[12] and due to I.Sag - suggest that conjunction in natural language can be used to conjoin phrases of the same syntactic category, to yield another phrase of this category:

$$\text{Pat is } ((\text{stupid})^{AP} \text{ and } (\text{healthy})^{AP})^{AP}. : S \qquad (1)$$

$$\text{Pat hired ((a Republican)}^{NP} \text{ and (a banker)}^{NP})^{NP}. : S \qquad (2)$$

If the conjuncts are of different syntactic types, the conjunction is ungrammatical:

$$\text{Pat hired (a Republican)}^{NP} \text{ and (proud of it)}^{AP}. : * \qquad (3)$$

$$\text{Pat hired (a Republican)}^{NP} \text{ and (at the office)}^{PP}. : * \qquad (4)$$

This may lead us to assume that *and* is of type

$$\text{and} : \alpha \times \alpha \rightarrow \alpha, \qquad \alpha = NP, AP, ... \qquad (5)$$

That is, *and* is *parameter polymorphic* in the sense that its type is a scheme with a free type parameter α - ranging over all or some specified set of types. But with some verbs, the conjuncts may be of *different* types:

$$\text{Pat has become (a banker)}^{NP} \text{ and (very conservative)}^{AP}. : S \qquad (6)$$

$$\text{Pat is (a Republican)}^{NP} \text{ and (proud of it)}^{AP}. : S \qquad (7)$$

$$\text{Pat is (healthy)}^{AP} \text{ and (of sound mind)}^{PP}. : S \qquad (8)$$

$$\text{That was (a rude remark)}^{NP} \text{ and (in very bad taste)}^{PP}. : S \qquad (9)$$

These examples pose two problems:

Problem 1 *When is a conjoined phrase with conjuncts of different type grammatical?*

Problem 2 *What is the type of the conjoined phrase, when the conjuncts are of different types?*

Some of the examples indicate that $((x)^{NP}$ and $(y)^{AP})$ or $((x)^{NP}$ and $(y)^{PP})$ are grammatical, while others indicate that they are not. The same applies to similar constructs like *or*.

S.Shieber proposed a solution to these questions that essentially assigns a more general type to conjunction, namely

$$\text{and} : \alpha_1 \times \alpha_2 \rightarrow \alpha, \qquad \alpha_1 \sqsubseteq \alpha,\ \alpha_2 \sqsubseteq \alpha, \qquad \alpha = NP, AP, ... \qquad (10)$$

where the side condition $\alpha_i \sqsubseteq \alpha$ poses a constraint on the types that may be substituted for the variables: $\tau \sqsubseteq \sigma$ means that type τ is *subsumed by* type σ. The subsumption relation between types should be effectively testable; hence we assume that it is given by some partial order between type *expressions*, inductively defined along the syntax of type expressions. Shieber alludes to one such relation $\tau \sqsubseteq \sigma$ - which he writes as $\sigma \leq \tau$ -, where τ is a substitution instance of σ.

A more accurate description of Shieber's solution is given in Section 2. We will point out some drawbacks of Shieber's solution, and in particular will discuss whether - in parsing conjoined phrases - it is necessary that we have to solve sets of inequations $\tau \sqsubseteq \sigma$ between type expressions.

The problem of deciding whether, given a set S of inequations, there is a solution or not, is called the ***semi-unification problem.***

Motivated by Shieber's proposal, W.Rounds and J.Dörre[1] have studied this problem for ***feature terms*** and shown that it is undecidable. The same problem for ***first order terms*** (or simple types) arose in type synthesis for functional programming languages [3, 7, 5], in proof theory [10] and in term rewriting [4]. A proof of undecidability for this case has been given by Kfoury e.a.[6]. Therefore, it seems worthwhile to have a careful study on whether Shieber's proposal can be modified so as to avoid raising instances of the undecidable semi-unification problem.

In Section 3, we will look at the corresponding work in type synthesis for programs, and introduce various notions of polymorphism studied in this context: parameter polymorphism, LET-polymorphism, and subtype polymorphism.

Using the basic idea behind LET-polymorphism, in Section 4 we introduce an improvement of Shieber's proposed analogy between AND- and LET-polymorphism. Section 5 demonstrates that by using the same ideas as in Milner's[8] typing rule for LET, a parsing strategy is possible that is based on a ***principal type property.*** This strategy avoids to raise instances of the semi-unification problem, and gives decidability of typability for a new typing rule for polymorphic conjunction.

We only deal with polymorphic conjunctions ***in object position***, and have to leave the case of ***subject position*** to further studies. Thus the main point of the paper is the refined correspondence between LET- and AND-polymorphism, and the demonstration that it is possible to add subsumption constraints in specifying (some aspects of) a grammar, while sticking to unification-based parsing.

Acknowledgement: I wish to thank Bill Rounds for making me aware of a possible connection between semi-unification and Shieber's subsumption constraints. Many thanks also to Fritz Henglein for discussions and e-mails that kept my interest in semi-unification alive.

2 Shieber's Solution

In his Thesis, S.Shieber[12] proposed to exploit subsumption constraints in specifying natural language grammars. The particular phenomenon mentioned in this connection was the ***polymorphic AND*** as presented by the examples in the Introduction. (Shieber's proposal was based on the LET-construct in the functional programming language ML, which will be explained in Section 3.) To describe Shieber's proposal, I will replace the notation of unification grammar by the more perspicious notation of type synthesis in programming. For simplicity, I also will replace feature terms by simple type expressions. Let ρ, σ, τ, etc. range over ***type expressions***, where

- Each type variable and each type constant is a type expression, and
- If G is a n-ary type constructor and $\sigma_1, \ldots, \sigma_n$ are type expressions, so is $G(\sigma_1, \ldots, \sigma_n)$.

We use α, β etc. as type variables. A ***typing statement*** is an expression $e : \sigma$, where e is an expression of our formal (or natural) language L as specified by some grammar. A ***type environment*** (or lexicon) is a set Σ of typing statements for variables or constants of L.

We assume that a calculus is given by means of which one can derive *judgements* $\Sigma \vdash e : \sigma$, asserting that under the assumptions Σ, expression e is of type σ. For natural language grammars, the relation $\Sigma \vdash e : \sigma$ (or rather the calculus for deriving such judgements) specifies when phrase e is of (syntactic) type σ relative to lexicon Σ. Algorithms that synthesize types to given programs then correspond to parsing algorithms in natural language processing. Hence the slogan 'parsing as type inference'.

The three ingredients to Shieber's proposal for analysing 'polymorphic conjunctions' are the following (S 1), (S 2) and (S 3). Points (S 1) and (S 2) address Problem 2, while (S 3) is concerned with Problem 1.

(S 1) Use subsumption constraints $\sigma \sqsubseteq \tau$, and the subtype axiom

$$\text{(sub)} \qquad \Sigma \cup \{e : \sigma\} \vdash e : \tau, \qquad \text{if } \sigma \sqsubseteq \tau.$$

(S 2) Use the following typing rule for polymorphic conjunction:

$$\text{(and)} \qquad \frac{\Sigma \vdash e_1 : \sigma_1, \quad \Sigma \vdash e_2 : \sigma_2}{\Sigma \vdash (e_1 \text{ and } e_2) : \sigma}, \qquad \text{if } \sigma_1 \sqsubseteq \sigma, \sigma_2 \sqsubseteq \sigma.$$

(S 3) Grammaticality of phrases $f(a \text{ and } b)$ depends on whether the argument type of the (unary) verb (-phrase) f is the same as the type of the conjunction, i.e. use

$$\text{(app)} \qquad \frac{\Sigma \vdash f : \sigma \rightarrow \tau, \quad \Sigma \vdash (e_1 \text{ and } e_2) : \sigma}{\Sigma \vdash f\,(e_1 \text{ and } e_2) : \tau}.$$

(S 1) and (S 2) have to be made more precise by specifying a notion of 'subtype', which may depend on the set Σ of typing assumptions. Two syntactical notions of subtyping are relevant: subtyping by instantiation of type variables (that are generic with respect to Σ), and subtyping of records by adding additional fields. (See Section 3)

(S 2) simply says that the types of the conjuncts have to be subsumed by the type of the conjunction. Thus the types of the conjuncts may be inconsistent with each other, but have to be consistent with the type of the conjunction.

(S 3), which is just the traditional application rule from typed λ-calculus, is less explicit in Shieber[12]. It tries to account for the context in explaining when conjunctions are grammatical, by noting that different verbs are more or less selective with respect to their arguments' types. This is somewhat of a restriction to conjunctions in object position, which we will follow below.

Example 1 (Shieber) Let the following types (motivated by Chomsky's 'X-bar theory') be given:

$$\begin{array}{ll} NP = [n = +, v = -, bar = 2], & AP = [n = +, v = +, bar = 2], \\ VP = [n = -, v = +, bar = 2], & PP = [n = -, v = -, bar = 2], \\ \sigma = [bar = 2], & \tau = [n = +, bar = 2]. \end{array} \tag{11}$$

Then by record-subtyping we have: NP and AP are subtypes of τ, and VP, PP, τ are subtypes of σ, but VP and PP are *not* subtypes of τ. Given an appropriate lexicon Σ, by (sub) and (and) we can derive

$$\Sigma \vdash (\text{healthy})^{\text{AP}} \text{and (of sound mind)}^{\text{PP}} : \sigma, \tag{12}$$

as AP $\sqsubseteq \sigma$ and PP $\sqsubseteq \sigma$. However, since PP $\sqsubseteq \tau$ is not the case, (and) does not allow to derive

$$\Sigma \vdash (\text{healthy})^{\text{AP}} \text{and (of sound mind)}^{\text{PP}} : \tau. \tag{13}$$

Suppose that in addition to the above, we can derive $\Sigma \vdash$ Pat is : $\sigma \rightarrow S$, $\Sigma \vdash$ Pat became : $\tau \rightarrow S$, and $\Sigma \vdash$ Pat hired : $NP \rightarrow S$. Then by (app) we can also derive

$$\Sigma \vdash (\text{Pat is})^{\sigma \rightarrow \text{S}} \text{ (healthy and of sound mind)}^{\sigma} : \text{S}, \tag{14}$$

$$\Sigma \vdash (\text{Pat became})^{\tau \rightarrow \text{S}} \text{ (a banker and very conservative)}^{\tau} : \text{S}, \tag{15}$$

$$\Sigma \vdash (\text{Pat hired})^{\text{NP} \rightarrow \text{S}} \text{ (a banker and a secretary)}^{\text{NP}} : \text{S}, \tag{16}$$

but not

$$\Sigma \vdash (\text{Pat became})^{\tau \rightarrow \text{S}} \text{ (healthy and of sound mind)}^{\sigma} : *, \quad \text{or} \tag{17}$$

$$\Sigma \vdash (\text{Pat hired})^{\text{NP} \rightarrow \text{S}} \text{ (a banker and very conservative)}^{\tau} : *. \tag{18}$$

The main objections to Shieber's solution are that his rule (and) leads to unnecessary and semantically dubious syntactic types - like 'NP and AP' -, and that it apparently does not even solve the parsing problem it was designed for.

Problem 3 *How can we* parse *polymorphic conjunctions, i.e. determine their types?*

Obviously, we first have to derive types for the constituents e_1 and e_2 of (e_1 and e_2), with results $e_i : \sigma_i$ say, and then find some type σ such that the applicability conditions $\sigma_i \sqsubseteq \sigma$ hold. But it is unclear

- whether we can decide if there is such σ,
- whether we can decide if there is a unique one, or
- how we can choose one in case there are many.

Of course there may be different answers for different subsumption relations $\sqsubseteq$. It does not seem obvious to me that the types that occur in linguistics form a nice lattice structure which would guarantee positive answers to such questions. In fact, for one such relation, Dörre and Rounds[1] have shown that solvability of sets of inequations $\tau \sqsubseteq \sigma$ over *feature algebras* is undecidable.

Also, it may well be the case that the subsumption relation depends on the set Σ of typing assumptions, and thus be not a global relation; it would have to be adjusted to changes of Σ in connection with the treatment of bound variables (as in the typing rule for λ-abstraction, which of course occurs only implicitly in natural language).

3 Notions of Polymorphism in Programming

In this Section we sketch three notions of polymorphism in programming languages, to provide the background for a new look at polymorphic conjunction in Sections 4 and 5.

a) Parametric polymorphism

The programmer can define functions f which have *schematic* types $\Sigma \vdash f : \sigma(\alpha) \to \tau(\alpha)$, α not in Σ, and the parameters α can be instantiated to different types in different uses of f. In fact, the programmer does not *declare* f to have a parametric polymorphic type, but rather, a most general type scheme is automatically inferred (synthesized) from the defining term for f, if the term is typable at all. This kind of typing discipline allows to take advantage of polymorphic functions f *in subject (predicate) position*, i.e. $(f[0,1,2,3], f[\text{"}ab\text{"}, \text{"}cde\text{"}])$ is a typable pair-expression, for $f = \text{reverse} : \alpha\text{-list} \to \alpha\text{-list}$, say.

b) LET-polymorphism

As the familiar typing rule for λ-abstraction demands that all occurrences of the abstracted variable in the defining term have the same type, a functional like

$$F := \lambda f.(f[0,1,2,3], f[\text{"}ab\text{"}, \text{"}cde\text{"}])$$

that takes polymorphic functions as arguments, cannot be defined in languages like ML[2]. Hence, useful expressions like $(F \cdot \text{reverse})$ cannot be typed either, due to the

Subterm Property: Any subterm of a typable term is typable,

which holds for the usual notions of typing for λ-terms. To take advantage of polymorphic functions *in object position*, as s in $(F \cdot s)$, ML introduced a new syntactic construct, LET, together with a new reduction rule

$$(let\ f = s\ in\ t) \to_{let} t[s/f] \tag{19}$$

and a new typing rule

$$\text{(let)} \qquad \frac{\Sigma \vdash f : \sigma, \quad \Sigma \cup \{f : \sigma_1, \ldots, f : \sigma_n\} \vdash t : \tau}{\Sigma \vdash\ (let\ f = s\ in\ t)\ :\ \tau}, \qquad \text{if } \sigma \leq_\Sigma \sigma_i.$$

In this rule, Σ must not contain a typing statement for f, and $\sigma \leq_\Sigma \sigma_i$ means that σ_i is obtained from σ by instantiating type variables not occurring in Σ, i.e. those that sometimes are called *generic with respect to* Σ.

Although $t[s/f]$, $(\lambda f.t) \cdot s$, and $(let\ f = s\ in\ t)$ all have the same meaning, note that

- $(let\ f = s\ in\ t)$ improves on $(\lambda f.t) \cdot s$ as it avoids the untypable subterm $\lambda f.t$, and
- $(let\ f = s\ in\ t)$ improves on $t[s/f]$ as it is syntactically more abstract, i.e. abstracts the various occurrences of s in $t[s/f]$ into *one* occurrence of s in $(let\ f = s\ in\ t)$, replacing the others by a bound variable f.

In particular, the only occurrence of s in (*let* $f = s$ *in* t) is in ***object position***, and although (*let* $f = s$ *in* t) is of the same type as $t[s/f]$, syntax analysis - including type inference - is easier (i.e. more efficient) with (*let* $f = s$ *in* t) than with $t[s/f]$.

c) **Subtype polymorphism**
This is relevant only when entering details of feature terms, which can be represented as recursive records (as opposed to simple types = trees). We do not go into details in this Preliminary Version, but only mention that we obtain a ***sub***record by ***adding*** additional components to a record. (This has already been used in Example 1.)

4 An Improved Analogy Between LET and AND

We now summarize the main features of a modified treatment of polymorphic conjunctions. Examples and motivation of the typing rule will be given in Section 5, as well as proofs of the technical claims concerning parsability.

By treating the polymorphic AND as a syntactic abstraction, in much the same way as LET is used as a syntactic abstraction, we can exploit the lessons from LET-polymorphism.

- The AND in object position can be understood as an abbreviating syntactic construct, whose meaning is captured by the implicit reduction rule

$$f(a \text{ and } b) \rightarrow_{\text{and}} ((fa) \text{ and } (fb)). \tag{20}$$

- The type of $f(a$ and $b)$ should be the type of $((fa)$ and $(fb))$, and Shieber's typing rule for conjunction should be replaced by

$$\text{(And)} \quad \frac{\Sigma \vdash f : \sigma \rightarrow \tau, \quad \Sigma \vdash e_1 : \sigma_1, \quad \Sigma \vdash e_2 : \sigma_2}{\Sigma \vdash f\ (e_1 \text{ and } e_2) : \rho}, \quad \begin{array}{l} \text{if } \sigma_1 \rightarrow \rho \sqsubseteq \sigma \rightarrow \tau, \\ \text{and } \sigma_2 \rightarrow \rho \sqsubseteq \sigma \rightarrow \tau. \end{array}$$

The rule (And) has the following advantages over Shieber's rule (and):

- We avoid to type the 'subexpression' (e_1 and e_2). In particular, there is no need to establish a semantics for (NP and PP)-phrases (etc., cf. the examples in the Introduction), which would lead to dubious semantic entities.
- The fact that the verb f strongly restricts the possible grammaticality of $f(a$ and $b)$, is built into the typing rule. We can thus allow for fine-grained context-dependency, whereas Shieber's rule allows to type (a and b) independent of its context.
- In typing (fe_1 and fe_2) we implicitly use ***parametric*** polymorphic conjunction only, which is more restrictive (hence less 'overgenerating') than Shieber's ***subtype*** polymorphic conjunction.

- The parsing problem of Shieber's rule vanishes, as - exactly as for LET - we *first* derive the type scheme $\sigma \rightarrow \tau$ of f, and *then* can find the appropriate instances $\sigma_i \rightarrow \rho$ for typing the conjuncts fe_i, simply by unifying into different copies of the scheme $\sigma \rightarrow \tau$ with 'fresh' generic variables.[1]

These advantages, and the typing rule, follow from the underlying analogy

$$f(a \text{ and } b) \quad \approx \quad (\mathit{let}\ x = f\ \mathit{in}\ ((xa)\ \mathit{and}\ (xb))), \tag{21}$$

viewing the left hand side as the 'natural language' (variable free) version of the right hand side, together with the restriction to parametric polymorphic conjunction.

Contrary to what seems to be suggested by Dörre/Rounds[1], by the rule given above we are not forced to solve instances of the (in general undecidable) *semi*-unification problem, when looking whether the applicability condition of the rule is satisfiable. As indicated, this can be done by unification, just as in the LET-rule.

However, parsing with 'reflexive' variants of polymorphic conjunction, which would be defined by a rule like

$$\text{(and-refl)} \qquad \frac{\Sigma \cup \{e_1 : \sigma_1, e_2 : \sigma_2\} \vdash (e_1 \text{ and } e_2) : \sigma}{\Sigma \vdash (e_1 \text{ and } e_2) \ : \sigma}, \qquad \text{if } \sigma_1 \sqsubseteq \sigma,\ \sigma_2 \sqsubseteq \sigma.$$

might indeed lead to the semi-unification problem, as does the typing rule for 'polymorphic recursion' of [7] and, similarly, [9, 3, 5]:

$$\text{(rec)} \qquad \frac{\Sigma \cup \{f : \sigma_1, \ldots, f : \sigma_n\} \vdash s : \sigma}{\Sigma \vdash \mathit{rec}\ f.s \ : \sigma}, \qquad \text{if } \sigma \leq_\Sigma \sigma_1, \ldots, \sigma \leq_\Sigma \sigma_n.$$

I do not know of linguistic phenomena that would need the 'reflexive polymorphic conjunction' kind of rule.

5 Type Inference for Polymorphic Conjunction

Having set up our typing rule (And) for polymorphic conjunction, two questions should be settled for a type inference calculus involving this rule:

- Is typability of expressions decidable? In other words, is there an algorithm that, given a set Σ of typing assumptions and a term e, tells us whether there is a type σ such that $\Sigma \vdash e : \sigma$ is provable?
- Does the principal type property hold? That is, does every typable term have a most general (or *principal*) type, such that all other types are instances thereof?

[1] This is true at least if we restrict ourselves to parametric polymorphism (cf. Section 5), i.e. use $\sigma \leq_\Sigma \sigma_i$ as subsumption relation $\sigma_i \sqsubseteq \sigma$. It is not quite clear whether it holds when parametric and subtype polymorphism are combined, which seems necessary if we take feature terms as type expressions.

To answer these questions, we will assume that a calculus is given that satisfies

(D 1) If $\Sigma \vdash e : \sigma$ is provable, and R is a substitution of types for type variables, then $R(\Sigma) \vdash e : R(\sigma)$ is provable, too.

(D 2) For the calculus without rule (And), decidability of typability and the principal type property hold. More precisely, we assume that there is an algorithm W such that:

1. $W(\Sigma, e)$ terminates for every finite set Σ of typing assumptions and every term e, with one of the following two results.
2. If $W(\Sigma, e) = (S, \sigma)$, then $S(\Sigma) \vdash e : \sigma$ is provable and any other typing of e under some instance of Σ is less general.
3. If $W(\Sigma, e) = \text{fail}$, then e is not typable under any specialisation $S(\Sigma)$ of Σ.

In particular, we want to see whether (And) can be added to Milner's[8] type inference algorithm W for (core-) ML, without loosing these properties.[2] In this Section, we also restrict our notion of polymorphism to parametric polymorphism. For a discussion on various combinations of parametric with sub(record)type polymorphism, see [11].

5.1 The Semi-Unification Problem Raised by the Naive Typing Procedure for Polymorphic AND

Suppose W is an algorithm with the above properties, which we want to extend to our rule (And). Let a set of typing assumptions Σ and an expression $e := f(e_1 \text{ and } e_2)$ be given. To see whether e can be typed, and find its most general typing in case it can, we would like to proceed as follows:

1. Use $W(\Sigma, f)$ to type the verb phrase (or function) f. If this does not fail and $(S_0, \sigma \rightarrow \tau)$ is the result, then $S_0(\Sigma) \vdash f : \sigma \rightarrow \tau$ is provable, and is the most general typing for f.

2. Use $W(S_0(\Sigma), e_1)$ to type e_1, and assume that this succeeds with result (S_1, σ_1), so that $S_1 S_0(\Sigma) \vdash e_1 : \sigma_1$ is the most general typing for e_1 we are looking for.

3. Proceed similarly with e_2, and let $S_2 S_1 S_0(\Sigma) \vdash e_2 : \sigma_2$ be the resulting most general typing for e_2.

4. To obtain the same set of typing assumptions in all three derivations, applying substitution $S_2 S_1$ and S_2 to the first and second proof, respectively. We now have derived most general typings $S_2 S_1 S_0(\Sigma) \vdash f : S_2 S_1(\sigma \rightarrow \tau)$, $S_2 S_1 S_0(\Sigma) \vdash e_1 : S_2(\sigma_1)$, and $S_2 S_1 S_0(\Sigma) \vdash e_2 : \sigma_2$ fitting to the top line of rule (And). For simplicity of notation, we assume from now on that $\Sigma \vdash f : \sigma \rightarrow \tau$, $\Sigma \vdash e_1 : \sigma_1$, and $\Sigma \vdash e_2 : \sigma_2$ are most general typings.

[2] Expressions like *reverse* ([0,1,2,3] *and* ["ab","cd","ef"]) might be useful in programming (similar to 'map'-constructs), perhaps with type *int*-list × *string*-list, which would need a modification of (And).

5. Choose a good candidate for the result type ρ, for example τ. (We will see later how to find a most general ρ as the common type of fe_1 and fe_2.)

6. If the side conditions $\sigma_i \to \rho \sqsubseteq \sigma \to \tau$ of (And) are satisfied, we can apply this rule and obtain a derivation of $\Sigma \vdash f(e_1 \text{ and } e_2) : \rho$, which is most general (for optimal choice of ρ).

7. However, if the side conditions are not satisfied we cannot yet conclude that there is no typing for $f(e_1 \text{ and } e_2)$, but have to determine ***whether there is a specialization of the given proof whose types do satisfy the side condition.*** A most general such specialization would provide the principal typing for $f(e_1 \text{ and } e_2)$ we were looking for.

Now, assume the subsumption relation $\tau_2 \sqsubseteq \tau_1$ is defined via instantiation of variables generic with respect to Σ, i.e. as $\tau_1 \leq_\Sigma \tau_2$. In order to find a specialization satisfying the side conditions, we have to solve the following problem. We use a fresh type variable α instead of ρ, in order not to put an unnecessary constraint on the solution:

Problem 4 *Given Σ, $\sigma \to \tau$, $\sigma_1 \to \alpha$ and $\sigma_2 \to \alpha$, is there a substitution R such that*

$$R(\sigma \to \tau) \leq_{R(\Sigma)} R(\sigma_i \to \alpha)$$

for $i = 1, 2$? And if so, what is the most general one, if that exists?

It is not hard to see that this problem is equivalent to the following one:

Problem 5 *Given Σ, $\sigma \to \tau$, σ_1 and σ_2, and a fresh type variable α, are there substitutions R, S_1 and S_2 such that, for $i = 1, 2$, we have $S_i R(\sigma \to \tau) = R(\sigma_i \to \alpha)$, and $S_i R(\beta) = R(\beta)$ for each type variable β free in Σ ?*

This, however, is a special case of the following *semi-unification problem*, which has recently been shown to be undecidable[6]:

Problem 6 *Given a some inequations $\rho_i \leq_i \sigma_i$, $i = 1, \ldots, n$, and some equations $\tau_j = \mu_j$, $j = 1, \ldots, k$, between first-order terms, do there exist substitutions $R, S_1, \ldots S_n$ such that $S_i R(\rho_i) = R(\sigma_i)$ and $R(\tau_j) = R(\mu_j)$ for all $i \leq n$ and $j \leq k$?*

The same problem occurred in trying to develop a type inference algorithm for the 'polymorphic recursion' rule (rec) [3, 7, 5], and showed that typability is undecidable in the presence of (rec) [6] - although the principal type property still holds [9].

Thus, in order to get a notion of polymorphic conjunction that admits decidability of typability (and principal type property), we have to avoid raising instances of the semi-unification problem.

5.2 How to Avoid the Semi-Unification Problem in Deciding Typability with Polymorphic LET

We recall that the rule (let) has similar side condition as (And) and (rec) have, but typability in the presence of (let) *is* decidable. Hence the subsumption conditions by themselves cannot cause undecidability.

So how can one derive most general typings for $(let\ x = f\ in\ e)$ without raising instances of the semi-unification problem to satisfy the side conditions of (let)? We can assume that Σ does not contain a typing statement for x and proceed as follows:

1. Suppose $W(\Sigma, f) = (S, \sigma)$, so $S(\Sigma) \vdash f : \sigma$ is the most general typing for f.
2. For the i-th occurrence of x in e, let σ_i be a copy of σ, where the generic variables are replaced by fresh ones. If $W(S(\Sigma) \cup \{x : \sigma_1, \ldots, x : \sigma_n\}, e) = (R, \tau)$, then the most general typing of e is $RS(\Sigma) \cup \{x : R(\sigma_1), \ldots, x : R(\sigma_n)\} \vdash e : \tau$.
3. Adjust the set of assumptions in the first proof, obtaining $RS(\Sigma) \vdash f : R(\sigma)$ as most general typing for f.
4. It can be shown that $R(\sigma) \leq_{RS(\Sigma)} R(\sigma_i)$ holds for each i, so we can apply (let) to obtain a proof of $RS(\Sigma) \vdash (let\ x = f\ in\ e) : \tau$.

It is not hard to see that $W(\Sigma, (let\ x = f\ in\ e)) = (RS, \tau)$ indeed gives the most general typing.

The essential difference between (let) and (rec) is that in typing $(let\ x = f\ in\ e)$, we can *first* derive the most general typing σ for f, *then* use copies of σ with fresh generic variables as (constraints on the) assumed types for x in typing e. No such trick is possible in deciding typability of a polymorphic recursive definition according to (rec), as we cannot generate a "pattern" that would constrain the assumed types for f in typing the defining term s of $rec\ f.s$. Similar problems arise with the rules (and) or (and-refl).

5.3 Deciding Typability with Polymorphic AND

We now show that typability with respect to our rule (And) for polymorphic conjunction is decidable, and principal types exist. We adopt the method used for LET, and avoid the semi-unification problems of the naive approach of Section 5.1.

The following tentative derivation is the motivation behind our rule (And), where we assume that Σ does not contain a typing statement for x, and of course x does not occur in e_1 and e_2.

$$\dfrac{\dfrac{\Sigma \vdash f : \sigma \to \tau, \qquad \dfrac{\dfrac{\Sigma \vdash e_i : \sigma_i}{\Sigma \cup \{x : \sigma_i \to \rho\} \vdash xe_i : \rho}\ \text{for } i = 1,2}{\Sigma \cup \{x : \sigma_1 \to \rho,\ x : \sigma_2 \to \rho\} \vdash (xe_1 \text{ and } xe_2) : \rho}\ (*)}{\Sigma \vdash (let\ x = f\ in\ (xe_1\ and\ xe_2)) : \rho}\ (**)}{\Sigma \vdash f(e_1 \text{ and } e_2) : \rho}\ (***)$$

To turn this into a correct derivation, using the familiar rules for assumptions and function application, we

- need and : $\rho \times \rho \to \rho$ in (*), which follows if we use parameter polymorphic conjunction, i.e. assume that for all α, 'and' has type and : $\alpha \times \alpha \to \alpha$,
- need $\sigma \to \tau \leq_\Sigma \sigma_i \to \rho$ in (**), in order to satisfy the applicability condition of (let), and
- use our motivating analogy $f(a \text{ and } b) \approx (let\ x = f\ in\ (xa\ and\ xb))$, as a definition of polymorphic conjunction, to justify (***).

Thus, if we assume the first and last of these, we can turn (And) into a derived rule.

Theorem 1 *With respect to rule (And), typability is decidable and the Principal Type Property holds.*

Proof: (Sketch) By the ideas sketched in discussing rule (let), we may assume that we already have most general typings

$$\Sigma \vdash f : \sigma \to \tau \quad \text{and} \quad \Sigma \cup \{x : \sigma_i \to \tau_i\} \vdash xe_i : \tau_i,$$

such that $\sigma \to \tau \leq_\Sigma \sigma_i \to \tau_i$ for $i = 1, 2$. However, we want the conjuncts fe_i (but not necessarily the e_i!) be of the same type. So, if they are not, we have to find a specialization of the derivations where they are. If τ_1 and τ_2 are not unifiable, this is impossible. So suppose R is the most general unifier of τ_1 and τ_2. Refine the derivations by applying R, to obtain derivations of

$$R(\Sigma) \vdash f : R(\sigma \to \tau) \quad \text{and} \quad R(\Sigma) \cup \{x : R(\sigma_i) \to \rho\} \vdash xe_i : \rho,$$

where ρ is $R(\tau_1) = R(\tau_2)$. The only subtle point now is to see that the applicability conditions $R(\sigma \to \tau) \leq_{R(\Sigma)} R(\sigma_i) \to \rho$ of (And) are satisfied. This follows from the fact that, as we can assume that none of the generic variables of $\sigma \to \tau$ occurs in τ_i, R does not operate on these variables, nor do they occur in a type substituted in by R. Hence from $\sigma \to \tau \leq_\Sigma \sigma_i \to \tau_i$ we obtain $R(\sigma \to \tau) \leq_{R(\Sigma)} R(\sigma_i) \to \rho$ and can apply (And) to get a derivation of

$$R(\Sigma) \vdash (let\ x = f\ in\ (xe_1\ and\ xe_2)) : \rho,$$

or $R(\Sigma) \vdash f(e_1 \text{ and } e_2) : \rho$, respectively. By the construction, it also seems clear that this is the most general typing. □

We now have described how W has can be extended to decide typability and derive most general typings for polymorphic conjunctions with respect to (And). Condition (D 1) remains true, if stated with some technical restriction concerning the generic variables (cf. a similar Lemma in [7]), and so W can be used for 'and's that are not at the 'top-level', too, as would be expected.

Here is a very simple example of an application of (And) in natural language. (There is no specializing of the result type τ to ρ.)

Example 2 Let NP[num = α] be a type expression for noun phrases, with number feature α ranging over {singular, plural}, and S be the category of sentences. In derivations we can use

$$\frac{\begin{array}{c}\Sigma \vdash \text{Pat hired} : \text{NP[num} = \alpha] \rightarrow \text{S}, \\ \Sigma \vdash \text{a banker} : \text{NP[num = singular]}, \\ \Sigma \vdash \text{two secretaries} : \text{NP[num = plural]}\end{array}}{\Sigma \vdash \text{Pat hired (a banker and two secretaries)} : \text{S}}, \tag{22}$$

since the side conditions

$$\begin{array}{c}\text{NP[num} = \alpha] \rightarrow \text{S} \ \leq_\Sigma \ \text{NP[num = singular]} \rightarrow \text{S}, \\ \text{NP[num} = \alpha] \rightarrow \text{S} \ \leq_\Sigma \ \text{NP[num = plural]} \rightarrow \text{S},\end{array} \tag{23}$$

of (And) are satisfied, assuming α does not occur in Σ.

In contrast, taking this subsumption relation $\leq_\Sigma$, Shieber's rule would force upon us the typing

$$\Sigma \vdash \text{(a banker and two secretaries)} : \text{NP[num} = \alpha]$$

with *variable* (generic) number feature, which is technically all right, but unacceptable: the conjunction *must* be of type NP[num = plural], for use in subject positions, for example. But then (and)'s subsumption condition for the first conjunct, NP[num = singular] $\sqsubseteq$ NP[num = plural], does not hold.

However, there is an alternative based on the notion of subrecord, which Shieber might think of: if NP, without a number-feature, is a type of its own, with subtypes NP[num = singular] $\sqsubseteq$ NP and NP[num = plural] $\sqsubseteq$ NP, by (and) we can derive

$$\Sigma \vdash \text{(a banker and two secretaries)} : \text{NP}.$$

Thus, if we also have $\Sigma \vdash$ Pat hired : NP $\rightarrow$ S, by rule (app) we get the same typing for the whole sentence, as with (And) above. Note that in this case, the unique least common supertype of NP[num = singular] and NP[num = plural] is easily found by dropping the number fields. In general, to compute the least common supertype (= feature structure) seems to be impossible, according to the undecidability results of [1].

Finally, let us look at Shieber's examples in the Introduction.

Example 3 Using our rule (And), together with subtyping by instantiation of generic variables, we get

$$\frac{\Sigma \vdash \text{Pat is} : \sigma \rightarrow S, \qquad \Sigma \vdash \text{healthy} : \text{AP}, \qquad \Sigma \vdash \text{of sound mind} : \text{PP}}{\Sigma \vdash (\text{Pat is})^{\sigma \rightarrow S}((\text{healthy})^{\text{AP}} \text{ and } (\text{of sound mind})^{\text{PP}}) : \text{S}}, \tag{24}$$

provided the side conditions $\sigma \rightarrow \text{S} \leq_\Sigma \text{AP} \rightarrow \text{S}$ and $\sigma \rightarrow \text{S} \leq_\Sigma \text{PP} \rightarrow \text{S}$ hold. In order to satisfy these, we need only modify σ and τ in Shieber's types of (11), using

$$\sigma = [n = \alpha, v = \beta, bar = 2] \qquad \text{and} \qquad \tau = [n = +, v = \gamma, bar = 2] \tag{25}$$

instead, with variables α, β, and γ not in Σ.[3] The reader may check that then we get the same partial order as in Example (1), and the same statements about grammaticality for all the example sentences.

From the semantic view, there seems to be an advantage in not having a 'type' σ and τ as in (11), and not assign a type to $((\text{healthy})^{\text{AP}}$ and $(\text{of sound mind})^{\text{PP}})$ etc. I would even tend to assume that there is no 'type' NP, but only a parameterized family of types NP[num = singular], NP[num = plural], etc. for other features. Clearly, *in the process of parsing* we have to deal with partial *knowledge* about the syntactic types of phrases. But this does not mean that to each such piece of knowledge there is a reasonable 'syntactic type' of phrases (far less: a corresponding class of semantic entities for these).

In particular, the partial order on types given by subrecords might lead to untractable computational problems - in determining sups of feature structures, for example, - that could perhaps be avoided by stressing parameter and LET-polymorphism, and thereby using another algebra of types. I hope the examples demonstrate that there is at least some room in this direction.

References

[1] Jochen Dörre and William Rounds. On subsumption ond semiunification in feature algebras. IWBS Report, IBM Deutschland, Stuttgart, December 1989.

[2] R. Harper, R. Milner, and M. Tofte. The Definition of Standard ML. Version 3. LFCS Report Series ECS-LFCS-89-81, Laboratory for Foundations of Computer Science, Dept. of Computer Science, University of Edinburgh, May 1989.

[3] Fritz Henglein. Type inference and semi-unification. In *Proceedings of the 1988 ACM Conference on LISP and Functional Programming. Snowbird, Utah, July 25-27*, pages 184–197, 1988.

[4] D. Kapur, D. Musser, P. Narendran, and J. Stillman. Semi-unification. In *Proc. Foundations of Software Technology and Theoretical Computer Science, December 1988, Pune, India.* Springer LNCS, 1988.

[5] A.J. Kfoury, J. Tiuryn, and P. Urcyczyn. Computational consequences and partial solutions of a generalized unification problem. In *Fourth IEEE Symposium on Logic in Computer Science. Asiloma, California, June 5–8, 1989.*

[6] A.J. Kfoury, J. Tiuryn, and P. Urcyczyn. The undecidability of the semi-unification problem. BUCS Report, Boston University, October 1989.

[3] More precisely, type expressions containing free variables are thought of as type schemata; hence we would only take the instances NP and AP of this schema and put them into the type hierarchy. One could also turn the schema into a new type $\sigma = \forall\alpha.[n = \alpha, v = \alpha, bar = 2]$ - where α ranges over $\{+, -\}$ -, by introducing a second universe of types, as it is done in the programming language case.

[7] Hans Leiß. Semi-unification and type inference for polymorphic recursion. Bericht ASE-5-89, Siemens AG, München, May 1989.

[8] Robin Milner. A theory of type polymorphism in programming. *Journal of Computer and System Sciences*, 17:348–375, 1978.

[9] Alan Mycroft. Polymorphic type schemes and recursive definitions. In *International Symposium on Programming. 6th Colloquium. Toulouse, April 17-19, 1984*, pages 217–228. Springer LNCS 167, 1984.

[10] Pavel Pudlák. On a unification problem related to Kreisel's conjecture. *Commentationes Mathematicae Universitatis Carolinae*, 29(3):551–556, 1988.

[11] Didier Rémy. Typechecking records and variants in a natural extension of ML. In *ACM Symposium on Principles of Programming Languages*, pages 77–89, 1989.

[12] Stuart Shieber. *Parsing and Type Inference for Natural and Computer Languages.* PhD thesis, Stanford University, March 1989.

Unification Grammar and the Language Theory of Automath

(Extended Abstract)

Urs Egli
Fachgruppe Sprachwissenschaft
Universität Konstanz
D–7750 Konstanz

Unification Grammar is a version of extended phrase structure grammar which was invented by Stuart Shieber. It continues the structuralist tradition of IC-Analysis, as done by R. Wells, Z. Harris and others. The Context Free Grammars of N. Chomsky can be viewed as a formalization of aspects of this tradition in the context of Post production systems. It proved equivalent to the Backus-Naur-Form of programming language description. Whereas Chomsky tried to modify phrase structure grammar by the introduction of transformations, there has been a tradition, begun by G. Harman, to replace transformations by the use of complex categories and rule schemata. The tradition of complex categories is even older than structuralism, as it goes back to the ancient distinction, made by Dionysios Thrax (2nd cent. b.d.) at the beginnings of traditional grammar, between μέρη λόγου (Word Classes like Noun and Verb) and παρεπόμενα (supplementary features of a word, like cases and numbers). This use of secondary categories has been made popular by G. Gazdar in his Generalized Phrase Structure Grammar. Shieber's Unification grammar is a continuation of this trend. He explicitly considers the formal structure of complex categories. His choice is the representation as graphs. But there is another possibility, i.e. to represent complex categories by first order terms. Unification of terms as known from the literature on automatic deduction can then be used instead of the conceptually more complicated graph unification.

(1) Unification algorithm for terms (Loveland p. 78)

(1) If $E1\ \Pi$ and $E2\ \Pi$ are identical, then the expressions are unifiable and Π is a most general unifier.

(2) Otherwise, let n be the leftmost point of disagreement, let $t1$ be the simple expression at position n in $E2\ \Pi$ if $t1$ is then a variable, otherwise let $t1$ be the simple exprassion at position n in $E1\ \Pi$, and let $t2$ be the simple expression at position n in the alternate $E_j\ \Pi$ ($j \in (1,2)$);

(a) if $t1$ is a variable and does not occur in $t2$ then $\Pi = \Pi \vee (t1 / t2)$; go to (1).

(b) otherwise, $E1$ and $E2$ are not unifiable.

Fast linear algorithms of first order term unification (with execution time bounded by a linear function of the length of terms to be unified) are available (Paterson and Wegman).

A simplified version of the so called most famous syntax rule which goes back to Plato's Sophistes can be rendered in the following form

(2) Sentence -> NP(case(nom), number(x)) VP(number(x) tense(y)).

In this rule schema important linguistic facts like agreement in number between verb and subject and nominative case assignment to the subject can be expressed quite naturally. A general concept of term rewriting grammars arises quite naturally. These grammars are much like context free grammars and differ only in that they use rule schemata formulated with terms containing free variables instead of the atomar categories (nonterminal symbols) of context free grammars. The definition of a derivation is quite straightforward. Every substitution instance of a rule schema is one of an infinity of rewriting rules. We may control instantiation by the concept of unification, which singles out useful substitution instances. We may define the language generated by such a grammar as in the case of context free grammars.
One of the advantages of this conceptualization is that it makes possible a comparison with old language theory schemes such as the language theory of Auto(mated) math(ematics) as developed and described by de Bruijn and his coworkers since 1968, which is one of several versions of a T(ype) A(ssignment system). TAs are connected with the theory of typed Lambda-Calculus. There exist other versions like the system TAP of Reynolds-Girard second order polymorphic types and the system of intuitionistic type theory by Martin-Löf, with which Automath may be compared. The relevant Automath rules of type assignment are the following (in Reynolds' notation):

(3) Rules for automath (Hindley & Seldin p. 232)

(te) $$\frac{M \in \lambda x \in \alpha.\beta \quad N \in \alpha}{MN \in (\lambda x \in \alpha.\beta)N}$$

(ti) $$\frac{\begin{array}{c}(x \in \alpha) \\ M \in \beta\end{array}}{\lambda x \in \alpha.\, M \in \lambda x \in \alpha.\, \beta}$$

Now we can compare a third tradition besides PSG and TA, vs. the C(ategorial) G(rammars). A wedding of Categorial grammar and Unification Grammar has already taken place in schemes like C(ategorial) U(nification) G(rammar) and U(nification C(ategorial) G(grammar), which explicate Montague´s dictum of the existence of a categorial translation of every syntactic category. The semantical functional category is the kernel of every syntactic category in these schemes. Besides this there exist modern versions of CG which are continuations of the Lambek calculus in the form of contemporary extended categorial grammars. An early attempt to integrate λ-calculus are

Cresswell's λ-categorial languages. The more formal problems of this tradition can best be studied by drawing on all these three traditions.

(4) the three traditions:

a) (extended) P(hrase) S(tructure) G(rammar)
b) (extended) C(ategorial) G(rammar)
c) (extended) T(ype) A(ssignment Systems).

This is all the more interesting, as the arrow of functionality can be defined in the language theory of Automath in the following way:

(5) $\lambda x \in \alpha.\beta = \alpha \text{->} \beta$, if x doesn't occur free in β.
(te) then becomes

$$\frac{M \in (\alpha \text{->} \beta) \quad N \in \alpha}{MN \in \beta}$$

I shall call the theory of the functionality arrow and first order terms the theory of variable polymorphic types + functionality or terms + functionality, as instead of admitting only constant basic types we allow every first order term as a basic category. The rule (3) can be expressed as a rule schema in the following way:

(6) rewriting rule schema
$\beta \Rightarrow \alpha \text{->} \beta \ \alpha$ (=> is the arrow of replacement, -> is the arrow of functionality) .

The systems must contain a lexicon as with ordinary categorial grammar in order to be linguistically applicable besides the ->-elimination rule. By admitting the empty word and assigning categories to them we can represent every term rewriting grammar as a term + functionality grammar.

(7) The rule $t \Rightarrow t_1 t_2 \dots t_n$ of a term rewriting grammar corresponds to the assignment of (t_1-> ... (t_n->t)...) to the empty word in a term + functionality grammar.

The other way round, a representation of term + functionality grammars as term rewriting grammars is also possible via rule schema (6). Term + functionality grammars are even special cases of term rewriting grammars, i.e. a binary term rewriting system.

(8) Binary term rewriting grammars with rule schemata of the type
t -> t_1 t_2, where the t_i and t are first order terms, which possibly share variables, and lexical rule schemata t-> w, where t is a first order term and w a non-empty word.

Whereas unrestricted term rewriting systems generate the recursively enumerable languages, binary term rewriting systems generate decidable languages. They are rather complex, as is shown by the fact, that a single grammar in the class generates the class of satisfiable formulae of the propositional calculus.

(9) A term rewriting grammar for boolean satisfiability
Satz-> S(w,x); S(w,x)-> N S(f,x); S(f,x)-> N S(w,x); S(w,x)-> K S(w,x) S(w,x); S(f,x)-> K S(w,x) S(f,x) / K S(f,x) S(w,x) / K S(f,x) S(f,x); S(x,y)-> Atomsatz(x,y); Atomsatz(x,y)-> p Indices(x,y); Indices(x,f(y,z)) -> ´ Indices(x,z); Indices(x,f(x,z))-> ´.

The grammar follows the syntax of propositional forms rather closely, but adds the information contained in the truth table associated with the connectives to the syntax of formation rules. The second argument of S(x,y) is a list of truth values of the variables in the order of the index of the variable, which is expressed by a series of primes. The list is decomposed by the two rules introducing the indices of the variables. The last rule imposes the truth value of the n^{th} variable on the n^{th} variable, when generating the n^{th} prime.
We can form a binary term rewriting grammar by applying standard techniques like preunification of the nonbranching rule with the branching rules and breaking rules with triple branching into two rules with an intermediate polymorphic category. Thus, the element problem for binary term rewriting grammars is NP-hard. There is, however, possibly a problem connected with this grammar, as there seems to be a fast algorithm for the membership problem of binary term rewriting grammars, suggested by the algorithm of Cocke, Younger, and Kasami.

Let us now address the question of whether the extra power of lambda types in Automath (which correspond to the product types of Martin-Löf, i.e. types of the form $\Pi x \in \alpha.\beta$) is useful in doing natural langage syntax. I will fist present an example of a grammar which uses generalized lexical rules and the term elimination rule (te) of Automath. I have to sketch the linguistic motivation of the lexikal rules for plural noun phrases. First, plural noun phrases like "Studenten" have the same form in every one of the four cases. This is the reason why we have to abstract from the case in the lexical rule. Second, the syntax presupposed contains a rule like

(10a) VP(X) -> NP(a) VP(Xa)

which generalizes the sentence expansion rule and the VP expansion rule of old-fashioned Syntactic Structure-like Grammars. a is the variable for the case of the noun phrase, which corresponds to the last entry in the list of cases (Xa) demanded by the verb. Third, we may accept the following slogan:

(10b) NPs are predicate functors (VP->VP)

We are now in a position to understand the example.

(11) deduction of the grammaticality of "Studenten lesen"

(1)+ $\lambda x \in$ Valenz.$\lambda a \in$ Kasus.$\lambda y \in$ Kopf.Studenten $\in$
$\lambda x \in$ Valenz.$\lambda a \in$ Kasus.$\lambda y \in$ Kopf.(VP(f(x,ford(a,pl,3ps)),y)->VP(x,y))

(2)+ ______________________________ nil $\in$ Valenz

(3) ($\lambda x \in$ Valenz.$\lambda a \in$ Kasus.$\lambda y \in$ Kopf.Studenten) nil $\in$
($\lambda x \in$ Valenz.$\lambda a \in$ Kasus.$\lambda y \in$ Kopf.(VP(f(x,ford(a,pl,3ps)),y)
->VP(x,y))) nil

(4) $\lambda a \in$ Kasus.$\lambda y \in$ Kopf.Studenten $\in$
.$\lambda a \in$ Kasus.$\lambda y \in$ Kopf.(VP(f(nil,ford(a,pl,3ps)),y)->VP(x,y))

(5)+ ______________________________ nom $\in$ Kasus

......

(6)+ ______________________________ kopf(pl,3ps) $\in$ Kopf

(7) Studenten $\in$ (VP(f(nil,ford(nom,pl,3ps)),kopf(pl,3ps))
->VP(nil,kopf(pl,3ps)))

(8)+ __________ lesen $\in$ (VP(f(nil,ford(nom,pl,3ps)),kopf(pl,3ps))

(9) Studenten lesen $\in$ VP(nil,kopf(pl,3ps)))

In this example, the lines which are marked with + are lexical rules. Underlining shows application of the rule (te). Besides, there are steps of lambda conversion. The features include head features of the verb (Kopf), case lists (Valenz), and cases(Kasus). The last line says that "Studenten lesen" is a zero place verb (with case list nil), i.e. a sentence.

The example shows that we can do syntax in the language theory of Automath. The use of extended TA for the analysis of natural language is thus quite natural. The assignment of rule schemata in term + functionality grammar corresponds to the assignment of Product- or Lambda-Types in Automath-like Type assignment. Syntax can thus be viewed as controlled deduction by the Π-elimination or the (te) rule. We have a choice between terms + functionality and product categories. This choice is somehow analogous to the difference between doing classical logic via free variable logic and doing it via quantification theory.

(12) Free variable logic : quantification theory =
Terms + Functionality : Π-Types

Both approaches should be pursued. An open question concerns the problem of deciding whether Automath-like grammar can always be replaced by free variable + functionality grammar. In every event, unification grammar should be studied in the context of extended type assignment systems, and vice versa extended type assignment systems should be studied in the context of mathematical linguistics.

Aristotle's πτῶσις as well as the Stoic concept of ἔγκλισις contain the germ of a theory of product categories. Aristotle knew that a linguistic category like the type of Verbs corresponds to a family of types. Each type of the family is a case of the verb (πτῶσις ῥήματος cp. Aristotle de interpret. 17a10 ed. Mignucci varia lectio), an instantiation of the product category. Product categories are pervasive in natural language. Perhaps, such a representation of the main categories will also contribute to the problem of giving a semantics of the παρεπόμενα in the classical sense. In ordinary Montague semantics and related schemes they are hardly treated semantically in a principled fashion.

Literature

N. G. DE BRUIJN. A Survey of the Project AUTOMATH. In: J. P. SELDIN & J. R. HINDLEY eds. *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism.* London, Academic Press, 1980.

M. A. HARRISON. *Introduction to Formal Language Theory.* Reading, Mass., Addison-Wesley, 1978.

R. HINDLEY & J. SELDIN. *Introduction to combinators and lambda-calculus.* Cambridge, Cambridge University Press, 1986.

J. E. HOPCROFT & J. D. ULLMAN. *Introduction to Automata Theory, Languages, and Computation.* Reading/Mass., Addison-Wesley, 1979. (Addison-Wesley Series in Computer Science)

M. JOHNSON. *Attribute-value logic and the theory of grammar.* Stanford, CSLI, 1988. (CSLI Lecture Notes)

D. LOVELAND. *Automated Theorem Proving: A Logical Basis.* Amsterdam, North-Holland, 1978. (Fundamental Studies in Computer Science 6)

PER MARTIN-LÖF. *Intuitionistic Type Theory.* Naples, Bibliopolis, 1984.

R. T. OEHRLE & E. BACH & D. WHEELER eds. *Categorial Grammars and Language Structures.* Dordrecht, Reidel,1988. (Studies in Linguistics and Philosophy 32)

M. S. PATERSON & M. N. WEGMAN. Linear Unification. *Journal of Computer and Systems Sciences* 16, 1978, 158-167.

S. SHIEBER. *An Introduction to Unification-based Approaches to Grammar*. Stanford, CSLI, 1986. (CSLI lecture notes)

RUDIMENTARY KRIPKE MODELS FOR THE HEYTING PROPOSITIONAL CALCULUS

Kosta Došen

Matematički Institut
Belgrade, Yugoslavia

The Heyting propositional calculus **H** is sound and complete with respect to Kripke models based on quasi-ordered frames. Besides this class of Kripke models there are many smaller classes of Kripke models with respect to which **H** is sound and complete. For example, we can require from the frames of Kripke models that in addition to being quasi-ordered they satisfy one or more of the following:

— the frame is partially ordered, i.e. we have added antisymmetry,
— the frame is generated, i.e. there is a point which is lesser than or equal to every point,
— the frame is a tree,
— the frame is a Jaśkowski tree,
— the frame is finite.

The propositional calculus **H** is also sound and complete with respect to classes of Kripke models which are all based on a single frame (for example, this frame may be the disjoint union of all finite quasi-ordered frames), or with respect to classes which contain a single Kripke model (like the class whose only member is the canonical model for **H**, familiar from the Henkin-style completeness proof for **H**).

For all these classes of Kripke models the class of all quasi-ordered Kripke models is the largest class, in which all are included. Here we will consider classes of models with respect to which **H** can be shown sound and complete in which the class of all ordinary quasi-ordered Kripke models is properly included. In producing models in these wider classes we will feel free to tamper as much as we can with the conditions on the underlying frames, while the conditions concerning valuations on these frames and the definition of holding in a model will be practically identical as in ordinary Kripke models for **H**. These new models are not meant to replace ordinary Kripke models for the investigation of **H**. Neither are they meant to be philosophically significant. We want to have them only as an instrument for the analysis of the inner mechanism of Kripke models. But they might also raise some interesting technical questions.

The mood of this paper will be close to the mood of correspondence theory (which started in modal logic [1], and was more recently extended to intuitionistic logic [9]). A number of our results will be of the form that a frame satisfies certain conditions concerning its relation iff it satisfies certain conditions concerning valuations on it, or something similar. However, this is not a paper at the level of correspondence theory, because it does not go far enough. It only introduces notions, and proves for them rather straightforward matters which perhaps could lead to a more advanced theory.

We will concentrate here only on propositional logic and leave aside a possible extension of our approach to predicate logic. The paper will be divided into three sections. In the first section we introduce our main generalization of Kripke models for **H**, called *rudimentary* Kripke models. The frames of these models must be only serial, but in the absence of reflexivity we assume for valuations on these frames a condition converse to the usual heredity condition of ordinary Kripke models for **H**. We prove that **H** is sound and complete with respect to rudimentary Kripke models, and consider questions related to completeness. We show that in a certain sense rudimentary Kripke models make the largest class of Kripke-type models with respect to which **H** is *strongly* sound and complete.

In the second section we present a canonical Kripke model for **H** which, though serial and transitive, is not reflexive, and is hence not an ordinary Kripke model, but rudimentary. We also consider briefly at the end of this section a representation for Heyting algebras which is in the background of our canonical model.

In the third section we consider rudimentary Kripke models where valuations are defined inductively. We find necessary and sufficient conditions on frames for the inductive character of rudimentary Kripke models of this type, which make a proper subclass of the class of all rudimentary Kripke models. We also consider at the end of this section some questions related to modal logic.

In a sequel to this paper [4] we shall consider three related topics. First, we shall introduce a very general notion of Beth models for **H**, such that rudimentary Kripke models may be conceived as a particular type of these models. These Beth models are interesting because for them we make another assumption analogous to the converse heredity of rudimentary Kripke models. We shall also consider such Beth models where valuations are defined inductively.

Next we shall present in [4] the correspondence between on the one hand conditions on frames of various types of rudimentary Kripke models and on the other hand the characteristic schemata of Dummett's logic, the logic of weak excluded middle and classical propositional logic.

Finally, we shall consider in [4] a generalization of rudimentary Kripke models which consists in restricting the conditions for rudimentary Kripke models only to those points of our frames which are accessible from some point. In a rather natural sense this makes the largest class of Kripke-type models with respect to which **H** is sound and complete, though even larger classes may be envisaged.

1. Rudimentary Kripke models

Our propositional language has infinitely many propositional variables, the propositional constant $\bot$, and the binary connectives $\rightarrow$, $\wedge$ and $\vee$. For propositional variables we use the schematic letters $p, q, r, \dots, p_1, \dots$, for formulae the schematic letters

$A, B, C, \ldots, A_1, \ldots$, and for sets of formulae the schematic letters $\Gamma, \Delta, \Theta, \ldots, \Gamma_1, \ldots$. As usual, $A \leftrightarrow B$ is defined as $(A \to B) \wedge (B \to A)$ and $\neg A$ as $A \to \bot$. We denote the set of all formulae by $\mathbf{L}$, the set of all formulae in which $\bot$ does not occur by $\mathbf{L}^+$, and the set of all formulae in which only propositional variables and $\to$ occur by $\mathbf{L}^{\to}$. In the metalanguage we use $\Rightarrow$, $\Longleftrightarrow$, & , *or*, *not*, $\forall$, $\exists$ and set-theoretical symbols, with the usual meaning they have in classical logic.

The Heyting propositional calculus $\mathbf{H}$ in $\mathbf{L}$ is axiomatized by the following usual axiom-schemata:

$$A \to (B \to A),\ \ (A \to (B \to C)) \to ((A \to B) \to (A \to C)),$$
$$(C \to A) \to ((C \to B) \to (C \to (A \wedge B))),\ \ (A \wedge B) \to A,\ \ (A \wedge B) \to B,$$
$$A \to (A \vee B),\ \ B \to (A \vee B),\ \ (A \vee B) \to ((A \to C) \to ((B \to C) \to C)),$$
$$\bot \to A,$$

and the rule *modus ponens*. It is well-known that this axiomatization is separative, in the sense that an axiomatization of a fragment of $\mathbf{H}$ involving some connectives, among which we must have $\to$, is obtained by assuming *modus ponens* and all those axiom-schemata from the list above in which the connectives of the fragment in question occur. So the positive Heyting propositional calculus $\mathbf{H}^+$ in $\mathbf{L}^+$ is axiomatized by rejecting $\bot \to A$, and the implicational fragment of $\mathbf{H}$, i.e. the system $\mathbf{H}^{\to}$ in $\mathbf{L}^{\to}$, is axiomatized by the first two axiom-schemata and *modus ponens*.

A *frame* is $\langle W, R \rangle$ where W is a nonempty set and R is a binary relation on W. Members of W, which in modal logic are called *worlds*, will here be called more neutrally *points*. We use $x, y, z, \ldots, x_1, \ldots$ for members of W, and $X, Y, Z, \ldots, X_1, \ldots$ for subsets of W. For a frame $\langle W, R \rangle$ a subset X of W will be called *hereditary* iff for every x

$$x \in X \ \Rightarrow\ \forall y (x\, R\, y \ \Rightarrow\ y \in X),$$

and it will be called *conversely hereditary* iff for every x

$$\forall y (x\, R\, y \ \Rightarrow\ y \in X) \ \Rightarrow\ x \in X.$$

For a frame $\langle W, R \rangle$ and $X, Y \subseteq W$ we have the binary operation $\to_R$ defined by:

$$X \to_R Y = \{x : \forall y (x\, R\, y \ \Rightarrow\ (y \in X \ \Rightarrow\ y \in Y))\}.$$

A *pseudo-valuation* v on a frame $\langle W, R \rangle$ is a function from $\mathbf{L}$ into $\mathbf{P}W$, i.e. the power set of W, which satisfies the following conditions for every $A, B \in \mathbf{L}$:

$(v\,\bot)$ $\ v(\bot) = \varnothing$,

$(v \to)$ $\ v(A \to B) = v(A) \to_R v(B)$,

$(v\,\wedge)$ $\ v(A \wedge B) = v(A) \cap v(B)$,

$(v\,\vee)$ $\ v(A \vee B) = v(A) \cup v(B)$.

A *valuation* v on a frame $\langle W, R \rangle$ is a pseudo-valuation which satisfies:

(*A-Heredity*) *for every formula A the set $v(A)$ is hereditary*, (*Converse A-Heredity*) *for every formula A the set $v(A)$ is conversely hereditary.*

A *rudimentary Kripke model* is $\langle W, R, v \rangle$ where $\langle W, R \rangle$ is a frame and v a valuation on this frame. A formula A ***holds in*** $\langle W, R, v \rangle$ iff $v(A) = W$.

With our usual experience with Kripke models, instead of A-Heredity and Converse A-Heredity we would expect only the following conditions:

(*p-Heredity*) *for every propositional variable p the set $v(p)$ is hereditary,*

(*Converse p-Heredity*) *for every propositional variable p the set $v(p)$ is conversely hereditary.*

Valuations would be defined by specifying v for propositional variables and using $(v\,\bot)$, $(v \rightarrow)$, $(v\,\wedge)$ and $(v\,\vee)$ as clauses in an inductive definition. That A-Heredity and Converse A-Heredity obtain would be derived by induction on the complexity of A. Rudimentary Kripke models whose frames have properties which guarantee that every v defined on them in such an inductive way satisfies A-Heredity and Converse A-Heredity make an important proper subclass of the class of all rudimentary Kripke models (we will study this subclass in the third section). Here, however, we deal first with models where there is no guarantee that a pseudo-valuation which satisfies p-Heredity and Converse p-Heredity will satisfy also A-Heredity and Converse A-Heredity. In arbitrary rudimentary Kripke models A-Heredity and Converse A-Heredity are not derived but stipulated; namely, we restrict ourselves to those pseudo-valuations where these two heredity conditions have somehow been secured.

The following proposition shows that frames of rudimentary Kripke models cannot be completely arbitrary:

Proposition 1. *For every rudimentary Kripke model $\langle W, R, v \rangle$, the relation R is* ***serial****, i.e. $\forall x\, \exists y\, (x\; R\; y)$.*

Proof. Since for every x we have $x \notin v(\bot)$, by Converse A-Heredity there is a y such that $x\; R\; y$ and $y \notin v(\bot)$. *q.e.d.*

The next proposition shows that we need not assume anything besides seriality for frames of rudimentary Kripke models:

Proposition 2. *If in the frame $\langle W, R \rangle$ the relation R is serial, then there is a valuation v on $\langle W, R \rangle$.*

Proof. On $\langle W, R \rangle$ where R is serial let $v(p)$ be either W or $\emptyset$ and let $v(\bot) = \emptyset$. Then using the conditions $(v \rightarrow)$, $(v\wedge)$ and $(v\,\vee)$ we define $v(A)$ for every formula A. It is easy to check by induction on the complexity of A that $v(A)$ is either W or $\emptyset$. (The only interesting case in this induction is when A is of the form $A_1 \rightarrow A_2$, and we have $v(A_1) = W$ and $v(A_2) = \emptyset$; then $v(A_1 \rightarrow A_2) = \{x : not\, \exists y\, (x\; R\; y)\} = \emptyset$, by using the seriality of R.) It is clear that A-Heredity and Converse A-Heredity obtain if $v(A) = W$. If $v(A) = \emptyset$, then A-Heredity is vacuously satisfied and Converse A-Heredity follows from the seriality of R. *q.e.d.*

This proof shows that for every serial frame $\langle W, R \rangle$ the set $\{W, \emptyset\}$ is closed under the operations $\rightarrow_R$, $\cap$ and $\cup$. (We call a frame $\langle W, R \rangle$ *serial* iff R is serial, and similarly with models and other properties.)

A *positive* valuation v^+ on a frame $\langle W, R \rangle$ is a function from $\mathbf{L}^+$ into $\mathbf{P}W$ which satisfies all the conditions for valuations except $(v \perp)$, which does not apply anymore. A *positive* rudimentary Kripke model is $\langle W, R, v^+ \rangle$. In a positive rudimentary Kripke model R can be completely arbitrary, even empty, as the following proposition shows:

Proposition 3. *For every frame $\langle W, R \rangle$ there is a positive valuation v^+ on $\langle W, R \rangle$.*

Proof. For every $A \in \mathbf{L}^+$ let $v^+(A) = W$, and check that v^+ is a positive valuation. *q.e.d.*

This proof is based on the simple fact that for an arbitrary frame $\langle W, R \rangle$ the set $\{W\}$ is closed under $\rightarrow_R$, $\cap$ and $\cup$. We can similarly show that in *implicative* rudimentary Kripke models $\langle W, R, v^{\rightarrow} \rangle$, where $v^{\rightarrow}$ maps $\mathbf{L}^{\rightarrow}$ into $\mathbf{P}W$ and satisfies $(v \rightarrow)$, A-Heredity and Converse A-Heredity, R can also be completely arbitrary.

The valuations defined in the proof of Proposition 2 and the positive valuation defined in the proof of Proposition 3 are trivial, since in every rudimentary Kripke model of the first proof every two-valued tautology holds, and in the positive rudimentary Kripke model of the second proof every formula of $\mathbf{L}^+$ holds. Of course, not all rudimentary, or positive rudimentary, Kripke models are trivial in this way.

In a *quasi-ordered* frame $\langle W, R \rangle$ the relation R is reflexive and transitive, and because of reflexivity these frames are serial. Rudimentary Kripke models based on such frames, which we will call *quasi-ordered Kripke models*, are the ordinary Kripke models for $\mathbf{H}$. The conditions for valuations which we have given above are necessary and sufficient for valuations in these ordinary Kripke models, though with ordinary Kripke models they are usually introduced in a different way. Namely, instead of A-Heredity we assume only p-Heredity, whereas Converse A-Heredity is not assumed in any form. The conditions $(v \perp)$, $(v \rightarrow)$, $(v \wedge)$ and $(v \vee)$ are assumed exactly as above. By induction on the complexity of A we can then demonstrate A-Heredity, whereas Converse A-Heredity is an immediate consequence of the reflexivity of R. So every quasi-ordered Kripke model is a rudimentary Kripke model, but not vice versa, as Proposition 2 shows. As we have remarked, the rudimentary Kripke models of the proof of Proposition 2 are trivial, but we will see below in Proposition 7 and in the next section that there are nontrivial rudimentary Kripke models which are not quasi-ordered.

We will now demonstrate that $\mathbf{H}$ is sound and complete with respect to rudimentary Kripke models. For soundness we have the following proposition:

Proposition 4. *If B is provable in* $\mathbf{H}$, *then B holds in every rudimentary Kripke model.*

Proof. We proceed by induction on the length of proof of B in $\mathbf{H}$. If B is an axiom, the only case where we must invoke Converse A-Heredity is when B is of the form $(B_1 \rightarrow (B_2 \rightarrow B_3)) \rightarrow ((B_1 \rightarrow B_2) \rightarrow (B_1 \rightarrow B_3))$, and this is why in the basis of the induction we will consider only this case as an example.

Suppose for B of the form above that for some x we have $x \notin v(B)$. We easily infer that we must have y, z and t such that:

$x\ R\ y$ and $y \in v(B_1 \rightarrow (B_2 \rightarrow B_3))$,

$y\ R\ z$ and $z \in v(B_1 \rightarrow B_2)$,

$z\ R\ t$, $t \in v(B_1)$ and $t \notin v(B_3)$.

So $t \in v(B_2)$ and by A-Heredity $t \in v(B_2 \rightarrow B_3)$. Then from $t \notin v(B_3)$, by Converse A-Heredity, it follows that there must be a u such that $t\ R\ u$ and $u \notin v(B_3)$. By A-Heredity $u \in v(B_2)$, but since $t \in v(B_2 \rightarrow B_3)$ we obtain a contradiction.

For the induction step suppose that $v(B_1) = W$ and $v(B_1 \rightarrow B_2) = W$. Next suppose $x\ R\ y$. Since $x \in v(B_1 \rightarrow B_2)$ and $y \in v(B_1)$ we obtain $y \in v(B_2)$. So $\forall y\,(x\ R\ y \ \Rightarrow\ y \in v(B_2))$, from which $x \in v(B_2)$ follows by Converse A-Heredity. *q.e.d.*

For completeness it is enough to appeal to the completeness of **H** with respect to quasi-ordered Kripke models. Indeed, if B holds in all rudimentary Kripke models, then B holds in all quasi-ordered Kripke models, and hence B is provable in **H**. So we have:

Proposition 5. *A formula B is provable in* **H** *iff B holds in every rudimentary Kripke model.*

We can similarly demonstrate the soundness and completeness of $\mathbf{H}^+$ with respect to positive rudimentary Kripke models, and of $\mathbf{H}^\rightarrow$ with respect to implicative rudimentary Kripke models.

In the background of the soundness of **H** with respect to rudimentary Kripke models is the following algebraic fact. For every rudimentary Kripke model $\langle W, R, v\rangle$, the set $\{v(A) : A \in \mathbf{L}\}$ contains $\emptyset$ and is closed under the operations $\rightarrow_R$, $\cap$ and $\cup$; the algebra $\langle\{v(A) : A \in \mathbf{L}\}, \rightarrow_R, \cap, \cup, \emptyset\rangle$ is a Heyting algebra. In terms of frames, for every frame $\langle W, R\rangle$, every set $\mathcal{A}$ of hereditary and conversely hereditary subsets of W which contains $\emptyset$ and is closed under the operations $\rightarrow_R$, $\cap$ and $\cup$ is a Heyting algebra. (For $\emptyset$ to be conversely hereditary our frame must be serial.) When we verify for $X, Y, Z \in \mathcal{A}$ that

$$X \cap Y \subseteq Z \iff X \subseteq Y \rightarrow_R Z$$

we use the hereditariness of X from left to right, whereas the hereditariness of Y and converse hereditariness of Z are used from right to left. Our frame may be such that $\mathcal{A}$ never coincides with the set of *all* hereditary and conversely hereditary subsets of W (this will become clear in the third section; see Proposition 19 and the comments following this proposition).

For a frame $\langle W, R\rangle$, let us define R^k, where $k \geq 0$, by the following recursive clauses:

$$x\,R^0\,y \iff x = y,$$
$$x\,R^{k+1}\,y \iff \exists z\,(x\,R^k\,z \ \&\ z\,R\,y).$$

It is clear that $x\,R^1\,y \iff x\;R\;y$. Note that A-Heredity is equivalent with the conditions that for every A and every x:

$$x \in v(A) \;\Rightarrow\; \forall y\,((\exists k \geq m)\,x\,R^k\,y \;\Rightarrow\; y \in v(A)),$$

where $m = 0$ or $m = 1$. On the other hand, Converse A-Heredity is not equivalent with the converse conditions, namely that for every A and every x the converse implication obtains, with either $m = 0$ or $m = 1$, though it implies these converse conditions, for both $m = 0$ and $m = 1$ (the converse condition where $m = 0$ is vacuously true). However, in the presence of A-Heredity, Converse A-Heredity is equivalent with the converse condition where $m = 1$.

The soundness of **H** with respect to rudimentary Kripke models can be inferred from the following proposition too:

Proposition 6. *For every rudimentary Kripke model $\langle W, R, v\rangle$ there is a quasi-ordered Kripke model $\langle W, R', v'\rangle$ with the same W such that for every A we have $v(A) = v'(A)$.*

Proof. For a rudimentary Kripke model $\langle W, R, v\rangle$ we define $\langle W, R', v'\rangle$ by stipulating that $x\,R'\,y$ iff $(\exists k \geq 0)x\,R^k\,y$, and $v'(A) = v(A)$. In other words, R' is the reflexive and transitive closure of R, and v and v' coincide. Then we verify that $\langle W, R', v'\rangle$ is indeed a quasi-ordered Kripke model.

The only part of this verification which is not quite straightforward is when in the verification that v' is a valuation on $\langle W, R'\rangle$ we have to check that v' satisfies $(v \rightarrow)$, i.e. when we show that:

$$\forall y\,((\exists k \geq 0)\,x\,R^k\,y \;\Rightarrow\; (y \in v(B) \;\Rightarrow\; y \in v(C))) \quad \textit{iff}$$
$$\forall y\,(x\;R\;y \;\Rightarrow\; (y \in v(B) \;\Rightarrow\; y \in v(C))).$$

From left to right we just appeal to the fact that $x\;R\;y \;\Rightarrow\; (\exists k \geq 0)x\,R^k\,y$. For the other direction suppose that for some y we have $(\exists k \geq 0)\,x\,R^k\,y$, $y \in v(B)$ and $y \notin v(C)$. If $k = 0$, then $x \in v(B)$ and $x \notin v(C)$. Hence, by the Converse A-Heredity of v in $\langle W, R, v\rangle$ we have a y such that $x\;R\;y$ and $y \notin v(C)$, and by the A-Heredity of v in $\langle W, R, v\rangle$ we also have $y \in v(B)$. If $k > 0$, then for some z we have $x\,R^{k-1}\,z$ and $z\;R\,y$. It follows that $z \notin v(B \rightarrow C)$. Either $x = z$, in which case $x \notin v(B \rightarrow C)$, or $x \neq z$, in which case by the A-Heredity of v in $\langle W, R, v\rangle$ it again follows that $x \notin v(B \rightarrow C)$. Hence there is a y such that $x\;R\,y$, $y \in v(B)$ and $y \notin v(C)$. *q.e.d.*

As a kind of converse of Proposition 6 we can demonstrate the following:

Proposition 7. *For every quasi-ordered Kripke model $\langle W, R, v\rangle$ there is a rudimentary Kripke model $\langle W^*, R^*, v^*\rangle$ which is not quasi-ordered such that for every A we have $v(A) = W$ iff $v^*(A) = W^*$.*

Proof. If $W' = \{x' : x \in W\}$ and $W' \cap W = \emptyset$, let $W^* = W \cup W'$. On W' we define R' by:

$$x'\,R'\,y' \iff (x\,R\,y \;\;\&\;\; x \neq y),$$

and we let $R^* = R \cup R' \cup \{\langle x', x \rangle : x \in W\}$. So the frame $\langle W^*, R^* \rangle$ consists of $\langle W, R \rangle$ plus an irreflexive copy $\langle W', R' \rangle$ of $\langle W, R \rangle$ such that for every $x' \in W'$ and $x \in W$ we have $x' R^* x$. The frame $\langle W^*, R^* \rangle$ is not reflexive since $\langle W', R' \rangle$ is irreflexive, and if there is a $y \in W$ distinct from $x \in W$ such that $x\, R\, y$, then $\langle W^*, R^* \rangle$ is not transitive, since though we have $x' R^* x$ and $x\, R^* y$, we don't have $x' R^* y$.

If $v'(A) = \{x' \in W' : x \in v(A)\}$, let $v^*(A) = v(A) \cup v'(A)$. It is straightforward to check that v^* is a valuation on $\langle W^*, R^* \rangle$ and that $\langle W^*, R^*, v^* \rangle$ is a rudimentary Kripke model such that our proposition is satisfied. *q.e.d.*

For W and W^* as in the proof of Proposition 7, let f be a function from W^* *onto* W defined by $f(x) = x$ and $f(x') = x$. Then f is a pseudo-epimorphism, or zigzag morphism, from $\langle W^*, R^* \rangle$ onto $\langle W, R \rangle$ (see [7], pp. 70–75; [1], pp. 174, 187; or [9], 2.4.2) since we have:

$$(\forall z, t \in W^*)(z\, R^*\, t \;\Rightarrow\; f(z)\, R\, f(t)),$$
$$(\forall z \in W^*)(\forall y \in W)(f(z)\; R\; y \;\Rightarrow\; (\exists t \in W^*)(f(t) = y \;\;\&\;\; z\, R^*\, t)).$$

We also have for every $z \in W^*$ and every formula A that:

$$z \in v^*(A) \iff f(z) \in v(A).$$

More generally, we can prove Proposition 7 by letting W^* be the disjoint union of two or more sets W_i, each in *one-one* correspondence with W. On W_i for every $x, y \in W$ such that $x \neq y$ we have $x_i\, R_i\, y_i$ iff $x\, R\, y$, but for some $x_i \in W_i$ we may lack $x_i\, R_i\, x_i$, which makes $\langle W_i, R_i \rangle$ nonreflexive. In the relation R^* on W^* is included the union of all the relations R_i and moreover for every $x_i \in W_i$ we have an $x_j \in W_j$ such that $x_i\, R^*\, x_j$. If one of the frames $\langle W_i, R_i \rangle$ is nonreflexive, then $\langle W^*, R^* \rangle$ is nonreflexive, whereas transitivity will fail if $x_i\, R^*\, x_j$ and $x_j\, R^*\, x_k$ but *not* $x_i\, R^*\, x_k$ (in the proof of Proposition 7 above, transitivity fails for a different reason). The set $v^*(A)$ is the union of all the sets $v_i(A) = \{x_i \in W_i \colon x \in v(A)\}$. The frame $\langle W, R \rangle$ is a pseudo-epimorphic image of $\langle W^*, R^* \rangle$ under $f : W^* \to W$ defined by $f(x_i) = x$, and we have $x_i \in v^*(A)$ iff $f(x_i) \in v(A)$.

Proposition 6 says that for every rudimentary Kripke model there is a quasi-ordered Kripke model in which the same formulae hold, and since the converse is trivially satisfied, it might seem that rudimentary Kripke models do not bring anything new. However, they may bring something new if instead of holding in a model we consider holding in a frame. We say that A *holds in a frame* $\langle W, R \rangle$ iff for every valuation v on $\langle W, R \rangle$ we have that A holds in $\langle W, R, v \rangle$. Our soundness and completeness result of Proposition 5 can equivalently be expressed in terms of holding in frames; namely B is provable in **H** iff B holds in every serial frame.

It is also true that B is provable in **H** iff B holds in *every* frame, for if a frame $\langle W, R \rangle$ is not serial, then, since there are no valuations on $\langle W, R \rangle$, it is vacuously satisfied that for every valuation v on $\langle W, R \rangle$ every formula A holds in $\langle W, R, v \rangle$. This is like accepting among our rudimentary Kripke models also models where W is empty, since if W is empty, every formula A holds vacuously in $\langle W, R, v \rangle$. The problem with these vacuous holdings is that every formula, even $\bot$, will have a model, or a frame in which it holds. So, we should exclude vacuous holdings from our considerations.

In [10] Šehtman gave an example of an intermediate propositional logic **S** incomplete with respect to any class of partially ordered frames, i.e. there is no class $\mathcal{C}$ of partially ordered frames such that B is provable in **S** iff B holds in every frame in $\mathcal{C}$. It follows easily that there is no class $\mathcal{C}$ of quasi-ordered frames such that B is provable in S iff B holds in every frame in $\mathcal{C}$, since for every quasi-ordered frame there is a partially ordered frame in which the same formulae hold (for the quasi-ordered frame $\langle W, R \rangle$ we take the partially ordered frame $\langle W', R' \rangle$ where with $[x] = \{y : x\ R\ y\ \ \&\ \ y\ R\ x\}$ we have $W' = \{[x] : x \in W\}$ and $[x]\,R'\,[y]$ iff $x\ R\ y$). However, for all we know, it seems possible that a logic like **S** be incomplete with respect to any class of quasi-ordered frames but nevertheless complete with respect to a class of serial frames, where the holding of formulae in serial frames is defined in terms of rudimentary Kripke models. In the proof of Proposition 6 we produced out of a rudimentary Kripke model $\langle W, R, v \rangle$ a quasi-ordered Kripke model $\langle W, R', v' \rangle$ in which the same formulae hold, but this does not mean that in $\langle W, R \rangle$ and $\langle W, R' \rangle$ the same formulae will hold. Every valuation v on $\langle W, R \rangle$ will induce an equivalent valuation v' on $\langle W, R' \rangle$, as in the proof of Proposition 6, but on $\langle W, R' \rangle$ we might have valuations to which no valuation corresponds on $\langle W, R \rangle$. For example, let in $\langle W, R \rangle$ *not* $x\ R\ x$; then for a v' on $\langle W, R' \rangle$ we can have

$$x \notin v'(A) \quad \text{and} \quad \forall y\,((\exists k \geq 1)\, x\, R^k\, y \ \Rightarrow\ y \in v'(A)),$$

but for no v on $\langle W, R \rangle$ we can have

$$x \notin v(A) \quad \text{and} \quad \forall y\,((\exists k \geq 1)\, x\, R^k\, y \ \Rightarrow\ y \in v(A)).$$

So we ask the following question:

(1) Is there an intermediate propositional logic incomplete with respect to any class of quasi-ordered frames but complete with respect to a class of serial frames?

When we show that a logic like Šehtman's **S** is incomplete with respect to any class of quasi-ordered frames we find a formula B which is not a theorem of **S** but which holds in every quasi-ordered frame in which all the theorems of **S** hold. In order to show that B is not a theorem of S, or of a similar logic, we can use a more general type of frames (like the *general*, or *first order*, frames in modal logic; see [7], pp. 62–67). Can serial frames be used for the same purpose, namely:

(2) Is there a set of formulae Γ and a formula B such that in every quasi-ordered frame in which all the members of Γ hold B holds too, whereas there is a serial frame in which all the members of Γ hold and B does not hold?

A positive answer to (1) entails a positive answer to (2), but (2) seems to be a weaker question.

A question related to (1) is:

(3) Is there an intermediate propositional logic incomplete with respect to any class of serial frames?

Let $\Gamma \vdash B$ mean as usual that there is a proof of B in **H** from hypotheses in Γ. A positive answer to (3) would show that it is impossible to prove for **H** the completeness direction of the following *strong* soundness and completeness for every Γ and B:

$\Gamma \vdash B$ iff for every serial frame $\langle W, R \rangle$, if all the members of Γ hold in $\langle W, R \rangle$, then B holds in $\langle W, R \rangle$.

Šehtman's result mentioned above shows that such a strong completeness fails when we replace serial by quasi-ordered or partially ordered frames. However, we can easily establish the following *strong* soundness and completeness for **H**:

Proposition 8. *For every* Γ *and* B*:*

$\Gamma \vdash B$ *iff for every rudimentary Kripke model* $\langle W, R, v \rangle$*, if all the members of* Γ *hold in* $\langle W, R, v \rangle$*, then* B *holds in* $\langle W, R, v \rangle$*.*

Before proving this proposition let us note that its right-hand side:

$(*)$ *for every* $\langle W, R, v \rangle$, $\bigcap_{C \in \Gamma} v(C) = W \;\Rightarrow\; v(B) = W$

is equivalent for rudimentary Kripke models with the seemingly stronger assertion:

$(**)$ *for every* $\langle W, R, v \rangle$, $\bigcap_{C \in \Gamma} v(C) \subseteq v(B)$.

That $(*)$ implies $(**)$ follows from the fact that for every rudimentary Kripke model $\langle W, R, v \rangle$ and $x \in W$, the *submodel generated by* x, i.e. $\langle W_x, R_x, v_x \rangle$ where:

$$W_x = \{y \in W : (\exists k \geq 0)\, x\, R^k\, y\},$$
$$(\forall y, z \in W_x)(y\, R_x\, z \iff y\, R\, z),$$
$$v_x(A) = v(A) \cap W_x,$$

is a rudimentary Kripke model. Suppose that in $\langle W, R, v \rangle$ we have $x \in \bigcap_{C \in \Gamma} v(C)$. Then by the A-Heredity of v in $\langle W, R, v \rangle$, in $\langle W_x, R_x, v_x \rangle$ we have $\bigcap_{C \in \Gamma} v_x(C) = W_x$, and by $(*)$ we obtain $v_x(B) = W_x$. So $x \in v(B)$. That $(**)$ implies $(*)$ follows immediately from the definitions and does not rely on either A-Heredity or Converse A-Heredity.

Proof of Proposition 8. The soundness direction is a simple corollary of Proposition 4. For if $\Gamma \vdash B$, then by the deduction theorem either B is provable in **H** or for some $n \geq 1$ and some $C_1, \ldots, C_n \in \Gamma$ we have that $C_1 \to (C_2 \to \ldots \to (C_n \to B) \ldots)$ is provable in **H**. In either case $(**)$ obtains (in the latter case we apply A-Heredity and Converse A-Heredity). The completeness direction follows immediately from the fact that this implication obtains when we replace rudimentary Kripke models by quasi-ordered Kripke models. *q.e.d.*

An analogous strong soundness and completeness can also be proved for $\mathbf{H}^+$ with respect to positive rudimentary Kripke models, and for $\mathbf{H}^\to$ with respect to implicative rudimentary Kripke models.

Let us say that $\langle W, R, v\rangle$ is a ***pseudo-Kripke model*** iff $\langle W, R\rangle$ is a frame and v a pseudo-valuation; A holds in $\langle W, R, v\rangle$ iff $v(A) = W$. (The logic in **L** sound and complete with respect to all pseudo-Kripke models is axiomatized in [2].) We can interpret Proposition 8 as saying that A-Heredity and Converse A-Heredity are sufficient conditions on pseudo-Kripke models for obtaining the strong soundness and completeness of **H**. That these two heredity conditions are also in a certain sense necessary will be inferred from the following two propositions, which are an immediate consequence of definitions:

Proposition 9. *A pseudo-valuation v on a frame $\langle W, R\rangle$ satisfies A-Heredity iff for every B and C we have $v(B) \subseteq v((C \to C) \to B)$.*

Proposition 10. *A pseudo-valuation v on a frame $\langle W, R\rangle$ satisfies Converse A-Heredity iff for every B and C we have $v((C \to C) \to B) \subseteq v(B)$.*

Now we can prove the following:

Proposition 11. *The class of all rudimentary Kripke models is the largest class of pseudo-Kripke models with respect to which* **H** *is strongly sound and complete in the sense that for every Γ and B, $\Gamma \vdash B$ iff* (**).

Proof. The sufficiency of A-Heredity and Converse A-Heredity follows from Proposition 8. Next we show their necessity. Since for **H**, for every B and C we have $\{B\} \vdash (C \to C) \to B$ and $\{(C \to C) \to B\} \vdash B$, for each of our pseudo-Kripke models $\langle W, R, v\rangle$, for every B and C we must have $v(B) = v((C \to C) \to B)$. Then we apply Propositions 9 and 10. *q.e.d.*

(Note that $\{B\} \vdash (C \to C) \to B$ is related to the deduction theorem, whereas $\{(C \to C) \to B\} \vdash B$ is related to *modus ponens*.)

Though (*) and (**) are equivalent for rudimentary Kripke models, (*) does not imply (**) for every pseudo-Kripke model. For example, that $v(B) = W$ implies $v((C \to C) \to B) = W$ is satisfied for every pseudo-Kripke model, but $v(B) \subseteq v((C \to C) \to B)$ may fail in the absence of A-Heredity. So we cannot replace (**) by (*) in Proposition 11. We will see in the last section of [4] that the class of rudimentary Kripke models is properly included in the largest class of pseudo-Kripke models with respect to which **H** is strongly sound and complete in the sense that for every Γ and B, $\Gamma \vdash B$ iff (*); and the latter class is properly included in the largest class of pseudo-Kripke models with respect to which we can prove the ordinary soundness and completeness of **H**.

The Kolmogorov-Johansson, or *minimal*, propositional calculus **J** in **L** is obtained by rejecting $\bot \to A$ from our axiomatization of **H**. This system does not differ essentially from $\mathbf{H}^+$, and it is not difficult to obtain a soundness and completeness result for **J** with respect to "rudimentary" Kripke models which differ from rudimentary Kripke models for **H** only in not requiring $v(\bot) = \varnothing$; the set $v(\bot)$ can be an arbitrary hereditary and conversely hereditary set. "Rudimentary" Kripke models for **J** need not be serial, and their frames may be completely arbitrary. (So Johansson may after all have been right in calling **J** *minimal*.)

2. A canonical rudimentary Kripke model

We shall now consider a nontrivial rudimentary Kripke model which is not quasi-ordered, but is analogous to the canonical partially ordered Kripke model familiar from the Henkin-style completeness proof for **H**.

A set of formulae Γ is *consistent* iff for some A *not* $\Gamma \vdash A$; the set Γ is *deductively closed* iff for every A we have that $\Gamma \vdash A$ implies $A \in \Gamma$; and Γ has the *disjunction property* iff for every A and B we have that $A \vee B \in \Gamma$ implies $A \in \Gamma$ or $B \in \Gamma$. A set of formulae which is consistent, deductively closed and has the disjunction property will be called a *prime theory*.

A set of formulae Γ will be called *A-maximal* iff $A \notin \Gamma$ and for every B, either $B \in \Gamma$ or $B \to A \in \Gamma$. It is easy to check that a prime theory Γ is maximal (in the sense that for every prime theory Δ if $\Gamma \subseteq \Delta$, then $\Gamma = \Delta$) iff Γ is $\bot$-maximal (the same holds when we replace prime theories by consistent deductively closed sets).

Let $W_c = \{\Gamma : \Gamma \textit{ is a prime theory}\}$, and let us define on W_c the relation R_c by:

$$\Gamma\, R_c\, \Delta \iff (\Gamma = \Delta \ \&(\exists A)\, \Gamma \textit{ is } A\textit{-maximal}\,) \quad \textit{or} \quad \Gamma \subset \Delta$$

where $\Gamma \subset \Delta$ means that Γ is a proper subset of Δ. Next let $v_c(A) = \{\Gamma \in W_c : A \in \Gamma\}$. We shall call $\langle W_c, R_c, v_c \rangle$ the *canonical rudimentary Kripke model* for **H**. This model differs from the usual canonical Kripke model for **H** only in the definition of R_c; in the usual canonical Kripke model $\Gamma\, R_c\, \Delta$ is defined as $\Gamma \subseteq \Delta$. Let us first prove the following proposition:

Proposition 12. *The canonical rudimentary Kripke model for* **H** *is a rudimentary Kripke model.*

Proof. It is clear that W_c is nonempty and that $\langle W_c, R_c \rangle$ is a frame. That the conditions $(v\,\bot)$, $(v\,\wedge)$ and $(v\,\vee)$ are satisfied for v_c follows immediately from the definition of prime theories. To verify $(v \to)$ for v_c we show that for every prime theory Γ and every A and B:

(I) $A \to B \in \Gamma \iff \forall \Delta (\Gamma\, R_c\, \Delta \Rightarrow (A \in \Delta \Rightarrow B \in \Delta))$.

From left to right this follows immediately from the fact that $\Gamma\ R_c\ \Delta$ implies $\Gamma \subseteq \Delta$. For the other direction suppose $A \to B \notin \Gamma$; hence $B \notin \Gamma$. If $A \in \Gamma$ and for some C the set Γ is C-maximal, we have $\Gamma\, R_c\, \Gamma$, $A \in \Gamma$ and $B \notin \Gamma$. If $A \in \Gamma$ and there is no C such that Γ is C-maximal, then for some D we have $D \notin \Gamma$ and $D \to B \notin \Gamma$, and we extend $\Gamma \cup \{D\}$ to a prime theory Δ such that $A \in \Delta$ and $B \notin \Delta$. If $A \notin \Gamma$, we extend $\Gamma \cup \{A\}$ to a prime theory Δ such that $A \in \Delta$ and $B \notin \Delta$.

If in (I) we let A be $C \to C$, then since in **H** we have $B \leftrightarrow ((C \to C) \to B)$ we immediately obtain A-Heredity and Converse A-Heredity for v_c. *q.e.d.*

To prove the strong completeness of **H** with respect to rudimentary Kripke models we could use the canonical rudimentary Kripke model instead of the usual canonical Kripke model for **H**. As for the usual canonical model, if *not* $\Gamma \vdash B$, then there is a prime theory Δ such that $\Gamma \subseteq \Delta$ and $B \notin \Delta$.

The canonical rudimentary Kripke model for **H** is serial, as it follows from Proposition 1 (and as can directly be proved by copying the argument in the proof of Proposition 12). This model is also transitive, but it is not reflexive. Let $\Gamma_{\mathbf{H}}$ be the set of theorems of **H**. The set $\Gamma_{\mathbf{H}}$ is a prime theory for which there is no A such that $\Gamma_{\mathbf{H}}$ is A-maximal. Otherwise there would be an A which is not a theorem of **H** such that for a propositional variable p foreign to A we would have that $p \vee (p \rightarrow A)$ is a theorem of **H**. So we don't have $\Gamma_{\mathbf{H}}\ R_c\ \Gamma_{\mathbf{H}}$.

The set $\Gamma_{\mathbf{H}}$ is not the only prime theory Γ for which there is no A such that Γ is A-maximal. Such are also all the prime theories $\Gamma_B = \{C : \{B\} \vdash C\}$ where B is a *Harrop formula* nonequivalent to $\bot$ in **H**; that Γ_B is a prime theory follows from the fact for Harrop formulae B we have that if $\{B\} \vdash C_1 \vee C_2$, then either $\{B\} \vdash C_1$ or $\{B\} \vdash C_2$ (see [6] or [8], p. 55). That there is no A such that Γ_B is A-maximal is shown as follows. Suppose $A \notin \Gamma_B$, i.e. *not* $\{B\} \vdash A$; then for a p foreign to both B and A we have that *not* $\{B\} \vdash p$ (otherwise we would have $\{B\} \vdash \bot$, and hence also $\{B\} \vdash A$) and *not* $\{B\} \vdash p \rightarrow A$ (otherwise we would have $\{B\} \vdash (C \rightarrow C) \rightarrow A$, and hence $\{B\} \vdash A$). The prime theory $\Gamma_{\mathbf{H}}$ is a particular case of a Γ_B where B is the Harrop formula $q \rightarrow q$.

When we build the canonical Kripke model for **H**, if we assume the definition of the canonical valuation v_c, then A-Heredity implies for every Γ and Δ:

(II) $\Gamma\ R_c\ \Delta \ \Rightarrow\ \Gamma \subseteq \Delta$,

Converse A-Heredity implies for every Γ and every B:

(III) $B \notin \Gamma \ \Rightarrow\ \exists\Delta(\Gamma\ R_c\ \Delta\ \ \&\ \ B \notin \Delta)$,

and one direction of the condition $(v \rightarrow)$ implies for every Γ and every A and B:

(I $\Leftarrow$) $A \rightarrow B \notin \Gamma \ \Rightarrow\ \exists\Delta(\Gamma\ R_c\ \Delta\ \ \&\ \ A \in \Delta\ \ \&\ \ B \notin \Delta)$.

The other direction of $(v \rightarrow)$ and the conditions $(v\,\bot)$, $(v\,\wedge)$ and $(v\,\vee)$, together with (II), (III) and the requirement that $\Gamma_{\mathbf{H}}$ be included in every Γ, imply the definition of prime theories. For prime theories Γ and Δ we have that (I $\Leftarrow$) implies (III), and (II) is equivalent with the converse of (I $\Leftarrow$). So (II) and (I $\Leftarrow$) are equivalent with (I) of the proof of Proposition 12, and as this proof shows, (I) is necessary and sufficient for verifying that $\langle W_c, R_c, v_c\rangle$ is a rudimentary Kripke model.

In the usual canonical Kripke model for **H** we take for R_c the largest relation possible, and we identify R_c with the subset relation $\subseteq$. But, as our canonical rudimentary Kripke model shows, we need not do that. We can take a relation on prime theories properly included in $\subseteq$ which will also satisfy (I). Our new relation R_c is serial, but it is not reflexive. It is also transitive, though this is not needed for rudimentary Kripke models.We leave open the following question:

(4) Can we define on W_c a relation R_c which satisfies (I) and is not transitive, or neither transitive nor reflexive?

Let now W_c be the set of all prime theories which for some A are A-maximal, and define on this W_c the relation R_c and v_c as for our canonical rudimentary Kripke model for **H**. Then R_c coincides with $\subseteq$, and $\langle W_c, R_c, v_c\rangle$ is a partially ordered Kripke model for **H**. To verify this we establish that if *not* $\Gamma \vdash A$, then there is a prime theory Δ such that $\Gamma \subseteq \Delta$ and Δ is A-maximal. So our canonical rudimentary Kripke model for **H** has a partially ordered canonical Kripke model for **H** as a proper submodel.

In the background of our canonical rudimentary Kripke model for **H** lies a representation theorem for Heyting algebras. For a Heyting algebra $\langle \mathcal{A}, \rightarrow, \wedge, \vee, \bot \rangle$, let $W_{\mathcal{A}} = \{x : x \textit{ is a prime filter of } \mathcal{A}\}$. For $a \in \mathcal{A}$, a prime filter x will be called *a-maximal* iff $a \notin x$ and $(\forall b \in \mathcal{A})(b \in x \textit{ or } b \rightarrow a \in x)$. Then we define on $W_{\mathcal{A}}$ the following relation analogous to our R_c:

$$x \, R_{\mathcal{A}} \, y \iff (x = y \;\; \& \;\; (\exists a \in \mathcal{A}) x \textit{ is a-maximal}) \textit{ or } x \subset y.$$

If $f(a) = \{x \in W_{\mathcal{A}} : a \in x\}$, then $\langle \{f(a) : a \in \mathcal{A}\}, \rightarrow_{R_{\mathcal{A}}}, \cap, \cup, \varnothing \rangle$ is a Heyting algebra isomorphic to our initial Heyting algebra $\mathcal{A}$ by the mapping f. As before, the relation $R_{\mathcal{A}}$, though it must be transitive, need not be reflexive. If $\mathcal{A}$ is the Lindenbaum algebra of **H** and x is the principal filter generated by the equivalence class of a Harrop formula nonequivalent to $\bot$, then x is a prime filter for which there is no $a \in \mathcal{A}$ such that x is a-maximal (which is shown by an argument analogous to what we had above for the prime theories Γ_B).

3. Inductive Kripke models

In this section we study frames for which it is enough to assume that pseudo-valuations on them satisfy p-Heredity and Converse p-Heredity in order to infer by induction on the complexity of A that A-Heredity and Converse A-Heredity are satisfied. These frames and the corresponding rudimentary Kripke models, which we will call *inductive*, will be more like ordinary frames and Kripke models for **H**, but we shall see that they need not be quasi-ordered.

In ordinary Kripke models for **H** besides the conditions for pseudo-valuations we assume only p-Heredity. That A-Heredity is satisfied in full generality is then proved by induction on the complexity of A. The transitivity of R is a sufficient condition for this induction to go through. (Actually, in the induction step we use the transitivity of R only for the case when A is of the form $A_1 \rightarrow A_2$, and then we don't need the induction hypothesis; see Proposition 23 below. The instance of A-Heredity where A is $\bot$ is satisfied vacuously.) Before showing that the necessary and sufficient condition on frames for this induction to go through is weaker than transitivity we introduce the following notions.

For a frame $\langle W, R \rangle$ and an arbitrary $X \subseteq W$ let:

$$\mathbf{Cone}\, X = \{y : (\exists x \in X)(\exists k \geq 0)\, x \, R^k \, y\},$$
$$\mathbf{Cone}^- X = \{y : \textit{not } (\exists x \in X)(\exists k \geq 0) y \, R^k \, x\}.$$

The operations **Cone** and **Cone**$^-$ are connected with hereditary sets by the following two propositions, whose straightforward proofs will be omitted:

Proposition 13. *For every frame $\langle W, R\rangle$ and every $X \subseteq W$, the set* $\mathbf{Cone}\, X$ *is the least hereditary superset of* X.

As a corollary of this proposition we obtain that X is hereditary iff $\mathbf{Cone}\, X = X$.

Proposition 14. *For every frame $\langle W, R\rangle$ and every $X \subseteq W$, the set* $\mathbf{Cone}^{-} X$ *is the greatest hereditary set disjoint from* X.

So, in particular, the sets:

$$\mathbf{Cone}\{x\} = \{y : (\exists k \geq 0)\, x\, R^k\, y\},$$
$$\mathbf{Cone}^{-}\{x\} = \{y : \mathit{not}\, (\exists k \geq 0)\, y\, R^k\, x\}$$

are hereditary. It is clear that $y \in \mathbf{Cone}\{x\}$ iff $x \notin \mathbf{Cone}^{-}\{y\}$. It is also clear that for every $X \subseteq W$ in a frame $\langle W, R\rangle$ the sets:

$$\mathbf{Cone}_m\, X = \{y : (\exists x \in X)(\exists k \geq m)\, x\, R^k\, y\},$$
$$\mathbf{Cone}_m^{-}\, X = \{y : \mathit{not}\, (\exists x \in X)(\exists k \geq m)\, y\, R^k\, x\},$$

where $m \geq 0$, are hereditary.

We shall say that a relation R in a frame $\langle W, R\rangle$ is *weakly transitive* iff

$$\forall x, z\, (x\, R^2\, z \;\Rightarrow\; \exists t\, (x\, R\, t \;\;\&\;\; t \in \mathbf{Cone}\{z\} \;\;\&\;\; z \in \mathbf{Cone}\{t\})).$$

Then we can prove:

Proposition 15. *In a frame $\langle W, R\rangle$ the relation R is weakly transitive iff for every pseudo-valuation v on $\langle W, R\rangle$, if v satisfies p-Heredity, then v satisfies A-Heredity.*

Proof. From left to right we proceed by induction on the complexity of A in order to show that v satisfies A-Heredity. The crucial case in the induction step is when A is of the form $A_1 \to A_2$. Suppose for some x and y that $x \in v(A_1 \to A_2)$, $x\ R\ y$ and $y \notin v(A_1 \to A_2)$. Then for some z we have $y\ R\ z$, $z \in v(A_1)$ and $z \notin v(A_2)$. So there is a t such that $x\ R\ t$, $(\exists k \geq 0)\, z\, R^k\, t$ and $(\exists m \geq 0)\, t\, R^m\, z$. By the induction hypothesis we get $t \in v(A_1)$, but since $x\ R\ t$ we also have $t \in v(A_2)$. Then again by the induction hypothesis $z \in v(A_2)$, which is a contradiction.

For the other direction suppose that for some x, y and z we have $x\ R\ y$, $y\ R\ z$ and

$$\forall t\, ((x\, R\, t \;\;\&\;\; t \in \mathbf{Cone}\{z\}) \;\Rightarrow\; z \notin \mathbf{Cone}\{t\}).$$

Then by Propositions 13 and 14 it is clear that there is a pseudo-valuation v which satisfies p-Heredity such that $v(p_1) = \mathbf{Cone}\{z\}$ and $v(p_2) = \mathbf{Cone}^{-}\{z\}$. We know that $z \notin \mathbf{Cone}\{t\}$ iff $t \in \mathbf{Cone}^{-}\{z\}$. It follows that $x \in v(p_1 \to p_2)$, but since $z \in v(p_1)$ and $z \notin v(p_2)$ we have $y \notin v(p_1 \to p_2)$. So v does not satisfy A-Heredity. *q.e.d.*

This proposition shows that in every weakly transitive frame $\langle W, R\rangle$ the set $\mathcal{A}$ of *all* hereditary subsets of W contains $\varnothing$ and is closed under the operations $\to_R$, $\cap$ and $\cup$. Moreover, the weak transitivity of R is not only sufficient but also necessary for that to be the case. The algebra $\langle \mathcal{A}, \rightarrow_R, \cap, \cup, \emptyset \rangle$ is a distributive lattice with zero which for every $X, Y, Z \in \mathcal{A}$ satisfies:

$$X \cap Y \subseteq Z \ \Rightarrow \ X \subseteq Y \rightarrow_R Z.$$

The converse implication may fail. For example, let $X = W \rightarrow_R \emptyset$, $Y = W$ and $Z = \emptyset$; then a ***dead end*** x, i.e. a point x such that there is no y for which $x \ R \ y$, will belong to $(W \rightarrow_R \emptyset) \cap W$ but cannot belong to $\emptyset$. So our algebra $\mathcal{A}$ is not necessarily a Heyting algebra.

Weak transitivity is satisfied by transitive frames, but it is clear that this is a weaker condition than transitivity. This condition is exclusively tied to the connective $\rightarrow$ and it is not invoked in any other part of the proof of Proposition 15, not involving $\rightarrow$.

However, in connexion with rudimentary Kripke models we are not interested in inferring A-Heredity from p-Heredity, as we did in the proof of Proposition 15, but we want to infer A-Heredity and Converse A-Heredity from p-Heredity and Converse p-Heredity. In other words, we want to infer that a pseudo-valuation which satisfies p-Heredity and Converse p-Heredity is a valuation. In order to show what are the necessary and sufficient conditions on frames for this, we shall introduce the following notions.

For a frame $\langle W, R \rangle$ a nonempty subset X of W will be called an ***ω-chain from x*** iff there is a mapping f from the ordinal ω ***onto*** X such that $f(0) = x$ and $(\forall n \in \omega) f(n) \ R \ f(n+1)$. Let $\omega(x) = \{X \subseteq W : X \textit{ is an } \omega\textit{-chain from } x\}$. An ω-chain from x makes an infinite sequence $x_0 x_1 x_2 \ldots$ such that $x_0 = x$ and for every $n \geq 0$ we have $x_n \ R \ x_{n+1}$. Since f in the definition of ω-chains need not be ***one-one***, there may be repetitions in the sequence $x_0 x_1 x_2 \ldots$, and an ω-chain need not be infinite; it may actually be the singleton $\{x\}$ if $x \ R \ x$. In arbitrary frames there may be points x such that $\omega(x)$ is empty; for example, x may be a dead end. For every x the set $\omega(x)$ is nonempty iff our frame is serial.

For a frame $\langle W, R \rangle$ and $X \subseteq W$ let

$$\mathbf{Cl}_\omega X = \{y : (\forall Y \in \omega(y)) \, Y \cap X \neq \emptyset\}.$$

The set $\mathbf{Cl}_\omega X$ contains all the points y such that every ω-chain from y intersects X. Every y such that $\omega(y)$ is empty will also belong to $\mathbf{Cl}_\omega X$, since for such a y it is vacuously satisfied that every ω-chain from y interesects X. The operation $\mathbf{Cl}_\omega$ satisfies for every $X, Y \subseteq W$:

$$X \subseteq \mathbf{Cl}_\omega X,$$
$$\mathbf{Cl}_\omega \mathbf{Cl}_\omega X = \mathbf{Cl}_\omega X,$$
$$\mathbf{Cl}_\omega X \cup \mathbf{Cl}_\omega Y \subseteq \mathbf{Cl}_\omega (X \cup Y),$$

and in serial frames we also have $\mathbf{Cl}_\omega \emptyset = \emptyset$, but we need not have:

$$\mathbf{Cl}_\omega (X \cup Y) \subseteq \mathbf{Cl}_\omega X \cup \mathbf{Cl}_\omega Y.$$

So $\mathbf{Cl}_\omega$ is not quite a topological closure operation. If X and Y are hereditary subsets of W, we also have:

$$\mathbf{Cl}_\omega(X \cap Y) = \mathbf{Cl}_\omega X \cap \mathbf{Cl}_\omega Y.$$

The operation $\mathbf{Cl}_\omega$ is analogous to an operation which naturally arises in connexion with Beth models (see [5], 3.2, and the section on rudimentary Beth modelsin [4]).

It is clear that if X is a conversely hereditary subset of W in a frame $\langle W, R\rangle$ and $y \notin X$, then $(\exists Y \in \omega(y))\, Y \cap X = \varnothing$. The operation $\mathbf{Cl}_\omega$ in connected with conversely hereditary sets by the following proposition:

Proposition 16. *For every frame $\langle W, R\rangle$ and every $X \subseteq W$, the set $\mathbf{Cl}_\omega X$ is the least conversely hereditary superset of X.*

Proof. To show that $\mathbf{Cl}_\omega X$ is conversely hereditary suppose $y \notin \mathbf{Cl}_\omega X$. Then there is a $Y \in \omega(y)$ such that $Y \cap X = \varnothing$ and a $z \in Y$ such that $y\ R\ z$. The set $Y \cap \mathbf{Cone}\{z\}$ is an ω-chain from z disjoint from X, i.e. $z \notin \mathbf{Cl}_\omega X$.

To show that $\mathbf{Cl}_\omega X$ is the least conversely hereditary superset of X suppose Y is conversely hereditary and $X \subseteq Y$, and let there be an x such that $x \in \mathbf{Cl}_\omega X$, i.e. $(\forall Z \in \omega(x))\, Z \cap X \neq \varnothing$, and $x \notin Y$. Since Y is conversely hereditary there is a $Z' \in \omega(x)$ such that $Z' \cap X = \varnothing$, which is a contradiction. So $\mathbf{Cl}_\omega X \subseteq Y$. *q.e.d.*

As a corollary of this proposition we obtain that X is conversely hereditary iff $\mathrm{Cl}_\omega X = X$.

Propositions 16 and 13 show that $\mathbf{Cl}_\omega$ is analogous to **Cone**. Is there an operation analogous to $\mathbf{Cone}^-$, which applied to X would give the greatest conversely hereditary set disjoint from X? The following example shows that such an operation need not exist. Lef $W = \{0,1,2\}$ and $R = \{\langle 0,1\rangle, \langle 0,2\rangle, \langle 1,1\rangle, \langle 2,2\rangle\}$. Then *the* greatest conversely hereditary set disjoint from $\{0\}$ does not exist ($\{1\}$ and $\{2\}$ are conversely hereditary, but $\{1,2\}$ is not).

The following proposition connects the operation $\mathbf{Cl}_\omega$, and conversely hereditary sets, with reflexivity:

Proposition 17. *In a frame $\langle W, R\rangle$ the relation R is reflexive iff for every $X \subseteq W$ we have $\mathrm{Cl}_\omega X = X$.*

Proof. ($\Rightarrow$) Suppose R is reflexive and $x \in \mathbf{Cl}_\omega X$. Then $\{x\} \in \omega(x)$, and hence $x \in X$.

($\Leftarrow$) Suppose for some x *not* $x\ R\ x$. Then for every $Y \in \omega(x)$ we have $Y \cap \{y : x\ R\ y\} \neq \varnothing$, i.e. $x \in \mathbf{Cl}_\omega\{y : x\ R\ y\}$, but $x \notin \{y : x\ R\ y\}$. *q.e.d.*

The following proposition connects the operation $\mathbf{Cl}_\omega$ with hereditary sets:

Proposition 18. *For every frame* $\langle W, R\rangle$ *and every hereditary* $X \subseteq W$, *the set* $\mathbf{Cl}_\omega X$ *is hereditary.*

Proof. Suppose $x \in \mathbf{Cl}_\omega X$, $x \; R \; y$ and $Y \in \omega(y)$. Then $\{x\} \cup Y \in \omega(x)$, and hence $(\{x\} \cup Y) \cap X \neq \emptyset$, i.e. for some $z \in \{x\} \cup Y$ we have $z \in X$. If $z \in Y$, then $Y \cap X \neq \emptyset$. If $z \notin Y$, then $z = x$, and $y \in Y \cap X$ since X is hereditary. So $y \in \mathbf{Cl}_\omega X$. *q.e.d.*

Hence, $\mathbf{Cone}\,\mathbf{Cl}_\omega\,\mathbf{Cone}\,X = \mathbf{Cl}_\omega\,\mathbf{Cone}\,X$. However, we don't always have $\mathbf{Cl}_\omega\,\mathbf{Cone}\,\mathbf{Cl}_\omega X \subseteq \mathbf{Cone}\,\mathbf{Cl}_\omega X$, i.e. we can have a conversely hereditary set X such that $\mathbf{Cone}\,X$ is not conversely hereditary. For example, let our frame have $W = \{a, b\} \cup \{0, 1, 2, \dots\}$ and let $R = \{\langle a, 0\rangle, \langle b, 0\rangle\} \cup \{\langle n, n+1\rangle : n \in \omega\}$. Then $\{a\}$ is conversely hereditary, but $\mathbf{Cone}\{a\}$ is not, since $b \notin \mathbf{Cone}\{a\}$ and $b \in \mathbf{Cl}_\omega\,\mathbf{Cone}\{a\}$.

We are now ready to show what are the necessary and sufficient conditions on frames for inferring that every pseudo-valuation which satisfies p-Heredity and Converse p-Heredity is a valuation. We shall say that in a frame $\langle W, R\rangle$ the relation R is *prototransitive* iff

$$\forall x, z\, (x\, R^2\, z \;\Rightarrow\; (\forall Z \in \omega(z)) \exists t\, (x\, R\, t \;\;\&\;\; t \in \mathbf{Cl}_\omega\,\mathbf{Cone}\{z\} \;\;\&\;\; t \notin \mathbf{Cl}_\omega\,\mathbf{Cone}^-\, Z)).$$

This condition says that if $x\, R^2\, z$, then for every ω-chain Z from z there is a t such that $x \; R \; t$, every ω-chain from t intersects $\mathbf{Cone}\{z\}$, and there is an ω-chain from t which is disjoint from $\mathbf{Cone}^-\, Z$. If for $X, Y \subseteq W$ we have that for every $x \in X$ there is a $y \in Y$ such that $(\exists k \geq 0)\, x\, R^k\, y$ (i.e. for every $x \in X$ we have $Y \cap \mathbf{Cone}\{x\} \neq \emptyset$) we shall say that X is a *shadow* of Y. The last conjunct above, which claims that there is an ω-chain from t which is disjoint from $\mathbf{Cone}^-\, Z$, says that this ω-chain from t is a shadow of the ω-chain Z. Note that the consequent of the condition of prototransitivity is satisfied vacuously if $\omega(z)$ is empty.

Every transitive relation is prototransitive. For suppose R is transitive, $x\, R^2\, z$ and Z is an ω-chain from z. Then we have $x \; R \; z$, $z \in \mathbf{Cl}_\omega\,\mathbf{Cone}\{z\}$ and there is an ω-chain from z, namely Z itself, which is a shadow of the ω-chain Z. We also have that every weakly transitive relation is prototransitive, but of course prototransitivity entails neither weak transivity nor transivity.

We shall say that in a frame $\langle W, R\rangle$ the relation R is ***protoreflexive*** iff

$$\forall x\, (\forall X_1, X_2 \in \omega(x))\, \exists y\, (x\; R\; y \;\;\&\;\; y \notin \mathbf{Cl}_\omega\,\mathbf{Cone}^-\, X_1 \;\;\&\;\; y \notin \mathbf{Cl}_\omega\,\mathbf{Cone}^-\, X_2).$$

This condition says that if X_1 and X_2 are ω-chains from x, not necessarily distinct, then there is a y such that $x \; R \; y$ and from y we have an ω-chain which is a shadow of X_1 and an ω-chain which is a shadow of X_2. Every reflexive relation is protoreflexive. For if R is reflexive, then for ω-chains X_1 and X_2 from x we have $x \; R \; x$, and X_1 is a shadow of X_1 and X_2 a shadow of X_2. Of course, protoreflexivity does not entail reflexivity.

We can now prove the proposition for which we have been preparing all along:

Proposition 19. *In a frame* $\langle W, R\rangle$ *the relation* R *is serial, prototransitive and protoreflexive iff for every pseudo-valuation* v *on* $\langle W, R\rangle$, *if* v *satisfies p-Heredity and Converse p-Heredity, then* v *satisfies A-Heredity and Converse A-Heredity.*

Proof. ($\Rightarrow$) We proceed by induction on the complexity of A in order to show that v satisfies A-Heredity and Converse A-Heredity. In the basis of this induction we use the *seriality* of R in order to demonstrate that $v(\bot)$ is conversely hereditary. That $v(\bot)$ is hereditary is satisfied vacuously.

In the induction step we first prove that $v(A_1 \rightarrow A_2)$ is hereditary. So suppose for some x and y that $x \in v(A_1 \rightarrow A_2)$, $x\ R\ y$ and $y \notin v(A_1 \rightarrow A_2)$. Then for some z we have $y\ R\ z$, $z \in v(A_1)$ and $z \notin v(A_2)$. By the Converse A-Heredity of the induction hypothesis there is a $Z \in \omega(z)$ such that $Z \cap v(A_2) = \varnothing$. So by the *prototransitivity* of R there is for this Z a t such that $x\ R\ t$, $t \in \mathbf{Cl}_\omega\,\mathbf{Cone}\{z\}$ and $t \notin \mathbf{Cl}_\omega\,\mathbf{Cone}^- Z$. If $t \notin v(A_1)$, then there is a $U \in \omega(t)$ such that $U \cap v(A_1) = \varnothing$, which contradicts $t \in \mathbf{Cl}_\omega\,\mathbf{Cone}\{z\}$ and $\mathbf{Cone}\{z\} \subseteq v(A_1)$; we used the Converse A-Heredity and A-Heredity of the induction hypothesis. So $t \in v(A_1)$, and since $x\ R\ t$, we obtain $t \in v(A_2)$. But there is a $U \in \omega(t)$ which is a shadow of Z, and $U \subseteq v(A_2)$ by the A-Heredity of the induction hypothesis. This is in contradiction with $Z \cap v(A_2) = \varnothing$ and the A-Heredity of the induction hypothesis.

For the converse hereditariness of $v(A_1 \rightarrow A_2)$ suppose $x \notin v(A_1 \rightarrow A_2)$, i.e. there is a y such that $x\ R\ y$, $y \in v(A_1)$ and $y \notin v(A_2)$. By the converse A-Heredity of the induction hypothesis there is a z such that $y\ R\ z$ and $z \notin v(A_2)$. By the A-Heredity of the induction hypothesis $z \in v(A_1)$, and so $y \notin v(A_1 \rightarrow A_2)$. Note that we did not appeal to any particular property of R in this paragraph.

For the hereditariness and converse hereditariness of $v(A_1 \wedge A_2)$, and for the hereditariness of $v(A_1 \vee A_2)$, we do not appeal to any particular properties of R, and we will omit these easy cases. It remains to consider the converse hereditariness of $v(A_1 \vee A_2)$. So suppose $x \notin v(A_1 \vee A_2)$, i.e. $x \notin v(A_1)$ and $x \notin v(A_2)$. By the Converse A-Heredity of the induction hypothesis there is an $X_1 \in \omega(x)$ such that $X_1 \cap v(A_1) = \varnothing$ and an $X_2 \in \omega(x)$ such that $X_2 \cap v(A_2) = \varnothing$. So by the *protoreflexivity* of R there is a y such that $x\ R\ y$, $y \notin \mathbf{Cl}_\omega\,\mathbf{Cone}^- X_1$ and $y \notin \mathbf{Cl}_\omega\,\mathbf{Cone}^- X_2$. If $y \in v(A_1)$, then for the $Y \in \omega(y)$ which is a shadow of X_1 we would have $Y \subseteq v(A_1)$, which is in contradiction with $X_1 \cap v(A_1) = \varnothing$; we used the A-Heredity of the induction hypothesis. So $y \notin v(A_1)$, and we obtain analogously $y \notin v(A_2)$, which means $y \notin v(A_1 \vee A_2)$.

($\Leftarrow$) If R is not *serial*, then $v(\bot)$ is not conversely hereditary. Suppose R is not *prototransitive*, i.e. for some x, y and z we have $x\ R\ y$, $y\ R\ z$ and there is a $Z \in \omega(z)$ such that

$$\forall t\,((x\ R\ t\ \ \&\ \ t \in \mathbf{Cl}_\omega\,\mathbf{Cone}\{z\}) \Rightarrow t \in \mathbf{Cl}_\omega\,\mathbf{Cone}^- Z).$$

Then by Propositions 13, 14, 16 and 18 it is clear that there is a pseudo-valuation v which satisfies p-Heredity and Converse p-Heredity such that $v(p_1) = \mathbf{Cl}_\omega\,\mathbf{Cone}\{z\}$ and $v(p_2) = \mathbf{Cl}_\omega\,\mathbf{Cone}^- Z$. It follows that $x \in v(p_1 \rightarrow p_2)$. We also have $z \in v(p_1)$ and $z \notin v(p_2)$, because $Z \cap \mathbf{Cone}^- Z = \varnothing$ (otherwise for some $t \in Z$ we would have $t \in \mathbf{Cone}^- Z$, but $t\ R^0\ t$). So $y \notin v(p_1 \rightarrow p_2)$, and A-Heredity fails.

Suppose R is not ***protoreflexive***, i.e. for some x there are $X_1, X_2 \in \omega(x)$ such that

$$\forall y\,(x \; R \; y \;\Rightarrow\; (y \in \mathbf{Cl}_\omega \, \mathbf{Cone}^- X_1 \;\; or \;\; y \in \mathbf{Cl}_\omega \, \mathbf{Cone}^- X_2)).$$

Then by Propositions 14, 16 and 18 it is clear that there is pseudo-valuation v which satisfies p-Heredity and Converse p-Heredity such that $v(p_1) = \mathbf{Cl}_\omega \, \mathbf{Cone}^- X_1$ and $v(p_2) = \mathbf{Cl}_\omega \, \mathbf{Cone}^- X_2$. It follows that $\forall y(x \; R \; y \;\Rightarrow\; y \in v(p_1 \vee p_2))$. However, $x \notin v(p_1)$, because $X_1 \cap \mathbf{Cone}^- X_1 = \emptyset$, and analogously $x \notin v(p_2)$. So $x \notin v(p_1 \vee p_2)$, and Converse A-Heredity fails. *q.e.d.*

Proposition 19 shows that in every frame $\langle W, R \rangle$ which is serial, prototransitive and protoreflexive, the set $\mathcal{A}$ of *all* hereditary and conversely hereditary subsets of W contains $\emptyset$ and is closed under the operations $\rightarrow_R$, $\cap$ and $\cup$. Moreover, the seriality, prototransitivity and protoreflexivity of R are not only sufficient but also necessary for that to be the case. As we know, $\langle \mathcal{A}, \rightarrow_R, \cap, \cup, \emptyset \rangle$ is a Heyting algebra.

The seriality of R is exclusively tied to $\bot$, so that if we restrict ourselves to pseudo-valuations v from $\mathbf{L}^+$, we can omit the requirement of seriality from the left-hand side of Proposition 19. Similarly, prototransitivity is exclusively tied to $\rightarrow$ and protoreflexivity to $\vee$. So, if we restrict ourselves to pseudo-valuations v from $\mathbf{L}^\rightarrow$, we need to keep only the requirement of prototransitivity on the left-hand side of Proposition 19. The same holds if pseudo-valuations are from the $(\rightarrow, \wedge)$ fragment of $\mathbf{L}$, and if they are from the $(\rightarrow, \bot)$ or $(\rightarrow, \wedge, \bot)$ fragment, we need seriality and prototransitivity. Seriality is equivalent with the condition that $\emptyset$ is conversely hereditary, prototransitivity with the condition that for every $X, Y \subseteq W$ the set $\mathbf{Cl}_\omega \, \mathbf{Cone}\, X \rightarrow_R \mathbf{Cl}_\omega \, \mathbf{Cone}\, Y$ is hereditary, and protoreflexivity with the condition that for every $X, Y \subseteq W$ the set $\mathbf{Cl}_\omega \, \mathbf{Cone}\, X \cup \mathbf{Cl}_\omega \, \mathbf{Cone}\, Y$ is conversely hereditary. This is shown as in the proof of Proposition 19.

We shall call frames $\langle W, R \rangle$ where R is serial, prototransitive and protoreflexive *inductive frames*. An *inductive Kripke model* is then defined as a $\langle W, R, v \rangle$ such that $\langle W, R \rangle$ is an inductive frame and v, called an ***inductive valuation***, is a pseudo–valuation which satisfies p-Heredity and Converse p-Heredity. Proposition 19 guarantees that inductive valuations on inductive frames are valuations, i.e. that inductive Kripke models are rudimentary Kripke models. In every inductive Kripke model $\langle W, R, v \rangle$, for every propositional variable p, the set $v(p)$ may be any hereditary and conversely hereditary subset of W we choose. In an arbitrary rudimentary Kripke model this is not the case, since the set of *all* hereditary and conversely hereditary subsets of W need not be closed under the operations $\rightarrow_R$, $\cap$ and $\cup$. A fortiori, it will not be a Heyting algebra with these operations.

If holding in frames is defined in terms of inductive valuations, instead of valuations of rudimentary Kripke models, then $\mathbf{H}$ is not sound with respect to serial frames but it is sound and complete with respect to inductive frames. Inductive frames do not make the largest class of frames which would give this soundness and completeness, because pseudo-valuations need not satisfy exactly A-Heredity and Converse A-Heredity in order to secure the soundnes of $\mathbf{H}$. As we will show in the last section of [4], somewhat weaker forms of these conditions will also do. However, we know that A-Heredity and Converse A-Heredity are necessary for the strong soundness and completeness of Proposition 11.

We may also envisage frames of the type $\langle W, R, \mathcal{A}' \rangle$, called *general inductive frames*, where $\langle W, R \rangle$ is an inductive frame and $\mathcal{A}'$ is a subalgebra of the Heyting algebra $\langle \mathcal{A}, \rightarrow_R, \cap, \cup, \varnothing \rangle$ of *all* hereditary and conversely hereditary subsets of W. These frames are analogous to the *general*, or *first-order*, frames in modal logic (see [7], pp. 62–67). If we restrict inductive valuations on $\langle W, R, \mathcal{A}' \rangle$ to those which take values in $\mathcal{A}'$, we obtain that **H** is sound and complete with respect to all general inductive frames. Ordinary inductive frames may be conceived as a particular type of general inductive frames where $\mathcal{A}'$ is the Heyting algebra of *all* hereditary and conversely hereditary subsets of W. At an even more general level we would have *general rudimentary frames* of the type $\langle W, R, \mathcal{A} \rangle$ where $\langle W, R \rangle$ is a serial frame and $\mathcal{A}$ a particular set of hereditary and conversely hereditary subsets of W which contains $\varnothing$ and is closed under the operations $\rightarrow_R$, $\cap$ and $\cup$. Valuations on these frames would be restricted to those which take values in $\mathcal{A}$.

If holding in frames is defined in terms of pseudo-valuations which satisfy only p-Heredity, as for ordinary Kripke models for **H**, then **H** is not sound with respect to inductive frames. We know that in this sense **H** is sound and complete with respect to quasi-ordered frames, but there is an interesting class of frames properly in between the class of inductive frames and the class of quasi-ordered frames with respect to which **H** is also sound and complete in this sense. This is the largest class of frames such that every pseudo-valuation on a frame in this class which satisfies p-Heredity will also satisfy A-Heredity and Converse A-Heredity. Frames in this class, called *weakly quasi-ordered frames*, satisfy *weak reflexivity*:

$$\forall x\, (\exists k \geq 1)\, x\, R^k\, x$$

and *weak transitivity* from Proposition 15:

$$\forall x, z\, (x\, R^2\, z \;\Rightarrow\; \exists t (x\, R\, t \;\&\; t \in \mathbf{Cone}\{z\} \;\&\; z \in \mathbf{Cone}\{t\})).$$

Reflexivity of course entails weak reflexivity but not vice versa. Also every quasi-ordered frame is weakly quasi-ordered but not vice versa.

The following proposition about weak reflexivity is analogous to Proposition 17:

Proposition 20. *In a frame $\langle W, R \rangle$ the relation R is weakly reflexive iff for every hereditary $X \subseteq W$ we have $\mathbf{Cl}_\omega X = X$.*

Proof. ($\Rightarrow$) Suppose R is weakly reflexive, $X \subseteq W$ is hereditary and $x \in \mathbf{Cl}_\omega X$. From weak reflexivity it follows that there is an ω-chain Z from x in which x is cyclically repeated. Since $Z \cap X \neq \varnothing$ and X is hereditary we get $x \in X$.

($\Leftarrow$) Suppose for some x *not* $(\exists k \geq 1)\, x\, R^k\, x$. Then the set $\mathbf{Cone}_1\{x\} = \{y : (\exists k \geq 1)\, x\, R^k\, y\}$ is hereditary and $x \in \mathbf{Cl}_\omega\, \mathbf{Cone}_1\{x\}$, but $x \notin \mathbf{Cone}_1\{x\}$. *q.e.d.*

As a corollary we obtain:

Proposition 21. *In a frame $\langle W, R\rangle$ the relation R is weakly reflexive iff for every pseudo-valuation v on $\langle W, R\rangle$, if v satisfies p-Heredity, then v satisfies Converse p-Heredity.*

Propositions 15 and 21 from right to left show that weak reflexivity and weak transitivity are necessary if we want to infer A-Heredity and Converse A-Heredity from p-Heredity. That these conditions are also sufficient follows from Propositions 15 and 20 from left to right.

We can easily verify that weakly quasi-ordered frames could alternatively be defined by assuming weak reflexivity and prototransitivity. So weak reflexivity, which entails seriality and protoreflexivity, is really the new assumption we make when we pass from the class of inductive frames to its proper subclass made of all weakly quasi-ordered frames.

We have already shown in Proposition 15 that the weak transitivity of R in a frame $\langle W, R\rangle$ is necessary and sufficient for the set $\mathcal{A}$ of *all* hereditary subsets of W to contain $\emptyset$ and be closed under the operations $\rightarrow_R$, $\cap$ and $\cup$. However, $\langle \mathcal{A}, \rightarrow_R, \cap, \cup, \emptyset\rangle$ need not have been a Heyting algebra. Proposition 20 shows that the weak reflexivity of R is necessary and sufficient to make every member of $\mathcal{A}$ conversely hereditary. So for every weakly quasi-ordered frame, $\langle \mathcal{A}, \rightarrow_R, \cap, \cup, \emptyset\rangle$ is a Heyting algebra.

We also prove the following opposite of Proposition 21:

Proposition 22. *In a frame $\langle W, R\rangle$ we have*

$$\forall x, y\,(x\ R\ y\ \Rightarrow\ (\forall Y \in \omega(y))\, x \in Y)$$

iff for every pseudo-valuation v on $\langle W, R\rangle$, if v satisfies Converse p-Heredity, then v satisfies p-Heredity.

Proof. ($\Rightarrow$) Suppose $x \in v(p)$, $x\ R\ y$ and $y \notin v(p)$. Then by Converse p-Heredity there is a $Y \in \omega(y)$ such that $Y \cap v(p) = \emptyset$. But since $x\ R\ y$ we have that $x \in Y$, which is a contradiction.

($\Leftarrow$) Suppose for some x and y that $x\ R\ y$ and there is a $Y \in \omega(y)$ such that $x \notin Y$. Then by Proposition 16 it is clear that there is a pseudo-valuation which satisfies Converse p-Heredity such that $v(p) = \mathbf{Cl}_\omega\{x\}$. We infer that $x \in v(p)$ and $y \notin v(p)$, i.e. p-Heredity fails. *q.e.d.*

This proposition shows what happens if we define holding in frames in terms of pseudo-valuations which satisfy only Converse p-Heredity and expect these pseudo-valuations to give rise to rudimentary Kripke models. We know that frames for rudimentary Kripke models must be serial, and with seriality the condition of Proposition 22 entails that if $x\ R\ y$, then $x \in \mathbf{Cone}\{y\}$, which with A-Heredity would give that for every A we have $x \in v(A)$ iff $y \in v(A)$. But with that, every theorem of the classical propositional calculus would hold. Of course, the condition of Proposition 22 need not be satisfied by quasi-ordered frames.

We have characterized inductive frames and weakly quasi-ordered frames by conditions on pseudo-valuations which indicate that we can define inductively valuations on these frames. Is there a similar characterization of quasi-ordered frames in terms of conditions necessary and sufficient to make valuations inductively definable in some way? The condition on pseudo-valuations which corresponds to reflexivity is contained in Proposition 17, which says that in reflexive frames $\langle W, R\rangle$, and only in reflexive frames, every subset of W is conversely hereditary. So reflexivity implies Converse A-Heredity. On the other hand, transitivity secures the hereditariness of $v(A_1 \rightarrow A_2)$, as the following proposition shows:

Proposition 23. *In a frame $\langle W, R\rangle$ the relation R is transitive iff for every pseudo-valuation v on $\langle W, R\rangle$ and every A_1 and A_2 the set $v(A_1 \rightarrow A_2)$ is hereditary.*

Proof. ($\Rightarrow$) Suppose $x \in v(A_1 \rightarrow A_2)$, $x\ R\ y$, $y\ R\ z$ and $z \in v(A_1)$. Then by the transitivity of R we have $x\ R\ z$, and hence $z \in v(A_2)$.

($\Leftarrow$) Suppose $x\ R\ y$, $y\ R\ z$ and *not* $x\ R\ z$. Let $v(p_1) = \{z\}$ and $v(p_2) = \emptyset$ (we may also take $v(p_2) = W - \{z\}$). It follows that $x \in v(p_1 \rightarrow p_2)$ and $y \notin v(p_1 \rightarrow p_2)$. *q.e.d.*

This proposition from left to right shows that when for transitive frames we prove by induction on the complexity of A that pseudo-valuations on them which satisfy p-Heredity satisfy A-Heredity, in the induction step we don't need the induction hypothesis for the case when A is of the form $A_1 \rightarrow A_2$.

Though the conditions corresponding to reflexivity and transitivity are sufficient for the inductive character of valuations, it is not clear what conception of this inductive character would make reflexivity and transitivity also necessary. No doubt, quasi-ordered Kripke models stand out by their simplicity and naturalness, and they are not very far from weakly quasi-ordered rudimentary Kripke models. But it is not clear how the exclusive concern with quasi-ordered Kipke models could be justified by saying that only these models would work.

The previous results show that the assumptions of reflexivity and transitivity for ordinary quasi-ordered Kripke models for **H** are not exactly in the same position. Transitivity secures prototransitivity and weak transitivity, which are tied to implication. Reflexivity secures protoreflexivity, which is tied to disjunction, but it secures also seriality and weak reflexivity, which are not tied to disjunction. Reflexivity also secures at one stroke Converse A-Heredity. With reflexivity we have reduced an assumption about valuations to an assumption purely about frames, which does not mention valuations.

If reflexivity is written as $R^0 \subseteq R$, the converse condition $R \subseteq R^0$ would be sufficient for A-Heredity as reflexivity is sufficient for Converse A-Heredity. However, though we can replace Converse A-Heredity by reflexivity, we cannot replace A-Heredity by $R \subseteq R^0$. By assuming $R \subseteq R^0$ we would immediately bring in classical propositional logic.

Among inductive frames we find frames in which R is serial, transitive and satisfies *branching density*:

$$\forall x, x_1, x_2\, ((x\ R\ x_1\ \ \&\ \ x\ R\ x_2) \Rightarrow \exists y (x\ R\ y\ \ \&\ \ y\ R\ x_1\ \ \&\ \ y\ R\ x_2)),$$

which is a stronger version of protoreflexivity. These inductive frames need not be weakly quasi-ordered. That **H** is sound and complete with respect to these frames, with inductive valuations, was shown in [3]. These frames are interesting because they are the frames with respect to which the normal modal propositional logic **K4N** can be shown sound and complete (of course, with usual modal valuations on these frames). The system **K4N** is axiomatized by adding to the weakest normal modal propositional logic **K** the following axiom-schemata:

(**s**) $\neg\Box\neg(A \to A)$,
(**t**) $\Box A \to \Box\Box A$,
(**bd**) $\Box(\Box A \vee \Box B) \to (\Box A \vee \Box B)$.

This system is the weakest normal modal propositional logic in which **H** can be embedded by the modal translation which prefixes $\Box$ to every proper subformula which is a propositional variable or an implication. (This is shown in [3]; the language in [3] has $\neg$ as primitive instead of $\bot$, but the modal translation just mentioned does not differ essentially from the translation considered there in connexion with the minimality of **K4N** since here $\Box$ is not prefixed to $\bot$.) The schema (**bd**) defines branching density on frames, in the sense that a frame satisfies branching density iff every instance of (**bd**) holds in this frame (with respect to usual modal valuations). That in the same sense (**s**) defines seriality, and (**t**) transitivity, are among the oldest examples in the correspondence theory of modal logic (see [1]).

It is not clear whether the other conditions we have met in connexion with inductive frames: weak reflexivity, weak transitivity, protoreflexivity and prototransitivity, may be defined by modal schemata. The sentences by which we have introduced these conditions are not first-order. The following first-order condition related to weak reflexivity:

$$\forall x\,(x\,R^k\,x),$$

where $k \geq 0$, is defined by the modal schema $\Box^k A \to A$, where $\Box^0 A$ is A and $\Box^{k+1} A$ is $\Box\Box^k A$. Similarly, the following first-order condition related to weak transitivity:

$$\forall x, z\,(x\,R^2\,z \;\Rightarrow\; \exists t\,(x\,R\,t \;\;\&\;\; z\,R^k\,t \;\;\&\;\; t\,R^m\,z)),$$

where $k, m \geq 0$, is defined by:

$$\Box(B \to \Box^m C) \to \Box^2(\Box^k B \to C).$$

The schema (**t**), i.e. $\Box A \to \Box^2 A$, is equivalent to this schema when $k = m = 0$. However, this does not yet solve the question:

(5) Are weak reflexivity, weak transitivity, protoreflexivity and prototransitivity definable by modal schemata?

A similar, but different, question is:

(6) Can we axiomatize sets of modal formulae which hold in inductive frames, or weakly quasi-ordered frames?

If a modal system **M** is sound and complete with respect to a class of frames C (via modal valuations) and **H** is also sound and complete with respect to C (via rudimentary Kripke model valuations), we cannot immediately conclude that **H** must be embeddable in **M** by a modal translation. For example, the modal system sound and complete with respect to serial frames is **D**, i.e. **K**+(**s**), but **H** cannot be embedded in **D** by the modal translation which embeds **H** in **K4N**; we obtain the same thing with several other natural modal translations (as will be shown in a paper devoted to modal translations in normal modal logics), and it is unlikely that any modal translation would work. To put it roughly, it is as if p-Heredity and Converse p-Heredity require an infinity of operators $\Box$ to be prefixed to every p, and in the absence of modality reduction principles like $\Box A \leftrightarrow \Box\Box A$, which is probavle in **K4N**, no finite amount of operators $\Box$ would do.

The canonical rudimentary Kripke model for **H** of the previous section, though serial and transitive, is not reflexive. It is clear that it is also not weakly reflexive. However, we leave open the following question:

(7) Does the frame of the canonical rudimentary Kripke model for **H** satisfy branching density, or at least protoreflexivity?

With this question we conclude our preliminary investigation of rudimentary Kripke models. As announced in the introduction, we shall consider some further topics related to rudimentary Kripke models in [4].

Acknowledgements

I would like to thank Djordje Čubrić, Kit Fine, Silvio Ghilardi, Valentin Šehtman, Predrag Tanović and Slobodan Vujošević for stimulating discussions about several points of this paper and of [4].

References

[1] Benthem, J. van, Correspondence theory, *Handbook of Philosophical Logic*, vol. II, D. M. Gabbay and F. Guenthner eds, Reidel, Dordrecht, 1984, pp. 167–247.

[2] Corsi, G., Weak logics with strict implication, *Z. Math. Logik Grundlag. Math.* 33 (1987), pp. 389–406.

[3] Došen, K., Normal modal logics in which the Heyting propositional calculus can be embedded, *Heyting: Mathematical Logic*, P. Petkov ed., Plenum, New York (to appear; abstract: *Polish. Acad. Sci. Inst. Philos. Sociol. Bull. Sect. Logic* 17 (1988), no 1, pp. 23–33).

[4] Došen, K., Rudimentary Beth models and conditionally rudimentary Kripke models for the Heyting propositional calculus (*to appear*).

[5] Dragalin, A. G., *Mathematical Intuitionism: Introduction to Proof Theory* (in Russian), Nauka, Moscow, 1979 (English translation: American Mathematical Society, Providence, 1988).

[6] Harrop, R., Concerning formulas of the types $A \to B \vee C$, $A \to (Ex)B(x)$ in intuitionistic formal systems, *J. Symbolic Logic* 25 (1960), pp. 27–32.

[7] Hughes, G. E., and M. J. Cresswell, *A Companion to Modal Logic*, Methuen, London, 1984.

[8] Prawitz, D., *Natural Deduction: A Proof-Theoretical Study*, Almqvist & Wiksell, Stockholm, 1965.

[9] Rodenburg, P. H., *Intuitionistic Correspondence Theory*, doctoral dissertation, University of Amsterdam, 1986.

[10] Šehtman, V. B., On incomplete propositional logics (in Russian), *Dokl. Akad. Nauk SSSR* 235 (1977), pp. 542–545 (English translation: *Soviet Math. Dokl.* 18 (1977), pp. 985–989).

Valuation Semantics for Modal Logics

(Abstract)

Cesar A. Mortari

In this paper we make a presentation of *valuation semantics* for some systems of modal logic, and of its main byproduct, the *generalized truth-tables* (GTTs for short).

Valuation semantics were first introduced by Andréa Loparić, in a 1977 paper, for the modal propositional logic **K** (see [Lo]). In order to give a brief description of what valuation semantics is, let us take as a starting point a semantics for the classical propositional logic **PL**: there we see that a model is nothing more than an assignment of truth-values to the propositional variables, since the value of complex formulas can be calculated if the value their subformulas have is known. We could also say, in other words, that a model for **PL** is a function f from wffs into truth-values obeying certain conditions (like $f(\neg A) \neq f(A)$, for instance).

If we now consider a possible-world semantics for some intensional logic, we notice that the structure of a model undergoes a deep change: one doesn't talk anymore about *only one* assignment (which, in a sense, describes a *possible world*), but about a whole set (a "universe") of them. The value of a formula whose main operator is an intensional one thus depends also on the value its subformulas get on various other worlds which are *accessible*. Here is where the famous *accessibility relations* come into the picture: formally, a model is now a triple $<W, R, V>$, where W denotes a set of worlds, R is a binary (acessibility) relation over W, and V is a function which takes arguments in formulas and worlds and goes into truth-values. The beauty of this construction is that one can get models for different modal logics by laying different conditions upon the relation R. (For instance, requiring of it to be *reflexive* singles out a class of models which characterizes the logic T.) On the other hand, in spite of models changing in this way, truth definitions for intensional operators like '$\diamond$' (for "it is possible that...") are still given as usual, namely by means of necessary and sufficient conditions (*iff*-conditions: "$\diamond A$ is true iff this-or-that holds").

Valuation semantics proceed the other way round: a model, which is called a *valuation*, is just one "world" (a function from wffs into $\{0,1\}$ having some special properties); that is, one doesn't have to introduce a set of worlds and an accessibility relation. The change comes with respect to truth definitions for intensional operators, which now appear in the form "if $\diamond A$ is true then such-and-such conditions hold; and if $\diamond A$ is false then such-and-such other conditions hold".

One could argue, of course, about the propriety of the statement "a model is just one world", since to evaluate a formula one also has to take other valuations (i.e.: other models) in consideration. More than that, when all is said and done a valuation ends up being proved to be the characteristic function of a maximal consistent set. In a sense, then, the whole could be like saying, in the setting of a possible-world semantics, that the only universe (model) you have to consider is the class of all MCSs and, besides, you don't have to bother about introducing accessibility relations. This can be a question of seeing things this or that way. In the paper we'll also prove a kind of equivalence between valuation and possible-world semantics — which is not surprising at all, since the same formulas have to come out as valid. To sum things up, the main difference lies on the fact that valuations are not declared *a priori* to be characteristic functions of MCSs; unlike possible-world models, they are defined inductively for certain sequences of formulas; it is only afterwards that they are generalized and proved to be characteristic functions of MCSs. And it is exactly because they are so defined that they generate in an easy way decision procedures, namely the GTTs, which allow us to examine *all relevant models* to some formula.

Back to historical matters, Loparić and I gave, some years after her original paper, a valuation semantics for the minimal tense logic **Kt** ([LM]; it was presented in 1980 as a short communication on the 4th Brazilian Conference on Mathematical Logic). In my master dissertation, under her supervision, I extended this semantics to several other tense logics as well, including here some naive logics combining time and modality. ([Mo1, Mo2]) In my dissertation there were also some problems left open, like to adequately define a valuation semantics for **S4**, still a tough and open case.

Now to GTTs. One can, of course, argue about the propriety of the name "truth-table". They certainly neither are, nor pretend to be, *connective-defining truth-tables* — as we have, for instance, the one defining the truth-function "conjunction":

$\wedge$	1	0
1	1	0
0	0	0

We already know that intensional operators like "it is necessary that ..." are not truth-functional (where the value a formula gets depends exclusively on the values of its subformulas). Thus, if one takes the expression "truth-table" in this narrow sense, as meaning something that defines a truth-function, then GTTs are not truth-tables, but something else ("truth-tableaux", maybe). On the other hand, we also talk (perhaps by abuse of the language) about the truth-table for some formula A, like the following one for $a \to (b \to a)$:

a	b	$b \to a$	$a \to (b \to a)$
1	1	1	1
0	1	0	1
1	0	1	1
0	0	1	1

If we thus understand "truth-table" as denoting this kind of construction, then certainly GTTs deserve the name. With GTTs the procedure is pretty much the same as in the classical, truth-functional case: we also build, for some wff A, a sequence $A_1, \ldots, A_n$ of its subformulas, where $A = A_n$ is the last element; next we assign values to the propositional variables, and after having done this we compute values for the remaining formulas of the sequence. The difference is that the value of a modalized A_i in a certain line j of the GTT now depends not only on the value in j of its subformulas, but also on the values which some other wffs can take in other lines. It should now not be surprising at all that through this construction one can also determine whether A is valid (meaning it is true on all lines) or not. In this way, we could obtain things like the following:

	1	2	3	4	5	6
	p	$\neg p$	$\neg\neg p$	$\Box p$	$\Box\neg\neg p$	$\Box\neg\neg p \to \Box p$
1)	1	0	1	1	1	1
2)	0	1	0	1	1	1
3)	1	0	1	0	0	1
4)	0	1	0	0	0	1

This is a truth-table (in **K**) for $\Box\neg\neg p \to \Box p$. Which, as one can see, is a valid formula.

In the paper we present, in a first part, valuation semantics for normal modal logics. This extends, with new results, Loparić's and my own work on the subject. A second part considers valuations for classical modal logics; and, in a third one, we define GTTs for some logics taken as examples. As a last part, we discuss, for some of the logics, how to implement (in C) the construction of GTTs.

References

[Ch] CHELLAS, B. *Modal Logic*. Cambridge, Cambridge U.P, 1980.

[HC] HUGHES, G.E. & CRESSWELL, M.J. *An Introduction to Modal Logic*. 2nd ed. London, Methuen, 1972.

[LM] LOPARIC, A. & MORTARI, C.A. "Valuations in temporal logic." *The Journal of Non-Classical Logic*, 2(1): 49-60, 1984.

[Lo] LOPARIC, A. "The method of valuations in modal logic." In ARRUDA, A.I.; DA COSTA, N.C.A; CHUAQUI, R. *MATHEMATICAL LOGIC: Proceedings of the First Brazilian Conference*. Marcel Dekker, New York, 1977.

[Mo1] MORTARI, C.A. *Valorações para alguns sistemas de lógica do tempo*. (Valuations for some tense logical systems) Master dissertation, State University of Campinas, Spring 1982.

[Mo2] MORTARI, C.A. "On a system of modal-temporal logic." 1982.

SNS
Biesingerstr. 10
D-7400 Tübingen 1
FRG

Departamento de Filosofia
UFSC
88000 Florianópolis SC
BRAZIL

A Dynamic Modal Semantics for Default Reasoning and Generics

Nicholas Asher & Michael Morreau
Version I, April 1990

1. Points of Departure

1.1. Patterns of Generic Reasoning

For many years linguists and philosophers have been interested in the meaning of generic sentences like *birds fly* and dispositional sentences like *copper conducts electricity.* For about ten years computer scientists have been interested too, but the meaning of such expressions and the reasoning appropriate to them remain elusive. A degree of consensus has however emerged about some simple patterns of generic reasoning to which any acceptable account of the semantics of generics must do justice. Here, by way of introduction, are some of these patterns. They appear in order of increasing complexity.

DEFEASIBLE MODUS PONENS

From *birds fly* and *Tweety is a bird,* it follows that *Tweety flies.* But from *birds fly, Tweety is a bird,* and *Tweety does not fly* it does not follow that *Tweety flies.*

Closely related to the defeasible modus ponens is what we call

DEFEASIBLE TRANSITIVITY

From *birds fly* and *sparrows are birds* it follows that *sparrows fly.* But from *those who eat slowly enjoy their food* and *those who are disgusted by their food eat slowly* together with the analytic fact that *those who are disgusted by their food do not enjoy their food* it does not follow that *those who are disgusted by their food enjoy their food.*

One pattern of generic reasoning which is now very familiar from the Artificial Intelligence literature is the

NIXON DIAMOND

From *Republicans are non-pacifists, Dick is a republican, Quakers are pacifists,* and *Dick is a Quaker* it intuitively neither follows that *Dick is a non-pacifist,* nor that *he is a pacifist..*

This scepticism dissappears if we substitute in the above *quakers who are republicans* for *quakers,* since we then want to draw the conclusion that *Dick is a pacifist.* This idea that defeasible information about subkinds should take precedence over defeasible information about the kinds which subsume them gives rise to a fourth pattern which we call the

PENGUIN PRINCIPLE

From *birds fly, Tweety is a bird, penguins do not fly, Tweety is a penguin,* and *penguins are birds* it follows that *Tweety does not fly.*

That penguins do not fly is a defeasible fact about penguins, a rule which admits exceptions. That penguins are birds, on the other hand, is a matter of taxonomical fact to which there are no exceptions. It is interesting that swapping this taxanomical fact for the weaker defeasible fact does not change our intuitions about the penguin principle:

WEAK PENGUIN PRINCIPLE

From *adults are employed, Sam is an adult, students are not employed, students are adults* and *Sam is a student* it follows that *Sam is not employed.*

Here is one last intuitively valid pattern of generic reasoning which does not fit into the above list of increasingly complex argument forms, the

DUDLEY DOORITE

From *republicans are politically motivated, quakers are politically motivated* and *Dudley Doorite is either a republican or a quaker,* it follows that *Dudley Doorite is politically motivated.*

We don't claim that these are the only argument forms; merely that any honest and sane person must agree to them. The consensus disappears however with more complicated examples. As soon as they involve more than three sorts and more complicated relations between these, intuitions about the validity of argument forms tend to degrade quite rapidly. We think that a theory of generic meaning which does descriptive justice to these relatively simple argument forms has pretty good credentials as a normative theory in more complicated cases where intuitions waver.

1.2. The Ghost in the Machine

If argument forms just cited are the semantic facts which any acceptable theory of generics must save, then none of the better known theories of nonmonotonic reasoning is an acceptable theory of generics. While circumscription, autoepistemic logic and - with some qualifications - Reiter's default logic all provide representations for generic sentences which save the simpler patterns of defeasible modus ponens and the nixon diamond, none of these formalisms throws much light on the penguin principle, where specific information takes precedence. The problem is by now familiar in the field of non-monotonic reasoning as the problem of multiple extensions. Coding the premises of the penguin principle up in the manner of circumscription by means of a multitude of "abnormality predicates," for example, we find that minimisation of abnormality results in two kinds of minimal models. There are models where Tweety is an abnormal bird but a normal penguin, and consequently does not fly. But in addition there are others where he is a normal bird but an abnormal penguin, and does fly. Because of these latter and if you will undesirable models it then does not follow that Tweety does not fly, and we see that circumscription does not handle the penguin principle adequately. Similar problems confront default logic and autoepistemic logic. The solution which proponents of these theories have suggested is as familiar as the problem: the order in which default rules fire needs to be constrained; the predicates to be minimized in the case of circumscription need to be prioritized. They thus commit themselves to the

HYPOTHESIS OF THE GHOST IN THE MACHINE

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

This brings us to our second motivational point. It remains unclear exactly what kind of reasoning it is that we do with generics, whether it belongs to logic or to pragmatics. But whatever kind of reasoning it is, that more specific information takes precedence is intrinsic to it, and the penguin principle should emerge naturally from the semantics of generic sentences without the intervention of a user. We want the ghost exorcised from the machine. The issue here is not just a methodological nicety, a question of whether logic should be set up this way or that. The problem with pushing some of the meaning of expressions up out of the formalism by delegating work to ghosts is that it makes it very unclear how such expressions could ever be nested. Generic expressions of various kinds, including habitual expressions like *smokes*, are however very often nested. Here are some examples: *If politicians are dishonest, their party is normally in decline, people who normally don't drive don't normally fly either, kinds of animals that normally have feathers are normally kinds of animals that fly.* And an interstellar traveler might be able to establish the truth or otherwise of our conjecture that *normally, if birds fly, the gravitational constant g is not higher than 10.5.* Here is final really complex example, in which genericity and counterfactuality combine with knowledge and belief: *Nicholas believes that if he were Michael, then maybe fast cars would interest him, while Michael knows that if he were Nicholas, then aeroplanes would most certainly intrigue him.*

1.3. The Intensionality of Generic Information

A necessary feature of any non-monotonic logic that seeks to give a semantics to natural language sentences expressing defaults-- like generic sentences, or sentences using adverbs like *typically* and *normally*, is that it be intensional. It is by now very familiar that, say, *birds fly* is not to be understood as a universal quantification over all individual birds, stating that they all fly. But it is a surprisingly common misconception that generics and statements about typical or normal cases are to be analysed by means of other extensional quantifiers, for example *most*. In fact it is easy to find examples showing that there is no simple relationship at all between most-sentences and the corresponding generics.

EXAMPLE
Take any natural number other than zero, say 10,000. Now there are infinitely many natural numbers larger than this, but just a finite number of smaller ones, so it is true that *most natural numbers are larger than 10,000.* But it is senseless or at best false to make the corresponding generic statement that *natural numbers are larger than 10,000.* And it sounds equally silly to say that *natural numbers are typically larger than 10,000,* or that they are normally so.

EXAMPLE
Suppose you have an urn in which there are 20 balls and 19 of them are white, while 1 is black. So it is true that *most of the balls in the urn are white.* It is senseless or false to say that *balls in the urn are white.*

These two examples make the same point. The extensional most-statement is true while the corresponding generic one is not. In the next example it is just the other way around, the generic statement being true while the corresponding most-statement is false:

EXAMPLE Here is an exerpt from our forthcoming book on turtles:

> Giant marine turtles normally live to reach a grand old age ... One hundred years or more is not exceptional. Most of them however, in fact the vast majority, have no such luck. They fall prey to preditors within the first hours of their lives, while scurrying ac̶cross the beach from their hatching grounds to the sea.

If as we believe the semantics of generics is closely tied up with defaults, then these examples suggest that the semantics of default logic, and perhaps of other forms of nonmonotonic reasoning too, are best thought of in a non-extensional way. Modal semantics is the best understood kind of intensional theory, so it seems worth while to see to what extent techniques from modal logic can be adapted to the semantics of generics.

1.3. Generic Information and Belief Revision

Traditionally semantics is concerned with the truth conditions of expressions, and it is their truth conditions which are taken to determine their logic. The search for truth conditions has been conspicuously unsuccessful in the case of generics, but there is a newer semantic paradigm which throws more light on them. According to this paradigm it is not the truth conditions of expressions which matter in explicating their meaning and logic, but the changes which they bring about when they are added to bodies of information or belief. On this epistemic approach, to believe that birds fly is, roughly speaking, to revise your beliefs in such a way that on learning that something is a bird, you assume that it can fly (unless you already believed that it cannot, of course, and until you learn that it cannot). This is the semantic intuition underlying the epistemic account of generic meaning given below.

2. Dynamic, Modal Semantics for Generics and Default Reasoning

2.1 The Language of Modal Default Logic

Consider a first-order language L augmented with a binary generalized quantifier $>_x$. Formulas of the language $L_>$ are the usual first'order formulas together with the following: if $\varphi(x)$ and $\psi(x)$ are first-order formulas with just x free, then $\varphi(x) >x\ \psi(x)$ is a sentence. Boolean combinations of such sentences with each other and with first-order formulas are allowed, but $>_x$ may not be nested. A treatment of the nested $>_x$ language is promised in the next version of this paper.
We will presently develop a non-monotonic entailment notion $\models\!\approx$ but we begin with its monotonic core, $\Vdash$, the part of $\models\!\approx$ which is not defeasible.

2.2. The Monotonic Core

The monotonic entailment notion $\Vdash$ is to be defined by means of *belief revision structures* $<\beta,+>$, of which β is a set of belief states and + a revision function mapping pairs comprising belief states and $L_>$ sentences onto other belief states. Thus for any state s and sentence $\varphi \in L_>$, s+φ stands for the the result of updating s with φ. We now turn to each of these two components in turn.

2.2.1. BELIEF STATES
Let W be the set of all the possible worlds for L, although for the sake of simplicity we assume a (non-empty) domain D of individuals which is constant accross them. The interpretation of L in these possible worlds is completely standard. Furthermore, it is assumed that for any such individual $d \in D$ the language contains a constant $\underline{d}$ such that for any possible world $w \in W$, $[\underline{d}]_W = d$. Now we have:

DEFINITION A *belief state* s, t_2, ... is a pair (P,*) where
i) $P \subseteq W$, and
ii) $*: P \times \wp W \rightarrow \wp W$, where for every $p \in \wp W$ and $w \in P$, $*(w, p) \subseteq p$.

Here $\wp W$ stands for the power set of W. Intuitively, P represents the beliefs of a state (P,*) about what is *actually* the case, a sentence being supported by this state just in case it is true at each of the worlds in P. The normality function *, on the other hand, represents the beliefs which (P,*) has about what is *normally* the case. Thus, for example, a state which believes that birds fly (represented Bx $>_x$ Fx) will be one where for each individual $d \in D$ and each world $w \in P$, $*(w, [B\underline{d}]) \subseteq [F\underline{d}]$. (Here $[\varphi]$ is the set $\{w \in W: w \Vdash \varphi\}$ of possible worlds where φ is true). This sort of function is familiar from conditional logic.[1] It provides every possible world w with "windows", by means of which are visable all those worlds in which p holds along with everything else which is normally associated with p. And this for any proposition p. To return to the example, in a state which believes that birds fly, every possible world w has, for any individual d, a B$\underline{d}$ window through which (only) worlds are visible where both B$\underline{d}$ and F$\underline{d}$ hold. Also borrowed from conditional logic is the following interpretation of $L_>$ sentences relative to possible worlds and selection functions:

DEFINITION For any possible world w and worlds selection function *:

[1] For a simple language with unembedded conditionals, we may simplify matters and consider a normality function that is constant across worlds.

for $>_x$-free φ:

$w, * \models \varphi$ iff $w \models \varphi$ in the normal sense of $\models$.

for $>_x$ -free φx and ψx:

$w, * \models \varphi x >_x \psi x$ iff for all $\delta \in D$, $*(w, \varphi\underline{\delta}) \subseteq [\psi\underline{\delta}]$

The usual clauses for disjunction, negation, conjunction and so on.

Fixing the above informal discussion into a definition, we can now define the following support relation $\models$ between belief states and sentences of $L_>$:

DEFINITION For any belief state (P,*) and sentence $\varphi \in L_>$:
$(P,*) \models \varphi$ just in case for all $w \in P$, $w, * \models \varphi$.

Among all belief states, there is one which has special significance. This state stands for the happy condition of an uncorrupted believer who is innocent of everything except logic. It is denoted:

This state is defined to be (W, =), where = is the function that maps, for every world w, every set of possible worlds onto itself. In this blissful state of ignorance any possible world at all might yet turn out to be the actual world, and any possible world where some sentence is true is a normal possible world where that sentence is true.[2]

2.2.2. UPDATES

We now go on to define the function by means of which beliefstates are updated with $L_>$ sentences. For the meantime the update function will be only partial, being defined for all belief states but only for sentences which are $>_x$-free, or of the form $\varphi x >_x \psi x$. Note that all of the patterns of generic reasoning described earlier are entirely within this fragment of $L_>$.

DEFINITION: For each belief state (P,*), define $(P,*)+\varphi$ accordingly as φ is $>_x$-free or of the form $\varphi x >_x \psi x$:

i. where φ is $>_x$-free
$(P,*)+\varphi =_{defn} (P \cap [\varphi],*)$, and

ii. where φ is of the form $\varphi x >_x \psi x$,
$(P,*)+\varphi =_{defn} (P, \oplus)$,
in which $\oplus$ is the normality function defined by:

for each $d \in D$ and for each $w \in P$: $\oplus(w, [\varphi\underline{d}]) =_{defn} *(w, [\varphi\underline{d}]) \cap [\psi\underline{d}]$,
and otherwise $\oplus$ coincides with *.

The first clause says that on learning that Sam is a dodo, the set of ones epistemic possibilities is reduced to those possible worlds where Sam is a dodo. The second clause says that on learning that Birds fly, the sets of worlds where Sam, Tommy, Ully etc. are normal birds are to be reduced to those worlds where Sam, Tommy, Ully etc. are birds which fly. The following facts about updates are worthy of note:

FACTS:

$s+\varphi \models \varphi$

$s \models \varphi$ iff $s+\varphi = s$

---

[2] This is one way of thinking about what it is to have no informaJtion about what is normally the case, but not the only way. Another is to think of ignorance, about say whether or not birds fly, as having among ones epistemic possibilities some possible worlds where it holds that birds fly, and others where it is not so that birds fly. We are coming to think that this second explication of ignorance is the better one, and will adopt it in the next version of this paper.

$(s+\varphi)+\psi \quad = \quad (s+\psi)+\varphi$ that is, updates are order-insensitive.

Now letting a *belief revision model* be any pair $<\beta,+>$ of which β is a set of belief states and + defined as above, we define the monotonic core entailment notion $\models$ as follows:
DEFINITION: For $\Gamma \subseteq L_>$, $\varphi \in L_>$

$\Gamma \models \varphi$ iff for all belief revision models $<\beta, +>$ and all $s \in \beta$, if $s \models \Gamma$ then $s \models \varphi$.

FACTS:

$\models \varphi$, where φ is a theorem of first-order logic.
$\models \varphi >_x \varphi$
$\models \varphi >_x \top$
$\models \varphi >_x \psi \,\&\, \varphi >_x \zeta \rightarrow \varphi >_x \psi \,\&\, \zeta$
$\models \varphi \,\&\, \psi >_x \varphi$

There are other desirable axiom schemes which are not validated by the semantics as it stands. The following one corresponds to the Dudley Doorite argument scheme:

(DD) $\varphi >_x \psi \,\&\, \zeta >_x \psi \rightarrow \varphi \vee \zeta >_x \psi$

The question immediately arises of whether there are extra static requirements which could be placed on belief states, and extra dynamic requirements which could be placed on the update function, which would validate the above scheme (DD). The answer is in both cases yes:

DEFINITION: A worlds selection function * *respects unions* just incase for all possible worlds w and all sets p and q of possible worlds: $*(w, p \cup q) \subseteq *(w,p) \cup *(w, q)$.

It is now not difficult to verify that the scheme (DD) expresses the requirement that selection functions respect unions, in the sense that this scheme is valid just incase all selection functions respect unions. Now we need a new definition before showing how a dynamic restriction can be placed on the revision function which guarantees that (the selection functions of) belief states resulting from revisions respect unions. We define the relation « on belief states. Intuitively, (P,*) « (Q,#) means that (P,*) is stronger than (Q,#) as far as information about normality is concerned, while agreeing with (P,*) about what is in fact the case. Formally, this relation is defined to be the following partial order:

DEFINITION: (P,*) « (P,#) just in case for all $w \in P$, $p \in \wp W$, $*(w,p) \subseteq \#(w, p)$.

Now starting from the update function + defined earlier, which is not guaranteed to deliever belief states which respect unions, we show how to define a new update function ⊕ which is guaranteed to do so. To this end consider, for any belief state (P,*) and sentence φ, the set {s:s « (P,*)+φ and s respects unions} of all belief states stronger than (P,*)+φ which verify the scheme (DD). We know that this set is non-empty. It contains a state using the selection function * such that for arbitrary q, $*(w, q) = \emptyset$. It is not difficult to very that any « chain of belief states within this set has an element, maximal on «, which is in this set. Zorn's lemma then informs us that it has a unique «-maximal element. This partially justifies the following definition:

DEFINITION: For each belief state (P,*) and sentence φ for which + is defined, let (P,*)⊕φ be the «-maximal element of {s: s « (P,*)+φ and s respects unions}.

What this definition gives us, intuitively, is the weakest state stronger than s + φ that validates (DD). From now on we shall take + to be defined as ⊕ is in the definition just given.

2.3. The Non-Monotonic Periphery

So far one might notice something odd about this system that purports to be non-monotonic. It is monotonic. To get non-monotonicity, we need to refine and make precise two intuitive notions. To conclude from the fact that birds fly and that Tweety is a bird that Tweety flies, we need to make two assumptions:

1) that we know no other facts than what we are given (in particular, we don't know that Tweety does not fly) and

2) that everyone and everything is as normal as these facts allow.

The first notion can be made precise using the notion of updating. "Knowing no more than $\{\gamma_1, \ldots, \gamma_n\}$" comes to being in the belief state

(...( $+ \gamma_1) + \gamma_2) + \ldots) + \gamma_n$

Finding a technical way of making the second assumption of maximal normality requires a bit more work. First we expand our belief revision models to triples $<\beta, +, \leq>$, adding a relation $\leq$ which is defined on all belief states (that is, not just on those in β). Intuitively, $s \leq \tau$ just in case τ strengthens s by assuming some individual to be more normal in some respect than s assumes him to be. This relation we define in terms of the following one, which is intended to say what it is for an individual δ to be considered no less normal a φ in a belief state (Q, #) than in a weaker belief state (P, *):

DEFINITION:
$(P, *) \leq_{\delta, \varphi x} (Q, \#)$ iff

i.	$Q = P \setminus ([\varphi\underline{\delta}] \setminus \bigcup_{w \in P} *(w, [\varphi\underline{\delta}]))$,	If $P \cap (\bigcup_{w \in P} *(w, [\varphi\underline{\delta}])) \neq \emptyset$, and
ii.	$Q = P$	Otherwise.

while for each $w \in P$ and for each $p \in \wp(W)$:

iii.	$\#(w, p) = *(w, p) \setminus ([\varphi\underline{\delta}] \setminus *(w, [\varphi\underline{\delta}]))$,	If $*(w, p) \cap *(w, [\varphi\underline{\delta}]) \neq \emptyset$
iv.	$\#(w, p) = *(w, p)$	Otherwise.

This definition is in need of some elaboration. What is it now for an individual δ to be considered no less normal a φ in a belief state (Q, #) than in a weaker belief state (P, *)? Clauses i. and ii. say what this means as far as epistemic possibilities are concerned. If, as we intend, $*(w, [\varphi\underline{\delta}])$ stands for the set of possible worlds in which, as far as an isolated w and * are concerned, an individual δ is a normal φ, then $\bigcup_{w \in P} *(w, [\varphi\underline{\delta}])$ is the set of possible worlds in which a whole belief state (P,*) considers δ to be a normal φ. Clause i., then, requires Q to "strengthen" P by removing from P those possible worlds in which δ, though a φ, is not considered by (P, *) to be a normal φ. The restriction on clause i. and clause ii. together express that if (P,*) already believes δ to be an abnormal φ, then the assumption that δ is a normal φ can no longer be made while keeping all the information in (P,*). Clauses iii. and iv. do for a belief state's beliefs about what normally holds exactly what i and ii. did for its beliefs about what actually holds (as encoded in the set of epistemic possibilities) and will not be given a separate gloss here.

FACT: If for all $w \in P$, $*(w, [\varphi\underline{\delta}]) = [\varphi\underline{\delta}]$, then $(P,*) \leq_{\delta, \varphi x} (Q, \#)$ only if $(P,*) = (Q, \#)$.
What this intuitively means is that if a state (P,*) carries no contingent information about what normally holds when δ is a φ, then assuming δ to be a *normal* φ does not change (P,*) in any way. It is worth verifying this fact by simplifying i. and iii. of the above definition on the assumption that for all $w \in P$, $*(w, [\varphi\underline{\delta}]) = [\varphi\underline{\delta}]$.

Now we define $s \leq \tau$ just in case there are δ, φx such that $s \leq_{\delta, \varphi x} \tau$. Because $\leq$ is everywhere defined, we can (and do) require belief revision models to be closed under it. That is, we require that for any belief revision model $<\beta, +, \leq>$ and belief states s, τ with $s \leq \tau$ and $s \in \beta$, we have $\tau \in \beta$. This relation is reflexive and (apart from these 1 element cycles) well-founded. Such relations may be called *capped*. It is easy to check that the reflexivity of $\leq$ is established by any δ together with an empty predicate, say $x \neq x$. Establishing that there are no $\leq$ cycles involving more than one belief state is left to the reader as a useful exercise. Likewise worth verifying is the following: Defining *positive sentences* to be those in whose prenex normal form only $>_x$-free sentences appear in the scope of the negation sign, we have:

FACT: Positive sentences are stable under the assumption of normality. That is, for all positive sentences φ and belief states s, τ with $s \leq \tau$, if $s \models \varphi$ then $\tau \models \varphi$.

In fact all of the premises of the argument forms we set out to capture fall within the positive fragment of the language. So this fact justifies using the relation to say what it is to assume that everyone and everything is as normal as is consistent with the hard (positive) beliefs which one has in some particular belief state. It is simply to move to a maximal element of a $\leq$ chain leading from that belief state.

Reasoning about generics intuitively inolves, we said, two things: assuming that the premises of an argument are all one knows, and assuming that everyone and everything is as normal as is consistent with these premises. Now that we have fixed the former notion into our semantics as updating ☺ with the premises, and the latter notion as following $\leq$ chains of the resulting belief state to their maximal elements, the following dynamic definition of the non-monotonic consequence notion $\mid\approx$ will not come as a surprise. Letting $s+\Gamma$ stand for the result of updating s with all the elements in Γ, we have:

DEFINITION

$\Gamma \mid\approx \varphi$ iff for all s such that ☺ $+ \Gamma \leq_{max}$ s, $s \models \varphi$.

Note that since $\leq$ is capped, any $\leq$ chain leading from a belief state (which verifies any set of sentences Γ) has a maximal element. As a result if Γ is monotonically consistent, then Γ is non-monotonically consistent. Also note that $\Gamma \models \varphi \Rightarrow \Gamma \mid\approx \varphi$, and that since the monotonic notion of consequence contains all the classical validities, our non-monotonic consequence relation is supraclassical. Now defining φx, ψx, and ζx to be *independent formulas* just in case for each $d \in D$, each formula of the form $(\neg)\varphi x \& (\neg)\psi x \& (\neg)\zeta x$ is true in some $w \in W$, we have for independent φx, ψx, and ζx the following facts:

FACTS

$\varphi >_x \psi, \varphi(\underline{d}) \mid\approx \psi(\underline{d})$, but not

$\varphi >_x \psi, \varphi(\underline{d}), \neg\psi(\underline{d}) \mid\approx \psi(\underline{d})$. (Defeasible Modus Ponens)

$\varphi >_x \psi, \psi >_x \zeta \mid\approx \varphi >_x \zeta$, but not

$\varphi >_x \psi, \psi >_x \zeta, \varphi >_x \neg\zeta \mid\approx \varphi >_x \zeta$. (Defeasible Transitivity)

not$\{\varphi >_x \psi, \zeta >_x \neg\psi, \varphi(\underline{d}), \zeta(\underline{d}) \mid\approx \psi(\underline{d})$ (or $\neg\psi(\underline{d})$)$\}$. (Nixon $\Diamond$)

$\varphi >_x \zeta, \psi >_x \zeta \models (\varphi \vee \psi) >_x \zeta$. (Dudley Dorite)

The Penguin Principle is for us, from a motivational point of view, a very important one: it is the pattern of generic reasoning in which specific takes precedence over general information which has been haunting most formalisms for non-monotonic reasoning. The restricted language in which we have been working does not allow us to state it in its strong form,, since we do not have strict implication at our disposal with which to state that penguins, strictly, are birds. We can however state it its weak form, which only makes use of default implications:

$\varphi >_x \psi, \psi >_x \zeta, \varphi >_x \neg\zeta, \varphi(\underline{d}) \mid\approx \neg\zeta(\underline{d})$ (Weak Penguin Principle)

The penguin principle is not validated by the semantics as it stands. So just as with the Dudley Doorite principle (DD), the question arises as to whether there are static and dynamic constraints which, when placed on belief states and the update function respectively, would validate it. And once again the answer is in both cases yes:

DEFINITION: A worlds selection function * *respects specificity* just in case for all possible worlds w and all sets p, q and r of possible worlds:
If $*(w, p) \subseteq q \setminus r$ and $*(w, q) \subseteq r$, then $*(w, q) \subseteq r \setminus p$

As with (DD), we appeal to « in showing how a dynamic restriction can be placed on the revision function which guarantees that (the selection functions of) belief states resulting from revisions respect specificity. The argument is completely analogous to that used in validating (DD). Starting from the update function + which is not guaranteed to deliever belief states which respect specificity, we define a new update function ⊕ which is guaranteed to do so. To this end consider, for any belief state (P,*) and sentence φ, the set {s:s « (P,*)+φ and s respects specificity}. We know that this set is non-empty. It contains a state using the selection function * such that for arbitrary q, $*(w, q) = \emptyset$. It is not difficult to very that any «-chain of belief states within this set has a element, maximal on «, which is in this set. Zorn's lemma then informs us that it has a unique «-maximal element. This partially justifies the following definition:

DEFINITION: For each belief state (P,*) and sentence φ for which + is defined, let (P,*)⊕φ be the «-maximal element of {s:s « (P,*)+φ and s respects specificity}.

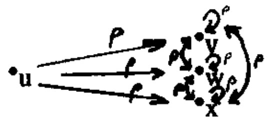
Taking + to be defined as ⊕ is in the definition just given we have, for independent φx, ψx, and ζx, the following fact:

FACT: $\varphi >_x \psi, \psi >_x \zeta, \varphi >_x \neg\zeta, \varphi(\underline{d}) \models\!\approx \neg\zeta(\underline{d})$

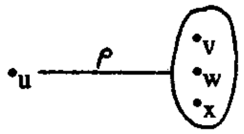
This and a selection of the other facts above are illustrated in the next section.

3. Some Worked Examples

We now illustrate the above facts by going through some of the patterns of generic reasoning which were our original motivation.. We will do a modus ponens, and then defeat it. We will look at transitivity, and see it defeated too. Finally, we will go through the weak penguin principle. To make things more perspicuous we introduce the following double simplification: Firstly, we restrict ourselves to a language containing just three (monadic) predicates. And secondly, we restrict ourselves to domains containing just a single element. The first restriction is insignificant in that no surprises are in store when the language is expanded to include other predicates. The second restriction is more significant since, while the illustrations given below do just as well for domains with more than one individual, some interesting new questions relating to the lottery paradox show up there. We turn to this matter in the second version of this paper.
A few words on how the diagrams below are to be read. What we see in each case is the evolution of a belief state as it is updated by means of the + function. Then, once all of the premises of the argument form in question have been read in, we follow ≤ chains to their maxima and check whether the conclusion of the argument is believed. Dots represent possible worlds, and circles the epistemic possibilities which are the first components of belief states. Worlds selection functions, the second components, are represented by means of arrows leading from possible worlds to (other) possible worlds, which arrows are marked p, q, r etc. A p-arrow leading from a possible world v to a possible world w means that $w \in *(v, p)$. Where many p-arrows lead from a world to an equivalence class of possible worlds, we replace them for clarity's sake by a p-balloon. Thus for example



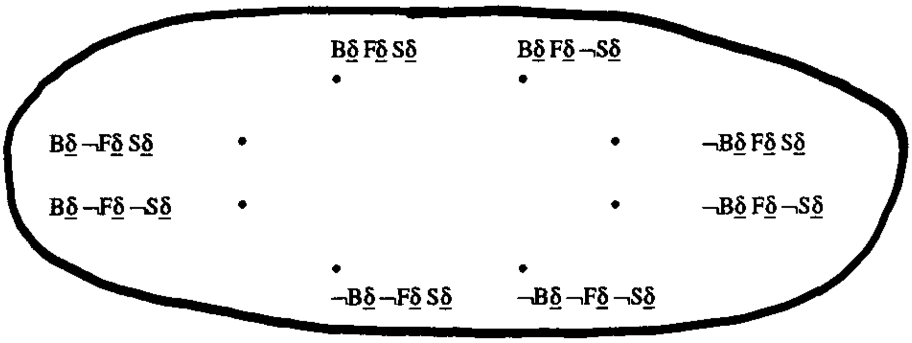
means the same as



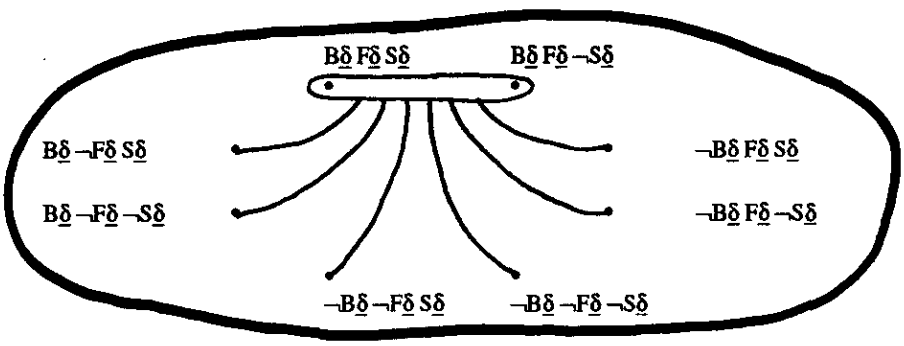
To distinguish these balloons which represent selection functions from the circles which represent sets of epistemic possibilities, the latter have been drawn with bold lines.
We adopt furthermore the convention that where p-arrows lead from every epistemic possibility of a state to every possible world in p (recall that this holds for every p in the naive state, and just means that the selection function in question carries no contingent generic information about p) none of the p-arrows have been drawn in. In order to justify this convention (and in order to follow the diagrams below) it is well worth noticing that where no [$\varphi\underline{\delta}$] -arrows have been drawn into a belief state (P,*), $(P,*)\leq_{\delta,\varphi x}(Q,\#)$ only if (P,*)=(Q, #). This simple consequence of the fact given directly after the definition of $\leq_{\delta,\varphi x}$ justifies the above convention, since it means that in finding our way along $\leq$ chains we can forget all about invisable arrows. To make things clearer we also leave out any arrows we are no longer interested in (say because we have already reached the maximum of a chain and they do not bear on whether or not the conclusion of the argument we are interested in is believed or not).

3.1. Modus Ponens ...

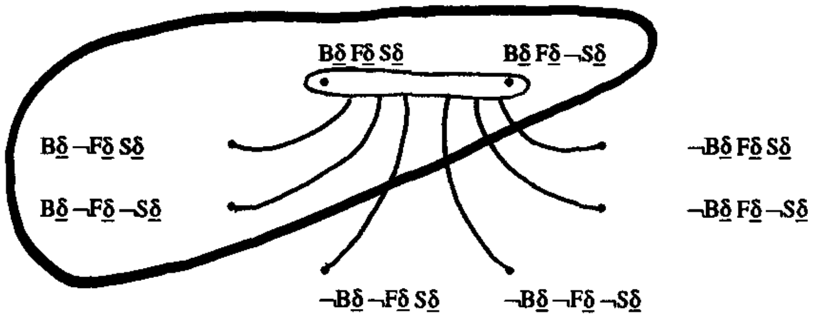
We verify that $B >_X F$, $B\underline{\delta} \models F\underline{\delta}$. In the language with non-logical symbols $\{S,B,F,\underline{\delta}\}$ we have in total eight possible worlds, so that  is the following state:



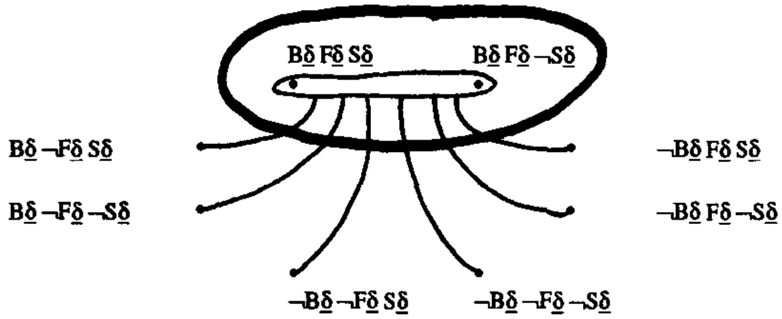
Updating this state with the premise $B >_X F$, the selection function is modified such that [B$\underline{\delta}$]-arrows run only to (all) worlds in [F$\underline{\delta}$]. The resulting state is as follows, in which all arrows are [B$\underline{\delta}$]-arrows.



Next we update with the second premise, B$\underline{\delta}$, which has the effect of shrinking the set of epistemic possibilities to those worlds where B$\underline{\delta}$ is true. This results in the following belief state:



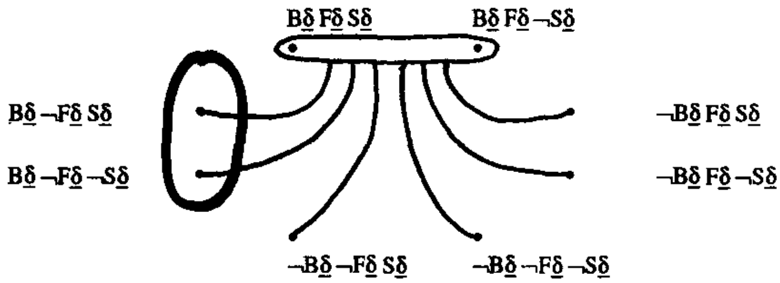
We now have to follow to their maximal elements all $\leq$ chains leading from this belief state, and there check to see whether the conclusion of the modus ponens is believed. In fact, as the reader can verify with pencil and paper, there is a unique such maximal element, and it is this:



Clearly this is a belief state which supports $F\underline{\delta}$. So we have just verified that $B >_X F, B\underline{\delta} \mid\!\approx F\underline{\delta}$.

3.2. ...and the defeat of Modus Ponens

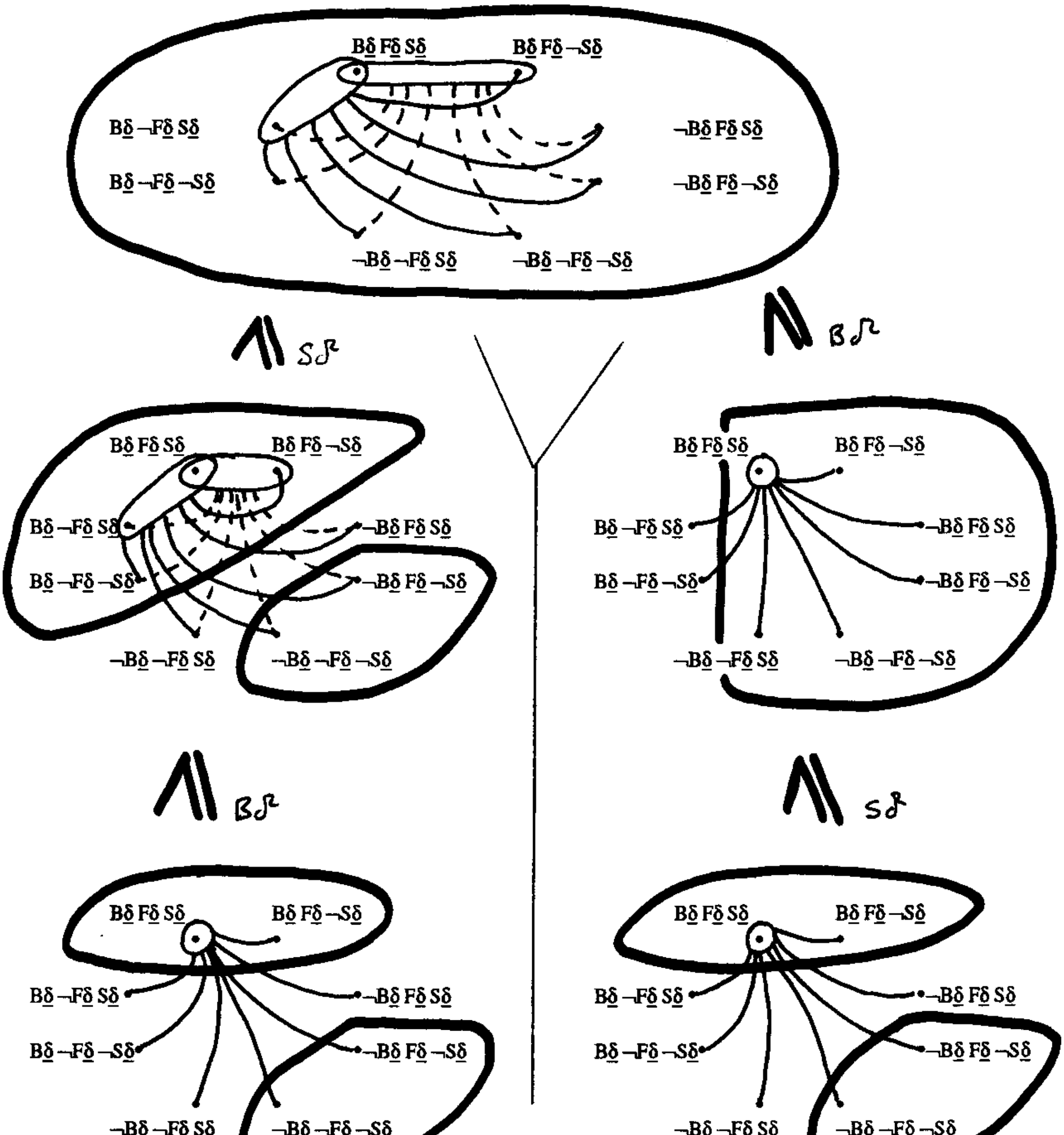
We want now to see how adding $\neg F\underline{\delta}$ to the premises of the previous argument defeats the conclusion that $F\underline{\delta}$. That is, we want to verify that not $B >_X F, B\underline{\delta}, \neg F\underline{\delta} \mid\!\approx F\underline{\delta}$. To this end, after updating with $B\underline{\delta}$ above we update with $\neg F\underline{\delta}$, thereby arriving at the following belief state:



It is not difficult to check, now, that this belief state is itself a maximal element of $\leq$ (since the only way for this state to bear the relation $\leq$ to any other state is in virtue of the failure clauses iii. and iv. of the definition of $\leq$, which leave everything as it is).
Clearly this state supports $\neg F\underline{\delta}$, and does not support $F\underline{\delta}$. So $B >_X F, B\underline{\delta}, \neg F\underline{\delta} \mid\!\approx \neg F\underline{\delta}$, and not $B >_X F, B\underline{\delta}, \neg F\underline{\delta} \mid\!\approx F\underline{\delta}$.

3.3. Transitivity...

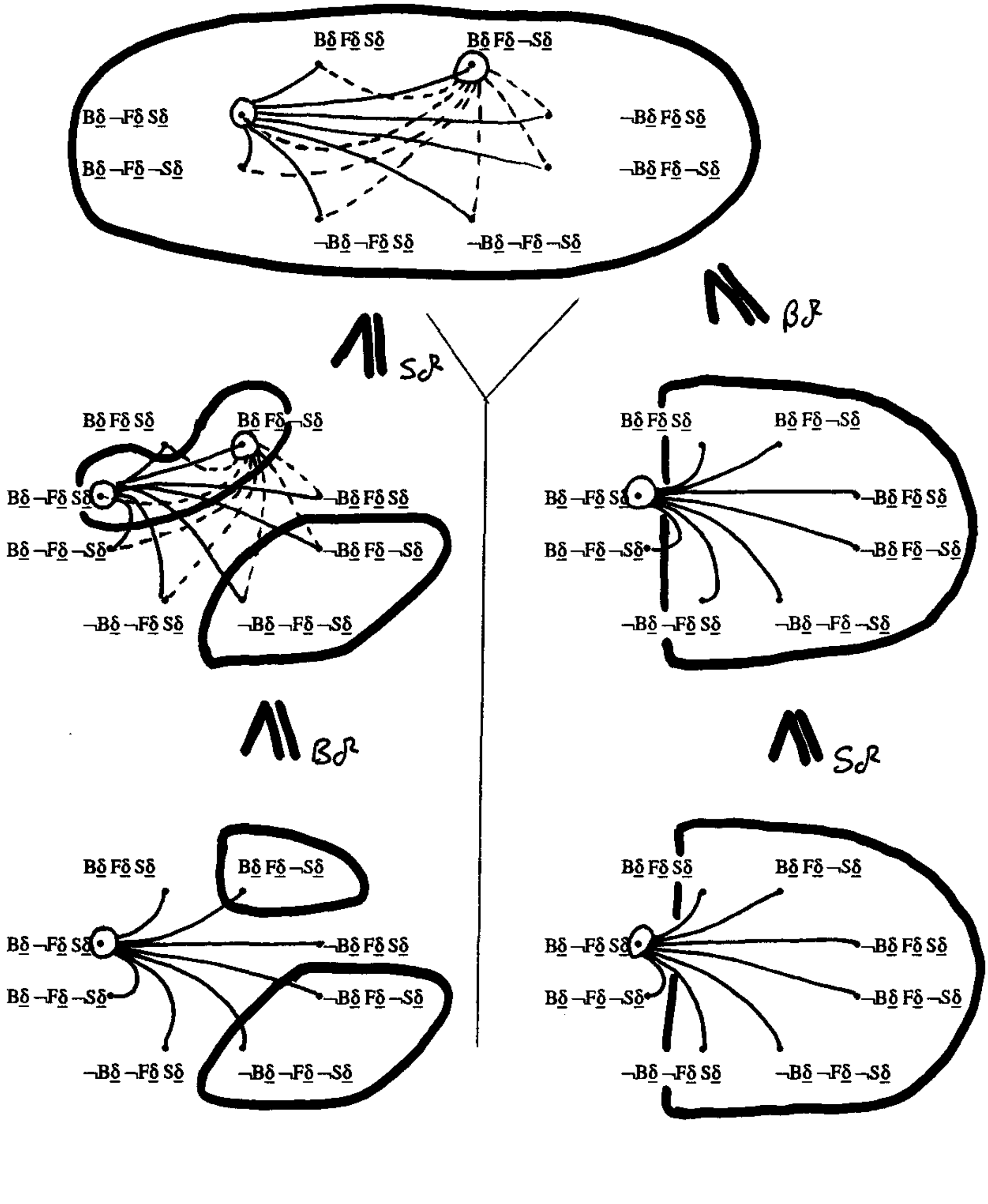
We now want to show that $S >_X B, B >_X F \mid\approx S >_X F$. To this end we update ☺ successively with $S >_X B$ and $B >_X F$, obtaining as a result the following belief state (here [S$\underline{\delta}$]-arrows are continuous, while the [S$\underline{\delta}$]-arrows are broken). Below this state have been drawn the only two nontrivial $\leq$ chains, at the end of each of which the conclusion of the argument, $S >_X F$, is believed. This validates the argument.



Note that as required to show that $S >_X B, B >_X F \mid\approx S >_X F$, each of these $\leq$ maximal states believes that $S >_X F$.

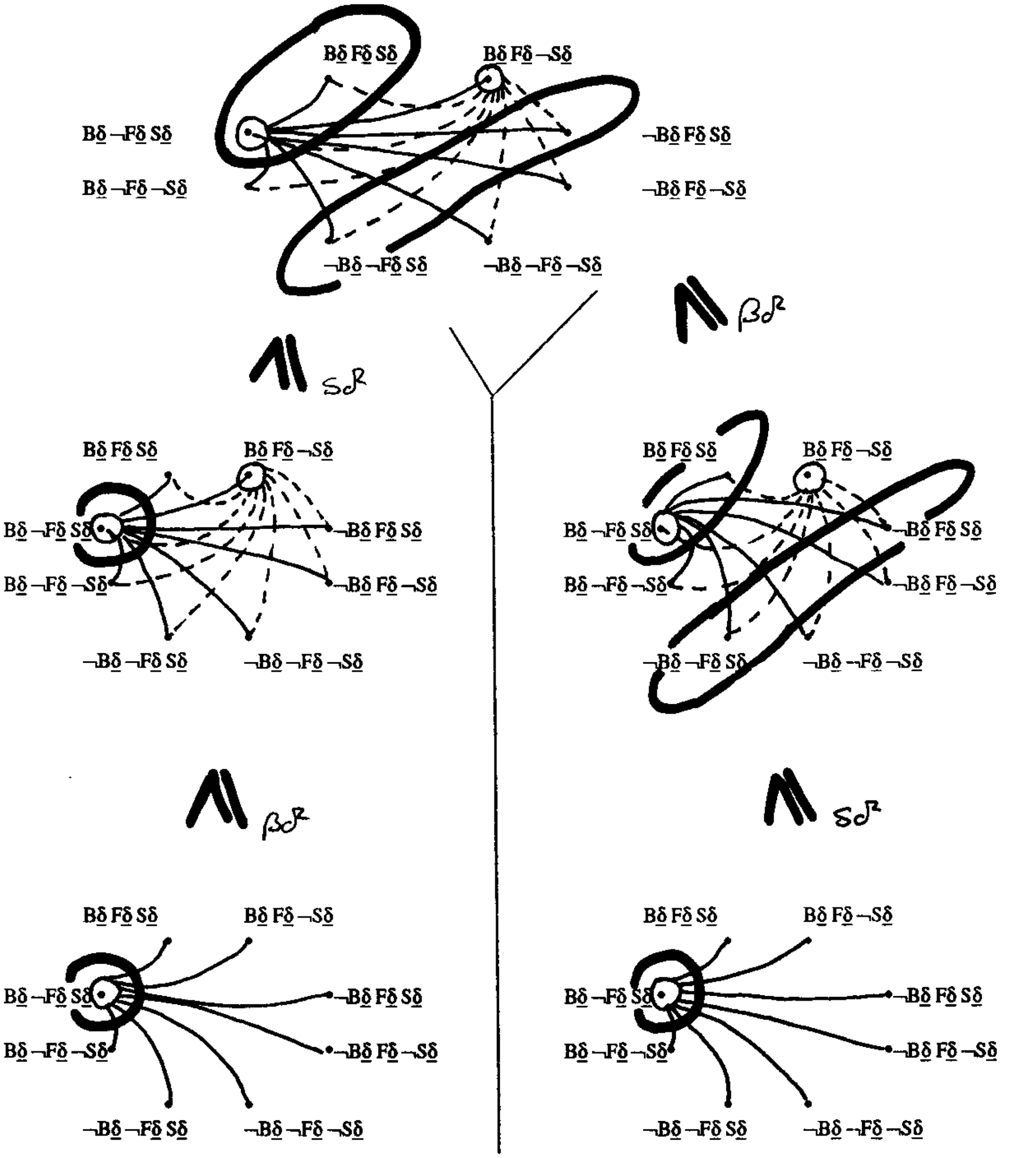
3.4. ...and Transitivity Defeated

We now want to see what happens if information is added to the effect that sparrows do not fly (this is of course not true, but that does not matter. Pretend for the moment that *sparrow* is our word for penguin). Then the conclusion that *sparrows fly* should no longer follow, so we now want to show that it is not so that $S >_X B, B >_X F, S >_X \neg F \mathrel{|\!\approx} S >_X F$. Updating the naive belief state ☺ with the premises $S >_X B$, $B >_X F$, $S >_X \neg F$, we arrive at the belief state represented below. Leading down from it are again (the only two non-trivial) $\leq$ chains, terminating at their maximal elements. At none of these maximal elements is $S >_X F$ believed, which is as we wanted it to be.



3.5. The Weak Penguin Principle

As a last worked example we now go on to verify an instance of the weak penguin principle. Again pretending that sparrows are penguins (so as to make good use of the work already done above), we show that $S >_X B$, $B >_X F$, $S >_X \neg F$, $S\underline{\delta} \mid\!\approx \neg F\underline{\delta}$. To this end, as in the example above is updated with successively $S >_X B$, $B >_X F$, $S >_X \neg F$, but this time $S\underline{\delta}$ is added too. This results in the following belief state, under which the (only two non-trivial) $\leq$ chains have once again been drawn. Clearly their two maximal elements believe $\neg F\underline{\delta}$, the conclusion of the weak penguin principle. Which is what was needed.



4. Comparisons with other Approaches to Nonmonotonic Reasoning

In a sense, the theory of nonmonotonic reasoning with which this one compares best is the theory of nonmonotonic semantic networks due to Touretzky, Horty and Thomason (1987). Their theory, like this one, takes defeasible modus ponens, defeasible transitivity and the penguin principle to be the argument forms which a theory of nonmonotonic reasoning must validate, and like this one their theory validates the penguin principle without recourse to a ghost. One important way in which this this approach differs from theirs, however, is its greater expressive power. With our modal semantics, boolean combinations of concepts are immediately at hand. This results in the validity in our theory, but not in theirs, of argument schemes in which composite concepts inherit properties from the parts of which they are composed-- e.g. Dudley Doorite. Another point where the greater expressive power of this theory shows up is in what might be called cyclic default theories, of which a simple example would be *birds fly* together with *flying things are birds* (presumably true before the advent of aeroplanes). Such theories cannot be represented in semantic networks for technical reasons having to do with the inheritance algorithm, but present no special difficulty for us.

Another theory of generics and defaults to which this one bears some resemblence at the level of technical realisation is that of Delgrande (1987). His, like ours, belongs to the tradition of possible worlds semantics, rebuilding the Stalnaker-Lewis semantics of conditionals as a semantics for generics. And his, like ours, takes a generic like Birds fly to mean more or less that any individual bird can under normal circumstances fly. In spite of coming from the same philosophical nest, the two theories diverge in important ways. First ours is dynamic; his is not. More significantly, Delgrande´s theory, like practically everybody elses, makes use of various mechanisms built on top of the basic semantics in order to get things working properly.

A final point of comparison is Veltman's (1989) theory, by which we were originally inspired. Although also a modal semantics for defaults and also dynamic, Veltman's theory differs from ours in at least the following ways: a) he restricts himself to a propositional language, whereas we give a semantics for a default quantifier with exceptions; b) in Veltman, default rules are not defeasible. Once they follow from some premises, they continue to follow no matter how the premises are added to. Clearly such a semantics cannot account for the pattern of defeasible transitivity.

5. References

Delgrande, J. (1988): "An Approach to Default Reasoning Based on a First-Order Conditional Logic" Artificial Intelligence vol 36 #1, 63-90.

Horty, J., Thomason, R., Touretsky, D. (1987): "A Skeptical Theory of Inheritance in Non-Monotonic Nets," CMU Tech Report CMU-CS-87-175.

Veltman, F. (1989): "Defaults in Update Semantics," DYANA deliverable 2.5.a

Veröffentlichungen des Seminars für natürlich-sprachliche Systeme der Universität Tübingen

Volger, H.: Theories which Admit Initial Structures. SNS-Bericht 85-1.

Guenthner, F. & H. Lehmann: Natürlich-sprachliche Systeme. Ein Überblick. SNS-Bericht 85-2. (Erschienen in *Informatik Spektrum*.)

Guenthner, F. & H. Lehmann & W. Schoenfeld: A Theory of Knowledge Representation. SNS-Bericht 85-3. (Erschienen im *IBM Journal of Research and Development* 1986, 30.1, 39-56).

Guenthner, F.: Zwei Typen von semantischen Relationen. SNS-Bericht 85-4. (= "Two Types of Semantic Relations", erscheint in H. Bunt (ed.), *Proceedings of the Symposion on Language Technology*, Tilburg 1988, Cambridge: Cambridge University Press.)

Guenthner, F.: Linguistic Meaning in Discourse Representation Theory. SNS-Bericht 85-5. (Erschienen in *Synthese* 73.3 (*Theories of Meaning*) 1987, 569-599.)

Wirth, R.: Halbautomatisierte Erweiterung eines Thesaurus. SNS-Bericht 85-6. (Erschienen in GWAI-85, Berlin: Springer, 204-212.)

Guenthner, F. & P. Sabatier: Formal Semantics and Knowledge Representation. SNS-Bericht 85-7. (Erschienen auf franz. in *Langages* 1986, 103-123).

Guenthner, F. & C. Sedogbo: Some Remarks on the Treatment of Errors in Natural Language Processing. SNS-Bericht 86-8. (Erschienen in *Programmation en Logique*, Actes du Séminaire 1987, Tregastel.)

Schulz, K.: An Exact Algorithm for Interval-based Temporal Information. SNS-Bericht 86-9.

Sedgbo, C.: AVAG. An Attribute-Value Grammar Tool. SNS-Bericht 86-10. (Erschienen in *Programmation en Logique*, Actes du Séminaire 1986, Tregastel.)

Nute, D.: LDR. A Logic for Defeasible Reasoning. SNS-Bericht 86-11.

Kamp, H.: Reading, Writing and Understanding. Comments on H. Sluga "Philosophy as a Kind of Writing". SNS-Bericht 86-12.

Kamp, H.: Belief Attribution and Context. SNS-Bericht 86-13.

Gabbay, D.: What is Negation in a Logical System? SNS-Bericht 86-14.

Zimmermann, Th.E.: Intensional Logic and Type Theory. SNS-Bericht 86-15. (Erscheint im *Journal of Symbolic Logic)*.

Forschungsstelle für natürlich-sprachliche Systeme: Jahresbericht. SNS-Bericht 86-16.

Krifka, M.: Nominal Reference and Temporal Constitution: Towards a Semantics of Quantity. SNS-Bericht 87-17. (Erschienen in J.J. Groenendijk e.a., *Proceedings of the 6th Amsterdam Colloquium* 1987, Institute for Language, Logic and Information, Amsterdam)

Schulz, K.: Event and Interval Structures. A Mathematical Comparison. SNS-Bericht 87-18.

Bäuerle, R.: Verben der propositionalen Einstellung. SNS-Bericht 87-19. (Erscheint in D.Wunderlich & A. v. Stechow: *Handbuch Semantik*. Kronberg: Athenäum.)

Bäuerle, R. & Th.E. Zimmermann: Fragesätze. SNS-Bericht 87-20. (Erscheint in D.Wunderlich & A. v. Stechow: *Handbuch Semantik*. Kronberg: Athenäum.)

Mayer, R.: Bewegung und Kohärenz. SNS-Bericht 87-21.

Mayer, R.: Partikelverben und Bewegung. SNS-Bericht 87-22.

Mayer, R.: An Exercise in Linguistic Robotics. SNS-Bericht 87-23.

Mayer, R.: Mentale Modelle und Sprache. SNS-Bericht 87-24.

Krifka, M.: An Outline of Genericity. SNS-Bericht 87-25.

Schulz, K.: Countable Models of Theories for Action and Interval-based Time. SNS-Bericht 87-26.

Bealer, G.: Fine-Grained Type-Free Intensionality. SNS-Bericht 88-27. (Erscheint in B. Partee, G. Chierchia & R. Turner (eds.), *Property Theory, Type Theory, and Semantics*. Dordrecht: Reidel)

Hudelmaier, J.: A PROLOG Program for Intuitionistic Logic. SNS-Bericht 88-28.

Sæbø, K.J.: Two Papers on Cause and Purpose. SNS-Bericht 88-29. (Erscheint in D. Wunderlich & A. v. Stechow: *Handbuch Semantik*. Kronberg: Athenäum, bzw. in J. Jacobs e.a. *Syntax: An International Handbook of Contemporary Research*. Berlin: Walter de Gruyter).

Wirth, R.: Completing Proofs by Inverting Resolution. SNS-Bericht 88-30.

Mayer, R.: Spatial Coherence. SNS-Bericht 88-31.
Mayer, R.: Imperatives and Assertions. SNS-Bericht 88-32.
Lambek, J.: On the Unity of Algebra and Logic. SNS-Bericht 88-33.
Schulz, K.: On the Categoricity of Countable Inverval Structures. SNS-Bericht 88-34.
Lambek, J.: Multicategories Revisited. SNS-Bericht 88-35.
Wirth, R.: Learning by Failure to Prove. SNS-Bericht 88-36.
Mayer, R.: Compactness, Compression and Relevance. SNS-Bericht 88-37.
Zimmermann, Th.E.: Die klassische Theorie der Kontextabhängigkeit. SNS-Bericht 88-38.
Schulz, K.: A Guide to Makanin's Algorithm. SNS-Bericht 88-39.
Guenthner, F.: Features and Values 1988. SNS-Bericht 88-40.
Truckenbrodt, H.: Zur Syntax der Koordination. SNS-Bericht 88-41.
Krifka, M. (ed.): Genericity in Natural Language. Proceedings of the 1988 Tübingen Conference. SNS-Bericht 88-42.
Krüger-Thielmann, K.: Möglichkeiten der intelligenten Fehlerkorrektur im computergestützten Sprachunterricht. SNS-Bericht 88-43.
Guenthner, F., F. Reitzig, D. Schnabel, K. Wenger: Ein Expertensystem für das Arbeitsförderungsgesetz. SNS-Bericht 89-44.
Bealer,G. & U. Mönnich: Property Theories. SNS-Bericht 89-45. (Erschienen in D. Gabbay & F. Guenthner (eds.). *Handbook of Philosophical Logic* Vol. IV. Dordrecht: Reidel
Krifka, M.: Four Thousand Ships Passed through the Lock: Object-Induced Measure Functions on Events. SNS-Bericht 89-46 (Erscheint in *Linguistics and Philosophy*).
Phillips, J. D.: Using Explicit Syntax for Disambiguation in Speech and Script Recognition. SNS-Bericht 89-47.
Morreau, M. (ed.): Proceedings of the Tübingen Workshop on Semantic Networks and Nonmonotonic Reasoning. Vol. 1. SNS-Bericht 89-48.
Morreau, M. (ed.): Proceedings of the Tübingen Workshop on Semantic Networks and Nonmonotonic Reasoning. Vol. 2. SNS-Bericht 89-49.
Pafel, J.: An welcher Repräsentationsebene sollte die Skopusbestimmung ansetzen? Eine kritische Darstellung der SRI "Core Language Engine". SNS-Bericht 89-50.
Pelletier, F.J.: The Philosophy of Automated Theorem Proving. SNS-Bericht 89-51.
Truckenbrodt, H.: Koordination und Top-Down-Parsing in Prolog. SNS-Bericht 89-52.
Schulz, K.U.: A Note on Makanin's Algorithm. SNS-Bericht 89-53.
Mayer, R.: Universal Quantifiers and Domain Restriction. SNS-Bericht 89-54.
Gabbay, D.M.: The Tübingen Lectures 1988. SNS-Bericht 90-55.
Mayer, R.: Abstraction, Context and Perspectivization. SNS-Bericht 90-56.
Dosen, K.: Modal Logic as Metalogic. SNS-Bericht 90-57.
Schroeder-Heister, P. (ed.): Proceedings of the SNS Logic Colloquium, March 1990. SNS-Bericht 90-58.

Kursmaterialien des Seminars für natürlich-sprachliche Systeme

Dezelski, H.-J.: Pascal für Geisteswissenschaftler. 1988.Kursmaterialien 1.
Referate vom Seminar "Natürlich-sprachliche Datenbankabfragesysteme", Franz Guenthner, WS 87/88. Kursmaterialien 2.
Maier, P.: Das natürlichsprachliche Datenbank-Abfragesystem Orbel. 1989. Kursmaterialien 3.
Phillips, J.D.: Course Notes for the Introduction to Lisp. 1989.
Krifka, M. & E. Zimmermann: Einführung in die Sprachwissenschaft. Skript zur Einführungsvorlesung am Studiengang "Allgemeine Sprachwissenschaft und Nebenfächer". 1989. Kursmaterialien 4.