MISCELLANEA LOGICA

TOM V

Kamila Bendová and Petr Jirků

editors

Charles University in Prague
The Karolinum Press
2003
http://logika.ff.cuni.cz/papers/misclogV.pdf

Reviewed by: RNDr. Antonín Sochor, DrSc.,
Mgr. Ondřej Tomala

Supported by Jan Hus Educational Foundation.

ISBN 80–246–0799–9

Contents

Modal Logic as Higher-Order Logic

Mircea Dumitru

Abstract Propositional modal logic is usually viewed as a generalization and extension of propositional classical logic. The main argument of this paper is that a good case can be made that modal logic should be construed as a restricted form of second order classical logic. The paper examines one aspect of this second order connection having to do with an incompleteness phenomenon. The leading concept is that modal incompleteness is to be explained in terms of the incompleteness of standard second order logic, since modal language is basically a second order language.

Some basic notions of modal metalogic; frames and validity

In the beginning I shall sketch the metalogic framework of my research. People acquainted with what follows in this section are advised to move to the next section.

A general interpretation for a language of sentential modal logic (LSML) has three component parts: a set of worlds W, an accessibility relation R, and for each world $w \in W$, an associated assignment of truth-values to sentence-letters. There is also a particular world designated the actual world; but in the investigation that follows the latter component plays no role.

It is useful to redefine our notion of *general interpretation* in two stages. A *Kripke-frame*[1] is a pair $F = (W, R)$, where W is a set of worlds, and R is a binary relation on W. A frame F is said to be reflexive (symmetric, transitive, etc.) if and only if (iff) R_F is reflexive (symmetric, transitive, and so on). A *Kripke model* for LSML is a pair $M = (F, V)$, where F is a frame and V is a function defined for each sentence-letter π of LSML. V assigns each such π a subset of W (intuitively, the worlds at which π is true); thus $V(\pi) \in \mathcal{P}(W)$. If $M = (F, V)$ and $F = (W, R)$, we may also write $M = (W, R, V)$ or $M = (W_M, R_M, V_M)$. If $M = (F, V)$, then M is said to be *based on* F.

We now define *three semantic concepts*, that of (i) a formula's being *true at a world* in a model M, (ii) of a formula's being *valid in* a model M, and (iii) of a formula's being *valid in* a frame F.

To define being true at a world in a model M, we recursively define a relation $\models$ (read: "verifies"), a subset of $(M, w) \times \mathrm{Prop}(L)$, as the least relation satisfying:

EA: $(M, w) \models \pi$ iff $w \in V_M(\pi)$, for each sentence-letter π in LSML,

E¬: $(M, w) \models \neg\Phi$ iff $(M, w) \not\models \Phi$,

E&: $(M, w) \models \Phi \,\&\, \Psi$ iff $(M, w) \models \Phi$ and $(M, w) \models \Psi$,

E∨: $(M, w) \models \Phi \vee \Psi$ iff $(M, w) \models \Phi$ or $(M, w) \models \Psi$,

E→: $(M, w) \models \Phi \to \Psi$ iff $(M, w) \not\models \Phi$ or $(M, w) \models \Psi$,

E□: $(M, w) \models \Box\Phi$ iff $\forall u \in W_M$(If $R_M(w, u)$ then $(M, u) \models \Phi$),

E◇: $(M, w) \models \Diamond\Phi$ iff $\exists u \in W_M(R_M(w, u)$ and $(M, u) \models \Phi)$.

We can define now our two notions of validity: A formula Φ is *valid in a model* $M = (W_M, R_M, V_M)$ iff $(M, w) \models \Phi$ for every $w \in W_M$; we may write this as $\models_M \Phi$. A formula Φ is *valid in a frame* $F = (W_M, R_M)$ iff for every model M based on F, the formula Φ is valid in M; we may write this as $\models_F \Phi$.

These definitions concern only single formula validity. However, we can define complementary notions of semantic consequence in the same way: $\Sigma \models_M \sigma$ iff for every $w \in W_M$, if every member of Σ holds at w

[1]Whenever ambiguity does not rear its ugly face, I shall suppress the qualifying phrase "Kripke" when I speak about frames or models. However, when the contrast between Kripke-frames or models, on the one hand, and general-frames or models, on the other hand, is crucial for my explanation and for understanding the issue, I will restore the phrase which marks the contrast.

then σ holds at w; in other words, if $(M, w) \models \Sigma$ then $(M, w) \models \sigma$. And $\Sigma \models_F \sigma$ iff for every M based on F, we have $\Sigma \models_M \sigma$. (I shall assume Σ finite, so that $\Sigma \models_M \sigma$ iff $\models_M \wedge\Sigma \rightarrow \sigma$, where $\wedge\Sigma$ is the conjunction of all members of Σ. The well-known modal systems $\mathsf{S5}$, $\mathsf{S4}$, B, T all have the finite semantic consequence property: if Σ is infinite and $\Sigma \models_F \sigma$, then there is a finite subset Σ_0 of Σ such that $\Sigma_0 \models_F \sigma$; so the restriction to finite Σ is not significant for these systems.)

We define now the main metalogical concepts of interest for this paper. A *deductive system* of modal logic is either the system K[2] or a proper extension of K obtained by adding a decidable collection of axiom-sequents to K at least one of which is not itself K-derivable. A deductive system S is *sound with respect to* a class of frames $\mathcal{K}$ iff: if $\Sigma \vdash_S \sigma$, then for every frame $F \in \mathcal{K}$, we have $\Sigma \models_F \sigma$. If $\Sigma \models_F \sigma$ for every frame $F \in \mathcal{K}$, then the sequent $\Sigma \vdash_S \sigma$ is said to be $\mathcal{K}$-valid. So S is sound with respect to $\mathcal{K}$ iff every S-provable sequent is $\mathcal{K}$-valid. A deductive system S is *complete with respect to* a class of frames $\mathcal{K}$ iff: if $\Sigma \models_F \sigma$ for every frame $F \in \mathcal{K}$, then $\Sigma \vdash_S \sigma$. Equivalently, S is complete with respect to $\mathcal{K}$ iff every $\mathcal{K}$-valid sequent is S-provable. A deductive system S is *characterized* by a class of frames $\mathcal{K}$ iff S is both sound and complete with respect to $\mathcal{K}$, i.e. the S-provable sequents and the $\mathcal{K}$-valid sequents are the same. Lastly, a deductive system S is *complete simpliciter* iff there is some class $\mathcal{K}$ of frames such that S is characterized by $\mathcal{K}$.

Since its inception, at the end of 1950s and the beginning of 1960s, the possible worlds semantics has become an enormously successful program. Due to this powerful and flexible formal tool many modal systems, which by then had been investigated only with axiomatic means, got a real and insightful semantic interpretation. The methodological success of characterizing modal systems motivated the reasonable hypothesis that every modal deductive system is complete in the absolute sense defined above, i.e. it is characterizable by a class of Kripke-frames.

Well, today we know that this hypothesis is false, and we owe this piece of knowledge to the research of some modal logicians, like Kit Fine,

[2]The system K, so-called in honor of Saul Kripke, one of the inventors of possible worlds semantics, is the simplest modal system. We get it axiomatically by adding the axiom-sequent K: $\Box(\Phi \rightarrow \Psi) \rightarrow (\Box\Phi \rightarrow \Box\Psi)$ and the rule of necessitation Nec: if $\vdash_{\mathsf{K}} \Phi$ then $\vdash_{\mathsf{K}} \Box\Phi$, to any sound and complete deductive system of non-modal sentential logic.

S. Thomason or Johan Van Benthem, who are in the forefront of this discipline. And my aim here is just to show, first, how a semantic incomplete system looks like, and then to look for a profound and fundamental explanation of this interesting and also curious semantic phenomenon.

One way of putting this fact of there existing *incomplete* propositional modal logics is to say that there is a class of frames $\mathcal{F}$ that characterize a logic L that is not axiomatizable. A similar phenomenon occurs in second order classical logic, where one quantifies over subsets of the domain as well as over individuals. Classical second order validity is not axiomatizable; it too displays incompleteness aspects. To that problem Henkin offered a solution, which he called *general models*. In these, set quantifiers are restricted to a designated collection of subsets of the domain, and do not range over all subsets. Validity with respect to general models is axiomatizable. Henkin's general models were, in fact, the inspiration for the introduction of general frames into modal semantics.

Against the background sketched above, the gist of this paper is to give a second-order-based-explanation of modal incompleteness. The leading concept is that modal incompleteness is to be explained in terms of the incompleteness of standard second order logic, since modal language is basically a second order language. That is, the paper shows that there is a close connection between modal incompleteness and the incompleteness of standard second order logic. Roughly, the connection goes as follows. A completeness proof for an axiomatization or for a natural deduction system of a modal logic can be formalized in second order logic with standard semantics. At a certain point in the formalized proof, we need the existence of a certain set of possible worlds. Of course that set is in the range of second order quantifiers in standard second order models, but might not be in the allowed quantifier range of some general (Henkin) second order models. Thus the argument can be carried out in standard second order logic, which is not axiomatizable, and cannot always be carried out in the axiomatizable logic corresponding to general second order models. In effect, modal incompleteness is seen as an aspect of classical second order incompleteness.

Consequently, what follows falls into three sections. In the next section I shall present a very simple incomplete system, which was discovered by Johan Van Benthem. Then, in the following sections, I shall sketch two semantic systems for the language of second-order logic, which are needed in the last section for building–in its essentials–an explanation which ties this semantic phenomenon with the more profound fact that every second-order deductive system which is sound with respect to the standard semantics for its language is bound to be incomplete with respect to that semantics.

Beyond the clarification of certain technical aspects, the net result of my approach is that it sheds light on some unexpected connections between important results, which prima facie seem to be unconnected. To my mind, such links are very instrumental in pushing forward our subject, which as far as logic is concerned, let's remember what Frege said, is nothing more and nothing less than the truth itself.

The Incomplete System VB

We will show following Johan Van Benthem that a certain system of modal logic, VB (to honor Van Benthem) is incomplete, i.e. it is a system which is characterized by no class of frames $\mathcal{K}$. So a better tag for such a system would be "uncharacterizable system". The real form of uncharacterizability results is that of a conditional: "if system S is sound with respect to $\mathcal{K}$ then S is not complete with respect to $\mathcal{K}$".

We first define the system K^* to be the system K plus an additional axiom-sequent, viz.

K: $\Box(A \to B) \to (\Box A \to \Box B)$ plus the axiom-sequent $\Diamond\Box A \vee \Box A$.

Then, I just state with no proof that K^* is characterized by the class of frames in which every world is either a dead end or else is one step removed from a dead end; w is a dead end if it can see no world: $\forall u \neg Rwu$; w is one step removed from a dead end iff it can see some world which is a dead end: $\exists v(Rwv \,\&\, \forall u \neg Rvu)$.

We define now the system VB to be the system K plus the axiom-sequent $\Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B)$. The proof that VB is incomplete proceeds in two steps. First we show *Step 1*: Every frame for VB is a frame for K^*. (A frame F is said to be a frame for a system S if whenever $\Sigma \vdash_S \sigma$ we have $\Sigma \models_F \sigma$.) Then we show *Step 2*: $\Diamond\Box A \vee \Box A$ is not a theorem of VB. We have to make clear why is this establishing the incompleteness of VB. The reason is as follows: Suppose $\mathcal{K}$ is a class of frames with respect to which VB is sound. Then, once we have established Step 1, we may conclude that there are no counter examples to K^*-sequents based on frames in $\mathcal{K}$. In that case, $\models_K \Diamond\Box A \vee \Box A$. Hence, there is a sequent valid in $\mathcal{K}$ which, granted Step 2, is not derivable in VB. So VB is incomplete with respect to $\mathcal{K}$ (the class of frames with respect to which it is sound). I give now a few details and hints concerning the proof of these two steps. Proof of Step 1 requires a single lemma. Proof of Step 2 requires a sequence of lemmas.

Lemma 1 *Every frame for* VB *is a frame for* K^**.*

Proof Suppose $F = (W, R)$ is not a frame for K^*. We will be able to show then that F is not a frame for VB either. Since $\mathsf{K}^* = \mathsf{K}+\Diamond\Box A\vee\Box A$, and every frame *is* a frame for K, it follows that $\not\models_F \Diamond\Box A \vee \Box A$. As one knows this means in turn that there is a model M based on F and a world $w \in W_M$, such that $(M, w) \not\models \Diamond\Box A \vee \Box A$. If w were a dead end or could see a dead end, $\Diamond\Box A \vee \Box A$ would hold at w. Hence w is neither a dead end nor can it see one. Let u be a world that w can see. For the language $\{A, B\}$ we define a new model M' based on F in the following way: $V(A) = \emptyset$, i.e., A is false at every world. $V(B) = W - \{u\}$, where u is the previously chosen world which w can see. The lemma is established by showing that $(M', w) \not\models \Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B)$. Clearly, $(M', w) \not\models \Diamond\Box A$: though w can see some worlds, none of them are dead ends, and $V(A) = \emptyset$ (so at any world v that w can see, $\Box A$ fails). To show that $(M', w) \not\models \Box(\Box(\Box B \to B) \to B)$ we argue that $(M', w) \models \Diamond(\Box(\Box B \to B)\&\neg B)$. This follows from $(M', u) \models \Box(\Box B \to B)\&\neg B$ since w can see u. For this latter claim, observe that since $V(B) = W - \{u\}$, $(M', u) \models \neg B$. Also, u is not a dead end. Let v be any world that u can see. If $v = u$, then since $u \notin V(B)$, $(M', v) \models \Box B \to B$. (The conditional has a false antecedent.) And if $v \neq u$, since $v \in V(B)$, so again $(M', v) \models \Box B \to B$. ($v$ makes the consequent B true.) Hence every world u can see makes $\Box B \to B$ true which gives us the required result that $(M', u) \models \Box(\Box B \to B)$ and the Lemma follows, since we have shown that if $\not\models_F \Diamond\Box A \vee \Box A$, then $\not\models_F \Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B)$. ■

It remains to establish Step 2, that $\Diamond\Box A \vee \Box A$ is not a theorem of VB. The obvious way to try to do this is to look for a frame *for* VB in which $\Diamond\Box A \vee \Box A$ is not valid. But according to Lemma 1, there are no such frames. A less obvious procedure is to characterize a class D of *general models* about which we can prove that (a) all the VB-derivable sequents are valid in (every M in) D, but (b) $\Diamond\Box A \vee \Box A$ fails in at

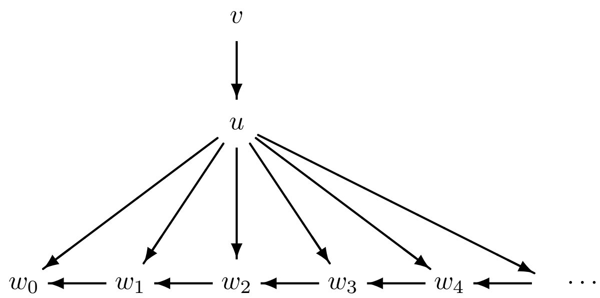


Figure 1: Recession frame

least one M in D. This suffices because to show that a sequent X is not provable in a system S, we only need to find some property or other which is possessed by all S- provable sequents but not by X. The class of general models D that we look for here is a subclass of the class of general models definable on a certain frame F, which is an example of a *recession* frame. F is defined as follows (see Fig. 1):

- ∘ $W = \{v, u, w_0, w_1, w_2, \dots\}$; that is, an infinite collection of worlds $\{w_i\}_{i \in \mathrm{N}}$ (N is the set of all natural numbers), and two additional worlds v and u.
- ∘ w_j can see w_i for every $i < j$; u can see every w_i; v can see u.

Next, we define the set G of subsets of W as the set of all finite subsets of $W - \{u\}$ and their complements in W:

- ∘ $G = \{\, X \subset W\, ;\, X$ is finite and $u \notin X$ or $\overline{X}$ is finite and $u \notin \overline{X}\,\}$. (Note that when $u \in X$, then X is infinite; the set G is closed under complements.)

More intuitively, the set G, which may be called the set of *allowable propositions in* W, can be generated by the following procedure:

> Take each finite subset X of W or infinite subset Y of W which has a finite complement in W;
> If X is finite and $u \notin X$, then put X in G;
> If Y is infinite and has a finite complement in W and $u \in Y$, then put Y in G;
> Nothing else will qualify as an element of G.

So, G is a set of sets each of which element being either finite, provided u is not a member of it, or infinite, provided it is the complement of a finite set and it (the infinite set) has u as one of its members. The class D of general models that we are interested in is the class of general models based on F with the valuation function V defined as:

- V is such that $V(\pi) \in G$ for every sentence-letter π.

Lemma 2 *Let M be a model (F, V), where F is the general recession frame defined above and V is the valuation function restricted as indicated above. If σ is a sentence whose sentence-letters are $\pi_1, \ldots, \pi_n$, and $V(\pi_i) \in G$ where $1 \leq i \leq n$, then $\{ w \in W \,;\, (M, w) \models \sigma \} \in G$.*

Proof Call the set of worlds at which a sentence is true the sentence's *worldset*. Then the Lemma says that the property of a sentence of having a worldset in G is preserved by the various ways of forming more complex from less complex sentences. The proof is by induction on the length of modal formulae. It suffices to show that $\neg$, $\vee$, $\Box$ preserve the property indicated in the lemma above, since the three operators form an expressively complete set. Let's observe that the Lemma 2 holds since the set G has the following properties:

(a) If $A \in G$ then $\overline{A} \in G$,

(b) If $A \in G$ and $B \in G$ then $A \cup B \in G$,

(c) If $A \in G$ then $\{ w \in W \,;\, \forall w' \in W (Rww' \Rightarrow w' \in A) \} \in G$. ∎

The role that this result plays in the overall strategy of the proof of VB's incompleteness is more clearly brought about by the following corollary: *in every model in D, the worldset of any sentence is an element of G.*

We can show now that every VB-sequent is D-valid, i.e. valid in every model $M \in D$.

Lemma 3 *If $\Sigma \vdash_{\mathsf{VB}} \sigma$ then $\Sigma \models_M \sigma$ for every $M \in D$.*

Proof Since every K-provable sequent is valid in the class of any model, we only have to show that for every $M \in D$, validity in M is preserved by the rule of theorem introduction (TI) using a substitution instance of the characteristic axiom-sequent of VB, say $\vdash \Diamond\Box p \vee \Box(\Box(\Box q \to q) \to q)$. We argue that in any model M in D, $(M, w) \models \Diamond\Box p \vee \Box(\Box(\Box q \to q) \to q)$ for every $w \in W_M$. Since w_0 is a dead end, $(M, w_0) \models \Box(\Box(\Box q \to q) \to q)$, and since every other world except v can see a dead end, $(M, w) \models \Diamond\Box p$ if $w \neq v$ and $w \neq w_0$. As for v itself, suppose aiming at absurdity that $(M, v) \not\models \Box(\Box(\Box q \to q) \to q)$. Then $(M, u) \models \Box(\Box q \to q)$ & $\neg q$ since u is the only world v can see (note that u is not therefore in the worldset of q). For each w_i then, $(M, w_i) \models \Box q \to q$. But trivially, $(M, w_0) \models \Box q$, w_0 being a dead end; so $(M, w_0) \models q$. In that case, $(M, w_1) \models \Box q$, so $(M, w_1) \models q$. By the same reasoning, for every i, $(M, w_i) \models q$. Thus the complement of the worldset of q is finite and contains u. Therefore the worldset of q is not a member of G, contradicting the corollary above. So $\Diamond\Box p \vee \Box(\Box(\Box q \to q) \to q)$ holds at all $w \in W_M$. ∎

We show now that the characteristic axiom of K^*, viz. $\Diamond\Box A \vee \Box A$, is not D-valid, for it gets refuted at some world in some model in D.

Lemma 4 *There is a model M in D such that $(M, w) \not\models \Diamond\Box A \vee \Box A$.*

Proof Let $M = (F, V)$ be a model based on the recession frame F defined by the valuation function $V(A) = \emptyset$ (the valuation function for every other sentence-letter, if any, is irrelevant for our purpose here of refuting the characteristic axiom-sequent of K^*). Then $M \in D$. Since $(M, w_i) \not\models A$ for all $i \in \mathrm{N}$, we have $(M, u) \not\models \Box A$, so $(M, v) \not\models \Diamond\Box A \vee \Box A$. And since $V(A) = \emptyset$, we have $(M, u) \not\models A$, so $(M, v) \not\models \Box A$.

Theorem 5 VB *is incomplete, i.e. it is not characterized by any class of frames.*

Proof Let's suppose that $\mathcal{K}$ is a class of frames for VB, i.e. with respect to which VB is sound. Then, by Lemma 1, all K^*-provable sequents are valid with respect to $\mathcal{K}$. In particular $\vdash \Diamond\Box A \vee \Box A$ is valid with respect to $\mathcal{K}$. By Lemmas 3 and 4 then, there is a $\mathcal{K}$-valid sequent, viz. $\models_{\mathcal{K}} \Diamond\Box A \vee \Box A$, which is not VB-provable. So VB is incomplete with respect to $\mathcal{K}$. Since $\mathcal{K}$ was arbitrary, this shows that VB is incomplete with respect to any class of frames for which it is sound, and thus VB is incomplete in an absolute sense. ∎

Semantics for the language of second-order logic (LSOL)

Basically, the concept of interpretation in second-order logic is similar to the one in first-order logic. Now I am going to present two distinct kinds of interpretation for the language of second-order logic: the *standard* interpretation for 'real' second-order logic, and then the Henkin (or general) interpretation for second-order logic.

Standard semantics for LSOL

A standard model for a language of second-order logic is basically the same kind of structure as a model for a first-order language, namely a pair $\langle D, I \rangle$ where D is the *domain* of the model (a set of objects), and I is an *interpretation* function that gives evaluation clauses for each logical connective (the same evaluation for every model) and assigns appropriate kinds of object constructed from objects that belong to D to each non-logical symbol in the language. To be more specific about this point it has to be added that standard second-order semantics sides with the semantics for the language of first-order logic in virtue of the fact that the domain of both is of the same type, viz. a set of individual objects. So, by setting a domain D the range of both first-order and second-order variables is thereby settled. The function I will take care, as it were, of the assignment of an appropriate object constructed from objects drawn from D to each non-logical symbol. On the other hand, standard second-order semantics differs essentially from Henkin semantics insofar as only in the case of the latter, and not in the case of the former, one should divide the domain of the interpretation in separate ranges: one for the first-order variables (individual variables) and one for second-order variables (sentential variables, n-place function variables, and n-place predicate variables), for any $n \in \mathrm{N}$.

The Tarski-style standard semantics for a second-order language will consist in an extension of the concept of first-order model for the language of first-order logic along the following lines. A standard model of a language of second-order logic, which contains at least one second-order variable, is a structure $\langle D, I \rangle$, where D is a set of objects, and I is an interpretation function. A *variable-assignment* is a function from each first- and second-order variable to elements drawn from D. Thus, a variable-assignment will assign a member of D to each first-order variable, a function from D^n to D to each n-place function variable, and a subset of D^n to each n-place relation variable. Let's observe that in the standard semantics a variable-assignment for an n-place predicate variable X^n in a language of second-order logic is a function from X^n to the set of *all* n-tuples drawn from D, i.e. the powerset of D^n.

Let now $\mathbf{M} = \langle D, I \rangle$ be a model and s an assignment on $\mathbf{M}$. The denotation of the n-place function variable $f(t_1, .., t_n)$ in $\mathbf{M}$ under assignment s is the value of the function $s(f^n)$ in $\mathbf{M}$ at the sequence of members of D denoted by each term t_i in $\mathbf{M}$ under s. (The denotation function for terms of the language of second-order logic is straightforwardly obtained from its first-order counterpart.)

Satisfaction will be the same kind of relation between models, assignments and formulae as in first-order logic, and we will get the proper inductive definition for a second-order formula's being true in a model $\mathbf{M}$ under an assignment s by adding the following three new clauses for an atomic second-order formula, a second-order universal quantification over function variables, and a second-order universal quantification over predicate variables, respectively. Thus,

I^S If X^n is an n-place predicate variable and $t_1, .., t_n$ is a sequence of n terms, then $\mathbf{M}, s \models X^n t_1, .., t_n$ iff the sequence of members of D denoted by each t_i under the assignment s is an element of $s(X^n)$.

II^S $\mathbf{M}, s \models \forall f \Phi$ iff $\mathbf{M}, s' \models \Phi$ for every assignment s' that is exactly like s at every variable except possibly f.

III^S $\mathbf{M}, s \models \forall X \Phi$ iff $\mathbf{M}, s' \models \Phi$ for every assignment s' that is exactly like s at every variable except possibly X.

In virtue of the inter-definability of $\exists$ and $\forall$ the corresponding clauses for $\exists f$ and $\exists X$ can be easily derived from the clauses (II^S) and (III^S) above.

Henkin-semantics for LSOL

The second semantics for the language of second-order logic is the Henkin semantics. The distinctive feature of it is that n-place predicate variables and n-place function variables can range over strict subsets of D^n and $D^n \times D$, respectively. In other words, the range of every predicate variable and function variable is a fixed subset of relations and functions on the domain, which may very well not include all the relations and all the functions on D^n and $D^n \times D$, respectively.

A Henkin model is a 4-tuple $\mathbf{M}^H = \langle D, D^*, F, I \rangle$ in which D and I are the domain of the model and an interpretation function for the non-logical vocabulary of the language, respectively. The new items in this new kind of model, viz. D^* and F, are a sequence of sets of relations on D^n, and a sequence of sets of functions on $D^n \times D$, respectively. Thus, for any finite $n \in \mathrm{N}$, $D^*(n)$ is a non-empty subset of the powerset of D^n, and $F(n)$ a non-empty subset of functions from D^n to D. The intuitive idea behind this construction of a Henkin-model is that the n-place predicate variables range over $D^*(n)$ and the n-place function variables range over $F(n)$.

A *variable-assignment* on a Henkin model differs significantly from its counterpart on a standard model. Although it is still a function that maps first-order variables into members of D, it varies essentially from what a variable-assignment is on a standard model with respect to predicate and function variables. Thus, a variable-assignment on a Henkin model maps each n-place predicate variable to a member of $D^*(n)$, which as we already remarked may be a proper subset of the powerset of D^n, and each n-place function variable to a member of $F(n)$, which likewise may be a proper subset of the collection of functions from D^n to D.

The remaining part of Henkin semantics is basically the same as the standard semantics, except of course for the new meaning that 'variable-assignment' gets in the Henkin semantics. There are then four new clauses:

- I^H Let $\mathbf{M}^H = \langle D, D^*, F, I \rangle$ be a Henkin model and s an assignment on $\mathbf{M}^H$. The denotation of $f(t_1, .., t_n)$ in $\mathbf{M}^H, s$ is the value of the function $s(f^n)$ at the sequence of members of D that are the references of each t_i, $1 \leq i \leq n$, on $\mathbf{M}^H$, under s.
- II^H If X^n is an n-place predicate variable, and $t_1, .., t_n$ a sequence of terms, then $\mathbf{M}^H \models X^n t_1, .., t_n$ if the sequence of members of D that are the references of each t_i, $1 \leq i \leq n$, on $\mathbf{M}^H$, under s, is an element that belongs to $s(X^n)$.
- III^H $\mathbf{M}^H, s \models \forall f \Phi$ if $\mathbf{M}^H, s' \models \Phi$ for every assignment s' on $\mathbf{M}^H$, which agrees with s at every variable except possibly at f.
- IV^H $\mathbf{M}^H, s \models \forall X \Phi$ if $\mathbf{M}^H, s' \models \Phi$ for every assignment s' on $\mathbf{M}^H$, which agrees with s at every variable except possibly at X.

The whole difference between standard semantics and Henkin semantics can be accounted for in terms of the different meanings that are attached to the phrase 'every assignment' in (II^S) and (III^S) on the one hand and (II^H) and (III^H) on the other hand. In the case of standard semantics an assignment to an n-place predicate variable and to an n-place function variable makes the variables range over the whole powerset of D^n, and over the collection of all functions from D^n to D, respectively. Whereas in the case of the Henkin semantics the collection of assignments may be *restricted* to those assignments only that assign members of different $D^*(n)$, where $D^*(n) \subseteq D^n$, and $F(n)$, where $F(n) \subseteq D^n \times D$, to the higher-order variables.

Explaining incompleteness in modal logic

Our semantics for modal logic is essentially a semantics for second-order monadic predicate logic (with a single binary relation constant R). If we inspect our definition of validity in a frame ($\models_F \Phi$) we see that for Φ to be valid in F it must be true in every world in every *model* based on F. The phrase "every model based on F" is a universal quantifier over assignments of subsets of W to the sentence-letters of the modal language. And since in the canonical translation of LSML into that language of second-order monadic predicate logic a sentence letter of the former becomes a monadic predicate of the latter, the force of "every model based on F" is intuitively—*no matter what subsets of W are assigned to the corresponding monadic predicates*. Hence, the quantification over models in the modal semantics can be captured by a second-order universal quantifier. For example, the statement

$$\models_F \Box(P \,\&\, Q)$$

says that for every $w \in W_F$, every world w can see satisfies P and satisfies Q, no matter what properties (subsets of W_F) are assigned to P and to Q. So, in second-order monadic logic, $\models_F \Box(P \,\&\, Q)$ can be written

$$F \models \forall P' \forall Q' \forall w \forall u (Rwu \rightarrow P'u \;\&\; Q'u)$$

in which we have changed "$\models_F$" into "$F \models$ " to indicate that the pair $\langle W, R \rangle$ is being regarded as an interpretation for a second-order language with a single binary relation constant R.

Here is the gist of my argument to the effect that the reduction of modal systems to second-order systems is effective for understanding modal incompleteness even though the same recursive procedure for reducing modal systems to second-order systems can be applied to *complete* modal systems which will also get mapped into second-order structures. Thus, when one says that second-order logic is incomplete what one has in mind mainly is that there is no second-order deductive system that is both sound and *complete* with respect to the standard semantics for second-order logic. Now, what we might expect to find in the case of translating *complete* modal systems into their second-order counterparts is something like this: after we translate the axioms and the transformation rules of such a modal system into corresponding second-order notations and mimic the complete modal system within its second-order counterpart, the *completeness* of that modal system boils down to the fact that *all* the second-order sentences which are *semantic consequences* of the second-order counterparts of the axioms of that modal system are *exactly* the second-order counterparts of the modal sentences which are deductive consequences within that complete modal system, i.e. are *exactly* the second-order counterparts of the theorems of that complete modal system. Roughly speaking, everything that can be inferred in a second-order logic system from the second-order translations of the axioms of a complete modal system using the second-order counterparts of the transformation rules of that modal system is a counterpart of a theorem of that complete modal system.

On the other hand, in the case of an *incomplete* modal system the situation that obtains is roughly of the following kind: one begins with the second-order translations of the axioms and the transformation rules of an incomplete modal system. Its *incompleteness* then consists in there being *logical (semantic) consequences* of the second-order counterparts of the modal axioms which can be inferred using second-order counterparts of the modal rules of transformation but which are such that they are not counterparts of any *modal theorems* of that incomplete modal system. Therefore, there is some second-order sentence which is the translation of some modal sentence under some recursive translation schemata from the language of that modal system into the language of a second-order system, and which is *valid* on the class of second-order structures which are the counterparts of the class of frames with respect to which the incomplete modal system is sound and which is such that its modal counterpart is *not derivable* within that incomplete modal system. Hence, that modal system is not complete with respect to the class of frames *for* that system, i.e. with respect to that class of frames with respect to which the system is *sound*. Thus the system is not characterizable and it is incomplete in an absolute sense.

To carry out the details of this reductive argument we have to show how the language of sentential modal logic can be mapped into the language of second-order logic. To this purpose we need a collection of recursive rules of translation (schemata) that will take formulae (wffs) of the language of sentential modal logic (LSML) as input and will yield the corresponding formulae of the language of second-order logic (LSOL) as output. What we look for here is a language in which the translation that is carried over is instrumental for the explanation that is sought here, viz. incompleteness in modal logic as a second-order phenomenon. And it turns out that what we need is a second-order language that for obvious reasons will be called the language of canonical translation (LCT). In a few words, what we are after here is the bringing about of a mechanism that will allow us to recast the whole apparatus needed to prove the incompleteness of the system VB into the terms that are proper to second-order logic.

The *lexicon of LCT*: One individual variable w; no individual constants; a sentence letter $\wedge$; for each sentence letter π of LSML except $\wedge$, the corresponding monadic-predicate letter λ_π; for each sentence letter π of LSML except $\wedge$, the corresponding monadic-predicate variable τ_π; sentential connectives, second-order and first-order quantifier symbols $\forall^2$, $\exists^2$, $\forall$, $\exists$, and parentheses. The *syntax of LCT*:

f-at: $\wedge$ is an atomic wff; if λ is any predicate letter and τ any predicate variable then λw, τw are atomic wffs;

f-con: If Φ and Ψ are wffs then so are $\neg\Phi$, $\Phi \,\&\, \Psi$, $\Phi \vee \Psi$, $\Phi \rightarrow \Psi$, and $\Phi \equiv \Psi$.

f-q^1: If Φ is a wff, then $\exists w\Phi$, and $\forall w\Phi$ are wffs.

f-q^2: If Φ is a wff, then $\exists^2\tau_\pi\Phi$, and $\forall^2\tau_\pi\Phi$ are wffs.

f!: Nothing is a wff unless it is certified as such by the previous rules.

The recursive schemata for translating modal formulae of LSML into LCT:

Trans^2-at: $\text{Trans}^2[\wedge, v] = \wedge$, where v is a fixed first-order variable;

$\text{Trans}^2[\pi, v] = \lambda_\pi v$, if π is a sentence-letter in LSML other than $\wedge$ and λ_π is the prime predicate corresponding to the sentence-letter π;

Trans^2-$\neg$: $\text{Trans}^2[\neg\Phi, v] = \neg\text{Trans}^2[\Phi, v]$;

Trans^2-&: $\text{Trans}^2[(\Phi \,\&\, \Psi), v] = (\text{Trans}^2[\Phi, v] \,\&\, \text{Trans}^2[\Psi, v])$;

Trans^2-$\vee$: $\text{Trans}^2[(\Phi \vee \Psi), v] = (\text{Trans}^2[\Phi, v] \vee \text{Trans}^2[\Psi, v])$;

Trans^2-$\rightarrow$: $\text{Trans}^2[(\Phi \rightarrow \Psi), v] = (\text{Trans}^2[\Phi, v] \rightarrow \text{Trans}^2[\Psi, v])$;

Trans^2-$\equiv$: $\text{Trans}^2[(\Phi \equiv \Psi), v] = (\text{Trans}^2[\Phi, v] \equiv \text{Trans}^2[\Psi, v])$;

Trans^2-$\Box$: $\text{Trans}^2[\Box\Phi, v] = \forall v'(Rvv' \rightarrow \text{Trans}^2[\Phi, v])$;

Trans^2-$\Diamond$: $\text{Trans}^2[\Diamond\Phi, v] = \exists v'(Rvv' \,\&\, \text{Trans}^2[\Phi, v])$.

To get the second-order sentence counterpart of a modal sentence we apply the schemata Trans^2 from outside in. Thus, where Φ_μ is any sentence in LSML, we start by an application of the appropriate Trans^2 to the main connective of Φ_μ, and then at every subsequent step we apply appropriate Trans^2 schemata to the main connectives of each formulae thereby obtained. We stop the translation after Trans^2 has been applied to every atomic sentence letter that occurs in Φ_μ. It is worth observing that in (Trans^2-$\Box$) and (Trans^2-$\Diamond$), a new meta-variable v' occurs. Just for getting a unique translation for a necessitate or possibilitate formula one can make the stipulation that there is a specific order in which such variables are to be picked up when those two Trans^2 clauses are applied, e.g. first u, then v, then u', then v', and so on.

The result of these applications of Trans^2 will be an open sentence of LCT, with the predicate variables that correspond to sentence letters in Φ_μ free. Thus, if Φ_μ is $\Diamond(A \vee B)$, and the predicate-variables that correspond to A and B are X, and Y, respectively, then after obvious successive applications of Trans^2 what we get is the open second-order sentence Φ_σ with X and Y free: $\exists u(Rwu \,\&\, (Xu \vee Yu))$.

Now, Φ^*_σ is the *full* second-order translation of Φ_μ, $\text{Fsot}[\Phi_\mu]$ for short, iff Φ^*_σ is the universal closure of Φ_σ with respect to all free first- and second-order variables of $\text{Trans}^2[\Phi_\mu, w]$, and $\Phi_\sigma = \text{Trans}^2[\Phi_\mu, w]$. In symbols,

$$\text{Fsot}[\Phi_\mu] = \Phi^*_\sigma = \forall p_1 \ldots \forall p_n \forall w \text{Trans}^2[\Phi_\mu, w],$$

where $p_1, \ldots, p_n$ are the monadic predicate variables (second-order variables) corresponding to the sentence letters $\pi_1, \ldots, \pi_n$ which occur in Φ_μ.

Using this recursive procedure we can get the Fsot of the characteristic axiom-sequents of VB and K^*, respectively.

$$\begin{aligned}
&\text{Fsot}[\Diamond\Box A \vee \Box(\Box(\Box B \rightarrow B) \rightarrow B)] = \forall X \forall Y \forall w\{\\
&\qquad \exists u(Rwu \,\&\, \forall v(Ruv \rightarrow Xv)) \vee\\
&\qquad \forall u(Rwu \rightarrow ([\forall v(Ruv \rightarrow (\forall v'(Rvv' \rightarrow Yv') \rightarrow Yv))] \rightarrow Yu))\}\\
&\text{Fsot}[\Diamond\Box A \vee \Box A] =\\
&\qquad \forall X \forall w[\exists u(Rwu \,\&\, \forall v(Ruv \rightarrow Xv)) \vee \forall u(Rwu \rightarrow Xu)].
\end{aligned}$$

However, it is not only formulae of LSML that have to be mapped into corresponding formulae of LSOL. For to carry out the attempted explanation of modal incompleteness we also need a way of reconfiguring the modal possible world semantics and the main metalogical modal notions definable within that frame as a second-order semantics, and second-order metalogical notions, respectively. It is worth keeping in mind that with respect to modal languages two different modal semantic systems can be constructed, viz. one which is based on the notion of *Kripke-frame*, and a second one which is based on the notion of *General-frame*. The main modal concept of interest for the issue of completeness vs. incompleteness, viz. the notion of a *modal formula's being valid in a frame*, gets the well-known definition "true in every world in every model based on a given frame". And of course, the definition will differ according to whether the frame in question is a *Kripke-frame* or a *General-frame*. Now we show how to reconfigure the modal semantics as second-order semantics.

If F_K is a *Kripke-frame* for the sentences of LSML, define S_2, the second-order model-structure corresponding to F_K, as the model-structure S_2 whose domain D is the domain W_F of K_F, which assigns to the n-place predicate letter (constant) Acc a set of n-tuples of objects drawn from D^n that corresponds exactly to the n-tuples of worlds that K_F assigns to R_{K_F}. (Hence Acc has the same degree as R_{K_F}.) Accordingly, for any *Kripke-model* which is based on a *Kripke-frame*, define the second-order model corresponding to the *Kripke-model*, as the interpretation that in addition to the correspondence defined above between *Kripke-frames* and *second-order model-structures* is such that for each sentence-letter π in LSML that is assigned a truth-value by each world in W_K under the evaluation function V, it assigns to the corresponding

$\text{Trans}^2[\pi]$ the extension which consists in exactly those $w \in W_K$ such that $w(\pi) = \top$. As readers can very easily check for themselves, the interpretation S_2 thereby obtained is a standard second-order model of the sort defined before in the subsection about the standard semantics for LSOL.

The same kind of maneuver allows us to make the transition from a *general-frame* and *model* for LSML to a *Henkin-frame* or *model* for LSOL. The telltale difference between the current case and that worked out one paragraph back is the following. In addition to what we had before, here we need to map the set G of sets of worlds drawn from the domain W of a *general-frame* into the set D_H^* of subsets of the domain D_H of the second-order Henkin-model-structure. Then, as before, we let the valuation-function I_H assign each $\text{Trans}^2[\pi]$ the extension which consists in exactly those $w \in W$ such that $w(\pi) = \top$. Obviously, this amounts to an assignment of a set of n-tuples that belongs to D_H^* to each $\lambda_\pi \in$ LSOL, which mirrors exactly the modal counterpart where sets of worlds drawn from G_G are assigned under V_G to every $\pi \in$ LSML.

I am now in the position to state a result, which is needed for the explanation that I seek in this paper.

Theorem 6 *For any sentence* $\Phi_\mu \in$ LSML *there exists a unique corresponding sentence* $\Phi_\sigma \in$ LSOL *such that* $\Phi_\sigma = \text{Fsot}[\Phi_\mu]$*, and for any* Kripke-frame $K_F = \langle W, R\rangle$ *there exists a corresponding standard second-order model-structure* $S_2 = \langle D, \text{Acc}\rangle$ *such that* $\langle W, R\rangle \models_F \Phi_\mu$ *iff* $\langle D, \text{Acc}\rangle \models_{S_2} \Phi_\sigma$.

Further, for any sentence $\Phi_\mu \in$ LSML *there exists a corresponding sentence* $\Phi_\sigma \in$ LSOL *such that* $\Phi_\sigma = \text{Fsot}[\Phi_\mu]$*, and for any* general-frame $F_G = \langle W, R, G\rangle$ *there exists a corresponding* second-order Henkin-model-structure $H_2 = \langle D, D^*, \text{Acc}\rangle$ *such that* $\langle W, R, G\rangle \models_G \Phi_\mu$ *if and only if* $\langle D, D^*, \text{Acc}\rangle \models_H \Phi_\sigma$.

Proof By Trans^2 recursive schemata, Fsot, and induction. ∎

It's time now to take again a look at our proof of the uncharacterizability of VB. The idea now is to represent the incompleteness of VB as a consequence of the incompleteness of second-order logic with standard interpretation, which allows $\text{Fsot}[\mathsf{VB}] \models_2 \text{Fsot}[\mathsf{K}^*]$, even though, in a sense to be made more precise, $\text{Fsot}[\mathsf{VB}] \not\vdash_2 \text{Fsot}[\mathsf{K}^*]$.

The apparatus of translation will allow us to reconfigure the modal semantic consequence relationship shown above as holding between VB and K* as the following claim of second-order semantic consequence.

Lemma 7 Fsot[VB] $\models_2$ Fsot[K*], *i.e.*

$$\forall X \forall Y \forall w \{\exists u(Rwu \,\&\, \forall v(Ruv \rightarrow Xv)) \vee \\ \forall u(Rwu \rightarrow ([\forall v(Ruv \rightarrow (\forall v'(Rvv' \rightarrow Yv') \rightarrow Yv))] \rightarrow Yu))\} \\ \models_2 \forall X \forall w[\exists u(Rwu \,\&\, \forall v(Ruv \rightarrow Xv)) \vee \forall u(Rwu \rightarrow Xu)].$$

Proof We show $\neg$Fsot[K*] $\models_2$ $\neg$Fsot[VB]. Unsurprisingly, the argument recapitulates the proof of the Lemma 1. Some familiarity with this proof, some feeling of déja vu is to be expected, for the second-order argument that follows here conveys the same general idea of the proof, which we gave before, of VB's modal incompleteness. Using quantifier shift, modality shift, and truth-functional equivalences, Fsot[$\neg$K*] is equivalent to

$$\exists P \exists w[\forall u(Rwu \rightarrow \exists v(Ruv \,\&\, \neg Pv)) \,\&\, \exists u(Rwu \,\&\, \neg Pu)]. \qquad \text{(i)}$$

while Fsot[$\neg$VB] is equivalent to

$$\exists P \exists Q \exists w\{\forall u(Rwu \rightarrow \exists v(Ruv \,\&\, \neg Pv)) \,\&\, \exists u(Rwu \,\& \\ \&\, ([\forall v(Ruv \rightarrow (\forall v'(Rvv' \rightarrow Qv') \rightarrow Qv))] \,\&\, \neg Qu))\} \qquad \text{(ii)}$$

To obtain (ii), we need a suitable instance, i.e., suitable P_0, Q_0, and w_0. Let P_0 and w_0 be any property and world which yield a true instance of (ii). Then straight away we have the first conjunct in the body of (ii). We now have to find a Q_0 and an u_0 such that

$$Rw_0u_0 \,\&\, ([\forall v(Ru_0v \rightarrow (\forall v'(Rvv' \rightarrow Q_0v') \rightarrow Q_0v))] \,\&\, \neg Q_0u_0) \qquad \text{(iii)}$$

From our true instance of (i) we have $\exists u(Rw_0u \& \neg P_0u)$, hence there must be a u_0 such that Rw_0u_0. This gives us the first conjunct of (iii). Let Q_0 be (a property whose extension is) the set $W - \{u\}$. Then evidently we have the last conjunct of (iii). To obtain

$$\forall v(Ru_0v \rightarrow (\forall v'(Rvv' \rightarrow Q_0v') \rightarrow Q_0v)) \qquad \text{(iv)}$$

assume Ru_0v_0. Then, either (a) $u_0 = v_0$ or (b) $u_0 \neq v_0$. If (a) then $\forall v'(Rv_0v' \rightarrow Q_0v')$ is false, since Rv_0u_0 but $\neg Q_0u_0$; so

$$\forall v'(Rv_0v' \rightarrow Q_0v') \rightarrow Q_0v_0$$

follows. If (b) then again $\forall v'(Rv_0v' \to Q_0v') \to Q_0v_0$, since Q_0v_0. Thus (iv) holds, so (iii) holds, so by first and second order $\exists$I, (ii) holds. ∎

However, this result does not really "explain" why every frame for VB is a frame for K^*, since it merely restates our earlier proof in a different language. But it allows us to relate the incompleteness of VB to the non-existence of a sound and complete set of inference rules for second-order logic.

We should pay careful attention here, though!

The point is not that in second-order logic $\mathsf{VB} \nvdash_2 \mathsf{K}^*$. For deductively, there is no one thing which is second-order logic—instead, there are various significantly different sound deductive systems. And though none of them is complete, there is certainly *some* collection of rules determining a deductive consequence relation $\vdash_2$ such that $\mathsf{VB} \vdash_2 \mathsf{K}^*$. For instance, trivially, we could introduce a second-order system of deduction in which the step from any instance of VB to a corresponding instance of K^* is a primitive rule. Or, less trivially, inspection of the second-order counterpart of Lemma 1, which is already close to a formal deduction, indicates that we will be able to derive Fsot[K^*] from Fsot[VB] in predicative monadic second-order logic.

So what, then, is the explanation of the incompleteness of VB?

To make some progress into this issue it is useful to think of modal systems other than K as theories of modality, and K as the logic. Thus, where before we would have written $A \vdash_{\mathsf{T}} \Diamond A$, treating the T-sequent as part of the logic, i.e. as part of the definition of a new deductive consequence relation $\vdash_T$, we will now write instead that

$$A, \Box\neg A \to \neg A \vdash_{\mathsf{K}} A.$$

This gives rise to a notion of *formula completeness* that is the counterpart of the notion of system completeness:

- ◦ A formula σ of LSML is said to be complete iff all the semantic consequences of it and its substitution instances (relative to auxiliary premises and the class of all frames) are derivable from it in K. In symbols: σ is complete iff whenever $\Gamma, \sigma^* \models \gamma$ then $\Gamma, \sigma^* \vdash_{\mathsf{K}} \gamma$, where σ^* is a substitution instance of σ.

The result that VB is an incomplete system becomes in this terminology the result that $\Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B)$ is an incomplete formula, because

$$\Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B) \models \Diamond\Box A \vee \Box A$$

but

$$\Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B) \nvdash_{\mathsf{K}} \Diamond\Box A \vee \Box A.$$

The explanation of the incompleteness of VB is then (not that there is no second-order logic in which VB entails K^*, but rather) that the deductive consequence relation $\vdash_{\mathsf{K}}$ of modal systems (of theories of modality) is significantly weaker than $\vdash_2^{\mathrm{p}}$ in a precise sense:

Definition 8 *Let L and L^* be two languages and let T (from 'translation') be a function $T : L \to L^*$, from the sentences of L into the sentences of L^*. Then $\vdash^*$ is an L^*-consequence relation which is said to be a* conservative extension of *an L-consequence relation $\vdash$, with respect to T, provided the following holds for any set of L-sentences Σ and any L-sentence σ:*

$$T(\Sigma) \vdash^* T(\sigma) \textit{ only if } \Sigma \vdash \sigma.$$

So a conservative consequence relation is one in the new language which does not allow sequents to be proven unless they are the translation of provable sequents from the original language. Since we already know that

$$\Diamond\Box A \vee \Box(\Box(\Box B \to B) \to B) \nvdash_{\mathsf{K}} \Diamond\Box A \vee \Box A$$

and we have just seen that

$$\mathsf{VB} \vdash_2^{\mathrm{p}} \mathsf{K}^*,$$

this shows that $\vdash_2^{\mathrm{p}}$ is non-conservative over $\vdash_{\mathsf{K}}$.

Its extra strength comes in part from the fact that the second-order variables can substitute for, and be substituted by, any first-order formula with one free variable. But it is perfectly conceivable that there should be interpretations with a first-order definable set of worlds that is not the worldset of any modal sentence; for example, in any transitive model where there are two worlds u and v which see and are seen by the same worlds and which make the same atomic sentences true, no modal sentence can have $W - \{u\}$ or $W - \{v\}$ as its worldset. But "$(\,.\,.\,) \neq u$" is still a perfectly acceptable first-order formula which is satisfied by all and only the members of $W - \{u\}$. In general, then, in predicative monadic second-order logic we can reason with statements that cannot even be *expressed* in LSML, which allows us to prove sequents in the former logic which are not provable modally.

The overall moral, then, is that the system VB is uncharacterizable because of the lack of expressive power of LSML as compared to the expressive power of LSOL. Thus, as the case that I worked out in this paper shows, we can reason in predicative second-order logic with formulae that LSML has no power to express. That is the main rationale for there being the case that the sequent Fsot[VB] $\vdash_2^{\mathrm{p}}$ Fsot[MV] is provable in second-order logic, whereas its modal version, viz. $\mathsf{VB} \vdash_{\mathsf{K}} \mathsf{MV}$ is not provable modally.

References

[1] J. F. A. K. van Benthem. Two simple incomplete modal logics. *Theoria*, XLIV(1):25–37, 1978.

[2] J. F. A. K. van Benthem. Syntactic aspects of modal incompleteness theorems. *Theoria*, XLV(1):63–77, 1979.

[3] J. F. A. K. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, 1983.

[4] G. Boolos and R. Jeffrey. *Computability and Logic*. Cambridge University Press, third edition, 1989.

[5] G. Boolos and G. Sambin. An incomplete modal logic. *J. Philosophical Logic*, 14:351–358, 1985.

[6] M. J. Cresswell. Incompleteness and the Barcan formula. *J. Philosophical Logic*, 24:379–403, 1996.

[7] R. Epstein. *The Semantic Foundation of Logic. Predicate Logic.* Oxford University Press, 1994.

[8] K. Fine. An incomplete logic containing S4. *Theoria*, XL(1):23–29, 1974.

[9] M. Fitting. Basic modal logic. In D. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and*

Logic Programming, volume 1: Logical Foundations, pages 365–448. Clarendon Press, Oxford, 1993.

[10] G. Forbes. *An Introduction to Modal Logic.* Tulane University, 1994.

[11] G. Hughes and M. Cresswell. *A Companion to Modal Logic.* Methuen & Co. Ltd, 1984.

[12] G. Hughes and M. Cresswell. *A New Introduction to Modal Logic.* Routledge, London, New-York, 1996.

[13] M. Manzano. *Extensions of First Order Logic.* Number 19 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1996.

[14] K. Segerberg. An essay in classical modal logic. Technical report, University of Uppsala, 1971. 3 volumes.

[15] S. Shapiro. *Foundations Without Foundationalism. A Case for Second Order Logic.* Clarendon Press, Oxford, 1991.

[16] S. K. Thomason. An incompleteness theorem in modal logic. *Theoria*, XL(1):30–34, 1974.

Mircea Dumitru
Faculty of Philosophy, University of Bucharest
dumitru@fil.unibuc.ro

Axiomatic Fuzzy Set Theories

Libor Běhounek[*]

Introduction

Fuzzy sets originated with Zadeh's paper [14] in which he proposed to investigate sets whose membership function is not confined to the usual two values $\{0, 1\}$, but can instead take values in the whole $[0, 1]$ interval. Zadeh's fuzzy sets found many applications in engineering and the theory of fuzzy sets grew quickly up, although it was often objected that it was just a theory of functions from some fixed domain to $[0, 1]$ and thus a part of real analysis.

Based upon the idea of fuzzy sets, fuzzy logic soon emerged: generalizing the obvious correspondence between intersection and conjunction, union and disjunction, and complementation and negation, fuzzy logical connectives were defined and used for the study of fuzzy predicates and statements. For a long time, however, fuzzy logic and fuzzy sets were rather engineering tools than a well-designed mathematical theory. Driven just by applications, it lacked (meta)theoretical grounding and general results; developed mostly by engineers for practical purposes, it suffered from arbitrariness in definitions and often even mathematical imprecision.

The development of t-norm based fuzzy logic (see the next section) can be viewed as an attempt to remedy these defects and set fuzzy logic on solid mathematical and methodological grounds. It enabled,

[*]The work on this paper was supported by the grant GA ČR 401/01/0218 'Modal and Many-Valued Logics' of the Grant Agency of the Czech Republic. Thanks are due to Petr Cintula for a minor correction.

i.a., a classification and metamathematical study of 'reasonable' systems of fuzzy logic, making them a legitimate part of the long tradition of many-valued logics equipped with Hilbert-style calculi and algebraic semantics. With [5], this effort seems to be completed to the point in which fuzzy logics can serve as ground theories for fuzzy mathematics—i.e., for constructing axiomatic theories over these logics, aimed to describe fuzzy arithmetic, fuzzy geometry or fuzzy analysis. Among these mathematical disciplines, perhaps the most prominent is fuzzy *set theory,* both because it would make Zadeh's fuzzy sets an axiomatic theory of its own right (and no longer one derived from classical analysis of real functions), and because a great deal of mathematics (if not all, as is the case with classical set theory) could hopefully be modeled within its framework.

This paper tries to summarize the 'state of art' in this field and show possible ways of its further development.

T-norm based fuzzy logic

The idea of t-norm based fuzzy logics is as follows (for details, see [5] or [6]):

We require that the semantics of fuzzy conjunction—a function from $[0,1]^2$ to $[0,1]$—satisfy a few natural conditions: commutativity, associativity, monotonicity in both arguments, the classical behavior on the classical truth values $\{0,1\}$, and continuity. Such functions had previously been studied in the context of probabilistic metric spaces and acquired the name *triangular norms,* or *t-norms* (continuous, as we require continuity).

Having fixed a continuous t-norm for the semantics of conjunction, we take as the semantics of implication the maximal function[3] that guarantees the validity of modus ponens. Negation can conveniently be defined as $\neg\varphi \equiv_{\mathrm{df}} \varphi \to 0$ and it turns out that the functions min and max are definable in terms of conjunction and implication. The latter is taken to be the semantics of disjunction, the former to be another ("weak") conjunction; since it might be found useful, we add it to the system of logical connectives and call *min-conjunction.* The equivalence connective can be defined as conjunction of both implications.

[3]Called the *residuum* of the t-norm; a t-norm and its residuum form the adjoint pair of a residuated lattice on $[0,1]$.

In this natural way, the choice of a particular t-norm $*$ for the fuzzy semantics of conjunction induces the fuzzy semantics of all other common propositional connectives. It is easily seen that these truth functions behave classically on the classical values $\{0, 1\}$. Every t-norm thus corresponds to a particular fuzzy propositional logic which generalizes the semantics (while retaining the syntax) of classical propositional logic.

Among t-norms, three are of special importance: the *minimum,* also called the *Gödel t-norm* $x * y = \min(x, y)$; the *product,* $x * y = x \cdot y$; and the *Łukasiewicz t-norm* $x * y = \max(0, x + y - 1)$.[4] Mostert-Shields' characterization theorem asserts that every continuous t-norm is a combination (a special kind of ordinal sum) of isomorphic copies of these three prominent t-norms.

Having fixed the fuzzy semantics of propositional connectives, one can define the evaluation of formulae in the obvious way (truth-functionally). A formula which in any evaluation gets the value 1 is called a 1*-tautology.* Given a t-norm $*$, the set of its 1-tautologies can be understood as the *logic* of $*$.

It turns out that some formulae (such as $\varphi \to \varphi$) are 1-tautologies for any continuous t-norm (we call them *t-tautologies*). There is a finite set of axioms, which (together with modus ponens) proved to be correct and complete with respect to t-tautologies; this axiomatics is called *basic fuzzy logic* or BL.

Any continuous t-norm logic (as well as classical logic) is an extension of BL. By adding axioms that express the specific properties of the Gödel, product or Łukasiewicz t-norms we get correct and complete axiomatizations of their respective sets of 1-tautologies. The resulting calculi are called Gödel, product, and Łukasiewicz propositional fuzzy logic and denoted by G, Π, and Ł respectively.

In fuzzy mathematics (including set theory), however, propositional calculus is not sufficient, as we need to talk about predicates (e.g., $=$ and $\in$). Fuzzy predicate calculus is therefore developed, extending the fuzzy propositional calculus in the following way:

The language of BL is enriched by predicate symbols, variables and the quantifiers $\exists$ and $\forall$ (and possibly also constants and function symbols). The usual Tarskian semantics is adjusted to handle the $[0, 1]$ truth

[4]The names refer to the fact that the corresponding many-valued logics were previously studied (for other purposes) by Gödel and Łukasiewicz.

values: individual variables take values in some fixed domain M (the universe of discourse), the interpretation of an n-ary predicate is a function from M^n to $[0,1]$ and the semantics of $\exists$ and $\forall$ is that of supremum and infimum. The axiomatics of BL (or any of its extensions) is supplemented with a finite number of axioms for quantifiers and the deduction rule of generalization. The resulting logic BL$\forall$ is not, however, complete with respect to all predicate t-tautologies in $[0,1]$ (the set of which is not recursively axiomatizable).[5]

The same construction yields the predicate calculi G$\forall$, $\Pi\forall$, and Ł$\forall$ for the three prominent extensions of BL.[6] Of these, only G$\forall$ is complete with respect to the standard [0,1]-semantics, while the sets of 1-tautologies of product and Łukasiewicz t-norms are not recursively axiomatizable. If we add LEM (the law of excluded middle, i.e., the schema $\varphi \vee \neg\varphi$) to BL (or any of its consistent extensions), we get the classical logic Bool.

Within the framework of t-norm based predicate calculi we are ready to develop fuzzy mathematics by specifying the list of predicates and extra-logical axioms of a *theory over* BL (or over any of its extensions). The elaborated metatheory of t-norm logics gives a solid background for such attempts.

In order to increase the expressive power of the underlying logic, one may wish to enrich its language with new truth-functional connectives. As the structure of truth values is a real interval, the most usual additional connectives are elementary arithmetical operations. Some of them are already available in Ł and Π,[7] and the resulting logic is usually an extension of Ł or Π (or both). Another common supplement is the unary operator Δ (Baaz delta) with the semantics $\|\Delta\varphi\| = 1$ iff $\|\varphi\| = 1$, and $\|\Delta\varphi\| = 0$ otherwise. The enriched logics BL$\forall\Delta$, G$\forall\Delta$, etc. allow to express the crisp[8] fact that the truth value of a formula is 1.

[5]It is complete with respect to a special class of linear residuated lattices called BL-algebras. In our present setting, we are not interested in such a general semantics, since the main motivation for fuzzy sets is $[0,1]$-semantics. For some fuzzy predicate logics, $[0,1]$-complete axiomatizations using infinitary rules are known.

[6]Since in fuzzy set theory we always work in predicate logic, we shall occasionally drop the symbol $\forall$ in the names of logics.

[7]E.g., the bounded sum is $\neg(\neg\varphi \mathbin{\&} \neg\psi)$ in Ł, the bounded fraction comes out to be the semantics of implication in Π, etc.

[8]In fuzzy sets and fuzzy logic it is traditional to call the 'classical' two-valued sets and formulae *crisp* (as opposed to *fuzzy*). A formula is crisp iff its value is always 0

Another option is to internalize truth values by means of nullary connectives (truth constants). The common choices are either to add just one truth constant h for .5 (the truth values 0 and 1 are definable), or to add some recursively representable dense subset of $[0, 1]$, usually the rational numbers.[9] Adding truth constants to the language has rather serious consequences—the resulting logic is no longer weaker than classical logic Bool; on the other hand, new concepts are available. An example of a system containing both arithmetical operations and truth constants is Takeuti-Titani TT∀ (see [11] or [5, 9.1]).

Unrestricted comprehension

The aspiration to develop set theory over many-valued logics is not new, though originally it was not connected with Zadeh's fuzzy sets (then not yet known). After the invention of many-valued logic there was a hope that Russell's paradox could be avoided and the comprehension scheme of naïve set theory

$$\exists z \forall x (x \in z \equiv \varphi(x))$$

retained if a suitable many-valued logic were used. Indeed, in Łukasiewicz 3-valued logic Russell's paradox cannot be reproduced: the formula $z \in z \equiv z \notin z$ is 1-satisfiable (if $z \in z$ takes the middle value, then so does $z \notin z$ and the equivalence perfectly holds). Though a modified version of Russell's paradox can be obtained in any finite-valued Łukasiewicz logic, it was conjectured that unrestricted comprehension would be consistent in infinite-valued Łukasiewicz logic.

The consistency of the comprehension scheme for *open* formulae (even in 3-valued Łukasiewicz logic) was proved by Skolem [9]. The result for infinite-valued logic was later extended to a wider class of formulae by Fenstad [2]. The full consistency result was finally proved by White [13] by means of proof-theoretical methods.[10]

or 1, otherwise it is fuzzy. It is easily seen that φ is crisp iff $\varphi \vee \neg\varphi$ holds.

[9] Adding truth constants for all numbers in $[0, 1]$ would make the language uncountable, which could have serious drawbacks; a dense subset makes all reals accessible by means of infima and suprema, i.e. ∀ and ∃. If the language is strong enough (e.g., contains the bounded sum and the product), adding just h suffices for the definability of a dense set of truth values.

[10] The logic used by White is even stronger than Ł∀: by means of an infinitary rule it completely axiomatizes the predicate logic of the Łukasiewicz t-norm on $[0, 1]$.

Unlike classical logic, Łukasiewicz logic is thus capable of harboring the comprehension scheme of naïve set theory. Since it postulates the existence of all sets definable by a formula, no additional axioms are required to secure the existence of unions, power sets, pairs, subsets, infinity, etc., nor is there any need for GB-style classes. The axiomatization ŁC of set theory by this single intuitive scheme is therefore very appealing.

It remains an open question whether a worthwhile set theory can be developed over ŁC. A negative result due to White prevents adding extensional identity, as

$$\forall q(q \in x \equiv q \in y) \rightarrow \forall q(x \in q \equiv y \in q)$$

is inconsistent with ŁC.[11]

The question is currently under study of the Prague workgroup on axiomatic fuzzy sets (P. Hájek, Z. Haniková, P. Cintula and the present author). If the theory proves worthwhile, it will have many interesting properties, viz. the existence of a universal set, definability of rational truth constants, and an apparatus for modeling the distinction between extensions and intensions of concepts represented by sets. If ever, the details will be published in a separate paper.

ZF-style fuzzy set theories

Another option for the development of set theory over fuzzy logic is to give some (possibly incomplete) set of axioms that allow to carry out the basic set-theoretical constructions, such as forming unions or intersections. As this is exactly the motivation of classical Zermelo-Fraenkel set theory, the choice of axioms is inevitably close to that of ZF, and any such theory has a strong ZF-flavor.

In contrast to comprehension-style set theory, ZF-style set theory can in principle be based on any t-norm logic.[12] It turns out, however,

[11] On the other hand, there are some interesting positive results in comprehension-based set theories over substructural logics (esp. BCK, which is weaker than Ł) that are directly transferable to ŁC. See [12].

[12] For most fuzzy logics, ZF-style is the only option, as the comprehension scheme is inconsistent with any logic in which a bivalent operator is definable, e.g. with any extension of G or Π (due to their negation), or any logic containing Baaz's Δ.

that the task is most easily accomplished over (the extensions of) Gödel logic. This is due to special properties of G∀, which is not only a t-norm fuzzy logic, but also an important extension of intuitionistic logic Int.[13] History proved that intuitionistic logic is strong enough for meaningful mathematics, including set theory; the enterprise is therefore destined to success. However, the results and methods of G∀-based set theory are hardly transferable to any other fuzzy logic as they heavily depend on the idempotence of Gödel conjunction.

In this approach it is common to use axioms identical in form to those of ZF. If the background logic is weaker than classical (as is the case with G∀), then besides consistency we should also be concerned with non-triviality: an inadvertent choice of axioms may make the theory become identical to classical ZF, which makes the whole effort meaningless. In particular, if the theory proves LEM, then it coincides with the same theory over classical logic (as Bool = BL + LEM = Int + LEM).[14] That this can easily happen is shown by the negative results of (even weaker) intuitionistic set theory (summarized e.g. in Grayson [4]). For instance, it can be shown that AC implies LEM in a very weak set theory over Int. Similarly, the axiom of well-foundedness (regularity) implies LEM, and many other axioms do so as well.

The non-triviality and consistency of a ZF-style theory can be proved by constructing a non-classical model. The usual method is that of sheaf models (i.e., analogues of Boolean-valued models for classical ZF). A $[0, 1]$-valued model for a ZF-style fuzzy set theory is presented by Takeuti and Titani [10] and [11][15] who adapted Heyting-valued models for intuitionistic set theory used by Powell [8] and Grayson [4]. Their models show, i.a., that suitable substitutes for AC and regularity (which lead to bivalence) are Zorn's lemma and $\in$-induction (both of which hold in these models). Moreover, the above authors construct an interpretation ('Powell inner model') of classical ZF within their theories by constructing a class of hereditarily stable sets ('hereditarily crisp', in fuzzy terminol-

[13] It extends Int by Dummet's axiom of prelinearity $(\varphi \to \psi) \vee (\psi \to \varphi)$ and the axiom of constant domains $\forall x(\chi \vee \psi(x)) \to (\chi \vee \forall x \psi(x))$. G∀ is therefore identical to Dummet predicate logic with constant domains. The interval $[0, 1]$ equipped with the Gödel t-norm (plus the derived operations) is a linear Heyting algebra.

[14] If the background logic contains non-trivial truth constants, then LEM even leads to a contradiction, since their existence is incompatible with bivalence.

[15] Their former system uses a logic which is essentially G∀ with truth constants, the latter (TT∀, see also [5, 9.1]) is enriched with additional arithmetical operations.

ogy). Along the way, they define and investigate several concepts interesting from the point of view of fuzzy set theory (e.g., natural numbers, ordinal numbers, power-set hierarchy V_α, etc.). However, their concern is mainly in the models and metamathematical properties of their systems, and less in the development of set theory within the system of axioms.

Recently, the method was successfully employed by Hájek and Haniková [7], who found a suitable ZF-style axiomatics over BL$\forall\Delta$ (N.B.: a much weaker logic with generally non-idempotent conjunction) and constructed a sheaf model for it, thus showing its non-triviality.

The initial results in set theory over pure G$\forall$ and a few general methodological remarks, which is the work of the present author, are discussed separately in the next section. In what follows, let any set theory over G$\forall$ be denoted by GST for short.

Set theory over Gödel logic

Even though the papers mentioned in the previous section contain a few advanced topics in set theory, such as ordinal numbers, the motivation for their choice of axioms is not discussed and seems a bit arbitrary. Even the very axioms of ZF are less motivated in another logic since the intuitions that justify them need not be as compelling as they are in Bool. In some systems, the axioms of ZF need technical adjustments for consistency's sake (e.g., Δ's occurring in strange places in the formulae,[16] or additional preconditions), which further complicate their justification. Some further axioms are added for technical reasons (e.g., to be able to prove the mutual interpretability result). The overall impression of such axiomatics is not very convincing and it is very doubtful that the consequences of such axioms should be *the* set theory for fuzzy logic.

Until it is clearer which axiomatics becomes canonical, a viable strategy for fuzzy set theory is to investigate consequences of the most simple and most intuitive axioms, such as identity axioms, extensionality, the existence of an empty set or singletons. This is the strategy pursued by the present author in [1].[17] The background logic is chosen to be pure G$\forall$

[16] Admittedly, the Δ's can partly be explained away by considerations regarding the deduction theorem and conservativity of defined functions in fuzzy logics with Δ.

[17] This section gives only the outline of the work done; the technical details will be published separately.

(without Δ or truth constants), not only because it is an important intermediary and t-norm logic, but also because it is both a restriction and extension of logics in which most relevant attempts to 'fuzzify' set theory take place, viz. TT and Int.

As stated above, when giving an axiomatics for fuzzy set theory, much attention should be paid to the motivation of the axioms. Neither should the choice of the defined predicates and functions blindly follow that of classical set theory. Some important concepts lack any classical motivation, because in the stronger logic Bool they are too trivial to be introduced. For instance, the *double complement* (or *support* in fuzzy terminology) of a set x, i.e. the set $\{ q ; \neg\neg(q \in x) \}$, coincides with the set x itself in Bool. The motivation for defined concepts should stem not only from classical theory, but also from engineering practice and applications of fuzzy sets.

Furthermore, within a weaker logic, classically equivalent definitions split into more concepts and we should decide which (one or more) of them correspond best to the classical counterpart and which to employ. A trivial example can be found at the very outset of GST, supposing the extensionality axiom $x = y \equiv \forall q(q \in x \equiv q \in y)$. Two classically equivalent definitions of inequality suggest themselves:

$$x \neq_1 y \equiv_{\mathrm{df}} \neg\forall q(q \in x \equiv q \in y)$$
$$x \neq_2 y \equiv_{\mathrm{df}} \exists q((q \in x \wedge q \notin y) \vee (q \notin x \wedge q \in y)).$$

It is easily shown that these concepts, a crisp 'negation of equality' and a fuzzy 'witnessed difference', are different, both are meaningful and not reducible to each other. GST should therefore employ both of them.

In some cases, neither of variant definitions of a concept preserves intuitions motivating it within Bool. The choice of an adequate definition should then not be guided by superficial analogies, but rather the *rôle* of the concept in set theory. For example ordinal numbers cannot be required both hereditarily transitive and well-ordered (nor even linearly ordered), as it would imply LEM. Though it may seem very important for ordinals to be linearly ordered, their key rôle in set theory resides in recursive definitions. The definability of ranks and the correctness of transfinite recursion are guaranteed by the definition of ordinals as hereditarily transitive sets (see [8] and [4]). This definition should therefore be adopted, although such ordinals need not be linearly ordered.

A robust method for defining concepts over weak logics is definition by abstraction (see [3]), since it is very resistent to the weakening of the underlying theory. For example, cardinal numbers can be defined in ZFC as a special kind of ordinals ("alephs"), but this definition relies heavily on the axiom of choice and it is not immediately clear whether there should be any connection between alephs and magnitudes of sets in a weaker logic. On the other hand, cardinalities can conveniently be defined in very weak theories over weak logics (including Int) as abstraction classes of the equipotence relation, as soon as the concepts of function and bijection have been defined. By this definition, the main features of the concept of cardinality are preserved, as they are embodied by the equipotence relation used for abstraction.

The easier part of building set theory over G∀ is class theory, i.e., a theory of fuzzy classes over some fixed domain of individuals. (In class theory, classes can contain individuals, but membership of classes in classes is not taken into account.) The most part of class theory is just reexpressing properties of fuzzy predicate calculus in set-theoretical language. Nevertheless, it is fuzzy class theory which is in fact used by engineers, rather than fuzzy set theory. Many concepts important in applied fuzzy sets can therefore be defined and studied already in class theory over G∀, for instance: the relations of equality and inclusion (and their negative variants, see above), compatibility and disjointness; the properties of crispness, fuzziness and normality; the usual binary class operations of union, intersection, difference and symmetric difference; the unary operations of support and complementation; empty and universal classes.

Fuzzy *set* theory proper begins with the concepts of singleton and pair. An incomplete list of set notions and constructions that can be defined and studied in GST follows: singletons, pairs, power sets, set unions and intersections, ordered pairs and tuples, Cartesian products, relations, functions and bijections, natural numbers (following [11]), ordinal numbers (following [8]), cardinal numbers (by abstraction), countability, continuum, etc.

As regards the general strategy for choosing the axioms, there are two options. Either only such axioms that are provable in classical set theory are chosen. In this approach, fuzzy sets are meant to *generalize* classical sets by admitting non-crisp sets: the classical theory with LEM correctly describes the universe of sets if by chance all sets under consideration are crisp. The consistency is guaranteed by the consistency of classical set theory; however, the non-triviality must be proved (usually the sheaf model from [11] or a similar one serves well). As shown above, such a theory is not just a selection of theorems valid in a weaker logic, as more non-equivalent notions can be defined and their relationship studied under the weaker logic.

The other option is to choose some axioms that are incompatible with classical theory,[18] for instance the *axiom of fuzziness*:[19]

$$\exists x \neg(\varphi(x) \vee \neg\varphi(x)).$$

The motivation for this kind of axioms is to secure that we are describing the universe not of crisp, but rather fuzzy sets. Classical axioms may possibly be required in the class of hereditarily crisp sets, which correspond to the crisp sets of the classical theory. In this approach, the non-triviality is automatic; it is the consistency that must be proved (usually also by sheaf models).

As outlined above, GST can be elaborated to a system which allows all common set-theoretical constructions. The thesis [1] shows, however, that G∀ has too little expressive power to serve as the background logic for a full-fledged *fuzzy* set theory. What it lacks most is a greater control over the system of truth values. It can be proved that in G∀, no proportional scaling of truth values is expressible. As a consequence, no equality relation corresponding to the 'similarity of characteristic functions' can be defined. Also some notions used in the applications of fuzzy sets, such as the kernel of a set, are undefinable. GST can therefore be regarded only as a *fragment* of a full-fledged fuzzy set theory, which must eventually be developed in a stronger system (probably one close to TT∀, though it will necessarily be more complicated both in theory and metatheory than G∀). The value of GST rests mostly in being the basic core of such systems. Nevertheless, in rare situations when only ordering of truth values is available and no scaling can be given at all, GST (possibly extended with Δ) seems to be the most adequate fuzzy set theory.

---

[18] Such axioms can even be classically inconsistent, and yet consistent in the weaker logic. Notice that the presence of non-trivial truth constants in the background logic falls under this option, since their existence is incompatible with bivalence.

[19] In G∀ the axiom entails that there is a non-crisp set.

Future work

An overview of the current state of art in the field of axiomatic fuzzy set theory has been given. What remains to be presented is a prospect of future research and directions of expected progress in the field.

Much can be done in set theories over logics based on G∀ and its extensions (incl. TT∀), as these are by far best known and most widely developed. Standard set-theoretical concepts of comparing magnitudes of sets could be investigated, especially the concept of finitude, which will probably be more complicated than it is already in Bool, but hopefully much simpler than it is in Int. Better understanding of functions is necessary for any such investigation. Various systems of subsets (such as filters, closure systems, or topologies), can readily be explored, as their definitions require only elementary apparatus. The independence of the axioms (esp. the axiom of double complementation) is not yet known. Other simple theories like fuzzy arithmetic should be developed and the possibility of their modeling within set theory explored. Attempts to internalize metamathematics can be undertaken.

As regards other fuzzy logics, class theory should be developed first. Probably the most important task (in set theory over any logic, as this has not yet been finished in GST, either) is to investigate simple relational structures, such as orderings, graphs, etc., as these are essential for any further development of set theory. A link to category theory should be made in order to make the definitions more plausible: a suitable definition of morphism between sets should be find (probably coinciding with that of mapping) and the definitions of categorially definable concepts (e.g., the Cartesian product) made consonant with it. A fuzzified category theory may well become necessary for this task.

References

[1] L. Běhounek. Teorie množin v Gödelově logice (Set theory over Gödel logic). Master thesis, Philosophical Faculty of Charles University, Department of Logic, 2002.

[2] J. E. Fenstad. On the consistency of the axiom of comprehension in the Łukasiewicz infinite valued logic. *Mathematica Scandinavica*, 14:65–74, 1964.

[3] G. Frege. *Die Grundlagen der Arithmetik.* W. Köbner, Breslau, 1884.

[4] R. Grayson. Heyting-valued models for intuitionistic set theory. In M. Fourman, C. Mulvey, and D. S. Scott, editors, *Application of Sheaves*, volume 743 of *Lecture Notes in Mathematics*, pages 402–414. Springer, Berlin, 1979.

[5] P. Hájek. *Metamathematics of Fuzzy Logic.* Kluwer, 1998.

[6] P. Hájek and L. Godo. Deductive systems of fuzzy logic. *Tatra Mountains Mathematical Publications*, 13:35–66, 1997.

[7] P. Hájek and Z. Haniková. A development of set theory in fuzzy logic. In M. Fitting and E. Orlowska, editors, *Beyond Two: Theory and Applications of Multiple-Valued Logic*, volume 114 of *Studies in Fuzziness and Soft Computing*, pages 273–285. Physica-Verlag, Heidelberg, 2003.

[8] W. C. Powell. Extending Gödel's negative interpretation to ZF. *J. Symbolic Logic*, 40(2):221–229, June 1975.

[9] T. Skolem. A set theory based on a certain 3-valued logic. *Mathematica Scandinavica*, 8:71–80, 1960.

[10] G. Takeuti and S. Titani. Intuitionistic fuzzy logic and intuitionistic fuzzy set theory. *J. Symbolic Logic*, 49(3):851–866, Sept. 1984.

[11] G. Takeuti and S. Titani. Fuzzy logic and fuzzy set theory. *Archive for Math. Logic*, 32:1–32, 1992.

[12] K. Terui. Light affine set theory: A naive set theory of polynomial time. *Studia Logica*, to appear, 2003.

[13] R. B. White. The consistency of the axiom of comprehension in the infinite-valued predicate logic of Łukasiewicz. *J. Philosophical Logic*, 8:509–534, 1979.

[14] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

Libor Běhounek
Philosophical Faculty, Charles University
Palachovo nám. 2, 116 38, Praha 1
libor.behounek@ff.cuni.cz

Do we have to deal with partiality?

Marie Duží[*]

Abstract We present a semantic theory based on the Transparent Intensional Logic, which takes into account the fact that some sentences of natural language may happen not to have any truth value. Using two-level semantics, where between an expression E and the denoted entity D there is the expressed meaning (perhaps Frege's sense of E or Church's concept of D) of the expression, we model the meaning of E by a TIL construction C. We show that the construction C can fail to construct anything, in other words, in natural language we sometimes use empty concepts; and the denotation D, being a function, can fail to return a value at an argument. Hence, presenting some sound objective philosophical arguments in favour of the need to handle partial functions, we look for adequate technical tools to accommodate them in our logic. Different perspectives of a logician, mathematician and computer scientist toward non-denoting expressions naturally lead to different strategies of handling partial functions by particular logical systems. A brief overview of the ways of handling non-denoting terms in quantification logic is presented, and the strategy of 'partiality being propagated up' accepted by TIL is examined in details.

Introduction

When dealing with semantics of a language, a logician has to take into account that there are 'holes in reality'. By these 'holes' we do not,

[*]The work on this paper was supported by the grant GA ČR 401/03/0403 'Principles of Logical Analysis' of the Grant Agency of the Czech Republic.

of course, mean 'holes in physical reality', but in abstract functions modelling particular functional/relational dependencies between entities. Thus, for instance, considering office of the President of the Czech Republic, we have to take into account that it may happen that the office is not occupied by any individual. When talking (*de re*) about the president in such a period when there is none, no reasonable sentence about such a non-existent president can be true. We are going to show that such a sentence cannot be false either, yet 'the President of CR' is a full-right term denoting an abstract function, and should not be reduced in Russell's way. We intend to show that if we want to conceptually depict reality adequately, these abstract modelling functions have gaps, they have to be partial. Where do these gaps come from? To explain this, we have to first introduce our basic philosophical approach to the analysis.

An expression E of a language denotes something beyond the language; let's call that something the *'denotation'* of E. Prevailing logical theories identify the meaning of E with this denotation (intension in possible world semantics—intensional logics, reference in extensional logics). But it is easy to show that 'denotational approach' is too coarse grained: theories based on standard logic conflate meanings of sentences (and generally of expressions) that are classically equivalent, even if these sentences are not strictly synonymous. To adduce an example, we can literally quote [15, p. 1]:

John walks. (1)

John walks and Bill talks or doesn't talk. (2)

[...] According to our intuition these are not synonymous sentences in the strictest sense, the first one does not even mention Bill or his talking, while the second one does. But no theory based on classical logic will be able to discriminate between the two. In Montague Grammar (see [14]), for example, the meaning of (1) will be a certain set of possible worlds (the set of worlds in which John walks) and the meaning of (2) will be the intersection of this set with the set that is the meaning of 'Bill talks or doesn't talk'. But the latter sentence is a classical tautology and so its meaning will be the set of *all* possible worlds. Hence (1) and (2) are predicted to be synonymous, which strictly speaking they are not.

At this point we agree with Muskens: this inaccuracy may seem not to be that important. After all, the theory gives a correct prediction about the relation of logical *entailment* here. The two sentences entail each other, and this fact correctly follows from Montague's theory. Can we not take the position that the relation of co-entailment, although being a more coarse-grained than the strict synonymy, will suffice as a good approximation to the theory of synonymy in natural language? Unfortunately, this strategy does not work. The natural languages are rich enough to express the differences in the meaning of co-entailing sentences. Well known are the examples of embedding the sentences in a context of propositional attitudes. Somebody can easily believe (1) without believing (2).

Here is another Muskens' example:

Mary sees John walk. (3)

Mary sees John walk and Bill talk or not talk. (4)

Since (1) and (2) are treated equivalent, by the Principle of Compositionality (3) and (4) must be too. But (4) entails that Mary sees Bill, while (3) clearly does not. We still do agree with Muskens: *we cannot content ourselves with an imperfect approximation of the relation of synonymy, since such an imperfection will immediately and necessarily lead to further imperfection in the way the relation of logical consequence is treated.*

But unlike Muskens, we do not believe and do not agree that 'again, the introduction of partiality helps' in this case. We do not follow Muskens' theory of *partial possible worlds* or Barwise & Perry 'Situation Semantics', and do not introduce partiality whenever it seems that it might help to further perfection. We strongly adhere to Tichý's Platonism and realism. Sentences of a given language (or rather propositions denoted by them) simply *are* true, false or neither, independently of our "allowing" them to be so, and they are never both true and false. Thus, for example, if Mary does not see Bill at all, then, of course, she cannot and does not see him talk or doing anything else, which does not mean that (as Muskens claims) "the sentence 'Bill talks' will be undefined, that is, neither true nor false, in the part of the world that is seen by her and as a consequence the sentence 'Bill talks or doesn't talk' and (2) are both undefined *in that situation as well*". Since (1) may still be true *in that situation*, in Muskens' partial logic (2) no longer follows from (1). Here we have to disagree. The sentence 'Bill talks' is either true or false[20], and we can at most admit that it is undefined *for Mary*. Note that Muskens uses 'classical entailment' to argue that (2) does not follow from (1). But (2) does follow from (1), independently of Mary's cognitive abilities and *independently of situations*. And (1) and (2) are true or false, dependently on situations, but independently of Mary's seeing that, there is no reason for introducing partiality here. According to Muskens, co-entailment in a partial theory will be a better approximation to synonymy than classical co-entailment is. In our opinion, Muskens is going to model our cognitive faculties, and his theory can be treated as an epistemic theory. New 'truth values' he introduces, namely N (undefined—neither true nor false) and B (overdefined—both true and false), are actually not (objective) truth-values of propositions, but, say 'subjective degrees of cognitive knowledge' in a particular situation. We can even introduce infinitely many such 'truth-values', for instance an interval between 0 and 1, which can map 'degrees of preciseness of measurement', or 'degrees of our conviction in the truth', or any other (subjective) degrees, and build up fuzzy logics, etc. We can even introduce new (objectively correct) *inference rules* of our logic that would better map the relation of logical consequence. Still, the relation of co-entailment, or co-denotation, will always be just an *approximation* to synonymy, and a counter-example would always be easy to find. Here is a simple one:

*Prague has one million forty-eight thousands five hundred and seventy six (*1.048.576*) inhabitants.* (5)

Prague has fifth power of sixteen (16^5*) inhabitants.* (6)

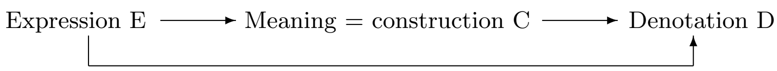
Prague has "hexa hundred thousands" (100.000_{16}*) inhabitants.* (7)

Dependently on 'situations', or states-of-affairs, these sentences are (simultaneously) either true or false. This means that they are equivalent, i.e. true (or false) in exactly same states-of-affairs, independently of anybody's knowing (believing, seeing) or even allowing that. Still, obviously, they are not synonymous, one can easily believe (5) without believing others.[21]

[20] We simplify the semantics of proper names here, taking 'Bill' simply as a label of an individual. Moreover, in TIL conception, there are no 'non-existing' individuals, we work with a stable, constant domain of individuals.

[21] Frankly, it took me a lot of effort to prove their equivalence, and my linguistic competence had nothing to do with that.

There are many theories that differ only minimally from Montague's original theory. Some of these theories fit the facts better than others, some are simpler and some are more sophisticated. But all these theories have one feature in common: they are 'denotational'. In other words, they neglect an important second (or rather primary) level of semantics, which is a never-ending source of troubles and imperfections. Between an expression E and its denotation D there is the *'mode of presentation'* of D (perhaps Frege's sense of E). We explicate this 'mode of presentation' by a precisely defined logical *construction* (or *Church's concept*), and identify this construction with the meaning of E. The whole conception can be illustrated by the following diagram[22]:



Now the question arises again: where do those (objective) truth-value gaps (or generally 'value gaps') come from? In what follows we show that the construction C can fail to construct anything, in other words, we do use empty concepts as well; and the denotation D, being a function, can fail to return a value at an argument. Hence, presenting some sound objective philosophical arguments in favour of the need to handle partial functions, we are going to look for adequate technical tools to accommodate them in our logic.

Transparent Intensional Logic (TIL)

Our underlying logical system is Transparent Intensional Logic (TIL) originated by late Pavel Tichý and developed by his followers, to name at least Pavel Materna. Omitting, for the sake of simplicity, the definition of Ramified Hierarchy of Types, we just briefly summarize[23]:

The expression E denotes ('talks about') an entity D (if any), which is either

(a) an intension, i.e. a mapping from possible worlds (type ω) to chronologies (type τ) of objects of some type α. Briefly, an α-in-

[22] For-runners of this conception were Frege [8], and, in particular, Church [3], see [13].

[23] For precise definitions, see, e.g., [23], [13].

tension is a partial function of type: $(\omega \rightarrow (\tau \rightarrow \alpha))$. The type of an α-intension is denoted by $((\alpha\tau)\omega)$, abbreviated $\alpha_{\tau\omega}$;

(b) an *extension*, which is not a function from possible worlds ...;

(c) a higher-order entity of type $*_n$, i.e. a construction (or an entity involving such a construction) that belongs to a higher-order type of order $n+1$, which constructs an object of a type of order n $(n \geq 1)$. We also say that the construction is of order n.

Note: entities we can talk about are organised in an infinite hierarchy. On the lowest level there are entities belonging to a first-order type. They are either basic non-functional entities (truth-values $o = \{\mathrm{True}, \mathrm{False}\}$, individuals—members of type ι, time-points/real numbers—members of type τ, and possible worlds—members of logical space ω), or (partial) functions from first-order types $\beta_1, \ldots, \beta_n$ to a first-order type α (the definition is inductive): members of types $(\alpha\beta_1 \ldots \beta_n)$. These 1^{st}-order entities are not structured from the "*algorithmic point of view*", though they may be wholes consisting of parts (like mathematical 'structures') and sets having elements. Entities belonging to the second-order types are (algorithmically) structured constructions of 1^{st}-order entities (and functions involving such constructions), 3^{rd}-order entities are constructions of 2^{nd} or 1^{st}-order entities, and so on; constructions are always more structured (belong to a higher-order type) then their products.

The meaning of E is conceived as the construction (of the denoted object D, if any) expressed by E, i.e. a higher-order object of type $*_n$ (order $n+1$), which constructs (identifies) a (lower-order) object of type of order n. Construction is a procedure, i. e. an 'instruction', 'recipe' of the way of arriving at the denoted entity. There are four basic kinds of constructions.[24] The first two (i and ii) are simple, one step instructions, they manage the direct contact with entities; the other two (iii and iv) are structured complexes:

i. *Variables* $x, y, z, \ldots$ construct objects of the respective types dependently on valuations v, they v-construct. Variables are not letters, they are constructions and particular 'x', 'y', 'z', ... are names of these constructions.

[24] Tichý in his [23] introduces also *execution* and *double-execution*. In conceptual modelling we use an *n-tuple* and *projection*. Since these additional kinds of constructions are not needed for our purpose, we don't introduce them here.

ii. *Trivialisation* of an entity X, denoted °X, is the most primitive concept of X (delivering X without any perspective). Where X is an entity whatsoever (even a construction), °X constructs X.

iii. *Closure:* If $x_1, x_2, \ldots, x_n$ are distinct variables that *v*-construct entities of types $\alpha_1, \alpha_2, \ldots, \alpha_n$, respectively, and Y is a construction that *v*-constructs an entity of type β, then $[\lambda x_1 \ldots x_n \, \mathrm{Y}]$ is a construction called closure, which *v*-constructs a function of type $(\beta\alpha_1 \ldots \alpha_n)$ mapping $\alpha_1 \times \ldots \times \alpha_n$ to type β.

iv. *Composition:* If X *v*-constructs a function F of a type $(\beta\alpha_1 \ldots \alpha_n)$, and $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ *v*-construct entities $\mathrm{A}_1, \ldots, \mathrm{A}_n$ being, respectively, of types $\alpha_1, \ldots, \alpha_n$, then the *composition* $[\mathrm{X}\ \mathrm{Y}_1 \ldots \mathrm{Y}_n]$ *v*-constructs the value (entity of type β, if any) of the (partial) function F on the arguments $\mathrm{A}_1, \ldots, \mathrm{A}_n$. Otherwise the composition $[\mathrm{X}\ \mathrm{Y}_1 \ldots \mathrm{Y}_n]$ does not construct anything, it is *v-improper.*

'Holes in reality'?

Now we have to return to the question raised above. Where do those 'holes in reality' come from, and how shall we accommodate them within our logical system? We already refused the theories of *partial possible worlds* modelling some subjective degrees of particular agents' cognition. Our logical space (basic type ω) is the set of possible worlds, each of which is conceived as the maximum set of possible consistent *facts.* According to Parmenides, we cannot talk about something that does not exist, and we agree. Yet, there as if were "non-existing entities" in reality we do talk about, like water sprites, round squares, Pegasus, the (present) King of France, etc. What then do we talk about? It has been convincingly shown by Pavel Tichý [20] that (non)existence is not a property of individuals but of intensions, namely the property of being instantiated in a given world w and time t. Hence intensions are *partial functions*, mappings, which do not associate each argument of the domain with *exactly* one value but with *at most* one value. We can state another important claim[25]: *empirical expressions always denote (talk about) non-trivial intensions, never extensions, and where it seems*

[25]This claim might be conceived as an explication of the epistemic category of empirical expressions: by an *empirical* we mean such an expression, the reference of which cannot be known *a priori*, but is always given *a posteriori.* Since logic is an *a*

that they denote extensions (values of the intensions in a world/time pair—the reference) they only possess de re supposition.[26]

When claiming, e.g., that Pegasus does not exist, we do not talk about any "non-existent individual" (there are none), but about an individual office, an intension of type $\iota_{\tau\omega}$, and we just claim that this function does not pick out any individual in the actual possible world now. Which object, if any, an empirical expression determines in the actual world at the present moment is a matter of contingent facts and cannot be "calculated" a priori. E.g., a language-user will be able to determine the *condition* for being the Mayor of Dunedin, but his linguistic competence won't suffice for fixing the individual, if any, who is actually and presently the Mayor of Dunedin. Similarly, a language-user will be able to calculate empirical truth-*conditions* of a sentence but never its actual and present truth-*value*. The contrast is between conditions and satisfiers, concepts and their instances. The definite description, however, may fail to have any value, and so will the whole sentence fail to have any truth-value.[27] Hence TIL 'language of constructions' must deal with 'non-denoting terms'.

But in mathematics we seem to talk about "non-existent" entities. In case of, e.g., 'round square' there is no problem; the denoted entity is an empty class (of geometrical figures). Yet, there was a lot of dispute about the greatest prime and mathematicians had to prove that the greatest prime does not exist. What then do we talk about? In this case we cannot say that the greatest prime denotes a non-instantiated intension. Mathematical (or generally analytical) expressions never denote (non-trivial) intensions, but extensions (if anything). The term 'the greatest prime' is a non-denoting term and in most of the classical logics it would not be a well-formed term, and thus would not have any meaning. But we can use this term so that to compose a reasonable (and true) sentence that the greatest prime does not exist. Hence in TIL 'the greatest prime' is a 'full-right term' and its meaning is simply a (composed) construction. To compose this construction, we use the (*partial*) function Singulariser ι^{α} of type $(\alpha(o\alpha))$, which associates each α-singleton with the only member of this set, and does not return any value at other

priori discipline, the *reference* is out of the scope of logic, and the *denotation* is an intension, not its value in a world / time.

[26]An extensive study on the *de dicto* / *de re* distinction is presented in [4].

[27]See [19].

sets (that have more than one members, or that are empty). In our case we use ι^τ and compose it with the set of prime numbers greater than or equal to any prime (x, y range over τ, prime / $(o\tau)$[28] is the set of primes): $\lambda x\,[[{}^{\circ}\text{prime}\ x] \wedge \forall y\,[[{}^{\circ}\text{prime}\ y] \supset [x \geq y]]]$. The resulting composition is

$$[{}^{\circ}\iota^\tau\ \lambda x\,[[{}^{\circ}\text{prime}\ x] \wedge \forall y\,[[{}^{\circ}\text{prime}\ y] \supset [x \geq y]]]]. \qquad \text{(GP)}$$

(read: 'the only x such that x is prime and for all y that are primes it holds that $x \geq y$).

But since the above set is empty, the function ι^τ cannot return any value and the whole construction GP fails to construct anything, we say that GP is *improper*. Hence the claim on (non)existence of mathematical entities is a claim about the whole construction, not its result, the construction has to be 'mentioned' (i.e. constructed by, e.g., trivialization). The analysis of the sentence 'The greatest prime does not exist' goes as follows: let Improper be the class of improper constructions of order 1 (/ $(o*_1)$). The sentence claims that the above construction (of the greatest prime) GP (/ $*_1$) belongs to this class, so we get:

$$[{}^{\circ}\text{Improper}\ {}^{\circ}[{}^{\circ}\iota^\tau\ \lambda x\,[[{}^{\circ}\text{prime}\ x] \wedge \forall y\,[[{}^{\circ}\text{prime}\ y] \supset [x \geq y]]]]].$$

Note that considering the four constructions defined above, only composition $[\mathrm{X}\ \mathrm{Y}_1 \ldots \mathrm{Y}_n]$ can be v-improper. It fails to construct anything if either

(a) the (partial) function F constructed by X does not have any value at the argument $(\mathrm{A}_1, \ldots, \mathrm{A}_n)$ constructed by constructions $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$, or

(b) any of the constructions X, $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$, is v-improper.[29]

Hence in our language of constructions a 'non-denoting' term (standing for an improper composition) is a well-formed term. The only case when the term '$[\mathrm{X}\ \mathrm{Y}_1 \ldots \mathrm{Y}_n]$' would not be well formed (actually $[\mathrm{X}\ \mathrm{Y}_1 \ldots \mathrm{Y}_n]$ would not be a construction at all) is the attempt to compose constructions not constructing entities of proper types, like $\lambda w \lambda t\,[{}^{\circ}\text{Student}_{wt}\ {}^{\circ}5]$.

[28] Sets are mapped by their characteristic functions, i.e., an α-set is an object of type $(o\alpha)$.

[29] We strongly adhere to Compositionality. Hence composing a "non-reasonable", i.e. improper, construction with any other (proper or improper) construction cannot result in a proper construction, because there is a gap in an argument. Anyway, we will return to this item later on.

This is in a good accordance with our linguistic intuitions, because the sentence '5 is a student' can hardly have any meaning.

Closure can never be improper, it always constructs a function, even it were a function not defined at any of its arguments, like for instance the function of dividing by zero, constructed by $\lambda x\,[{}^{\circ}\!:\, x\ {}^{\circ}0]$.

(Empirical) sentences of natural language denote propositions, i.e., intensions of type $o_{\tau\omega}$. They are also partial functions, in some w, t pairs these functions do not have to have any truth value. Sentences can be sometimes neither true nor false.[30] How can it happen that a sentence is neither true nor false in some states-of-affairs, and the denoted proposition is not defined in these w, t pairs?[31] Some sentences have presuppositions. For instance a simple sentence

Charles stopped smoking. (SP)

cannot be true in a situation when Charles never smoked. Nobody can stop doing something that he/she has never done. It might seem that the sentence SP is false in such a situation. But then the negated sentence

Charles did not stop smoking. (SN)

would be true, which means that he still does smoke[32], and that is not plausible either. Hence, if Charles never smoked, then both SP and SN are neither true nor false[33]. This is due to the fact that these sentences have a *presupposition* that Charles smoked.

We can define: a proposition Q is a *presupposition* of a proposition P iff Q's being true (in a world/time pair w, t) is a necessary condition for P to have any truth-value (in w, t). In other words, both P and negated P entail Q.

To introduce the notion of a presupposition rather more formally, we define three properties of propositions, objects of type $(o\, o_{\tau\omega})_{\tau\omega}$, $\mathrm{True}^{\mathrm{P}}$, $\mathrm{False}^{\mathrm{P}}$, and $\mathrm{Und}^{\mathrm{P}}$, as follows: let P be a proposition $(/\ o_{\tau\omega})$. Then

[30] Note that we do not introduce the third 'truth-value', like Muskens' N (neither). Our logic is a two-valued logic.

[31] We mean *objectively* not defined, so to say 'not defined for anybody', not only undefined for Mary not seeing that.

[32] See [19].

[33] Note that 'not true' does not mean the same as 'false'. A sentence can be 'not true' without being false.

- $[{}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}]$ returns value True iff ${}^{\circ}\mathrm{P}_{wt}$, otherwise False.
- $[{}^{\circ}\mathrm{False}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}]$ returns value True iff $[\neg\ {}^{\circ}\mathrm{P}_{wt}]$, otherwise False.
- $[{}^{\circ}\mathrm{Und}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}]$ returns True iff $[[\neg\ {}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}] \wedge [\neg\ {}^{\circ}\mathrm{False}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}]]$, otherwise False.

Now Q is a pressuposition of P if the following conditions hold:

$$\forall w \forall t\, [[{}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}] \supset [{}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{Q}]],$$
$$\forall w \forall t\, [[{}^{\circ}\mathrm{False}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}] \supset [{}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{Q}]].$$

In other words, Q is logically implied by both propositions that P and that not P.[34] Hence, necessarily, if Q is not true, then P is not true and P is not false:

$$\forall w \forall t\, [[\neg[{}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{Q}]] \supset [{}^{\circ}\mathrm{Und}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}]].$$

Note that, e.g., $[\neg[{}^{\circ}\mathrm{True}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{Q}]]$ is not equivalent to $[{}^{\circ}\mathrm{False}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{Q}]$, though our system is a two-valued logic. We do not work with a third truth--value: if $[{}^{\circ}\mathrm{Und}^{\mathrm{P}}_{wt}\ {}^{\circ}\mathrm{P}]$ v-constructs true, then ${}^{\circ}\mathrm{P}_{wt}$ is v-improper, the proposition P does not have *any* truth-value (at w, t).

A very frequent case is an *existential presuppostion* of a sentence. Whenever a sentence talks about an α-office 'the F' using the definite description 'the F' *de re*, the F has to exist so that the sentence had any truth-value. Examples:

The F is a G.
The F is believed by Charles to be a G.
The F is necessarily a G.

Logic should help to find the objective structures underlying expressions of a language, and it is now clear how 'value gaps' can be accommodated via improper constructions and partial functions, and it's also very clear why we must accept improperness and partiality: when

[34] We use the classical relation of entailment preserving truth 'from left to right' (from antecedent to consequent). The so-called 'strong entailment' introduced by Muskens, relation that not only preserves truth 'from left to right' but also preserves falsity 'from right to left' would not work here. Partiality is being propagated up, once a sentence does not have any truth-value, negating it or composing with any other sentence cannot result in a true or false statement.

modelling entities the empirical expressions talk about by intensions, functions from possible worlds, these functions have to be partial, for there are 'non-existing' entities we talk about. TIL handling partiality is determined by the above principles that turn on the same conception of language. A piece of language serves to point to a logical construction (concept) beyond itself, its sense. In order to reflect 'holes in reality' quite faithfully, to obtain a counterpart of Bolzano's *Gegenstandslosigkeit*, *TIL adopts partial functions and improper constructions.*

There have to be few restrictions on constructions, so that also those are included that construct partial functions. And there have to be even those constructions that do not construct anything, because such improper constructions are empty concepts expressed by some mathematical expressions. Our semantics runs smoothly even with partial functions and improper constructions, unlike more primitive theories. We opt for partiality for philosophical (rather than technical) reasons. Why then almost all the classical theories avoid handling partiality?

In first order and higher-order classical logic there are two kinds of expressions: terms and formulas. Terms are used to denote values in a mathematical structure or model, and formulas are used to make assertions about these values concerning their properties and relations to other values. The semantics of classical logic employs the *existence commitment:* terms always have a denotation. In other words, functional symbols are interpreted as total functions so that combining these symbols with argument symbols always denotes the value of the function at its argument. If the function were partial, then the term would not be well defined and thus would not have a natural denotation.

The reason of this state may consist in the fact that working with partial functions is rather incommodious. Many commonly accepted classical laws do not hold in case of logic involving partial functions: for instance, $(A \vee \neg A)$ is generally not a tautology and $(A \wedge \neg A)$ is not a contradiction (they may fail to have any truth value); de Morgan laws of negation for quantified formulas, the equivalence of β-transformation (λ-conversion), the reduction of n-ary functions to unary ones[35], the equivalence of formulas entailing each other, the equivalence of a relational and a functional view, etc. Thus it is easier to simply claim that 'the greatest prime' is a sense-less term, and violate thus the principle of

[35] See [21].

compositionality (how then can this 'sense-less term' be composed into a reasonable true proposition that the greatest prime does not exist?), than to solve the technical difficulties connected with partial functions. In our opinion, the task of a logician is not to "violate reality", but to adequately model semantic features of a language even at the price of some technical difficulties. We are not also going into the way of 'playing with logical symbols', i.e. defining new *ad hoc* connectives and entailment relations, so that particular commonly accepted laws would hold. In the remaining part of the paper we briefly examine these technical peculiarities and the way TIL handles them.

Non-denoting terms in quantification logic

The problem of how to deal with non-denoting terms in quantification logic has been extensively studied by logicians since the appearance of Russell's 'On denoting' paper [16]. Interest in non-denoting definite descriptions has lead to the formulation of so-called *free logics* in which terms of the logic are *free from existential assumptions*.[36] It might seem that the perspective of a mathematician who would like to reason about partial functions is different from that of the (philosophical) logician who is interested in non-denoting definite descriptions. Mathematical concepts may be strictly empty, like, e.g., the greatest prime that does not identify any object, whereas empirical definite descriptions always denote an intension, an office. Whereas a mathematician is not much interested in the *meaning* of, say, 5 : 0, or the greatest prime, he is content with the proof of their non-existence, the logician does worry what 'the present King of France', 'the president of CR (in February)' *mean*, why we can truly predict not only non-existence of them, but also some necessary requisites, like e.g., 'being a ruler of France', 'being the highest representative of CR', respectively. A computer scientist may have a third perspective on non-denoting terms. For a computer scientist an expression like 5 : 0 usually denotes a special value 'undefined: $\perp$', or an 'error' value, which is treated very much the same as an ordinary value. In particular, functions are strict with respect to 'error' values that are "propagated up" the functions. If a takes an 'error' value, then $f(a)$ returns an 'error' value as well. However, some computationally very useful functions are

[36] See [11].

not strict in this sense. The extensively used 'if then else' conditional function is not strict. Using classical logic, such a function f might be defined as follows: $[\phi(x) \supset f(x,y,z) = y] \wedge [\neg\phi(x) \supset f(x,y,z) = z]$.

If $\phi(x)$ is true, an efficient computer evaluating f returns value of y even if z takes an 'error' value, because it is never necessary to evaluate both y and z when computing function f. In this sense the function 'if then else' is not strict, the 'error' value is not strictly propagated up.

The different perspectives of a logician, mathematician and computer scientist toward non-denoting expressions naturally lead to different strategies of handling partial functions by particular logical systems. Axiomatic theories that are based on *generally quantified* formulas (sentences–axioms) usually develop the semantics that is derived from a mixed partial/total valuation approach with non-denoting terms and total valuation of formulas. Such an approach requires few changes to the framework of classical logic and is very closed to handling non-denoting terms in mathematical practice. William Farmer presents in his [7] a PF system that takes exactly this mathematician's perspective toward non-denoting expressions. In this paper Farmer also provides a nice survey of the major approaches to partial functions in quantificational formalistic logic, namely from a mathematician's perspective. The exhaustive study on all the approaches to the problem would be out of the scope of present paper. Since we will show that TIL handling partial functions does not deviate from mathematician's perspective as much as it might seem, we recapitulate briefly:

(a) Non-denoting expressions as non-well-formed terms This is one of the simplest ways of handling the problem[37], but such an approach has some serious shortcomings. First, whether or not a term is well formed would depend on the context in which it is interpreted. Second, in such a system there could be no recursive way of deciding whether or not a given term is well formed in the given context. Last, non-well-formed terms cannot be manipulated and reasoned about, though in standard mathematics we need to employ them *before* they are known to be non-denoting.

(b) Functions represented as relations In principle, an n-ary function, total or partial, can be represented as an $(n+1)$-ary relation. Rea-

[37]For instance constructivist logic (see [12]) takes this approach. A proof of the existence is thus a part of a constructivist construction.

soning about functions using relations is neither natural nor efficient, since functional application must be represented in an unnatural indirect fashion. This is just a pragmatic shortcoming. Which is more serious, such an approach does not allow for a formula without any truth-value, and thus does not provide an adequate semantics for natural language propositions. Russell's elimination of definite operators as semantically self-contained expressions is an example of its inadequacy.[38]

(c) Total functions with unspecified values This is essentially the approach a classical logic deals with definite descriptions[39], because no change has to be made to classical logic: all the functions are *formally* total. However, such an approach is completely inadequate in natural language semantics. Models of such theories are unnatural and do not "respect reality".

(d) Many-sorted logic A partial function can 'technically' be made total when reducing its domain. This is the approach often adopted in mathematics. We simply claim that division is defined on the '*sort*' $\mathcal{R} \times \mathcal{R}^-$, where $\mathcal{R}^-$ is the set of rationals $\mathcal{R} - \{0\}$. This simple idea works, however, only for *some* partial functions. For example, the binary function f over the reals defined by $f(x, y) = \sqrt{x - y}$, cannot be represented as a binary total function in many-sorted logic. We might try to do so using the function g_a defined by $g_a = f(x, a)$. This time a new sort would have to be used for each real number a. Unless there is a *recursive* way of forming *new sorts*, the approach leads to a proliferation of sorts and thus it is not usable in semantics of definite descriptions, for in case of propositions there is not a recursive way of forming (subsets of) the type ω (possible worlds).

(e) Error values This approach has been discussed above from the computer-scientist's perspective. It can be very useful in computer science when *evaluating* partial functions and non-strict functions, because quantification is over ordinary values as well as 'error' values. If particular rules handling functions with 'error values' are correctly defined, this approach can be conceived as an adequate modelling tool of 'partiality'. However, we have to investigate functions with value-gaps first, to be able to state such correct rules.

[38] See [16], [19].

[39] See [1].

(f) Non-existent values The above approach of 'error' values has a serious defect from a mathematician point of view: the fact that quantification is over all values, including error values. A way to escape this drawback is a simple modification of 'error' value approach: quantification only over existent values, non-existent values have an inferior status. On the other hand, free variables range over both kinds of values. This approach has been motivated by intuitionistic logic (see [18], [12]), where a non-existent value might be interpreted as something like "it is not known, but there may be a value". Since quantification is only over existent values, only few of the rules of classical logic must be modified. For example the substitution rule for generally quantified formulas has to be enriched by the "existence E guard": $\forall x\, \phi(x) \wedge \mathrm{E}(t) \vdash \phi(t)$. This approach is far superior for a mathematician. However, it violates some good mathematical practice, and for the same reasons is not suitable for logic handling (natural) language semantics. First, we would like to quantify over all the values. Second, bound and free variables do not have the same status as for their range, which is untenable. We deal with cases where there simply is no value, and no variable can range over 'no-values'.

(g) Partial valuation for terms but total for formulas This approach adds the following two rules to the standard evaluation:

I. A term denotes a value only if all its subterms denote values.

II. An atomic formula is *false* if any term occurring in it is non-denoting.

This approach is not plausible in natural language semantics: we have seen that there are sentences that cannot have any truth-value.

(h) Partial valuation for terms and formulas This very simple approach adds the following two rules to the standard rules of valuation:

I. A term denotes a value *only if all* its subterms denote values.

II. A formula denotes a truth-value *only if all* the terms occurring in it denote values.

In principal, this is the approach adopted by P. Tichý in his [21]. In this study we are going to analyse this approach in more details. We adhere to Compositionality, and we are not going to accept any *ad hoc* rules without a deep philosophical / logical justification. For the sake of convenience we will, rather non-precisely, use the notion of 'non-denoting term' for a v-improper composition (like $[[^{\circ}\text{Pegasus } w]\ t]$, or shortly $^{\circ}\text{Pegasus}_{wt}$), and denote the absence of a truth-value by $\perp$.

TIL approach to 'non-denoting terms'

We have seen that due to the principle of Compositionality the construction of composition is defined in such a way that partiality is being 'propagated up'. Once a construction C is v-improper, composing it with any other construction D cannot arise in a proper construction, for C does not supply an argument for D. This seems to be plausible. Since propositional connectives $\neg$, $\vee$, $\wedge$, $\supset$ and $\equiv$ are functions of types (oo) or (ooo), the propositional part of our logic behaves classically. Particular functions are defined as shown by Table 1 (it is not necessary to define $\neg$).

p	q	$p \wedge q$	$p \vee q$	$p \supset q$	$p \equiv q$	
1	1	1	1	1	1	
1	0	0	1	0	0	
0	1	0	1	0	0	
0	0	0	0	1	1	
1	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$*$
0	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$*$
$\perp$	1	$\perp$	$\perp$	$\perp$	$\perp$	$*$
$\perp$	0	$\perp$	$\perp$	$\perp$	$\perp$	$*$
$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	

Table 1: Propositional connectives

Rows marked by $*$ might seem to be peculiar. Aren't we used to take disjunction true iff at least one disjunct is true? And aren't we used to take implication true iff the antecedent is false?

Imagine a situation when Charles does not smoke. It seems that in such a situation we may truly claim the following:

Charles stopped smoking or he never smoked. (SPD)

But *alas*, when analysing SPD in TIL using disjunction $\vee$ (according to Table 1) we obtain construction of a proposition not defined at such w, t pairs when Charles never smoked! (StopSmoking / $(o\iota)_{\tau\omega}$), NeverSmoked / $(o\iota)_{\tau\omega}$)

$$\lambda w\lambda t\,[[\lambda w\lambda t\,[^{\circ}\mathrm{SS}_{wt}\ ^{\circ}\mathrm{Ch}]]_{wt} \vee [\lambda w\lambda t\,[^{\circ}\mathrm{NS}_{wt}{}^{\circ}\mathrm{Ch}]]_{wt}] \qquad (\mathrm{SPD}')$$

The problem is created by the proposition $\lambda w\lambda t\,[^{\circ}\mathrm{SS}_{wt}\ ^{\circ}\mathrm{Ch}]$ that does not have any truth value in the situation which is described above, hence $[\lambda w\lambda t\,[^{\circ}\mathrm{SS}_{wt}\ ^{\circ}\mathrm{Ch}]]_{wt}$ is v-improper. There are, of course, other possibilities to analyse the sentence SPD. For instance, as meaning 'the only proposition that is true and that is equal either to the proposition that Charles stopped smoking or that he never smoked'. Isn't it rather a wicked reformulation of the simple sentence SPD? The other possibility would be to analyse potentially neither true nor false sentences using the 'totalising' property of propositions $\mathrm{True}^{\mathrm{p}}$ / $(o\,o_{\tau\omega})_{\tau\omega}$, which returns true (at w, t) if the proposition is true (at w, t) *otherwise* false. The respective paraphrasing of SPD is *either it is true that Charles stopped smoking or he never smoked*, and the respective analysis[40]:

$$\lambda w\lambda t\,[[^{\circ}\mathrm{True}^{\mathrm{p}}_{wt}\ \lambda w\lambda t\,[^{\circ}\mathrm{SS}_{wt}\ ^{\circ}\mathrm{Ch}]] \vee [\lambda w\lambda t\,[^{\circ}\mathrm{NS}_{wt}\ ^{\circ}\mathrm{Ch}]_{wt}]]. \qquad (\mathrm{SPD}'')$$

Still other possibility would be to define the meaning of 'stopped smoking' as 'previously smoked and now does not smoke'; then the problem of the presupposition disappears (negated sentence would claim that he did not smoke or still smokes).

Anyway, to avoid the above peculiarities, we might define the connectives as follows: (But, if we accepted that, we would have to stand an unpleasant consequence: *We would have to give up the principle of Compositionality at the very bottom of our logic.* Don't forget that the sign '$\perp$' does not denote any 'non-existing truth-value' but the *absence* of truth-value argument.)

Conjunction $\wedge^2$:	`True` iff both the conjuncts are true
	`False` iff at least one of the conjunts is false
	$\perp$ otherwise
Disjunction $\vee^2$:	`True` iff at least one of the disjuncts is true

[40]Note that the incriminated proposition $\lambda w\lambda t\,[^{\circ}\mathrm{SS}_{wt}\ ^{\circ}\mathrm{Ch}]$ is mentioned now; its construction occurs *de dicto*.

	`False` iff both the disjuncts are false
	$\perp$ otherwise
Implication $\supset^2$:	`True` iff the consequent is true or the antecedent is false
	`False` iff the antecedent is true and the consequent is false
	$\perp$ otherwise

The resulting truth-value function would be defined by Table 2.

p	q	$p \wedge^2 q$	$p \vee^2 q$	$p \supset^2 q$	$p \equiv^2 q$	
1	1	1	1	1	1	
1	0	0	1	0	0	
0	1	0	1	0	0	
0	0	0	0	1	1	
1	$\perp$	$\perp$	1	$\perp$	$\perp$	*
0	$\perp$	0	$\perp$	1	$\perp$	*
$\perp$	1	$\perp$	1	1	$\perp$	*
$\perp$	0	0	$\perp$	$\perp$	$\perp$	*
$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	

Table 2: Modified propositional connectives

Quantifiers $\forall^\alpha$, $\exists^\alpha$ are also functions of types $(o(o\alpha))$, they are sets of sets. In type theories, quantifiers are defined as follows ($x \to \alpha$, A a construction $\to o$):

- $[^o\forall^\alpha\ \lambda x\,\mathrm{A}]$ constructs `True` iff the set constructed by $\lambda x\,\mathrm{A}$ is identical to the whole type α, otherwise `False`.
- $[^o\exists^\alpha\ \lambda x\,\mathrm{A}]$ constructs `True` iff the set constructed by $\lambda x\,\mathrm{A}$ is a nonempty subset of the type α, otherwise `False`.

Now it might seem that there is no problem of partiality, because quantifiers are composed with the construction of a *set* ($\lambda x\,\mathrm{A}$ can never be improper). But the construction A can itself be v-improper. Hence quantifiers are, so to say, also 'totalising'. An unpleasant consequence we have to pay attention to is the non-validity of de Morgan laws. To illustrate it, consider a simple sentence

All the (current and previous) Presidents of CR could speak Czech.

Providing there was no president of CR who was not able to speak Czech, the sentence seems to be perfectly true, because it is not true that some presidents of CR could not speak Czech. But, again, using quantifiers defined above, we obtain the analysis constructing a *false* proposition, PC / $\iota_{\tau\omega}$—President of CR, C / $(o\iota)_{\tau\omega}$—can speak Czech:

$$\lambda w \lambda t\, [{}^{\circ}\forall^{\tau}\, \lambda t'\, [[t' \leq t] \supset [{}^{\circ}\mathrm{C}_{wt'}\ {}^{\circ}\mathrm{PC}_{wt'}]]]$$

The problem is created by the fact that at such w, t' pairs when there was no president of CR the composition ${}^{\circ}\mathrm{PC}_{wt'}$ is v-improper, and by Compositionality the whole composition $[[t' \leq t] \supset [{}^{\circ}\mathrm{C}_{wt'}\ {}^{\circ}\mathrm{PC}_{wt'}]]$ is v-improper. Thus the set constructed by $\lambda t'\, [[t' \leq t] \supset [{}^{\circ}\mathrm{C}_{wt'}\ {}^{\circ}\mathrm{PC}_{wt'}]]$ is not the whole type τ (there are 'gaps'), and the ${}^{\circ}\forall^{\tau}$ returns false. Moreover, de Morgan law is not valid, because the set v-constructed by $\lambda t'\, [[t' \leq t] \wedge \neg[{}^{\circ}\mathrm{C}_{wt'}\ {}^{\circ}\mathrm{PC}_{wt'}]]$ is not nonempty (the characteristic functions is either false or undefined), and thus the composition

$$[{}^{\circ}\exists^{\tau}\ \lambda t'\, [[t' \leq t] \wedge \neg[{}^{\circ}\mathrm{C}_{wt'}\ {}^{\circ}\mathrm{PC}_{wt'}]]]$$

v-constructs also false, and the proposition constructed by the following construction is false.

$$\lambda w \lambda t\, [{}^{\circ}\exists^{\tau}\ \lambda t'\, [[t' \leq t] \wedge \neg[{}^{\circ}\mathrm{C}_{wt'}\ {}^{\circ}\mathrm{PC}_{wt'}]]].$$

It might seem, that a possible recovery would consist in using other kind of ('non-totalising') quantifiers, which would be a generalisation of the connectives $\wedge^2, \vee^2$. Then $\forall^2$ / $(o(o\alpha))$ is a function that returns

`False` iff the characteristic function of the set (/ $(o\alpha)$) returns value `False` at at least one of its arguments (/α),

`True` iff the characteristic function of the set (/ $(o\alpha)$) returns value `True` at all its arguments (/α),

$\perp$ otherwise,

while $\exists^2$ / $(o(o\alpha))$ is a function that returns

`False` iff the characteristic function of the set (/ $(o\alpha)$) returns value `False` at all its arguments (/α),

`True` iff the characteristic function of the set (/ $(o\alpha)$) returns value `True` at at least one of its arguments (/α),

$\perp$ otherwise.

Using these quantifiers might be more natural, because de Morgan laws would be valid. Anyway, the analysis of the above sentence (and of its negation) would construct the proposition that is undefined (in current w, t), which is again rather strange: negation of our sentence, namely

Some presidents of CR could not speak Czech.

is surely false in the described situation (in actual w, t). Similarly, a 'classical' analysis of the claim

There is no greatest natural number,

namely $\neg[{}^{\circ}\exists^{\tau}\lambda x\,[x = [{}^{\circ}\iota^{\tau}\lambda x\,[[{}^{\circ}\mathrm{Nat}\,x] \wedge \forall y\,[[{}^{\circ}\mathrm{Nat}\,y] \supset (x \geq y)]]]]]$, should construct true (as it does using quantifier of the former kind—${}^{\circ}\exists^{\tau}$ returns false here). Quantifiers should be 'totalising'.

There are two recoveries. We might reformulate the sentence as talking only about existing presidents:

Whenever the president of CR existed then whoever occupied the office could speak Czech,

which is certainly not synonymous to the original one, but this rephrasing could be taken as explicating the original sentence (E / $(o\iota_{\tau\omega})_{\tau\omega}$)—the property of an office 'to be occupied'):

$$\lambda w\lambda t\,[{}^{\circ}\forall^{\tau}\lambda t'\,[[t' \leq t] \supset [[{}^{\circ}\mathrm{E}_{wt'}\,{}^{\circ}\mathrm{PC}] \supset \\ \supset [{}^{\circ}\forall^{\iota}\lambda x\,[[x = {}^{\circ}\mathrm{PC}_{wt'}] \supset [{}^{\circ}\mathrm{C}_{wt'}\,x]]]]]],$$

which constructs a true proposition in our case.

Other two possibilities using the properties of propositions $\mathrm{True}^{\mathrm{p}}$, $\mathrm{Und}^{\mathrm{p}}$ are:

$$\lambda w\lambda t\,[{}^{\circ}\forall^{\tau}\lambda t'\,[[t' \leq t] \supset [{}^{\circ}\mathrm{True}^{\mathrm{p}}_{wt'}\lambda w\lambda t\,[{}^{\circ}\mathrm{C}_{wt}\,{}^{\circ}\mathrm{PC}_{wt}]] \vee \\ \vee [{}^{\circ}\mathrm{Und^{p}}_{wt'}\lambda w\lambda t\,[{}^{\circ}\mathrm{C}_{wt}\,{}^{\circ}\mathrm{PC}_{wt}]]]],$$

$$\lambda w\lambda t\,[{}^{\circ}\forall^{\tau}\lambda t'\,[[t' \leq t] \supset [[{}^{\circ}\mathrm{E}_{wt'}\,{}^{\circ}\mathrm{PC}] \supset [{}^{\circ}\mathrm{True}^{\mathrm{p}}_{wt'}\lambda w\lambda t\,[{}^{\circ}\mathrm{C}_{wt}\,{}^{\circ}\mathrm{PC}_{wt}]]]]].$$

(Read: previously and/or now, providing the president of CR existed then it was true that he could speak Czech.) Now the proposition that the president of CR could speak Czech is used *de dicto*, and no partiality problem arises.

A good solution is usually obtained by using more general quantifiers `All` and `Some`, both of type $((o(o\alpha))\,(o\alpha))$ defined as follows: `All` is the function that associates a set S of individuals with all those sets that contain S as its subset; while `Some` is the function that associates a set S with all those sets that have a non-empty intersection with S. These quantifiers are suitable for the analysis of 'subject-predicate' sentences of the form "All As are Bs", "Some As are Bs", (A / $(o\alpha)_{\tau\omega}$, B / $(o\alpha)_{\tau\omega}$):

$$\lambda w \lambda t\,[[{}^{\circ}\texttt{All}\,{}^{\circ}\mathrm{A}_{wt}]\,{}^{\circ}\mathrm{B}_{wt}], \qquad \lambda w \lambda t\,[[{}^{\circ}\texttt{Some}\,{}^{\circ}\mathrm{A}_{wt}]\,{}^{\circ}\mathrm{B}_{wt}].$$

(Of course, there is still a problem of correct defining an intersection, subset, etc., for "partial sets", i.e. sets with a characteristic *partial* function, solution of which is straightforward. However, defining a complement is not that simple.) The analysis of the original sentence would now v-construct true, `All` of the type $((o(o\iota))(o\iota))$:

$$\lambda w \lambda t\,{}^{\circ}\forall \lambda t'\,[[t' \leq t] \supset [[{}^{\circ}\texttt{All}\,\lambda x\,[x = {}^{\circ}\mathrm{PC}_{wt'}]]\,{}^{\circ}\mathrm{C}_{wt'}]].$$

(Read: previously or now, all who occupied the office of the president of CR could speak Czech.) However, 'iterated quantifiers' like 'for all x there are some $y\ldots$' cannot be analysed in this way, so that the problem of quantifiers is still in a way open.

Entailment, Equivalence and Synonymy

The correct, adequate analysis is a necessary condition for inferring all the logical consequences of our statements. Now we can easily infer that, e.g., if Václav Havel used to be the president of CR then he can speak Czech. Logical consequence is defined classically even in the case of partial functions as the truth preserving relation between (sets of) constructions and constructions:

A set $\{\mathrm{C}_1, \ldots, \mathrm{C}_n\}$ of constructions (constructing propositions) *entails* a construction C, denoted $\mathrm{C}_1, \ldots, \mathrm{C}_n \models \mathrm{C}$, if in all the w, t pairs where all the $\mathrm{C}_1, \ldots, \mathrm{C}_n$ construct true propositions the construction C constructs a proposition true at w, t as well. Formally,

$$\frac{[{}^{\circ}\texttt{True}^{\mathrm{p}}_{wt}\mathrm{C}_1], \ldots, [{}^{\circ}\texttt{True}^{\mathrm{p}}_{wt}\mathrm{C}_n]}{[{}^{\circ}\texttt{True}^{\mathrm{p}}_{wt}\mathrm{C}].}$$

Note that it does not hold that if $C_1 \models C_2$ and $C_2 \models C_1$ then C_1 and C_2 are equivalent. Propositions constructed are true in exactly the same world/time pairs, but can be false or undefined in distinct pairs. To define an equivalence of constructions in terms of entailment, we would need a stronger relation, preserving not only truth from left to right, but also falsity from right to left[41] (from consequent to antecedent). But the equivalence between constructions can be easily and more generally defined as follows:

Constructions C_1, C_2 are *equivalent* iff they construct one and the same entity. This definition can be generalized for expressions. Expressions E_1 and E_2 are *equivalent* (or co-denoting) iff constructions C_1, C_2 expressed by E_1, E_2, respectively, are equivalent. If, moreover, constructions (concepts) C_1, C_2 expressed by E_1, E_2, respectively, are identical[42], then E_1 and E_2 are *synonymous*. Now it is clear that in our logic sentences like (5), (6), (7) will not be analysed as synonymous, but only equivalent. The underlying constructions of the (one and the same) denoted proposition are distinct.

A guarded plea for the logic of partial functions

Up to now we have dealt with the technical peculiarities of the logic of partial functions and improper constructions (non-denoting terms). The question arises again: *Do we have to deal with partial functions?* We have advocated for logic of partial functions, because of the need to semantically model sentences with existential presupposition. In what follows we give some other sound logical arguments in favour of dealing with partial functions. We will show that *ad hoc* rules claiming that β-reduction is not allowed in cases of *de re* modalities, temporalities and attitudes can be precisely justified using our logic accommodating non-denoting terms due to explicit intensionalization of TIL (involving variables w, t ranging over ω, τ).

[41] See [15].

[42] It can be proved that relatively to a given conceptual system there is *the* (only one most adequate analysis of an expression) concept expressed by the expression. For details, see [6].

TIL makes it possible (unlike more primitive theories) to distinguish between the so-called *using* and *mentioning* constructions and functions. Constructions are primarily *used* to obtain the entity that is constructed. Functions are used to obtain the value assigned at an argument. But we often need to *mention* constructions as well; in such a case we use a higher-order construction of the mentioned construction. As an example, consider the sentence

> *Dividing x by zero is improper (i.e. does not yield any result for any x).*

The respective analysis (Improper / $(o*_1)$—the class of improper constructions of order 1):

$$[{}^{\circ}\mathrm{Improper}\ {}^{\circ}[{}^{\circ}\!:\, x\ {}^{\circ}0]]$$

Note that when mentioning constructions, their improperness does not result in a "non-reasonable" sentence without a truth-value. Such sentences can be easily true or false, because a proper (higher-order) construction of the mentioned improper one is used. Constructions are mentioned in the so-called *propositional attitudes*, because our knowledge, beliefs, etc., are primarily connected with concepts, i.e. constructions. Thus, e.g., Charles can easily believe that the King of France is bald. Using for the sake of simplicity the simple concept ${}^{\circ}$KF to construct the office KF of the King of France, a $\iota_{\tau\omega}$-object, K(ing) / $(o\iota)_{\tau\omega}$, our analysis is (Bel / $(o\iota*_1)_{\tau\omega}$):

$$\lambda w \lambda t\, [{}^{\circ}\mathrm{Bel}_{wt}\ {}^{\circ}\mathrm{Ch}\ {}^{\circ}[\lambda w \lambda t\, [{}^{\circ}\mathrm{Bald}_{wt}\ {}^{\circ}\mathrm{KF}_{wt}]]] \qquad (\mathrm{KF}\, \textit{dicto})$$

Or, Charles can believe that dividing five by zero gives zero:

$$\lambda w \lambda t\, [{}^{\circ}\mathrm{Bel}_{wt}\ {}^{\circ}\mathrm{Ch}\ {}^{\circ}[{}^{\circ}\!=[{}^{\circ}\!:\, {}^{\circ}5\ {}^{\circ}0]\ {}^{\circ}0]]$$

The improperness of the non-denoting terms ${}^{\circ}\mathrm{KF}_{wt}$ and $[{}^{\circ}\!:\, {}^{\circ}5\ {}^{\circ}0]$ does not cause any problems here, the respective concepts are just mentioned.

The improperness of non-denoting terms acts an important part when an improper construction is used to obtain an entity when there is none. If the construction $[{}^{\circ}\!=[{}^{\circ}\!:\, {}^{\circ}5\ {}^{\circ}0]]$ were not mentioned (by trivialisation) the latter construction above would relate Charles to a "non-existing constructed" value, which is, of course, impossible, the proposition would not have any truth-value. On the other hand, if we analysed Charles' belief in King's baldness as a relation to the denoted proposition (implicit possibly unconscious attitude[43]), the resulting proposition might be perfectly true, because the embedded clause always denotes an intension (proposition in this case), and this proposition is *mentioned* (Bel′ / $(o\iota o_{\tau\omega})_{\tau\omega}$):

$$\lambda w\lambda t\,[^\circ\mathrm{Bel}'_{wt}\ {}^\circ\mathrm{Ch}\ \lambda w\lambda t\,[^\circ\mathrm{Bald}_{wt}\ {}^\circ\mathrm{KF}_{wt}]]. \qquad \text{(KFd)}$$

Now Charles is not related to a "non-existing value", but to the whole proposition, the improperness of the composition $^\circ\mathrm{KF}_{wt}$ does not play any role. Using medieval terminology, we say that construction $^\circ$KF (and the term 'the King of France') is used *de dicto*.[44] The respective function (intension, proposition) is mentioned, partiality (non-existing value) does not come into play.

Consider, on the other hand, the *de re* variant of the sentence:

Charles believes of the King of France that he is bald. (KF*re*)

In the usual notation of doxastic logics the distinction is characterised as the contrast between

$$\mathrm{Bel}_{\mathrm{Charles}}\mathrm{Bald}[\mathrm{kf}] \qquad \textit{(de dicto)}$$
$$\exists x(x = \mathrm{kf} \wedge \mathrm{Bel}_{\mathrm{Charles}}\mathrm{Bald}[x] \qquad \textit{(de re)}$$

But there are worrisome questions concerning this analysis. Where does the existential quantifier come from in the *de re* case? There is no trace of it in the original sentence. And why the axiom of substituting the identicals cannot be used here to reduce *de re* to *de dicto*?

Hintikka and Sandu propose in their [10] a remedy by means of the Independence Friendly (IF) first order logic:

> Independence Friendly (IF) first-order logic deals with a frequent and important feature of natural language semantics. Without the notion of *independence*, we cannot fully understand the logic of such concepts as belief, knowledge, questions and answers, or the *de dicto* vs. *de re* contrast.

[43] See [5].

[44] For details on the contrast between *de dicto / de re*, see [4].

They solve the *de dicto* case as above, and propose the *de re* solution with the independence indicator '/ ':

$$\mathrm{B_{Charles}Bald[kf/B_{Charles}]}.$$

This is certainly a more plausible analysis, closer to the syntactic form of the original sentence, and the independence indicator indicates the essence of the matter; there are two *independent* questions: "Who is the King" and "What does Charles think of that person". Of course, Charles has to have a relation of an "epistemic intimity" [2] to a certain individual, but he does not have to connect this person with the office of the King (only the ascriber must do so). Still, the semantics of "/ $\mathrm{B_{Charles}}$" is not pellucid, and we will show that the informational independence can be precisely captured by means of TIL explicit intensionalization without using any new non-standard operators.

In Montague's like systems, the *de re* analysis might be presented in the form:

$$\lambda x\, \mathrm{Bel(Charles, Bald}(x))(\mathrm{kf}).$$

And there is an ad hoc rule: though β-reduction is an equivalent transformation, it must not be used within the scope of the "attitude operator".[45] And we have to ask, *why not*? To give a sound argument for this law, we have to logically justify it. Before doing that, we will prove an important claim[46] ($\rightarrow$ means constructs, ranges over):

Statement 1 *Performing β-reduction on the construction having the form $[\lambda x \lambda y\, [\mathrm{F}\ x\ y]\ \mathrm{C}]$, where $x \rightarrow \alpha$, $y \rightarrow \beta$, $\mathrm{F} \rightarrow (\gamma\alpha\beta)$, $\mathrm{C} \rightarrow \alpha$, is not an equivalent transformation in case of C being v-improper (even if F constructs a total function).*

Proof According to the definition of Composition, the construction $[\lambda x \lambda y\, [\mathrm{F}\ x\ y]\ \mathrm{C}]$ does not construct anything, it is v-improper in case of C being v-improper. Performing β-reduction, we obtain: $[\lambda y\, [\mathrm{F\ C}\ y]]$. But this closure can never be v-improper. In case of C being v-improper, the closure $\lambda y\, [\mathrm{F\ C}\ y]$ does construct an entity, namely a 'degenerated function', i.e. the function that is undefined at any argument y. ∎

[45] Worse even, Montague's IL does *not* have the Church-Rosser property, though an ordinary type logic (incl. TIL) has this property. This IL failure is the consequence of the lack of explicit intensionalization.

[46] Tichý in his [21] proved the equivalence of β-reduction on the assumption of the F not being improper.

Example Construction $[\lambda x \lambda y\, [{}^{\circ}\!> y\, x]\, {}^{\circ}0]$ is equivalent to $\lambda y\, [{}^{\circ}\!> y\, {}^{\circ}0]$. They both construct the class of positive numbers. But construction $[\lambda x \lambda y\, [{}^{\circ}\!> y\, x]\, [{}^{\circ}\!:{}^{\circ}5\, {}^{\circ}0]]$ is improper, whereas $\lambda y\, [{}^{\circ}\!> y [{}^{\circ}\!: {}^{\circ}5\, {}^{\circ}0]]$ constructs a degenerated class.

Construction $[\lambda x \lambda y\, [{}^{\circ}\!+ y\, x]\, {}^{\circ}1]$ is equivalent to $\lambda y\, [{}^{\circ}\!+ y\, {}^{\circ}1]$. They both construct the function 'adding 1'. But $[\lambda x \lambda y\, [{}^{\circ}\!+ y\, x]\, [{}^{\circ} : {}^{\circ}5\, {}^{\circ}0]]$ is improper, whereas $\lambda y\, [{}^{\circ}\!+ y\, [{}^{\circ} : {}^{\circ}5\, {}^{\circ}0]]$ constructs a degenerated function.

Note: an empty α-set and a degenerated α-set are not identical. Whereas the former is a subset of every α-set, the latter is *not* a subset of any α-set.

Analysing the *de re* attitude in TIL, we obtain:

$$\lambda w \lambda t\, [\lambda x\, [{}^{\circ}\mathrm{Bel}_{wt}\, {}^{\circ}\mathrm{Ch}\, [\lambda w \lambda t\, [{}^{\circ}\mathrm{Bald}_{wt}\, x]]]\, {}^{\circ}\mathrm{KF}_{wt}] \qquad \text{(KFr)}$$

(Read: the King of France is believed by Charles to be bald.)

To prevent a collision of variables, i.e. to correctly perform β-reduction, we have to "rename variables":

$$\lambda w \lambda t\, [\lambda x\, [{}^{\circ}\mathrm{Bel}_{wt}\, {}^{\circ}\mathrm{Ch}[\lambda w^* \lambda t^*\, [{}^{\circ}\mathrm{Bald}_{w^*t^*}\, x]]]\, {}^{\circ}\mathrm{KF}_{wt}] \qquad \text{(KFr}'\text{)}$$

and we get

$$\lambda w \lambda t\, [{}^{\circ}\mathrm{Bel}_{wt}\, {}^{\circ}\mathrm{Ch}[\lambda w^* \lambda t^*\, [{}^{\circ}\mathrm{Bald}_{w^*t^*}\, {}^{\circ}\mathrm{KF}_{wt}]]]. \qquad \text{(KFr}\beta\text{)}$$

Note that KFrβ is *not equivalent* to the *de dicto* case KFd, because ${}^{\circ}$KF still occurs *de re*; the ${}^{\circ}$KF has been subjected to the intensional descent with respect to the *reporter's perspective* w, t. Correctly performed β-reduction did not reduce the *de re* case to the *de dicto* one, which is correct: *de re* case and *de dicto* case are logically independent. But, KFrβ is also not equivalent to KFr', the resulting proposition P^{β} is distinct. Yes, it "behaves" in the same way in those w, t pairs where the original proposition P (constructed by KFr) is true. But in those w, t pairs where the original proposition P does not have any truth value (if the King of France does not exist, then he cannot be believed by anybody to be anything), the proposition P^{β} can be easily true or false, because Charles is now related to a 'degenerated proposition' that is undefined in all w^*, t^* pairs, which is not correct any more. The non-denoting term ${}^{\circ}\mathrm{KF}_{wt}$ is incorrectly drawn in *de dicto* context—the construction of a proposition, which cannot be improper. We can make another important claim:

Statement 2 *When a construction of an intension the F is used de re, there is an existential presupposition on the existence of the F; β-reduction that substitutes the F for the argument of a function and would thus abolish this presupposition is not an equivalent transformation, and therefore is not allowed.*

Now we examine another example of a non-equivalent β-reduction, namely the case of combining partial functions and 'totalising' quantifiers. Consider the sentence

The King of France might not have been a king of France. (FK)

A standard analysis using the modal operator $\Diamond$ ('possibly') might be, e.g., [Dummett 1981]: $(\lambda x\, \Diamond\neg \mathrm{K}x)(\iota y\, \mathrm{K}y)$ (the class of those who are possibly not a king contains the King).

Now there is a logical problem: Applying the β-rule of λ-calculi we get $\Diamond\neg \mathrm{K}(\iota y\, \mathrm{K}y)$ thus obtaining a contradiction, whereas FK is a meaningful sentence (that was true in some period before 1789). Dummett states an *ad hoc* principle that gets no support from standard logic: *when a 'modal expression' is applied to a definite description, the β-rule cannot be applied,* and he is not able to explain why.

There are four possible readings of the sentence FK, two *de dicto* and two *de re*, but only the *de re* readings are plausible (and are probably those intended).

(DD1) *It is possible that the King of France is not a king (of France).*

(DD2) *It is not necessary that the King of France is a king (of France).*

(DR1) *The King of France is possibly not a king (of France).*

(DR2) *The King of France is not necessarily a king (of France).*

Our analyses are:

$$\lambda w \lambda t\, {}^{\circ}\exists \lambda w^{*}\, {}^{\circ}\exists \lambda t^{*}\, [{}^{\circ}\neg\, [{}^{\circ}\mathrm{K}_{w^{*}t^{*}}\ {}^{\circ}\mathrm{KF}_{w^{*}t^{*}}]], \qquad (\mathrm{DD1}')$$

(°K, °KF *de dicto*)

constructs the impossible proposition, false in all w, t pairs (it is impossible that the King of France were not a king *"in the same world, time"*).

$$\lambda w \lambda t\, [{}^{\circ}\neg\, {}^{\circ}\forall \lambda w^{*}\, {}^{\circ}\forall \lambda t^{*}\, [{}^{\circ}\mathrm{K}_{w^{*}t^{*}}\ {}^{\circ}\mathrm{KF}_{w^{*}t^{*}}]], \qquad (\mathrm{DD2}')$$

(°K, °KF *de dicto*)

constructs the necessary proposition, true in all w, t pairs (the composition $[{}^{\circ}\mathrm{K}_{w^*t^*}\ {}^{\circ}\mathrm{KF}_{w^*t^*}]$, the King is a king, is *almost* true so to speak, i.e. true in all those w^*, t^* pairs where ${}^{\circ}\mathrm{KF}_{w^*t^*}$ is not v-improper, but in the other w^*, t^* pairs it is v-improper, so that ${}^{\circ}\forall\lambda w^*\ {}^{\circ}\forall\lambda t^*\ [{}^{\circ}\mathrm{K}_{w^*t^*}\ {}^{\circ}\mathrm{KF}_{w^*t^*}]$ is false).

$$\lambda w\lambda t\,[\lambda x\,[{}^{\circ}\exists\lambda w^*\ {}^{\circ}\exists\lambda t^*\ [{}^{\circ}\neg\,[{}^{\circ}\mathrm{K}_{w^*t^*}\ x]]]\ {}^{\circ}\mathrm{KF}_{wt}], \qquad (\mathrm{DR1}')$$

(${}^{\circ}$K *de dicto*, ${}^{\circ}$KF *de re*)

constructs a properly partial proposition P which was true in the actual world in some period before the year 1789, but which does not have any truth-value now (because ${}^{\circ}\mathrm{KF}_{wt}$ is improper). If P had any truth-value, the King of France would have to exist. Hence DR1 is, unlike DD1 and DD2, an empirical sentence. Whenever the King exists, the individual that occupies the office has the property of possibly not being the King.

$$\lambda w\lambda t\,[\lambda x\,[{}^{\circ}\neg\,[{}^{\circ}\forall\lambda w^*\ {}^{\circ}\forall\lambda t^*\ [{}^{\circ}\mathrm{K}_{w^*t^*}\ x]]]\ {}^{\circ}\mathrm{KF}_{wt}], \qquad (\mathrm{DR2}')$$

(${}^{\circ}$K *de dicto*, ${}^{\circ}$KF *de re*)

constructs the same proposition P as DR1$'$.

We can see that DR1$'$, as well as DR2$'$ are adequate analyses of our sentence FK. They are equivalent, but not identical, i.e. FK is *weakly homonymous.* But performing a β-reduction, we obtain

$$\lambda w\lambda t\,[{}^{\circ}\exists\lambda w^*{}^{\circ}\exists\lambda t^*\ [{}^{\circ}\neg\,[{}^{\circ}\mathrm{K}_{w^*t^*}\ {}^{\circ}\mathrm{KF}_{wt}]]], \qquad (\mathrm{DR1}\beta)$$

(${}^{\circ}$K *de dicto*, ${}^{\circ}$KF *de re*)

It constructs a *total* proposition P$'$, different from P. The proposition P$'$ "behaves" in the same way as P in those w, t pairs where ${}^{\circ}\mathrm{KF}_{wt}$ is a proper construction by returning true. But P$'$ is simply *false* in those w, t pairs where ${}^{\circ}\mathrm{KF}_{wt}$ is v-improper (for instance, in the actual possible world now). The class of those w^*, t^* for which $[{}^{\circ}\neg\,[{}^{\circ}\mathrm{K}_{w^*t^*}\ {}^{\circ}\mathrm{KF}_{wt}]]$ holds is not non-empty, it is a 'degenerated' class, because $[{}^{\circ}\neg\,[{}^{\circ}\mathrm{K}_{w^*t^*}{}^{\circ}\mathrm{KF}_{wt}]]$ is v-improper, and ${}^{\circ}\exists$ returns false.

$$\lambda w\lambda t\,[{}^{\circ}\neg\,[{}^{\circ}\forall\lambda w^*\ {}^{\circ}\forall\lambda t^*\ [{}^{\circ}\mathrm{K}_{w^*t^*}\ {}^{\circ}\mathrm{KF}_{wt}]]], \qquad (\mathrm{DR2}\beta)$$

(${}^{\circ}$K *de dicto*, ${}^{\circ}$KF *de re*) constructs another *total* proposition P″, the necessary one this time, different from P and P′. The proposition P″ "behaves" in the same way as P in those w, t pairs where $^\circ\mathrm{KF}_{wt}$ is a proper construction, it returns true. But P″ is simply also true in those w, t pairs where $^\circ\mathrm{KF}_{wt}$ is v-improper (for instance, in the actual possible world now).

We can see that neither DR1β nor DR2β is correct as a semantic analysis. Now, how shall we analyse the sentence

The King of France is a king? (FKis)

There are again two readings of the sentence, namely *de dicto* and *de re*. On its *de re* reading the sentence has no truth-value now (there is no current King of France). On its *de dicto* reading the sentence is true, expressing the fact that being a king is a necessary condition for being the King of France, or in other words, it is a *requisite* of the King's office. Being a requisite is a relation-in-extension between intensions:

$$[^\circ\mathrm{Requisite}\ ^\circ\mathrm{King}\ ^\circ\mathrm{King\text{-}of\text{-}France}].$$

Note that the true *de dicto* reading does not entail the existence of the King, the office is just *mentioned.*

The definition of a *requisite* P / $(o\alpha)_{\tau\omega}$ of an office O / $\alpha_{\tau\omega}$ is given by the following (E / $(o\alpha_{\tau\omega})_{\tau\omega}$, an 'existence property of the α-office', where x ranging over α of any type):

$$[^\circ\mathrm{Requisite}\ ^\circ\mathrm{P}\ ^\circ\mathrm{O}] =_{\mathrm{df}}$$
$$=_{\mathrm{df}}\ ^\circ\forall\lambda w ^\circ\forall\lambda t\,[[^\circ\mathrm{E}_{wt}\ ^\circ\mathrm{O}] \supset {}^\circ\forall\lambda x\,[[x = {}^\circ\mathrm{O}_{wt}] \supset [^\circ\mathrm{P}_{wt}\ x]]]$$

(read: whenever the office is occupied the respective occupier has to be a P).

Conclusion

We have shown that when adequately modelling natural language semantics, partial functions cannot be avoided. But this fact has nothing to do with our subjective insufficient cognitive abilities. We presented sound objective arguments for the need of using partial functions in our logic. Sentences talking about such partial functions (intensions) in the *de re* way cannot have any truth-value in case the function is *used* to obtain its value at an argument when there is none. On the other hand, mentioning such partial functions, i.e., talking about them in a *de dicto* way, does not cause any problems. Thus we conclude with the positive answer to the question raised in the title of this study. We have to deal with partiality, and though accommodating partial functions into a logic is rather incommodious, we have to look for technical tools enabling us not to violate natural language semantics.

References

[1] P. B. Andrews. *An Introduction to Mathematical Logic and Type Theory: to Truth through Proof.* Academic Press, Orlando, Florida, 1986.

[2] R. Chisholm. Knowledge and belief: *de dicto* and *de re*. *Philosophical Studies*, 29:1–20, 1976.

[3] A. Church. A formulation of the logic of sense and denotation. In P. Henle et al., editors, *Structure, Method, and Meaning. Essays in Honor of Henry M. Sheffer*, New York, 1951.

[4] M. Duží. Intensional logic and the irreducible contrast between *de dicto* and *de re*. Online http://www.cs.vsb.cz/duzi/, 2002.

[5] M. Duží and P. Materna. Propositional attitudes revised. In T. Childers, editor, *The Logica Yearbook 2000*, pages 163–174. Filosofia, Prague, 2001.

[6] M. Duží and P. Materna. Parmenides principle. In T. Childers, editor, *The Logica Yearbook 2003*, pages 159–178. Filosofia, Prague, 2003.

[7] W. Farmer. A partial version of Church's simple theory of types. *J. Symbolic Logic*, 55(3):1269–1291, 1990.

[8] G. Frege. Über Sinn und Bedeutung. *Zeitschrift f. Philosophie und philosophische Kritik*, 100:25–50, 1892.

[9] J. Hintikka and G. Sandu. Informational independence as a semantical phenomenon. In J. E. Fenstad et al., editors, *Logic, Methodology and Philosophy of Science*, pages 571–589, Amsterodam, 1989. Elsevier.

[10] J. Hintikka and G. Sandu. A revolution in logic? *Nordic J. Philosophical Logic*, 1(2):169–183, 1996.

[11] K. Lambert. Existential import revisited. *Notre Dame J. Formal Logic*, 4:288–292, 1963.

[12] P. Martin-Löf. Analytic and synthetic judgements in type theory. Talk given at the workshop on Kant and Contemporary Epistemology, 1992.

[13] P. Materna. Concepts and objects. *Acta Philosophica Fennica*, 63, 1974.

[14] R. Montague. *Formal Philosophy.* Yale University Press, New Haven and London, 1974.

[15] R. Muskens. *Meaning and Partiality.* Dissertation, University of Amsterdam, Amsterdam, 1989.

[16] B. Russell. On denoting. *Mind (New Series)*, 14:479–493, 1905.

[17] B. Russell. The philosophy of logical atomism. In *Logic and Knowledge*, pages 177–281, London, 1956.

[18] D. S. Scott. Identity and existence in intuitionistic logic. *LNM*, 753:660–696, 1979.

[19] P. F. Strawson. On referring. *Mind*, 59:320–344, 1950.

[20] P. Tichý. Existence and God. *J. Philosophy*, 76:403–420, 1979.

[21] P. Tichý. Foundations of partial type theory. *Reports on Mathematical logic*, 14:59–79, 1982.

[22] P. Tichý. Two fallacies in formal semantics. In M. Fricke, editor, *Essays in Honour of Bob Durrant*, pages 156–177. Otago University Philosophy Department, 1986.

[23] P. Tichý. *The Foundations of Frege's Logic.* De Gruyter, 1988.

[24] P. Tichý. Constructions as the subject matter of mathematics. In W. DePauli-Schimanovich et al., editors, *The Foundational Debate*, pages 175–185. Kluwer, 1995.

Marie Duží
VSB—Technical University of Ostrava
17. listopadu 15, Ostrava
marie.duzi@vsb.cz

On Language Levels of a Theory

Tomáš Holeček

Introduction

A theory is a set of first-order sentences. From the basic logical point of view a theory is well characterized by the set of all its deductive consequences. Now imagine our predicates are sorted into mutually disjoint groups. I am going to discuss three ways of logical characterization of a theory with respect to one of these groups. The philosophical reasons for doing this are explicated in my dissertation. The groups of predicates can symbolize different language levels and a theory can be viewed from these levels. The language levels are constituted by meaning differences.

Some prerequisites

For the sake of simplicity I consider first order language without functions and equality. As a syntactical tool for deduction I use Gentzen calculus LK (cf. [2]) in the form presented in [5] (i.e. in the form where sequents are pairs of sets of formulae rather than pairs of sequences of formulae). However, I need some minor changes and additions to the terminology: *Initial sequents* are sequents of the form $\langle \Gamma, \alpha \Rightarrow \Delta, \alpha \rangle$ where α may be an arbitrary atomic formula. If there is an initial sequent on a branch the atomic formula α from this sequent *closes* the branch. A branch is *saturated* if every non-atomic formula on the branch is the principal formula of some inference on the branch, and all possible specifications of generally quantified formulae in antecedents and all possible specifications of existentially quantified formulae in succedents occur on the branch.

The predicates of the language in question are sorted into mutually disjoint groups. I will use a meta-variable σ for these groups. If Γ is a set of formulae, then $\Gamma \restriction \sigma$ is the subset of Γ containing all the formulae from Γ which are composed of predicates from the group σ only; this subset is *the restriction of* Γ *to the group* σ. Then σ*-initial sequents* are sequents of the form $\langle \Gamma, \alpha \Rightarrow \Delta, \alpha \rangle$ where α may be an arbitrary atomic formula containing a predicate from the group σ.

In the proofs of my theorems I use the following two properties of the calculus:

Lemma 1 *(a) Let* Γ *and* Δ *be non-empty sets of formulae. If the sequent* $\langle \Gamma \Rightarrow \Delta \rangle$ *is provable then there exists a disjunction* φ *of some formulae from* Δ *such that the sequent* $\langle \Gamma \Rightarrow \varphi \rangle$ *is also provable.*

(b) Let P *be a predicate and let* Γ *and* Δ *be sets of formulae such that there is just one occurrence of* P *in the sequent* $\langle \Gamma \Rightarrow \Delta \rangle$*. Then no atomic formula containing* P *can close any branch of any tree of the sequent* $\langle \Gamma \Rightarrow \Delta \rangle$*.*

Proof (a) Follows immediately from the compactness of first order logic.

(b) Let's choose the single formula from the sequent $\langle \Gamma \Rightarrow \Delta \rangle$ in which P occurs. For all its subformulae containing P (or their specifications) φ: in any inference in which φ is the principal formula, it is unambiguously stated on which side of the upper sequent can the auxiliary formulae come. So P cannot appear on both sides of any sequent. ∎

Deductive consequences

Recall that a sequent is provable in Gentzen calculus iff it is the final sequent of some proof-tree composed of inferences with initial sequents at the end of all of its branches. Similarly, σ*-provable* sequent is the final sequent of some proof-tree with σ-initial sequents at the end of all of its branches. Trivially, all σ-provable sequents are provable.

If Γ is a theory then $\mathrm{Cn}(\Gamma)$ is the set of all sentences φ such that $\langle\, \Gamma \Rightarrow \varphi \,\rangle$ is provable sequent. Similarly, $\mathrm{Cn}_\sigma(\Gamma)$ is the set of all sentences φ such that $\langle\, \Gamma \Rightarrow \varphi \,\rangle$ is a σ-provable sequent.

With the use of these two concepts and the concept of restriction I can show the tree ways of logical characterization of the theory Γ:

1. $\mathrm{Cn}(\Gamma) \restriction \sigma$,
2. $\mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$,
3. $\mathrm{Cn}(\Gamma \restriction \sigma) \restriction \sigma$.

The following lemma shows interrelationships between these sets of deductive consequences of Γ:

Lemma 2 *(a) Let* Γ *be a set of sentences and let* σ *be a group of predicates. Then* $\mathrm{Cn}(\Gamma \restriction \sigma) \restriction \sigma \subseteq \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma \subseteq \mathrm{Cn}(\Gamma) \restriction \sigma$.

(b) There can exist a set of sentences Γ *and a group of predicates* σ *such that* $\mathrm{Cn}(\Gamma \restriction \sigma) \restriction \sigma \subset \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma \subset \mathrm{Cn}(\Gamma) \restriction \sigma$.

Proof (a) Let $\varphi \in \mathrm{Cn}(\Gamma \restriction \sigma) \restriction \sigma$. Then there exists a proof-tree of the sequent $\langle\, \Gamma \restriction \sigma \Rightarrow \varphi \,\rangle$ such that all branches end with initial sequents. All predicates on this tree belong to the group σ and therefore the initial sequents are σ-initial as well. Then $\varphi \in \mathrm{Cn}_\sigma(\Gamma \restriction \sigma) \restriction \sigma$. Since $\Gamma \restriction \sigma \subseteq \Gamma$, it is clear that $\varphi \in \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$.

Let $\varphi \in \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$. Then the sequent $\langle\, \Gamma \Rightarrow \varphi \,\rangle$ is σ-provable and therefore it is provable. Since all predicates in φ are from the group σ, it follows that $\varphi \in \mathrm{Cn}(\Gamma) \restriction \sigma$.

(b) Consider the set $\{Pa,\ Pa \vee Sa \rightarrow Pb,\ Sb,\ Sb \rightarrow Pc\}$ and the group containing the predicate P only. Then $Pc \in \mathrm{Cn}(\Gamma) \restriction \sigma$ but it is not in $\mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$. The proof-tree

$$\frac{\langle\, Pa, Pa \vee Sa \rightarrow Pb, Sb \Rightarrow Pc, Sb \,\rangle \qquad \langle\, Pa, Pa \vee Sa \rightarrow Pb, Sb, Pc \Rightarrow Pc \,\rangle}{\langle\, Pa, Pa \vee Sa \rightarrow Pb, Sb, Sb \rightarrow Pc \Rightarrow Pc \,\rangle}$$

has initial sequents at the end of both branches but there is no tree with the same final sequent and σ-initial sequents at the end of all branches.

$Pb \in \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$ but it is not in $\mathrm{Cn}(\Gamma \restriction \sigma) \restriction \sigma$. The proof-tree

$$\frac{\dfrac{\langle\, Pa, Sb, Sb \rightarrow Pc \Rightarrow Pb, Pa, Sa \,\rangle}{\langle\, Pa, Sb, Sb \rightarrow Pc \Rightarrow Pb, Pa \vee Sa \,\rangle} \qquad \langle\, Pa, Pb, Sb, Sb \rightarrow Pc \Rightarrow Pb \,\rangle}{\langle\, Pa, (Pa \vee Sa) \rightarrow Pb, Sb, Sb \rightarrow Pc \Rightarrow Pb \,\rangle}$$

has σ-initial sequents at the end of both branches. But the sequent $\langle Pa \Rightarrow Pb \rangle$ (i.e. the sequent $\langle \Gamma \restriction \sigma \Rightarrow Pb \rangle$) is not provable at all. ∎

Second order characterization

The three sets of consequences contain predicates from the group σ only. Theory Γ can contain other predicates as well. But within second order language (cf. [6]), Γ can be (in the way expressed in the following theorem) replaced by a set of sentences containing predicates from the group σ only with the same or lesser cardinality and the same consequences.

Let φ be a first order sentence. Then let $\sigma(\varphi)$ be the second order sentence obtained from φ by replacing each predicate form any other group than σ by a second order variable and adding existential quantifier for it at the beginning of the sentence (i.e. all occurrences of a predicate are replaced by the same variable), and let $\acute{\sigma}(\varphi)$ be the second order sentence obtained from φ by replacing each occurrence of any predicate form any other group than σ by different second order variable and adding existential quantifier for it at the beginning of the sentence (i.e. different occurrences of a predicate are replaced by different variables).

Theorem 3 *(a) Let Γ be a finite theory and let σ be a group of predicates. Then there exists a second order sentence δ containing predicates from the group σ only such that* $\mathrm{Cn}(\Gamma) \restriction \sigma = \mathrm{Cn}(\delta) \restriction \sigma$.

(b) Let Γ be a theory and let σ be a group of predicates. Then there exists a set of second order sentences Δ containing predicates from the group σ only with the same cardinality as Γ and such that $\mathrm{Cn}_\sigma(\Gamma) \restriction \sigma = \mathrm{Cn}_\sigma(\Delta) \restriction \sigma = \mathrm{Cn}(\Delta) \restriction \sigma$.

Proof The proofs of both (a) and (b) are based on transformations of proof-trees.

In (a), let φ be the conjunction of all sentences from Γ and let $\delta = \sigma(\varphi)$. If $\psi \in \mathrm{Cn}(\Gamma) \restriction \sigma$ then there exists a proof-tree of the sequent $\langle \Gamma \Rightarrow \psi \rangle$. Add new inferences under the final sequent, following the rule for conjunction in the antecedent, and get a proof-tree of the sequent $\langle \varphi \Rightarrow \psi \rangle$. Then add new inferences following the rule for second-order existential quantifier in the antecedent (for all predicates from other groups than σ; these predicates become second-order eigenvariables on the new proof-tree) and get a proof-tree of the sequent $\langle\, \sigma(\varphi) \;\Rightarrow\; \psi\,\rangle$. It follows that $\psi \in \mathrm{Cn}(\delta) \restriction \sigma$.

On the other hand, if $\psi \in \mathrm{Cn}(\delta) \restriction \sigma$ then there exists a proof-tree of the sequent $\langle\, \sigma(\varphi) \;\Rightarrow\; \psi\,\rangle$. This proof-tree can be transformed to the form mentioned in the previous paragraph. Now take as the new final sequent the first sequent above inferences decomposing second-order existential quantifiers and conjunctions from the formula $\sigma(\varphi)$, and replace all second-order eigenvariables by the corresponding predicates to get a proof-tree of the sequent $\langle\, \Gamma \;\Rightarrow\; \psi\,\rangle$. It follows that $\psi \in \mathrm{Cn}(\Gamma) \restriction \sigma$.

(b) Let Δ be the set $\{\, \delta\,;\, \exists\varphi \in \Gamma(\delta = \acute{\sigma}(\varphi))\,\}$. Trivially, Δ has the same cardinality as Γ.

If $\psi \in \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$ then there exists a proof-tree of the sequent $\langle\, \Gamma \;\Rightarrow\; \psi\,\rangle$ with σ-initial sequents at the end of all branches. In the final sequent, replace all occurrences of predicates from other groups than σ by new second-order variables. This replacement can be correctly extended to the entire proof-tree (track down the path of all predicates). Now, add under the final sequent new inferences following the rule for second-order existential quantifier in the antecedent (for all new second-order variables) and get a proof-tree of the sequent $\langle\, \Delta \;\Rightarrow\; \psi\,\rangle$. It follows that $\psi \in \mathrm{Cn}_\sigma(\Delta) \restriction \sigma$.

If $\psi \in \mathrm{Cn}_\sigma(\Delta) \restriction \sigma$ then there exists a proof-tree of the sequent $\langle\, \Delta \;\Rightarrow\; \psi\,\rangle$ with σ-initial sequents at the end of all branches. This proof-tree can be transformed to the form mentioned in the previous paragraph. Now take as the new final sequent the first sequent above inferences decomposing second-order existential quantifiers. Since no second-order variable can occur in any atomic formula closing any branch of this proof-tree the variables can be replaced by predicates in order to get a proof-tree of the sequent $\langle\, \Gamma \;\Rightarrow\; \psi\,\rangle$. It follows that $\psi \in \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$.

If $\psi \in \mathrm{Cn}_\sigma(\Delta) \restriction \sigma$ then trivially $\psi \in \mathrm{Cn}(\Delta) \restriction \sigma$.

If $\psi \in \mathrm{Cn}(\Delta) \restriction \sigma$ then there exists a proof-tree of the sequent $\langle\, \Delta \;\Rightarrow\; \psi\,\rangle$. Since there is no predicate from other group than σ in ψ and in Δ and all second-order eigenvariables can occur in the antecedent just once, it follows from the Lemma 1(b) that the sequent $\langle\, \Delta \;\Rightarrow\; \psi\,\rangle$ is σ-provable. Therefore $\psi \in \mathrm{Cn}_\sigma(\Delta) \restriction \sigma$. ∎

Model theoretic characterization

Recall that the set $\mathrm{Cn}(\Gamma)$ is valid in a structure $\mathbf{U}$ iff Γ is valid in $\mathbf{U}$. In the following theorem I present an analogous description of the sets $\mathrm{Cn}(\Gamma) \restriction \sigma$ and $\mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$.

Let $\langle \Gamma \Rightarrow \Delta \rangle$ be a sequent which is not provable. It follows that on some of its trees, there exists a saturated branch with no initial sequent. For this branch it is possible to construct a structure $\mathbf{U}$ satisfying (in a suitable valuation of free variables) all formulae from all antecedents including Γ and negations of all formulae from all succedents including Δ. This is well known from the proof of completeness of Gentzen calculus (cf. [4]).

Let $\mathbf{U}$ and $\mathbf{U}'$ be structures such that for any sentence φ containing predicates from the group σ only: $\mathbf{U} \models \varphi$ iff $\mathbf{U}' \models \varphi$. Then $\mathbf{U}$ and $\mathbf{U}'$ are *σ-equivalent*. Let Γ be a theory. Then Γ^σ is the set of all sentences from Γ with all occurrences of predicates from other groups than σ replaced by a new predicate (i.e. all predicates from any other group than σ can occur in Γ^σ just once).

Theorem 4 *(a) Let Γ be a theory, let σ be a group of predicates and let $\mathbf{U}$ be a structure. Then $\mathbf{U} \models \mathrm{Cn}(\Gamma) \restriction \sigma$ iff there exists a structure $\mathbf{U}'$ σ-equivalent with $\mathbf{U}$ such that $\mathbf{U}' \models \Gamma$.*

(b) Let Γ be a theory, let σ be a group of predicates and let $\mathbf{U}$ be a structure. Then $\mathbf{U} \models \mathrm{Cn}_\sigma(\Gamma) \restriction \sigma$ iff there exists a structure $\mathbf{U}'$ σ-equivalent with $\mathbf{U}$ such that $\mathbf{U}' \models \Gamma^\sigma$.

Proof (a) Let $\mathbf{U}'$ be a structure such that it is σ-equivalent with $\mathbf{U}$ and $\mathbf{U}' \models \Gamma$. Then $\mathbf{U}' \models \mathrm{Cn}(\Gamma)$ and $\mathbf{U}' \models \mathrm{Cn}(\Gamma) \restriction \sigma$. Since there are no predicates from other groups than σ in $\mathrm{Cn}(\Gamma) \restriction \sigma$, it follows that $\mathbf{U} \models \mathrm{Cn}(\Gamma) \restriction \sigma$.

Suppose there is no such a structure and let $\mathrm{Neg}(\mathbf{U})$ be a set of all sentences φ with predicates from the group σ only such that $\mathbf{U} \not\models \varphi$. The sequent $\langle \Gamma \Rightarrow \mathrm{Neg}(\mathbf{U}) \rangle$ is provable or else it would be possible to construct a structure $\mathbf{U}'$ satisfying Γ and negations of all sentences from $\mathrm{Neg}(\mathbf{U})$ and this structure would be σ-equivalent with $\mathbf{U}$. Therefore by Lemma 1(a) there exists a sentence ψ such that the sequent $\langle \Gamma \Rightarrow \psi \rangle$ is provable and ψ is a disjunction of some sentences from $\mathrm{Neg}(\mathbf{U})$. Since $\psi \in \mathrm{Cn}(\Gamma) \restriction \sigma$ and $\mathbf{U} \not\models \psi$ it follows that $\mathbf{U} \not\models \mathrm{Cn}(\Gamma) \restriction \sigma$.

(b) It is enough to show that $\mathrm{Cn}_\sigma(\Gamma)\restriction\sigma = \mathrm{Cn}(\Gamma^\sigma)\restriction\sigma$ and then apply (a). If $\psi \in \mathrm{Cn}_\sigma(\Gamma)\restriction\sigma$ then there exists a proof-tree of the sequent $\langle\,\Gamma \Rightarrow \psi\,\rangle$ with σ-initial sequents at the end of all branches. In the final sequent, replace all occurrences of predicates from other groups than σ by new predicates and get the sequent $\langle\,\Gamma^\sigma \Rightarrow \psi\,\rangle$. This replacement can be correctly extended to the entire proof-tree (track down the path of all predicates, as in the proof of Theorem 3(b)). It follows that $\psi \in \mathrm{Cn}(\Gamma^\sigma)\restriction\sigma$. If $\psi \in \mathrm{Cn}(\Gamma^\sigma)\restriction\sigma$ then there exists a proof-tree of the sequent $\langle\,\Gamma^\sigma \Rightarrow \psi\,\rangle$. Since there is no predicate form other groups than σ in ψ and all predicates from other groups than σ occur in Γ^σ just once, by Lemma 1(b) follows that all initial sequents on the proof-tree are σ-initial. Thus all predicates from other groups than σ can be correctly replaced to get a proof-tree of the sequent $\langle\,\Gamma \Rightarrow \psi\,\rangle$. It follows that $\psi \in \mathrm{Cn}_\sigma(\Gamma)\restriction\sigma$. ■

Language levels

I deal with questions of definability studied in [3] and [1]. Frank Ramsey in his treatise on theories used the second order characterization similar to my Theorem 3(a), and William Craig in his treatise on implicit and explicit definitions used a proof based on interpolation quite similar to the proof of my Theorem 4(a). These questions are of a great importance for the problem of language levels.

With Theorems 3 and 4 in mind, I can verbally describe the three sets of deductive consequences of a theory:

1. $\mathrm{Cn}(\Gamma)\restriction\sigma$ is the set of those consequences of Γ that can be expressed on a specific language level,
2. $\mathrm{Cn}_\sigma(\Gamma)\restriction\sigma$ is the set of those consequences of Γ that can be expressed on a specific language level and that are derived without the use of logic in other language levels,
3. $\mathrm{Cn}(\Gamma\restriction\sigma)\restriction\sigma$ is the set of consequences of that part of Γ that can be expressed on a specific language level.

References

[1] W. Craig. Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory. *J. Symbolic Logic*, 22(3):269–285, 1957.

.

[2] G. Gentzen. Investigations into logical deduction. In M. E. Szabo, editor, *Collected Papers of Gerhard Gentzen*, pages 68–131. North-Holland, 1969.

[3] F. P. Ramsey. Theories. In D. H. Mellor, editor, *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, pages 101–125. Routledge, London, 1978.

[4] H. Rasiowa and R. Sikorski. On the Gentzen theorem. *Fundamenta Mathematicae*, 48:57–69, 1960.

[5] V. Švejdar. *Logika: neúplnost, složitost a nutnost (Logic: Incompleteness, Complexity, and Necessity)*. Academia, Praha, 2002.

[6] G. Takeuti. *Proof Theory*. North-Holland, Amsterdam, 1975.

Tomáš Holeček
Faculty of Humanities, Charles University
holecek@ojrech.cz

Three-Element Implicative Matrices

Milan Matoušek, Petr Jirků*

Abstract The class of three-element implicative posets is defined. Unlike the classical theory, the identity $x \to x = 1$ is not assumed. We will show for which three-element implicative posets an analogy of *deductive law* exists, and for which is possible to describe the consequence operator with the *modus ponens* rule only.

Introduction

Recently we can observe an enormous growth of the number of publications on fuzzy logic in which truth values are taken from the interval [0, 1] of real numbers. From this aspect various well-known three-valued logics can be viewed only as special cases of fuzzy logic systems. We can present a little different viewpoint. We want to point out that the area of three-valued logics is still an ambitious and independent field of logical investigations which is not so strongly determined by the structure of the real interval of truth values as in fuzzy logic. Large variability of algebraic properties is stressed and mutual connections among three-valued logic systems are discussed with respect to such concepts as implicative algebras, MV algebras, equational systems, and some modal connectives posing the question how these concepts relate to each other with re-

*This paper was supported by grant 401/01/0218 of the Grant Agency of the Czech Republic.

spect to their prime logical and/or philosophical motivations. Special attention will be paid to three-valued logics.

We assume the basic notions of the theory of logical matrices, universal algebra and model theory to be known; the reader can find the necessary information in e.g. [11], [3], [4]. For the convenience of the reader, let us briefly review the basic notions of the theory of logical matrices as we will use them in the sequel.

Let Σ be a first order language, i.e. a set of operational and relational symbols. Let us denote $\mathrm{V} = \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ the set of all variables. Let us further denote $\mathrm{Term}(\Sigma)$ the set of all Σ-terms.

A Σ-*matrix* is any couple $M = \langle A, F\rangle$, where A is a Σ-structure (i.e. a structure in language Σ) and $\emptyset \neq F \subseteq \dot{A}$ (where $\dot{A}$ is the underlying set of structure A). The elements of set F are called *designated values* for matrix M. A *valuation* in matrix $M = \langle A, F\rangle$ is every map $e : \mathrm{V} \to \dot{A}$. It is obvious that any valuation M can be extended into map $\mathrm{Term}(\Sigma) \to A$ in natural way.

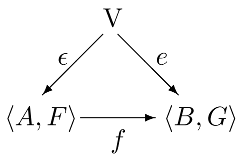
Let $M = \langle A, F\rangle$ be a Σ-matrix. Let us denote by $\mathrm{Taut}(M)$ the set of all Σ-terms s such that for any valuation v in M there is $v(s) \in F$. Let T be a set of Σ-terms, s being Σ-term. We will write $T \models_M s$, if for any valuation v in M such that $v(t) \in F$ for every $t \in T$, we have also $v(s) \in F$. Let us denote $\mathrm{Cn}_M(T) = \{\, s \in \mathrm{Term}(\Sigma)\ ;\ T \models_M s \,\}$. The map $\mathrm{Cn}_M : \mathcal{P}(\mathrm{Term}(\Sigma)) \to \mathcal{P}(\mathrm{Term}(\Sigma))$ is called the *consequence operator* determined by matrix M. Obviously, $\mathrm{Taut}(M) = \mathrm{Cn}_M(\emptyset)$.

Definition 1 *Let $M = \langle A, F\rangle$, $N = \langle B, G\rangle$ be Σ-matrices, f a map $\dot{A} \to \dot{B}$. Let us call f a* reduction *if the following conditions are satisfied:*
(i) *f is a surjective homomorphism from A to B;*
(ii) *$\forall x \in A : x \in F$ iff $f(x) \in G$ (equivalently: $F = f^{-1}(G)$).*
Let us call matrix N a reduct *of matrix M (or, equivalently, matrix M is* reducible *onto matrix N), if there is a reduction $M \to N$.*

Theorem 2 *Let matrix N be a reduct of matrix M. Then $\mathrm{Cn}_N = \mathrm{Cn}_M$.*

Proof Let $T \subseteq \mathrm{Term}(\Sigma)$ and $s \in \mathrm{Term}(\Sigma)$. We will prove $T \models_M s$ iff $T \models_N s$.

$(\to)$ At first let us assume $T \models_M s$. Let $e : \mathrm{V} \to B$ and $e(T) \subseteq G$. Since f is surjective, there is a map $\epsilon : \mathrm{V} \to A$ such that $\epsilon \circ f = e$:



Then $f(\epsilon(T)) \subseteq G$. Since f is reduction, it has to be $\epsilon(T) \subseteq F$. Since $T \models_M s$, it holds $\epsilon(s) \in F$. Thus, $f(\epsilon(s)) \in G$, i.e. $e(s) \in G$.

($\leftarrow$) Conversely, let $T \models_N s$. Let $\epsilon : \mathrm{V} \to A$ and $\epsilon(T) \subseteq F$. Then $f(\epsilon(T)) \subseteq G$, i.e. $(\epsilon \circ f)(T) \subseteq G$. Since $T \models_N s$, it also holds that $(\epsilon \circ f)(s) \in G$, i.e. also $f(\epsilon(s)) \in G$. Since f is a reduction, it has to be $\epsilon(s) \in F$. ■

Definition 3 *Let us denote* $\Sigma_{\mathrm{IP}} = \{\leq, \mathbf{0}, \mathbf{1}, \longrightarrow\}$, *where* $\leq$ *is a binary relational symbol,* $\mathbf{0}$ *and* $\mathbf{1}$ *are nullary operational symbols (i.e. constants) and* $\longrightarrow$ *is a binary operational symbol. Instead of* $\mathrm{Term}(\{\longrightarrow\})$ *we will write only* $\mathrm{Term}(\longrightarrow)$. *An* implicative poset *is any structure* P *of signature* Σ_{IP} *such that* $P \downarrow \{\leq, \mathbf{0}, \mathbf{1}\}$ *is a poset with the least and the greatest element satisfying the following laws:*

$$\mathbf{0} \longrightarrow \mathbf{0} \approx \mathbf{1}, \quad \mathbf{0} \longrightarrow \mathbf{1} \approx \mathbf{1}, \quad \mathbf{1} \longrightarrow \mathbf{0} \approx \mathbf{0}, \quad \mathbf{1} \longrightarrow \mathbf{1} \approx \mathbf{1},$$
$$\mathbf{x} \leq \mathbf{y} \Rightarrow \mathbf{z} \longrightarrow \mathbf{x} \leq \mathbf{z} \longrightarrow \mathbf{y},$$
$$\mathbf{x} \leq \mathbf{y} \Rightarrow \mathbf{x} \longrightarrow \mathbf{z} \geq \mathbf{y} \longrightarrow \mathbf{z},$$
$$\mathbf{x} < \mathbf{y} \Rightarrow \mathbf{x} \longrightarrow \mathbf{y} \approx \mathbf{1},$$
$$\mathbf{x} < \mathbf{1} \Rightarrow \mathbf{1} \longrightarrow \mathbf{x} < \mathbf{1}.$$

Let P be an implicative poset, $F \subseteq P$. Let us call the set F an *implicative filter* in the poset P if the following hold:

$1_P \in F$,
$\forall x, y \in P$: if $x \in F$ and $x \to y \in F$ then $y \in F$.

Lemma 4 *Let* P *be an implicative poset. Then sets* $F = \{1_P\}$, $F = \dot{P}$ *are always implicative filters in* P.

Proof For set $F = \dot{P}$ the statement is obvious. Let now $F = \{1_P\}$, $x \in F$, and $x \to y \in F$. Thus $1_P \to y = 1_P$. From the fourth condition in the Definition 3 it follows $y = 1_P$. ■

Lemma 5 *Let F be an implicative filter in the implicative poset P. Then F is also a filter in P.*

Proof Let $x \in F$, $x \leq y$. If $x = y$, then naturally also $y \in F$. Let now $x < y$. Then $x \rightarrow y = 1_P$. Now $x \in F$, $x \rightarrow y \in F$. Since F is an implicative filter, it also holds that $y \in F$. ■

An *implicative matrix* is any couple $M = \langle P, F \rangle$, where P is an implicative poset, F is an implicative filter in P. If $F \neq P$, we call the matrix M *nontrivial.* If $F = \{1_P\}$, we call the matrix M *algebraic.* The proof of the following theorem is easy:

Theorem 6 *In any implicative matrix $M = \langle P, F \rangle$ the rule* modus ponens *holds, i.e. for any Σ_{IP}-terms s, t we have: $s,\ s \longrightarrow t \models_M t$.*

By symbol **2** we will denote the *classical implicative matrix* (in which the set of designated elements is $\{1\}$).

Three-element implicative matrices

The implicative poset P is called *three-element*, if the underlying set of P has exactly three elements. The implicative matrix $M = \langle P, F \rangle$ is called *three-element*, if P consists of three elements.

Let us make the convention that for the underlying set of all three-element posets we take the set $0, 1/2, 1$ and in any three-element poset the inequality $0 < 1/2 < 1$ holds.

Theorem 7 *There are exactly eleven three-element implicative posets.*

Proof Let P be a three-element implicative poset. Then it has to be $1 \rightarrow 1 = 1$, $1 \rightarrow 0 = 0$, $1/2 \rightarrow 1 = 1$, $0 \rightarrow 1 = 0 \rightarrow 1/2 = 0 \rightarrow 0 = 1$. Unknown are the values $1 \rightarrow 1/2$, $1/2 \rightarrow 1/2$ a $1/2 \rightarrow 0$. Schematically:

P	1	1/2	0
1	1	?	0
1/2	1	?	?
0	1	1	1

From the definition of the implicative poset follows, that in P it has to hold:

$$1 \rightarrow 1/2 \leq 1/2 \rightarrow 1/2, \quad 1/2 \rightarrow 0 \leq 1/2 \rightarrow 1/2.$$

For the value $1/2 \to 1/2$ we have got following three possibilities:
• $1/2 \to 1/2 = 0$. Then it has to be $1 \to 1/2 = 0$, $1/2 \to 0 = 0$. In this way we get the poset denoted in the following table by symbol P_1.
• $1/2 \to 1/2 = 1/2$. Then it has to be $1 \to 1/2 \leq 1/2$, $1/2 \to 0 \leq 1/2$, i.e. $1 \to 1/2 \in \{0, 1/2\}$, $1/2 \to 0 \in \{0, 1/2\}$. In this way we get four posets, P_2–P_5.
• $1/2 \to 1/2 = 1$. Since $1 \to 1/2 < 1$, it has to be $1 \to 1/2 \in \{0, 1/2\}$, $1/2 \to 0 \in \{0, 1/2, 1\}$. In this way we get six posets, P_6–P_{11}.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}
$1 \to 1$	1	1	1	1	1	1	1	1	1	1	1
$1 \to 1/2$	0	0	1/2	0	1/2	0	1/2	0	1/2	0	1/2
$1 \to 0$	0	0	0	0	0	0	0	0	0	0	0
$1/2 \to 1$	1	1	1	1	1	1	1	1	1	1	1
$1/2 \to 1/2$	0	1/2	1/2	1/2	1/2	1	1	1	1	1	1
$1/2 \to 0$	0	0	0	1/2	1/2	0	0	1/2	1/2	1	1
$0 \to 1$	1	1	1	1	1	1	1	1	1	1	1
$0 \to 1/2$	1	1	1	1	1	1	1	1	1	1	1
$0 \to 0$	1	1	1	1	1	1	1	1	1	1	1

Finally, it is easy to show that all P_1–P_{11} are implicative posets. ∎

If P_i, where $i \in \{1, \ldots, 11\}$, is one of the above implicative posets, then by the symbol $\to_i$ we will denote the realisation of the operation $\longrightarrow$ in poset P_i. For the convenience of the reader, we show the tables of functions $\to_i$ for every three-element poset in Fig. 1.

Let us establish all three-element implicative matrices now. Let P be a three-element implicative poset, F an implicative filter in P. According to Lemma 5 F is a filter in P. Therefore exactly one of the following three possibilities will take place:

$$F = \{1\}, \quad F = \{1/2, 1\}, \quad F = \{0, 1/2, 1\}.$$

The sets $\{1\}$ and $\{0, 1/2, 1\}$ are implicative filters in P automatically. For the set $\{1/2, 1\}$ the following statement holds:

Theorem 8 *Let P be a three-element implicative poset. Then the set $\{1/2, 1\}$ is an implicative filter in P iff $1/2 \to 0 = 0$.*

P_1	1	1/2	0
1	1	0	0
1/2	1	0	0
0	1	1	1

P_2	1	1/2	0
1	1	0	0
1/2	1	1/2	0
0	1	1	1

P_3	1	1/2	0
1	1	1/2	0
1/2	1	1/2	0
0	1	1	1

P_4	1	1/2	0
1	1	0	0
1/2	1	1/2	1/2
0	1	1	1

P_5	1	1/2	0
1	1	1/2	0
1/2	1	1/2	1/2
0	1	1	1

P_6	1	1/2	0
1	1	0	0
1/2	1	1	0
0	1	1	1

P_7	1	1/2	0
1	1	1/2	0
1/2	1	1	0
0	1	1	1

P_8	1	1/2	0
1	1	0	0
1/2	1	1	1/2
0	1	1	1

P_9	1	1/2	0
1	1	1/2	0
1/2	1	1	1/2
0	1	1	1

P_{10}	1	1/2	0
1	1	0	0
1/2	1	1	1
0	1	1	1

P_{11}	1	1/2	0
1	1	1/2	0
1/2	1	1	1
0	1	1	1

Figure 1: Implicative posets

Proof Let us denote $F = \{1/2, 1\}$. At first, let F be an implicative filter in P. If $1/2 \to 0 \neq 0$, then $1/2 \to 0 \in F$. Since it also holds that $1/2 \in F$ and F is an implicative filter, we would obtain $0 \in F$, which is a contradiction.

Conversely, let $1/2 \to 0 = 0$. Further let $x, y \in P$, $x \in F$, $x \to y \in F$. If $y = 0$, then $x \to y = x \to 0 = 0$, since $1 \to 0 = 0$ holds always and $1/2 \to 0 = 0$ according to the assumption. It means $x \to y \notin F$, which is a contradiction. From the precceding follows $y > 0$, i.e. $y \in F$. ∎

From the posets P_i for $i = 1, \dots, 11$ we can construct implicative matrices of the form $\langle P_i, F\rangle$, where F is one of the sets $\{1\}$, $\{1/2, 1\}$ or $\{0, 1/2, 1\}$. Matrix $M = \langle P_i, \{0, 1/2, 1\}\rangle$, naturally, is not of interest, since we have in this case $\mathrm{Taut}(M) = \mathrm{Term}(\Sigma_{\mathrm{IP}})$. The couple $\langle P_i, \{1/2, 1\}\rangle$ is an implicative matrix if and only if the set $\{1/2, 1\}$ is an implicative filter in P_i. Let us recall that according to Theorem 8 the set $\{1/2, 1\}$ is an implicative filter in P_i iff $1/2 \to_i 0 = 0$, i.e. for $i = 1, 2, 3, 6, 7$. Let us denote therefore

$$L_i = \langle P_i, \{1\}\rangle, \quad i = 1, \dots, 11,$$
$$M_j = \langle P_j, \{1/2, 1\}\rangle, \quad j = 1, 2, 3, 6, 7.$$

At first we will show that some of matrices constructed in this way lead to the same consequence operator as matrix **2**.

Theorem 9 *Matrices L_{10} and L_{11} are reducible onto matrix* **2**.

Proof Let us consider the map $f : \{0, 1/2, 1\} \to \{0, 1\}$ defined in this way: $f(0) = f(1/2) = 0$, $f(1) = 1$. We can now easily prove that f is a reduction of matrices L_{10}, L_{11} onto matrix **2**. ∎

Lemma 10 *If $\{1/2, 1\}$ is a subuniverse in P_i, matrix M_i is reducible onto* **2**.

Proof Let us define the map $f : \{0, 1/2, 1\} \to \{0, 1\}$ in this way: $f(1) = f(1/2) = 1$, $f(0) = 0$. Then f is a homomorphism $P_i \to \mathbf{2}$:

$$\text{If } x, y \in \{0, 1\}, \text{ then } f(x \to_i y) = x \Rightarrow y = f(x) \Rightarrow f(y).$$

If one of x and y is $1/2$ and the other belongs to the set $\{1/2, 1\}$ then, since $\{1/2, 1\}$ is a subuniverse in P_i, it holds $x \to_i y \in \{1/2, 1\}$, and therefore $f(x \to_i y) = 1$ and $f(x) \Rightarrow f(y) = 1 \Rightarrow 1 = 1$.

At last, let one of the elements x, y be equal to $1/2$ and the other is 0. Then

$$f(0 \to_i 1/2) = f(1) = 1, \quad f(0) \Rightarrow f(1/2) = 0 \Rightarrow 1 = 1,$$
$$f(1/2 \to_i 0) = f(0) = 0, \quad f(1/2) \Rightarrow f(0) = 1 \Rightarrow 0 = 0$$

(we use the fact that $1/2 \to_i 0 = 0$!). ∎

Theorem 11 *Matrices M_3 a M_7 are reducible onto* **2**.

We will further deal only with the matrices L_i, $i = 1, \dots, 9$, and matrices M_j, $j = 1, 2, 6$. We will prove that these matrices determine pairwise different consequence operators:

Theorem 12 *The operators $\mathrm{Cn}_{L_1}, \dots, \mathrm{Cn}_{L_9}$, Cn_{M_1}, Cn_{M_2}, Cn_{M_6} are pairwise different.*

Proof We will make the proof in several steps.

Step 1. The operators $\mathrm{Cn}_{L_1}, \mathrm{Cn}_{L_6}, \mathrm{Cn}_{L_7}, \mathrm{Cn}_{L_8}, \mathrm{Cn}_{L_9}$, Cn_{M_2} are pairwise different. In this case even the sets $\mathrm{Taut}(L_1)$, $\mathrm{Taut}(L_6)$, $\mathrm{Taut}(L_7)$, $\mathrm{Taut}(L_8)$, $\mathrm{Taut}(L_9)$ and $\mathrm{Taut}(M_2)$ are pairwise different. This follows from the following two tables:

		L_1	L_6	L_7	L_8	L_9	M_2
t_1	$\mathsf{x} \to \mathsf{x}$	n	y	y	y	y	y
t_2	$\mathsf{x} \to (\mathsf{y} \to \mathsf{y})$	n	y	y	y	y	n
t_3	$\mathsf{x} \to (\mathsf{y} \to \mathsf{x})$	n	n	y	n	y	n
t_4	$(\mathsf{x} \to \mathsf{y}) \to (\mathsf{z} \to (\mathsf{x} \to \mathsf{y}))$	y	y	y	n	y	n
t_5	$(\mathsf{x} \to (\mathsf{x} \to \mathsf{y})) \to (\mathsf{x} \to \mathsf{y})$	y	y	y	n	n	y

	L_6	L_7	L_8	L_9	M_2
L_1	t_1	t_1	t_1	t_1	t_1
L_6	$\times$	t_3	t_4	t_3	t_2
L_7	$\times$	$\times$	t_3	t_5	t_2
L_8	$\times$	$\times$	$\times$	t_3	t_2
L_9	$\times$	$\times$	$\times$	$\times$	t_2

Step 2. Operators Cn_{L_2}, Cn_{L_3}, Cn_{L_4}, Cn_{L_5} are pairwise different and each of them is also different from all operators Cn_{L_1}, Cn_{L_6}, Cn_{L_7}, Cn_{L_8}, Cn_{L_9}, Cn_{M_2}. At first, we will show that $\mathrm{Taut}(L_i) \cap \mathrm{Term}(\longrightarrow) = \emptyset$ for every $i = 2, 3, 4, 5$.

Let $i \in \{2, 3, 4, 5\}$. Let us consider the valuation $e : \mathrm{V} \to L_i$ such that $e(x) = 1/2$ for any variable $x \in \mathrm{V}$. Then for every term $t \in \mathrm{Term}(\longrightarrow)$ holds $e(t) = 1/2$. Thus, there is no term $t \in \mathrm{Term}(\longrightarrow)$ such that $e(t) = 1$. Therefore $\mathrm{Taut}(L_i) \cap \mathrm{Term}(\longrightarrow) = \emptyset$.

From the previous it follows that the operators $\mathrm{Cn}_{L_2}, \mathrm{Cn}_{L_3}, \mathrm{Cn}_{L_4}, \mathrm{Cn}_{L_5}$ are different from all operators $\mathrm{Cn}_{L_1}, \mathrm{Cn}_{L_6}, \mathrm{Cn}_{L_7}, \mathrm{Cn}_{L_8}, \mathrm{Cn}_{L_9}$, Cn_{M_2}.

Further, from this follows that the operators Cn_{L_2}, Cn_{L_3}, Cn_{L_4}, Cn_{L_5} can be distiguished only by means of rules of consequence. For this see the following two tables:

		L_2	L_3	L_4	L_5
r_1	$\mathsf{x} \to \mathsf{y} \vdash (\mathsf{z} \to \mathsf{x}) \to (\mathsf{z} \to \mathsf{y})$	y	y	n	n
r_2	$\mathsf{x} \to \mathsf{y} \vdash (\mathsf{y} \to \mathsf{z}) \to (\mathsf{x} \to \mathsf{z})$	y	n	y	n

	L_3	L_4	L_5
L_2	r_2	r_1	r_1
L_3	$\times$	r_1	r_1
L_4	$\times$	$\times$	r_2

Step 3. Operators Cn_{M_1} and Cn_{M_6} are pairwise different and each of them is different from all operators $\mathrm{Cn}_{L_1}, \dots, \mathrm{Cn}_{L_9}$ and Cn_{M_2}.

First, let us prove that $\mathrm{Taut}(M_1) = \mathrm{Taut}(L_1)$, $\mathrm{Taut}(M_6) = \mathrm{Taut}(L_6)$. If $t \in \mathrm{Taut}(M_1)$, then term t cannot be a variable and from the table of the operation $\rightarrow_1$ we can see that term t can take only value 1. Thus, $t \in \mathrm{Taut}(L_1)$. The inclusion $\mathrm{Taut}(M_1) \subseteq \mathrm{Taut}(L_1)$ is proved. The converse inclusion is trivial. The identity $\mathrm{Taut}(M_6) = \mathrm{Taut}(L_6)$ can be proved in similar way.

Since the sets $\mathrm{Taut}(L_1)$ a $\mathrm{Taut}(L_6)$ are pairwise different, also the operators Cn_{M_1} and Cn_{M_6} are pairwise different. Further, the operator Cn_{M_1} is different from all operators $\mathrm{Cn}_{L_2}, \dots, \mathrm{Cn}_{L_5}$, $\mathrm{Cn}_{L_6}, \dots, \mathrm{Cn}_{L_9}$, Cn_{M_2} and Cn_{M_6}, the operator Cn_{M_6} is different from all operators Cn_{L_1}, $\mathrm{Cn}_{L_2}, \dots, \mathrm{Cn}_{L_5}$, Cn_{L_7}, Cn_{L_8}, Cn_{L_9}, Cn_{M_1} and Cn_{M_2} (we can distinguish them by means of tautologies).

It remains to prove $\mathrm{Cn}_{M_1} \neq \mathrm{Cn}_{L_1}$ and $\mathrm{Cn}_{M_6} \neq \mathrm{Cn}_{L_6}$. For this purpose we can take the following rules:

$$\mathbf{x} \models_{L_1} \mathbf{x} \rightarrow \mathbf{x}, \quad \mathbf{x} \not\models_{M_1} \mathbf{x} \rightarrow \mathbf{x},$$
$$\mathbf{x} \models_{L_6} \mathbf{y} \rightarrow \mathbf{x}, \quad \mathbf{x} \not\models_{M_6} \mathbf{y} \rightarrow \mathbf{x}.$$

■

Deductive terms

Definition 13 *Let $M = \langle P, F \rangle$ be an implicative matrix. The term $d(\mathbf{x}, \mathbf{y}) \in \mathrm{Term}(\rightarrow)$ is called a* deductive term *for the matrix M, if for any set of Σ_{IP}-terms X and for any Σ_{IP}-terms s, t we have:*

$$X, s \models_M t \quad \textit{iff} \quad X \models_M d(s, t).$$

The term $d(\mathbf{x}, \mathbf{y})$ is called a classical deductive term *for matrix M, if*

$$\forall a, b \in M : d_M(a, b) \in F \quad \textit{iff} \quad a \notin F \textit{ or } b \in F. \qquad (*)$$

Note that the condition $(*)$ can be equivalently written in this way:

$$\forall a, b \in M : d_M(a,b) \notin H \quad \text{iff} \quad a \in H \text{ and } b \notin H.$$

Lemma 14 *If* $d(\mathsf{x},\mathsf{y})$ *is a classical deductive term for matrix* M*, then* $d(\mathsf{x},\mathsf{y})$ *is also a deductive term for* M*.*

Proof Let X be a set of terms, s,t terms. We have to prove

$$X, s \models_M t \quad \text{iff} \quad X \models_M d(s,t).$$

First, let $X, s \models_M t$. We prove that $X \models_M d(s,t)$. Let $e : \mathrm{V} \to M$, $e(X) \subseteq F$. We have to prove $e(d(s,t)) \in F$, i.e. $d_M(e(s), e(t)) \in F$. We distinguish two cases:
• $e(s) \in F$. Then, since $X, s \models_M t$, it has to be also $e(t) \in F$. According to our condition now $d_M(e(s), e(t)) \in F$ holds.
• $e(s) \notin F$. Then according to our condition $d_M(e(s), e(t)) \in F$.

Conversely, let $X \models_M d(s,t)$. We will demonstrate that under this assumption it holds that $X, s \models_M t$. Let $e : \mathrm{V} \to M$, $e(X) \subseteq F$, $e(s) \in F$. Since $X \models_M d(s,t)$, $e(X) \subseteq F$, it has to be $e(d(s,t)) \in F$. Thus, $d_M(e(s), e(t)) \in F$. Since moreover $e(s) \in F$, from the condition $(*)$ it follows that $e(t) \in F$ holds. ■

Theorem 15 *Let us denote*

$$d^{(1)} = (\mathsf{x} \to \mathsf{x}) \to (\mathsf{x} \to \mathsf{y}),$$
$$d^{(6)} = ((\mathsf{x} \to \mathsf{x}) \to \mathsf{x}) \to \mathsf{y},$$
$$d^{(8)} = d^{(9)} = \mathsf{x} \to (\mathsf{x} \to \mathsf{y}).$$

Then $d^{(i)}$ *is classical deductive term for matrix* L_i*,* $i = 1, 6, 8, 9$*.*

Proof It is sufficient to construct tables of values for the functions d_i, where $d_i = (d^{(i)})_{L_i}$, for $i = 1, 6, 8, 9$. The left one of the following two tables is a table for d_1, d_6, and d_8, while the right one is a table for d_9:

L_{10}	1	1/2	0
1	1	0	0
1/2	1	1	1
0	1	1	1

L_{11}	1	1/2	0
1	1	1/2	0
1/2	1	1	1
0	1	1	1

■

Theorem 16 *Let us denote* $d^{(7)} = \mathbf{x} \to \mathbf{y}$. *Then the term* $d^{(7)}$ *is a deductive term, but not classical deductive term in matrix* L_7.

Proof First, let us prove deductivity of term $d^{(7)}$. Let X be a set of terms, s, t terms. We will demonstrate that

$$X \models_{L_7} s \to t \quad \text{iff} \quad X, s \models_{L_7} t.$$

If $X \models_{L_7} s \to t$, then from (MP) it follows that $X, s \models_{L_7} t$. Conversely, let $X, s \models_{L_7} t$. Let e be the valuation in L_7 such that $e(X) \subseteq \{1\}$. We will prove now that $e(s \to t) = 1$. There are three possibilities:
• $e(s) = 0$. Then evidently $e(s) \to_7 e(t) = 1$.
• $e(s) = 1$. Then $e(X) \subseteq \{1\}$ and $X, s \models_{L_7} t$. In this case it must hold that $e(t) = 1$, and again $e(s) \to_7 e(t) = 1$.
• $e(s) = 1/2$. Let us consider the map $h : \{0, 1/2, 1\} \to \{0, 1/2, 1\}$ defined in this way: $h(0) = 0$, $h(1/2) = h(1) = 1$. Then h is an endomorphism $L_7 \to L_7$. Now

$$h(e(X)) \subseteq \{1\}, \quad h(e(s)) = 1,$$

i.e.

$$(e \circ h)(X) \subseteq \{1\}, \quad (e \circ h)(s) = 1.$$

Since $X, s \models_{L_7} t$, it must hold that $(e \circ h)(t) = 1$, i.e. $h(e(t)) = 1$, i.e. $e(t) \in \{1/2, 1\}$. Again $e(s) \to_7 e(t) = 1$. Thus, $d^{(7)}$ is a deductive term in L_7.

The table of map $d^{(7)}_{L_7}$ is directly the table P_7. Since $d^{(7)}_{L_7}(1/2 \to_7 0) = 0$, the term $d^{(7)}$ cannot be classical. ∎

Theorem 17 *There is no deductive term in* L_2, L_3, L_4 *and* L_5.

Proof Let us conversely assume that $d(x, y)$ is a deductive term in L_k ($k = 2, \dots, 5$). Since $x \models_{L_k} x$, from the definition of deductive term it follows that $\models_{L_i} d(x, x)$. However, the last statement is not possible, since $\mathrm{Taut}(L_k) \cap \mathrm{Term}(\longrightarrow) = \emptyset$. ∎

Theorem 18 *Let us denote*

$$\begin{aligned} m^{(1)} &= (\mathbf{y} \to \mathbf{y}) \to (\mathbf{x} \to \mathbf{y}), \\ m^{(2)} &= ((\mathbf{x} \to \mathbf{x}) \to (\mathbf{y} \to \mathbf{y})) \to (\mathbf{x} \to \mathbf{y}), \\ m^{(6)} &= (\mathbf{y} \to (\mathbf{x} \to \mathbf{y})) \to (\mathbf{x} \to \mathbf{y}). \end{aligned}$$

Then $m^{(i)}$ *is a classical deductive term in* M_i, $i = 1, 2, 6$.

Proof The table of the values for functions $m_i = (m^{(i)})_{L_i}$, $i = 1, 2, 6$, is

	1	1/2	0
1	1	1	0
1/2	1	1	0
0	1	1	1

∎

The derivational version of the consequence operator

Definition 19 *Let M be an implicative matrix, T a set of Σ_{IP}-terms, $t_1, \ldots, t_n$ terms, $n \geq 1$. Let us call the sequence $t_1, \ldots, t_n$ an M-*proof from assumptions T *if for each k such that $1 \leq k \leq n$ one of the folowing three conditions holds:*

- $t_k \in \mathrm{Taut}(M)$,
- $t_k \in T$,
- *the term t_k has risen from some previous terms by means of the modus ponens rule (MP), i.e. there exist indices $1 \leq i, j < k$ such that $t_j = t_i \longrightarrow t_k$.*

If $t_1, \ldots, t_n$ is M-proof (from assumptions T), then number n is called the length of the proof $t_1, \ldots, t_n$.

If moreover t is a term, then M-proof of the term t from assumptions T *will be understood any M-proof $t_1, \ldots, t_n$ from assumptions T such that $t_n = t$.*

Let M be an implicative matrix, T a set of terms. Let us denote $\mathrm{Cn}^+_M(T)$ the set of all terms t such that there exists M-proof of term t from assumptions T. Let us denote Cn^+_M the map $\mathcal{P}(\mathrm{Term}(\Sigma_{\mathrm{IP}})) \rightarrow \mathcal{P}(\mathrm{Term}(\Sigma_{\mathrm{IP}}))$ defined in this way: $T \mapsto \mathrm{Cn}^+_M(T)$.

Theorem 20 *Map Cn^+_M is an algebraic closure operator on the set* $\mathrm{Term}(\Sigma_{\mathrm{IP}})$. *Moreover, $\mathrm{Cn}^+_M \leq \mathrm{Cn}_M$.*

Proof The fact that Cn^+_M is an algebraic closure operator on the set $\mathrm{Term}(\Sigma_{\mathrm{IP}})$ can be proved by standard way. Further, let $T \subseteq \mathrm{Term}(\Sigma_{\mathrm{IP}})$. We will prove that $\mathrm{Cn}^+_M(T) \subseteq \mathrm{Cn}_M(T)$. So assume that $t \in \mathrm{Cn}^+_M(T)$. Let $t_1, \ldots, t_n$ M-proof of the term t from assumptions T. We will prove that for every index $1 \leq k \leq n$ it holds that $t_k \in \mathrm{Cn}_M(T)$. If $k = 1$, then $t_k \in \mathrm{Taut}(M)$ or $t_k \in T$; in both cases it naturally holds that $t_k \in \mathsf{Cn}_M(T)$. Let $1 < k$ and for every index $1 \leq l < k$ we already know $t_l \in \mathrm{Cn}_M(T)$. According to Definition 19 for the term t_k at least one of the following possibilities will take place:

- $t_k \in \mathrm{Taut}(M)$ or $t_k \in T$; in both cases we have again $t_k \in \mathrm{Cn}_M(T)$;
- there exist indices $1 \leq i, j < k$ such that $t_j = t_i \to t_k$. By the induction assumption $t_i, t_j \in \mathrm{Cn}_M(T)$. Thus, $t_i, t_i \to t_k \in \mathrm{Cn}_M(T)$. By the rule (MP) it now also holds that $t_k \in \mathrm{Cn}_M(T)$.

Thereby we have shown, that for every index $1 \leq k \leq n$ it holds that $t_k \in \mathrm{Cn}_M(T)$. By choosing $k = n$ we get $t_n \in \mathrm{Cn}_M(T)$, i.e. $t \in \mathrm{Cn}_M(T)$. ∎

Operator Cn_M^+ is called the *derivational version* of operator Cn_M.

Theorem 21 *Assume that $d(\mathbf{x}, \mathbf{y})$ is a deductive term in* finite *implicative matrix M. Then $\mathrm{Cn}_M^+ = \mathrm{Cn}_M$ if and only if $\mathbf{y} \in \mathrm{Cn}_M^+(\mathbf{x}, d(\mathbf{x}, \mathbf{y}))$.*

Proof First, let $\mathrm{Cn}_M^+ = \mathrm{Cn}_M$. Since $d(\mathbf{x}, \mathbf{y})$ is a deductive term, it holds that $\mathbf{y} \in \mathrm{Cn}_M(\mathbf{x}, d(\mathbf{x}, \mathbf{y})) = \mathrm{Cn}_M^+(\mathbf{x}, d(\mathbf{x}, \mathbf{y}))$.

Conversely, let $\mathbf{y} \in \mathrm{Cn}_M^+(\mathbf{x}, d(\mathbf{x}, \mathbf{y}))$. Let us prove first the following claim. *Let X be a set of terms, s, t terms. Let $d(s, t) \in \mathrm{Cn}_M^+(X)$. Then $t \in \mathrm{Cn}_M^+(X, s)$.*

Proof. According to our assumption $t \in \mathrm{Cn}_M^+(s, d(s, t))$. Further it holds that $d(s, t) \in \mathrm{Cn}_M^+(X, s)$, $s \in \mathrm{Cn}_M^+(X, s)$. So $\{s, d(s, t)\} \subseteq \mathrm{Cn}_M^+(X, s)$, and thus $\mathrm{Cn}_M^+(s, d(s, t)) \subseteq \mathrm{Cn}_M^+(\mathrm{Cn}_M^+(X, s)) = \mathrm{Cn}_M^+(X, s)$. Since t is an element of $\mathrm{Cn}_M^+(s, d(s, t))$, must be valid also $t \in \mathrm{Cn}_M^+(X, s)$. So the claim is proved.

Let us come back to our statement. Let X be a set of terms. We will prove that $\mathrm{Cn}_M^+(X) = \mathrm{Cn}_M(X)$. For this it is sufficient to prove the inclusion $\mathrm{Cn}_M(X) \subseteq \mathrm{Cn}_M^+(X)$.

- If $X = \emptyset$, then $\mathrm{Cn}_M^+(X) = \mathrm{Cn}_M^+(\emptyset) = \mathrm{Taut}(M) = \mathrm{Cn}_M(\emptyset) = \mathrm{Cn}_M(X)$.
- Let $X \neq \emptyset$. Let $t \in \mathrm{Cn}_M(X)$. Since M is a finite matrix, the operator Cn_M is algebraic. Consequently, there exists a finite set $Y \subseteq X$ such that $t \in \mathrm{Cn}_M(Y)$. Let $Y = \{s_1, \ldots, s_n\}$. Thus $t \in \mathrm{Cn}_M(\{s_1, \ldots, s_n\})$, i.e. $\{s_1, \ldots, s_n\} \models_M t$. From deductivity of the term d it now follows that $\models_M d(s_1, d(\ldots d(s_n, t) \ldots))$, i.e. that $d(s_1, d(\ldots d(s_n, t) \ldots)) \in$

$\mathrm{Cn}_M(\emptyset) = \mathrm{Cn}^+_M(\emptyset)$. By the n-times use of the previous Claim we get $t \in \mathrm{Cn}^+_M(\{s_1, \ldots, s_n\}) \subseteq \mathrm{Cn}^+_M(X)$. ∎

Theorem 22 *For* $i \in \{7, 8, 9\}$ *the identity* $\mathrm{Cn}^+_{L_i} = \mathrm{Cn}_{L_i}$ *is valid.*

Proof The statement follows from the previous Theorem and from the shape of deductive terms $d^{(7)}$, $d^{(8)}$ and $d^{(9)}$. ∎

Theorem 23 *For* $i \in \{2, 3, 4, 5\}$ *it holds that* $\mathrm{Cn}^+_{L_i} < \mathrm{Cn}_{L_i}$.

Proof Since $\mathrm{Taut}(L_i) = \emptyset$, it holds that $\mathrm{Cn}^+_{L_i} = \mathrm{Cn}^+_{L_j}$, $i, j \in \{2, 3, 4, 5\}$. If for example $\mathrm{Cn}^+_{L_2} = \mathrm{Cn}_{L_2}$, then the identity $\mathrm{Cn}_{L_2} = \mathrm{Cn}^+_{L_3} \leq \mathrm{Cn}_{L_3}$ would be valid, and this is contradiction. ∎

The identity $\mathrm{Taut}(L_i) = \emptyset$ is substantial for the previous proof. How can we prove the inequality $\mathrm{Cn}^+_M < \mathrm{Cn}_M$ in a general case? For this purpose it is necessary to find a set of terms T and term s such that $s \notin \mathrm{Cn}^+_{L_i}(T)$ and $T \models_{L_i} s$ (in the case, when M is finite and $d(\mathbf{x}, \mathbf{y})$ is a deductive term in M, according to Theorem 20 it is sufficient to take $T = \{\mathbf{x}, d(\mathbf{x}, \mathbf{y})\}$, $s = \mathbf{y}$). The validity of $s \notin \mathrm{Cn}^+_{L_i}(T)$ will be proved such that we will find some invariant for the elements from the set $\mathrm{Cn}^+_M(T)$, which is not fulfilled by term s. This invariant will be founded on the following statement.

Corollary 24 *Let* G *be an implicative filter in the implicative matrix* $M = \langle P, F \rangle$. *Let* T *be a set of terms. Further, let* $e : \mathrm{V} \to M$ *be a valuation such that* $e(T \cup \mathrm{Taut}(M)) \subseteq G$. *If now* $s \in \mathrm{Cn}^+_M(T)$, *then also* $e(s) \in G$.

Proof Let us denote $X = \{\ t \in \mathrm{Term}(\Sigma_{\mathrm{IP}})\ ;\ e(t) \in G\ \}$. We will prove that for set X it holds that $\mathrm{Cn}^+_M(X) = X$. Naturally, it will be if and only if $\mathrm{Taut}(M) \subseteq X$ and set X is closed on the rule (MP). The inclusion $\mathrm{Taut}(M) \subseteq X$ follows immediately from our assumptions. Let now $s, s \to t \in X$. It means, $e(s) \in G$, $e(s \to t) \in G$. Thus $e(s) \to_M e(t) \in G$. Since G is an implicative filter, it has to be $e(t) \in G$. The equality $\mathrm{Cn}^+_M(X) = X$ is proved.

Let now $s \in \mathrm{Cn}^+_M(T)$. From $T \subseteq X$ we have $\mathrm{Cn}^+_M(T) \subseteq \mathrm{Cn}^+_M(X)$. It should also be true that $s \in \mathrm{Cn}^+_M(X) = X$, i.e. $e(s) \in G$. ∎

Corollary 25 *Let us assume that in finite implicative matrix* $M = \langle P, F\rangle$ *there exists a deductive term* $d(\mathbf{x},\mathbf{y})$*, an implicative filter* G *and a valuation* $e : \mathrm{V} \to M$ *such that*

$$e(\mathbf{x}) \in G, \quad e(d(\mathbf{x},\mathbf{y})) \in G, \quad e(\mathrm{Taut}(M)) \subseteq G, \tag{1}$$

$$e(\mathbf{y}) \notin G. \tag{2}$$

Then $\mathrm{Cn}^+_M < \mathrm{Cn}_M$.

Proof Put $T = \{\mathbf{x}, d(\mathbf{x},\mathbf{y})\}$. From (1) we have $e(T \cup \mathrm{Taut}(M)) \subseteq G$. From Corollary 24 it then follows that $\mathbf{y} \notin \mathrm{Cn}^+_M(\mathbf{x}, d(\mathbf{x},\mathbf{y}))$. This further means according to Theorem 20 that $\mathrm{Cn}^+_M < \mathrm{Cn}_M$. ∎

Let us remark that if the assumptions of the previous statement are fulfilled, then it must hold $F \neq G \subset P$ (where $G \subset P$ means $G \subseteq P$ and $G \neq P$). Namely, if $F = G$, then from the fact that $\mathbf{x}, d(\mathbf{x},\mathbf{y}) \models_M \mathbf{y}$, it should follow $e(\mathbf{y}) \in F = G$.

Let now $M = \langle P, F\rangle$ be three-element implicative matrix, where $F \subset P$. It should be either $F = \{1\}$, or $F = \{1/2, 1\}$. In the case $F = \{1\}$ (i.e. $M = L_i$) we can take into account for filter G only $G = \{1/2, 1\}$, in the case $F = \{1/2, 1\}$ (i.e. when $M = M_j$) we can take into account only $G = \{1\}$. Here

- $\{1/2, 1\}$ is an implicative filter in L_i iff $1/2 \to_{L_i} 0 = 0$,
- $\{1\}$ is an implicative filter in M_j iff $1 \to_{M_j} 0 < 1$.

Theorem 26 *For* $i \in \{1, 6\}$ *the following* $\mathrm{Cn}^+_{L_i} < \mathrm{Cn}_{L_i}$ *holds.*

Proof Let us set $G = \{1/2, 1\}$. Then G is an implicative filter in L_i, because in algebras L_1, L_6 holds the identity $1/2 \to 0 = 0$. Take the valuation $e : \mathrm{V} \to L_i$ such that $e(\mathbf{x}) = 1/2$ and $e(\mathbf{y}) = 0$. Then

$$\begin{aligned} e(d^{(1)}) &= e([((\mathbf{x}\to\mathbf{x}) \to (\mathbf{x}\to\mathbf{x})) \to \mathbf{x}] \to \mathbf{y}) = (1 \to_1 e(\mathbf{x})) \to_1 e(\mathbf{y}) = \\ &= (1 \to_1 1/2) \to_1 0 = 0 \to_1 0 = 1, \\ e(d^{(6)}) &= e(((\mathbf{x}\to\mathbf{x}) \to \mathbf{x}) \to \mathbf{y}) = ((1/2 \to_6 1/2) \to_6 1/2) \to_6 0 = \\ &= (1 \to_6 1/2) \to_6 0 = 0 \to_6 0 = 1. \end{aligned}$$

At the same time $e(\mathrm{Taut}(L_i)) = \{1\} \subseteq G$. Our statement follows now directly from Corollary 25. ∎

Theorem 27 *For* $i \in \{1, 6\}$ *the following* $\mathrm{Cn}^+_{M_i} < \mathrm{Cn}_{M_i}$ *holds.*

Proof Let us set $G = \{1\}$. Then G is an implicative filter M_i. Let us take the valuation $e : \mathrm{V} \to M_i$ such that $e(\mathbf{x}) = 1$, $e(\mathbf{y}) = 1/2$. Then

$$\begin{aligned} e(m^{(1)}) &= e((\mathbf{y} \to \mathbf{y}) \to (\mathbf{x} \to \mathbf{y})) = (1/2 \to_1 1/2) \to_1 (1 \to_1 1/2) = \\ &= 0 \to_1 0 = 1, \\ e(m^{(6)}) &= e((\mathbf{y} \to (\mathbf{x} \to \mathbf{y})) \to (\mathbf{x} \to \mathbf{y})) = \\ &= (1/2 \to_6 (1 \to_6 1/2)) \to_6 (1 \to_6 1/2) = \\ &= (1/2 \to_6 0) \to_6 0 = 0 \to_6 0 = 0. \end{aligned}$$

Our proposition follows now from Corollary 25. ■

Theorem 28 $\mathrm{Cn}^+_{M_2} < \mathrm{Cn}_{M_2}$.

Proof We will prove $\mathbf{y} \notin \mathrm{Cn}^+_{M_2}(\mathbf{x}, m^{(2)}(\mathbf{x}, \mathbf{y}))$. By contradiction. Let $\mathbf{y} \in \mathrm{Cn}^+_{M_2}(\mathbf{x}, m^{(2)}(\mathbf{x}, \mathbf{y}))$. Let $\varepsilon : \mathrm{V} \to M_2$ be a valuation such that $\varepsilon(\mathbf{x}) = 1$, $\varepsilon(\mathbf{y}) = 1/2$ and $\varepsilon(z) = 0$ for $z \in \mathrm{V} \setminus \{\mathbf{x}, \mathbf{y}\}$. Let further ξ be a valuation in M_2 identically equal to 0.

Claim 1 If $s \in \mathrm{Cn}^+_{M_2}(\mathbf{x}, m^{(2)}(\mathbf{x}, \mathbf{y}))$, then $\varepsilon(s) \geq 1/2$.
Proof. Let $G = \{1/2, 1\}$. Then G is an implicative filter in M_2. And the following holds

$$\begin{aligned} \varepsilon(m^{(2)}(\mathbf{x}, \mathbf{y})) &= \varepsilon(((\mathbf{x} \to \mathbf{x}) \to (\mathbf{y} \to \mathbf{y})) \to (\mathbf{x} \to \mathbf{y})) = \\ &= ((1 \to_2 1) \to_2 (1/2 \to_2 1/2)) \to_2 (1 \to 1/2) = \\ &= (1 \to_2 1/2) \to_2 0 = 0 \to_2 0 = 1. \end{aligned}$$

Thus, $\varepsilon(\mathbf{x}) \in G$, $\varepsilon(m^{(2)}(\mathbf{x}, \mathbf{y})) \in G$. From the Corollary 24 it also follows that $\varepsilon(s) \in G$, i.e. $\varepsilon(s) \geq 1/2$, qed.

Claim 2 If $\varepsilon(t) = 1/2$, then $\mathrm{var}(t) = \{\mathbf{y}\}$.
Proof. By induction of the length of term t. If $l(t) = 1$, then the term t must be a variable, and therefore $t = \mathbf{y}$ must hold. Let $l(t) = n \geq 2$ and for every term with the length $\leq n - 1$ the statement holds. Let $t = r \to s$. Then $\varepsilon(t) = \varepsilon(r) \to_2 \varepsilon(s)$. Thus, $\varepsilon(r) \to_2 \varepsilon(s) = 1/2$, and therefore must $\varepsilon(r) = \varepsilon(s) = 1/2$ hold. Now we can use the induction assumption.

Claim 3 For every term t holds $\xi(t) \in \{0, 1\}$.
Proof. Obvious.

Since $\mathbf{y} \in \mathrm{Cn}^+_{M_2}(\mathbf{x}, m^{(2)}(\mathbf{x}, \mathbf{y}))$, there exists M_2-proof $t_1, \ldots, t_n$ from assumptions $\{\mathbf{x}, m^{(2)}(\mathbf{x}, \mathbf{y})\}$ such that $\mathrm{var}(t_n) = \{\mathbf{y}\}$, $\xi(t_n) = 0$. Let

$t_1, \ldots, t_n$ be such proof of the shortest length. Following the definition of the proof then for term t_n some of the following possibilities must take place:
• $t_n \in \mathrm{Taut}(M_2)$. However this is excluded since $\xi(t_n) = 0$.
• $t_n = \mathbf{x}$ or $t_n = m^{(2)}(\mathbf{x}, \mathbf{y})$. But in both these cases $\mathbf{x} \in \mathrm{var}(t_n)$, and therefore this case also cannot take place; thus the third case must take place, i.e.
• term t_n rose from some previous terms by means (MP). Let these be terms

$$t_i,\ t_j = t_i \to t_n, \qquad (1 \le i, j < n).$$

From Claim 1 it follows that $\varepsilon(t_i) \ge 1/2$, $\varepsilon(t_j) \ge 1/2$. At the same time $\varepsilon(t_j) = \varepsilon(t_i) \to_2 \varepsilon(t_n) = \varepsilon(t_i) \to_2 1/2$. If $\varepsilon(t_i) = 1$, then the following should hold $\varepsilon(t_j) = 1 \to_2 1/2 = 0$, which is a contradiction. It must be therefore $\varepsilon(t_i) = 1/2$. Then also $\varepsilon(t_j) = 1/2$. From Claim 2 it follows that $\mathrm{var}(t_i) = \mathrm{var}(t_j) = \{\mathbf{y}\}$. Finally, from Claim 3 it follows that $\xi(t_i) \in \{0,1\}$, $\xi(t_j) \in \{0,1\}$. If $\xi(t_i) = \xi(t_j) = 1$, then the following should hold $1 = \xi(t_j) = \xi(t_i) \to_2 \xi(t_n) = 1 \to_2 0 = 0$, and it would be a contradiction. It must therefore be $\xi(t_i) = 0$ or $\xi(t_j) = 0$. In both cases we get contradiction with the minimality of the proof $t_1, \ldots, t_n$. ∎

Let us finally formulate one open question related to our investigation. Let $\mathcal{L}$ denote the class of all implicative posets L such that there exists a (classical) deduction term for the matrix $M_L = \langle L, \{1_L\} \rangle$. Is the class $\mathcal{L}$ axiomatizable?

References

[1] A. Avron. Relevant entailement: semantics and formal systems. *J. Symbolic Logic*, 49:334–342, 1984.

[2] A. Avron. On an implication connective of RM. *Notre Dame J. Formal Logic*, 27, 1986.

[3] S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra.* Springer, New York, 1873.

[4] C. C. Chang and H. J. Keisler. *Model Theory.* North-Holland, 1973.

[5] R. Meyer and Z. Parks. Independent axioms for the implicational fragment of Sobociński's three-valued logic. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, 18:291–295, 1972.

[6] Z. Parks. Dependence of some axioms of Rose. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, 18:189–192, 1972.

[7] Z. Parks. A note on R-mingle and Sobociński's three-valued logic. *Notre Dame J. Formal Logic*, 13:227–228, 1972.

[8] A. Rose. A formalization of Sobociński's three-valued implicational propositional calculus. *J. Comp. Syst.*, 1:165–168, 1953.

[9] A. Rose. An alternative formalization of Sobociński's three-valued implicational propositional calculus. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, 2:166–172, 1956.

[10] B. Sobociński. Axiomatization of a partial system of three-value calculus of propositions. *J. Comp. Syst.*, 1:23–55, 1952.

[11] R. Wójcicki. *Theory of Logical Calculi.* Kluwer, 1888.

Milan Matoušek, Petr Jirků
Philosophical Faculty, Charles University
Palachovo nám. 2, 116 38, Praha 1
milan.matousek@ff.cuni.cz, petr.jirku@ff.cuni.cz

Some Properties of Kripke Semantics for Intuitionistic Logic

Michal Peliš

Abstract Kripke semantics is the most popular possible-world semantics for many non-classical logics. The text contains a historical overview of semantical analysis for intuitionistic logic, but the very aim is an illustration of two important truth-preserving operations of Kripke semantics, generation and p-morphism, that allow us to narrow down classes of frames and models for intuitionistic logic. We will follow the way back to tree-frames and special classes of finite tree-frames, in particular Jaśkowski's frames.

Introduction

Semantics based on the notion of *possible world* (the Kripke semantics is one of them) are the latest semantics for the non-classical logics even if the word-connection *possible world* was appearred in Gottfried Leibniz's ideas and papers about propositions with necessity and possibility.[47]

The very beginning of intuitionistic logic was a bit complicated. Some principles of intuitionism were formulated by L. E. J. Brouwer (1907)[48] as a reaction to the crisis in mathematics in the end of 19th century. Brouwer refused to speak about any formal calculus because he did not ascribe any importance to logic as a "language of (intuitionistic)

[47] Gottfried Leibniz (1646–1716), a German philosopher and mathematician.
[48] When we use an exact year, we mean the date of publication.

mathematics". The first intuitionistic calculi were presented in the end of 1920s following the classical logic in the natural way (Kolmogorov, Orlov, Glivenko, Heyting) ([1], p. 56). Brouwer's continuator A. Heyting repeated the intuitionist's standpoint in his introductory sentence in [4], which was the first paper of the Dutch school about formal calculus of intuitionistic logic:

> Die intuitionistische Mathematik ist eine Denktätigkeit, und jede Sprache, auch die formalistische, ist für sie nur Hilfsmittel zur Mitteilung. Es ist prinzipiel unmöglich, ein System von Formeln aufzustellen, das mit der intuitionistischen Mathematik gleichwertig wäre, (...)

The beginning of 1930s is the period of the development of algebraical structures often called *valuation systems* or *matrices*. Especially the Polish school (Tarski, Łukasiewicz, Jaśkowski) and its interest in the many-valued logics produced works in this field ([2], p. 120). They were focused on the structures $\langle W, I, f_{\rightarrow}, f_{\vee}, f_{\wedge}, g_{\neg}\rangle$, where W is the set of values, $I \subset W$ is the set of chosen values, $f_{\rightarrow}, f_{\vee}, f_{\wedge}$ are binary function from W^2 to W and $g_{\neg}$ is a unary function from W to W. What type of structure would be suitable as a semantics for intuitionistic logic was sketched by Tarski at the study of the deduction systems. Some properties were known from Kurt Gödel's works as well. Those structures are usually called *Heyting algebras.*[49] Jaśkowski proved the completeness of the intuitionistic propositional calculus for a special class of finite Heyting algebras (1936). In the 1940s McKinsey and Tarski studied the topological interpretation of intuitionistic propositional logic, which was originally only a secondary result of M. H. Stone's work. Some authors understand in this step the way toward the possible-world semantics. They interpret elements of some space $\mathcal{S}$ as "worlds" and every formula is connected with a subset of elements in which is valid (cf. [2], p. 121).

Kripke semantics

Under the term "founder of possible-world semantics" we almost always understand two names: S. A. Kripke and his predecessor E. W. Beth.

[49] The oldest name is probably *Brouwerian algebras* and the later one *pseudo--Boolean algebras.*

Beth's *semantic tableaux* from 1950s were inspired by Gerhard Gentzen's (1934) and S. C. Kleene's calculus (see [9]).

Kripke published his semantical analysis for intuitionistic logic in 1965.[50] The main inspiration was in Beth's semantic tableaux, his tree-semantics and the well known relation between intuitionistic logic and modal logic S4.

Kripke speaks about the motivation for introducing possible-world semantics in [6] (p. 19, footnote 18):

> The main and the original motivation for the "possible worlds analysis"—and the way it clarified modal logic—was that it enabled modal logic to be treated by the same set theoretic techniques of model theory that proved so successful when applied to extensional logic.

The original Kripke's semantics is the tree-type semantics but usually used definition is more general (as it is done in the following Definitions 1 and 2). Let us call this definitions *Kripke semantics*, in a usual way. It is possible to see "didactic" advantages, Kripke semantics is taught in courses of non-classical logics and we meet it in most (new) text-books. We see the advantages in its "philosophical backround" and applicability to many modal and intermediate logics.

In the following text we will only speak about propositional logic. In particular, the term *formula* is an abbreviation for a formula in the propositional language. We will not introduce axioms of intuitionistic propositional calculus (IPC), only the term *intuitionistic tautology* will be used (see the Definition 3).

Definition 1 (Kripke frame) *An* (intuitionistic) Kripke frame *is a pair* $\mathbf{K} = \langle W, R \rangle$ *consisting of a nonempty set (of possible worlds)* W *and a partial order* R *on* W *(i.e.* R *satisfies three condidtions: reflexivity, transitivity and antisymmetry).*

The elements of W are (informally) called *possible worlds* (or only *worlds*) and the relation R *relation of accessibility.* If xRy for $x, y \in W$ then we say that "y is accessible from x" or "x sees y".

[50] See [5]. His semantics for modal logic S5 was published already in 1959. Kripke's thoughts connected with the topic of possible worlds originate from the beginning of 1960s (see [6], p. 3).

Definition 2 (Kripke model) *An* (intuitionistic) Kripke model *is a pair* $\mathbf{M} = \langle \mathbf{K}, \Vdash \rangle$ *where* $\mathbf{K}$ *is a Kripke frame and* $\Vdash$ *is a relation (between elements of* W *and formulas) which satisfies the following conditions:*

- *If* p *is an atomic formula,* $x \Vdash p$*, and* xRy *then* $y \Vdash p$*,*
- $x \Vdash \varphi \wedge \psi \Leftrightarrow x \Vdash \varphi$ *and* $x \Vdash \psi$*,*
- $x \Vdash \varphi \vee \psi \Leftrightarrow x \Vdash \varphi$ *or* $x \Vdash \psi$*,*
- $x \Vdash \varphi \rightarrow \psi \Leftrightarrow \forall y \in W($*if* xRy *and* $y \Vdash \varphi$ *then* $y \Vdash \psi)$*,*
- $x \Vdash \neg\varphi \Leftrightarrow \forall y \in W($*if* xRy *then* $y \nVdash \varphi)$*.*

We often shorten the words *Kripke frame* and *Kripke model* to *frame* and *model*.

A formula φ is *satisfied* in a world $x \in W$ of a model $\mathbf{M}$ if $x \Vdash \varphi$. If a formula φ is satisfied in all $x \in W$, i.e. if $\{ x \in W \,;\, x \Vdash \varphi \} = W$, then we say that φ is *valid in the model* $\mathbf{M}$ and write $\mathbf{M} \models \varphi$.

If a formula φ is valid in every model $\mathbf{M}$ based on a frame $\mathbf{K}$ then φ is *valid in the frame* $\mathbf{K}$ (notation: $\mathbf{K} \models \varphi$). Thus $\mathbf{K} \not\models \varphi$ if there is a model $\mathbf{M} = \langle \mathbf{K}, \Vdash \rangle$ (a *counter-example*) such that $\mathbf{M} \not\models \varphi$.

Definition 3 *A formula valid in every Kripke frame is an* intuitionistic tautology.

Truth-preserving operations

The following text is focused on some properties of Kripke frames and models. One of the reasons why the Kripke semantics is so popular (not only as a semantics for intuitionistic logic) could be in its properties we are going to study. Two operations will be introduced which allow us to narrow down the classes of frames and models for intuitionistic propositional logic. It is a well known fact that it is sufficient to investigate only finite Kripke models for the decision whether a formula is a (propositional) intuitionistic tautology (Finite Model Property, FMP), but our main point is in the type of structure of frames and models. The last step will be at Jaśkowski's result in the language of Kripke semantics.

Generation

Let us have two Kripke frames $\mathbf{F} = \langle V, S\rangle$ and $\mathbf{G} = \langle W, R\rangle$. The frame $\mathbf{F}$ is a *subframe* of the frame $\mathbf{G}$ if $V \subseteq W$ and $S = R \cap V^2$. In case the set V is R-closed (i.e. V forms upward closed subset of W with respect to the relation R),[51] the frame $\mathbf{F}$ is called *generated subframe* of $\mathbf{G}$ (notation: $\mathbf{F} \sqsubseteq \mathbf{G}$).

For every two frames $\mathbf{F}$ and $\mathbf{G}$, where $\mathbf{F} \sqsubseteq \mathbf{G}$, and every formula φ,

$$\mathbf{G} \models \varphi \Rightarrow \mathbf{F} \models \varphi. \tag{8}$$

It is clear that the implication from right to left does not hold.

Similar notions could be introduced for models. Let $\mathbf{M} = \langle \mathbf{F}, \Vdash\rangle$ and $\mathbf{N} = \langle \mathbf{G}, \Vdash_{\mathbf{N}}\rangle$ be Kripke models. The model $\mathbf{M}$ is a *submodel* of the model $\mathbf{N}$ (in symbols: $\mathbf{M} \subseteq \mathbf{N}$), if $\mathbf{F} \subseteq \mathbf{G}$ and the relation $\Vdash$ is a restriction of the relation $\Vdash_{\mathbf{N}}$ to the set V. If $\mathbf{F} \sqsubseteq \mathbf{G}$ then the model $\mathbf{M}$ is called *generated submodel* (in symbols: $\mathbf{M} \sqsubseteq \mathbf{N}$). It is useful to mention that for every formula φ and every element $x \in V$,

$$x \Vdash \varphi \Leftrightarrow x \Vdash_{\mathbf{N}} \varphi. \tag{9}$$

According to Definition 3, all formulas valid in every Kripke frame are intuitionistic tautologies. In the following Theorem 4 this class will be restricted to only *rooted frames.* A frame $\mathbf{F} = \langle V, S\rangle$ is rooted if there is an element (*root*) $x \in V$ such that $\forall y \in V(xRy$ and $(yRx \Rightarrow y = x))$.

Theorem 4 *A formula is an intuitionistic tautology iff it is valid in all rooted frames.*

To verify Theorem 4 it suffices to prove the following lemma.

Lemma 5 *For every frame $\mathbf{G} = \langle W, R\rangle$ and every formula φ, $\mathbf{G} \models \varphi$ iff for every rooted subframe $\mathbf{F} \sqsubseteq \mathbf{G}$ it is the case that $\mathbf{F} \models \varphi$.*

Proof Thanks to (8) only the implication from right to left is to be proved. Let $\mathbf{G} \not\models \varphi$, then there is a model $\mathbf{N} = \langle \mathbf{G}, \Vdash_{\mathbf{N}}\rangle \not\models \varphi$, i.e. there is an element $x \in W$ satisfying $x \nVdash_{\mathbf{N}} \varphi$. Let us take the set of all elements $y \in W$ such that xRy (notation: $x\!\uparrow$).[52] We have just gained a rooted subframe $\mathbf{F} = \langle x\!\uparrow, S\rangle \sqsubseteq \mathbf{G}$ and a model $\mathbf{M} = \langle \mathbf{F}, \Vdash\rangle$ where $\mathbf{M} \not\models \varphi$ (thanks to (9)), whence $\mathbf{F} \not\models \varphi$. ∎

[51] In symbols: $\forall x, y \in W((xRy \wedge x \in V) \Rightarrow y \in V)$.

[52] It is an R-closed set generated by the one-point set $\{x\}$.

P-morphism

Let $\mathbf{F} = \langle V, S\rangle$ and $\mathbf{G} = \langle W, R\rangle$ be Kripke frames and f a map from V onto W. The map f is called *p-morphism*[53] iff for every $x, y \in V$:

1. $xSy \Rightarrow f(x)Rf(y)$,
2. $f(x)Rf(y) \Rightarrow \exists z \in V(xSz \text{ and } f(z) = f(y))$.

The first condition means that a map f preserves "accessibility", f is monotonous. The second one prohibits a change in a validity of formulas in frames. Only a change in the structure of frames is possible, but not in valid formulas. It is easy to verify a preservation of closed sets concerning a relation of accessibility (in particular, a map of an S-closed set is an R-closed set) and a composition of two p-morphisms is again a p-morphism.

Let us have a p-morphism f of a frame $\mathbf{F}$ to $\mathbf{G}$ and models $\mathbf{M} = \langle \mathbf{F}, \Vdash\rangle$ and $\mathbf{N} = \langle \mathbf{G}, \Vdash_{\mathbf{N}}\rangle$. The map f is a p-morphism of $\mathbf{M}$ to $\mathbf{N}$ if for every atomic formula p and every $x \in V$,

$$x \Vdash p \Leftrightarrow f(x) \Vdash_{\mathbf{N}} p.$$

If f is a p-morphism of $\mathbf{M}$ to $\mathbf{N}$, it is easy to prove by induction on the construction of φ that

$$x \Vdash \varphi \Leftrightarrow f(x) \Vdash_{\mathbf{N}} \varphi.$$

Theorem 4 restricted the class of all intuitionistic models to rooted frames. Now the aim is to continue on that restriction to frames which are *trees*. Tree frames are rooted and the set of all predecessors of any possible world is finite and linearly ordered by the relation of accessibility. We will show that for every rooted frame $\mathbf{G}$ there is a tree frame $\mathbf{F}$ and p-morphism

$$\mathbf{F} \xrightarrow{f} \mathbf{G}.$$

The situation would be following:

$$\text{TREE FRAME} \longrightarrow \text{ROOTED FRAME.}$$

The first step is to construct a tree frame and a p-morphism for any rooted frame.

[53] Or *pseudo-epimorphism* and *reduction* as well.

Let $\mathbf{G} = \langle W, R\rangle$ be a rooted frame and w_0 its root. Define all finite chains $\langle w_0, \ldots, w_n\rangle$ such that $w_0 R \ldots R w_n$. There are only chains whose left elements see right elements. Let $\mathbf{F} = \langle V, S\rangle$ where V is the set of all such chains and the relation of accessibility S is defined by the following way: $\langle w_0, \ldots, w_n\rangle S \langle v_0, \ldots, v_m\rangle$ iff $n \leq m$ and $w_i = v_i$ (for $i = 0, \ldots, n$). Clearly S is a partial order on V, $\langle w_0\rangle$ is a root of $\mathbf{F}$ and this frame is a tree.

In case a map f is defined such that $f(\langle w_0, \ldots, w_n\rangle) = w_n$, we get a p-morphism.

The second and last step is about validity of formulas. Suppose there is a p-morphism of $\mathbf{F}$ to $\mathbf{G}$. It is easy to verify

$$\mathbf{F} \models \varphi \Rightarrow \mathbf{G} \models \varphi.$$

If $\mathbf{G} \not\models \varphi$, i.e. if there is a counter-example for a formula φ among models on frame $\mathbf{G}$, then $\mathbf{F} \not\models \varphi$.

As a conlusion of these steps we have Theorem 6.

Theorem 6 *A formula is intuitionistic tautology iff it is valid in all tree frames.*

Sure, nothing surprising for anyone who is familiar with the original Kripke's tree-type semantics. We can do one step more and think about restrictions to special classes of trees.

Algebraic semantics (based on Heyting algebras) and Kripke semantics for intuitionistic propositional logics are equivalent in a special sense. It is possible to transform Kripke semantics to algebraic and conversely (see [3] and [7]). We mentioned Jaśkowski's result about completeness of IPC for a special class of Heyting algebras (see the Introduction above). If this class of Heyting algebras is transformed to a Kripke-type semantics, we get the following class of tree frames.

$\mathbf{J_1}$ is a tree with one node, and $\mathbf{J_{n+1}}$ is the result of adding a root to the disjoint union of n copies of $\mathbf{J_n}$ (see Figure 1). Let us call this class of tree frames *Jaśkowski's frames*.

In the connection with the notion of p-morphism it is useful to state the following lemma.

Lemma 7 *For every two positive integers k and j, where $k \geq j$, there is a p-morphism of $\mathbf{J_k}$ to $\mathbf{J_j}$.*

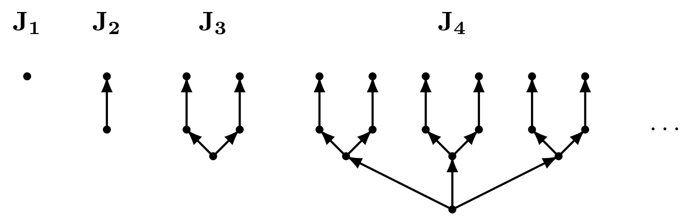


Figure 1: Jaśkowski's frames

For proving this lemma it is necessary and easy to show that there is a p-morphism of $\mathbf{J_k}$ to $\mathbf{J_{k-1}}$, for each positive integer k (it is done by induction on the number of nodes) and that a composition of p-morphisms is a p-morphism.

Our aim is to prove Theorem 8 which is Jaśkowski's result in the language of the possible-world semantics.

Theorem 8 *A formula is an intuitionistic tautology iff it is valid in all Jaśkowski's frames.*

The class of Jaśkowski's frames contains only finite frames. If there is a formula φ, which is not an intuitionistic tautology, it is known (thanks to FMP and Theorem 6) that there is a finite tree frame $\mathbf{T}$ and $\mathbf{T} \not\models \varphi$. We obtain a proof of Theorem 8 if there will be $\mathbf{J_n}$ such that $\mathbf{J_n} \not\models \varphi$ and the following lemma says that it is possible to find an appropriate frame from the class of Jaśkowski's frames to every finite tree frame and their relation given by a p-morphism.

Lemma 9 *For every finite tree frame* $\mathbf{T}$ *there exists an* n *and a p-morphism of* $\mathbf{J_n}$ *to* $\mathbf{T}$.

Proof By induction on the number of nodes in $\mathbf{T}$.

Evidently there is a p-morphism of $\mathbf{J_1}$ to $\mathbf{T_1} = \langle W, R \rangle$ where $|W| = 1$. (Similar situation is with a two-nodes tree $\mathbf{T_2}$.)

Let us suppose that there is a Jaśkowski's frame $\mathbf{J_n}$ and a p-morphism f_k of $\mathbf{J_n}$ to $\mathbf{T_k}$ for $k \geq 1$. We want to prove that there is a $\mathbf{J_j}$ and a p-morphism f_{k+1} of $\mathbf{J_j}$ to $\mathbf{T_{k+1}}$ as well. Figure 2 shows the shape of the

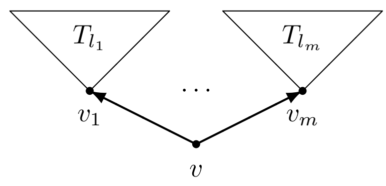


Figure 2: To the proof of Lemma 9

frame $\mathbf{T_{k+1}}$. The induction hypothesis says that there is a p-morphism of some Jaśkowski's frame $\mathbf{J_{n_i}}$ to $\mathbf{T_{l_i}}$ for each $i \in \{1, \dots, m\}$. Put $n_{\max}$ for the maximum from the set $\{n_1, \dots, n_m\}$. Then Lemma 7 justifies the existence of p-morphisms of $\mathbf{J_{n_{\max}}}$ to each $\mathbf{T_{l_i}}$. There are two cases left:

- ◦ $n_{\max} \geq m$, easy to find a p-morphism f_{k+1} of $\mathbf{J_{n_{\max}+1}}$ to $\mathbf{T_{k+1}}$.

  If $n_{\max} = m$, both trees $\mathbf{J_{n_{\max}+1}}$ and $\mathbf{T_{k+1}}$ have the same number of branches going out from their roots and f_{k+1} is an extension of f_k in the following sense: the root of $\mathbf{J_{n_{\max}+1}}$ is maped onto v (Figure 2) and then it is only a use of f_k.

  If the $\mathbf{J_{n_{\max}+1}}$ has more branches going out from the root than $\mathbf{T_{k+1}}$, i.e. if $n_{\max} > m$, then the situation is similar to the previous extention of f_k but some subtrees of $\mathbf{J_{n_{\max}+1}}$ will be maped onto one subtree of $\mathbf{T_{k+1}}$.

- ◦ $n_{\max} < m$, by using Lemma 7 we get a p-morphism of $\mathbf{J_m}$ to $\mathbf{T_{l_i}}$ for each $i \in \{1, \dots, m\}$ and a p-morphism f_{k+1} is defined from $\mathbf{J_{m+1}}$ to $\mathbf{T_{k+1}}$ similar to the first case in the previous item.

It is easy to verify that all f_{k+1}'s defined above fulfil both conditions from the definiton of a p-morphism. ∎

In [8] (p. 351) there is a more general version of Theorem 8. Suppose there is a class $[\mathbf{T_n}]$ of finite frames and for every finite tree $\mathbf{T}$ there is an element $\mathbf{T_n}$ of this class and p-morphism f_n of $\mathbf{T_n}$ to $\mathbf{T}$. A formula is an intuitionistic tautology iff it is valid in all tree frames from the class $[\mathbf{T_n}]$.

"The road we took"

It is nice and didacticaly pleasant to find out that Kripke semantics can show its historical stages by some operations based on its general definition. On the "road" we met some results about intuitionistic logic often known from other and older semantical analyses and we introduced two truth-preserving operations which enable us to study some properties of Kripke semantics and all logics they use this type of semantics (modal and intermediate logics). More about truth-preserving operations could be found in the book [1].

The overview of a semantical analysis history of intuitionistic logic is in [7]. There are some relations among semantics, some of their properties and a good list of references.

References

[1] A. Chagrov and M. Zakharyaschev. *Modal Logic.* Clarendon Press, Oxford, 1997.

[2] M. A. E. Dummett. *Elements of Intuitionism.* Clarendon Press, Oxford, 2000. Second Edition.

[3] M. C. Fitting. *Intuitionistic Logic, Model Theory and Forcing.* North-Holland, 1969.

[4] A. Heyting. Die formalen Regeln der intuitionistischen Logik. In *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*, Math. Kl., pages 42–56. Preussische Akademie der Wissenschaften, Berlin, 1930.

[5] S. A. Kripke. Semantical analysis of intuitionistic logic I. In *Formal Systems and Recursive Functions*, pages 92–130. North-Holland, Amsterdam, 1965.

[6] S. A. Kripke. *Naming and Necessity.* Harvard University Press, 1999. Eleventh Edition.

[7] M. Peliš. Počátky sémantiky intuicionistické logiky (Origins of Intuitionistic Semantics). Master thesis, Philosophical Faculty of Charles University, Department of Logic, 2001.

[8] A. S. Troelstra, editor. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis.* Springer, 1973.

[9] A. S. Troelstra and P. van Ulsen. The Discovery of E. W. Beth's Semantics for Intuitionistic Logic. Online http://www.illc.uva.nl/j50//contribs/troelstra.

Michal Peliš
Philosophical Faculty, Charles University
Palachovo nám. 2, 116 38, Praha 1
michal.pelis@ff.cuni.cz

On Structural Proofs of Gödel First Incompleteness Theorem

Vítězslav Švejdar

Abstract Besides the classical proof, Gödel First Incompleteness Theorem also has a proof which does not use self-reference and which I propose to call structural. We review the classical proof and present a version of a structural one. Then we discuss the role of Σ_1-completeness.

Introduction

The usual formulation of Gödel First Incompleteness Theorem says that *if T is a Σ_1-sound recursively axiomatized extension of Robinson arithmetic* Q *then T is incomplete.* Σ_1-soundness means that all Σ_1-sentences provable in T are valid in the structure $\mathbf{N}$ of natural numbers. The theory Q plays the role of a weak base theory: the weaker the base theory, the larger the class of theories T to which the theorem is applicable. The theorem is applicable to Peano arithmetic PA, and to all theories resulting by adding finite (or recursive) set of true (i.e. valid in $\mathbf{N}$) additional axioms to the set of axioms of Peano arithmetic.

There is a stronger version of Gödel First Incompleteness Theorem saying that under the same assumptions, the theory T is incomplete *and undecidable.* Yet stronger versions are more specific about both incompleteness and undecidability: the theory is incomplete to the extent that there are independent sentences which are as simple as Σ_1 or Π_1, and it is undecidable to the extent that the set $\mathrm{Thm}(T)$ of all sentences provable in T is RE-complete, i.e. creative. A very important Rosser version of Gödel First Incompleteness Theorem replaces the assumption of Σ_1-soundness by mere consistency.

Traditional ingredients of the proof of Gödel First Incompleteness Theorem are self-reference, definability in $\mathbf{N}$, and Σ_1-completeness of Q, together with some knowledge of properties of logical calculi. We outline the traditional proof below. By a *structural proof* I mean one which uses an argument of the following form: if two sets of sentences, e.g. $\mathrm{Thm}(T)$ and $\mathrm{Th}(\mathbf{N})$ (= the set of all arithmetical sentences valid in $\mathbf{N}$) are of different arithmetical classification, then they cannot be the same; so there exists a sentence in $\mathrm{Th}(\mathbf{N}) - \mathrm{Thm}(T)$, i.e. a sentence which is valid in $\mathbf{N}$ but unprovable in T. So a structural proof does not use self-reference, it uses some knowledge from the theory of recursive functions instead. Structural proofs appear e.g. in [4]. I show my version of a structural proof, which avoids even the use of Σ_1-completeness. It will give me an opportunity to make some remarks about the precise role of Σ_1-completeness.

Peano arithmetic, arithmetization

Arithmetical language is the language $L_0 = \{+, \cdot, 0, \mathrm{S}, \leq, <\}$ having two binary function symbols, one constant, one unary function symbol and two binary predicates. *Structure of natural numbers*, or *standard model of arithmetic*, is the structure $\mathbf{N} = \langle \mathrm{N}, +^{\mathbf{N}}, \cdot^{\mathbf{N}}, 0^{\mathbf{N}}, \mathrm{s}, \leq^{\mathbf{N}}, <^{\mathbf{N}} \rangle$, where s is the function $x \mapsto x + 1$, $0^{\mathbf{N}}$ is the number zero, $+^{\mathbf{N}}$ and $\cdot^{\mathbf{N}}$ are the usual addition and multiplication, and $\leq^{\mathbf{N}}$ and $<^{\mathbf{N}}$ are the usual unstrict and strict ordering on the set $\mathrm{N} = \{0, 1, 2, \dots\}$ of all natural numbers. In the sequel we omit the superscript $\mathbf{N}$; so e.g. $+$ is both a symbol of the arithmetical language and its "standard" realization, i.e. addition of natural numbers. An example of an arithmetical sentence is

$$\forall u \forall v (u \cdot v = \mathrm{S}(\mathrm{S}(\mathrm{S}(0))) \rightarrow u = \mathrm{S}(0) \vee v = \mathrm{S}(0)) \qquad (1)$$

which can be read $\mathsf{three\ is\ prime}$. The formula

$$\mathrm{S}(0) < x \mathbin{\&} \forall u \forall v (u \cdot v = x \rightarrow u = \mathrm{S}(0) \vee v = \mathrm{S}(0)) \qquad (2)$$

has one free variable x and can be read $\mathsf{the\ number}$ x $\mathsf{is\ prime}$. To avoid complicated symbolic expressions I use an informal reading of formulas and sentences typeset in sans serif font. So $\mathsf{three\ is\ prime}$ and (1) are identical sentences. The sentence (1) is *valid* in the structure **N**; in symbols, $\mathbf{N} \models \mathsf{three\ is\ prime}$. We write $\overline{3}$ and $\overline{1}$ instead of the terms S(S(S(0))) and S(0). In general, $\overline{n}$ is the term $\mathrm{S}(\mathrm{S}(\,.\,.\,\mathrm{S}(0)\,.\,.\,)$ with n occurences of the symbol S. These terms are called *numerals*. Numeral $\overline{n}$ represents the (metamathematical) number n in the arithmetical language, i.e. inside some formal theory. As an excercise the reader may try to design an arithmetical formula $\mathsf{Pow}(x)$ saying that the number x is a power of two, i.e. saying that the number x is one of the numbers $\overline{1}$, $\overline{2}$, $\overline{4}$, $\overline{8}$, . . . The point is that a symbol for exponentiation cannot be used in the construction of the formula $\mathsf{Pow}(x)$; we have no such symbol in the arithmetical language. Let me remark that the example with three being a prime and some other examples and notation are borrowed from [5].

Robinson arithmetic Q is formulated in the arithmetical language L_0 defined above, and it has seven axioms, e.g. $\forall x(x + 0 = x)$, stating some simple properties of natural numbers. Robinson arithmetic was defined in [7]. In some sources it has eight or nine axioms, see [1] or [6]. An example of a sentence provable in Q is $\overline{3} + \overline{2} = \overline{5}$, while an example of a sentence unprovable in Q is $\forall x \forall y(x + y = y + x)$. In general, Q is a weak theory: universal sentences are more likely to be unprovable than provable in Q. *Peano arithmetic* PA has the same language L_0 as Robinson arithmetic. Its axioms are the axioms of Q plus the *induction scheme*: if $\varphi(x, y_1, \ldots, y_n)$ is an arithmetical formula then the sentence

$$\forall y_1 \ldots \forall y_n(\varphi(0, \underline{y}) \,\&\, \forall x(\varphi(x, \underline{y}) \to \varphi(\mathrm{S}(x), \underline{y})) \to \forall x \varphi(x, \underline{y})), \qquad (3)$$

where $\underline{y}$ is an abbreviation for $y_1, \ldots, y_n$, is an axiom of PA. An example of a sentence provable in PA is for each z, there exists a prime greater than z, which can also be written as there are infinitely many primes.

If T is a theory then $\mathrm{Thm}(T)$ denotes the set of all sentences provable in T, i.e. of all sentences φ such that $T \vdash \varphi$. A theory T with a language $L \supseteq L_0$, where L_0 is as above, is *sound* if every arithmetical sentence provable in T is valid in **N**. A theory T with a language $L \supseteq L_0$ is an *extension of Robinson arithmetic* Q if every arithmetical sentence provable in Q is provable also in T. Peano arithmetic is one of the many extensions of Q, and it is a sound extension.

Formulas, sentences, or even proofs are finite sequences of symbols. Since the nature of symbols is irrelevant, these syntactical objects can be identified with finite sequences of natural numbers. Any finite sequence of natural numbers can be coded by a single natural number.

Thus formulas and other syntactical objects can be identified with natural numbers: they *are* natural numbers, and numerals like $\overline{0 = \mathrm{S}(0)}$ or $\overline{\textsf{three is prime}}$ are possible (we keep hiding arithmetical formulas under their informal reading typeset in sans serif font). Thus the sentence (1) above equals some natural number n (i.e. has numerical code n), and $\overline{\textsf{three is prime}}$ is a numeral containing exactly n occurrences of the symbol S.

An important property of all extensions T of Q is the *self-reference theorem*: for each formula $\psi(x)$ in the language of T there exists an arithmetical sentence φ such that the equivalence $\varphi \equiv \psi(\overline{\varphi})$ is provable in T. Put otherwise, the self-reference theorem says that an equation of the form $T \vdash \varphi \equiv \psi(\overline{\varphi})$, for an unknown arithmetical sentence φ, has a solution for every formula $\psi(x)$ in the language of the theory T. If $T \vdash \varphi \equiv \psi(\overline{\varphi})$ then, inside T, we know that φ is equivalent to the statement the sentence $\overline{\varphi}$ has the property ψ. So the solution φ of the equation $T \vdash \varphi \equiv \psi(\overline{\varphi})$ can be viewed as a sentence saying I have the property ψ.

A formula $\theta(x)$ *defines* a set $A \subseteq \mathrm{N}$ in $\mathbf{N}$ if $\forall n(n \in A \Leftrightarrow \mathbf{N} \models \theta(\overline{n}))$. A set $A \subseteq \mathrm{N}$ is *definable* in $\mathbf{N}$ if some formula $\theta(x)$ defines it. Probably the most difficult part of the proof of Gödel First Incompleteness Theorem is the fact that if T is recursively axiomatizable then the set $\mathrm{Thm}(T)$ of all (numerical codes of) its theorems is definable in $\mathbf{N}$. Thus we can fix an arithmetical formula $\mathsf{Pr}_T(x)$ that defines it. The fact that the formula $\mathsf{Pr}_T(x)$ defines the set $\mathrm{Thm}(T)$ is expressed by the following condition

$$T \vdash \varphi \Leftrightarrow \mathbf{N} \models \mathsf{Pr}_T(\overline{\varphi}), \tag{Def}$$

where φ is any sentence in the language of T. The formula $\mathsf{Pr}_T(x)$ can be read the sentence x is provable in T. The formula $\mathsf{Pr}_T(x)$ can be chosen so that, besides the condition Def, it has other useful properties, namely those expressed by the so called *derivability conditions* D1–D3 ([3], see also [1] or [6]). Out of these conditions we only need the first one:

$$T \vdash \varphi \Rightarrow \mathsf{Q} \vdash \mathsf{Pr}_T(\overline{\varphi}). \tag{D1}$$

A straightforward use of the condition Def shows that the converse implication in D1 is also true. An example on the use of the condition D1 is this: the sentence $\mathsf{Pr}_{\mathsf{PA}}(\overline{\textsf{there are infinitely many primes}})$ is provable in Q. Thus Q does not know that there are infinitely many primes, but it knows that PA knows this fact. With the self-reference theorem and conditions Def and D1 at hand, we are ready to outline the proof of Gödel First Incompleteness Theorem, in the version that *if T is a sound recursively axiomatized extension of* Q *then T is incomplete.* So let T be a sound recursively axiomatized extension of Q. By the self-reference theorem there exists an arithmetical sentence ν satisfying the equation $T \vdash \nu \equiv \neg \mathsf{Pr}_T(\overline{\nu})$. Assume that $T \vdash \nu$. Then from ν and $\nu \equiv \neg \mathsf{Pr}_T(\overline{\nu})$ we can, inside T, infer $\neg \mathsf{Pr}_T(\overline{\nu})$. On the other hand, the assumption that $T \vdash \nu$, the condition D1 and the fact that T is an extension of Q yield $T \vdash \mathsf{Pr}_T(\overline{\nu})$. So T is contradictory, which is a contradiction with the assumption that T is sound. So one half of the conclusion of Gödel First Incompleteness Theorem, saying that $T \nvdash \nu$, is proved. This fact and the condition Def yield $\mathbf{N} \not\models \mathsf{Pr}_T(\overline{\nu})$. From $\mathbf{N} \models \nu \equiv \neg \mathsf{Pr}_T(\overline{\nu})$ we have $\mathbf{N} \models \nu$ and $\mathbf{N} \not\models \neg\nu$. A consequence of soundness is $T \nvdash \neg\nu$.

A structural proof

Bounded quantifiers are quantifiers of the form $\forall v {\le} x$, $\exists v {\le} x$, $\forall v {<} x$, and $\exists v {<} x$. The meaning of bounded quantifiers is obvious, e.g. $\forall v {\le} x \varphi$ is a shorthand for $\forall v(v \le x \rightarrow \varphi)$. A *bounded formula*, or a Δ_0*-formula*, is a formula all quantifiers of which are bounded. A simple example of a bounded formula is $\overline{3} + \overline{2} = \overline{5}$. The formula (2) above expressing that the number x is prime is not bounded as it stands; it is however PA-equivalent to a bounded formula because the quantifiers $\forall u$ and $\forall v$ can be equivalently written as $\forall u {\le} x$ and $\forall v {\le} x$. A Σ_1*-formula* is a formula of the form $\exists v \psi$ where ψ is a bounded formula, whereas a Π_1*-formula* is a formula of the form $\forall v \psi$ where again ψ is a bounded formula. Thus a negation of a Π_1-formula is T-equivalent to a Σ_1-formula and vice versa, and this fact is true regardless of the theory T. Two extremely important properties of Σ_1-formulas are these: (i) a set $A \subseteq \mathrm{N}$ is recursively enumerable if and only if it is definable in $\mathbf{N}$ by a Σ_1-formula, and (ii) any Σ_1-sentence valid in $\mathbf{N}$ is provable in Q. The property (i) can be called *definability theorem* while (ii) is the Σ_1*-completeness theorem.* Both theorems are nontrivial and their proofs can be found in various sources including [1], [2], and [6].

We know that the set $\mathrm{Thm}(T)$ of a recursively axiomatized theory T is recursively enumerable. So the existence of a formula $\mathsf{Pr}_T(x) \in \Sigma_1$

satisfying the condition Def directly follows from the definability theorem. Then Σ_1-completeness implies that any such formula satisfies the condition D1. Thus the existence of a provability predicate $\mathsf{Pr}_T(x)$ satisfying the conditions Def and D1 directly follows from the definability and Σ_1-completeness theorems. My opinion is that the proofs of these two theorems are not more difficult than the direct construction of the provability predicate and proof of its properties.

A theory T with a language $L \supseteq L_0$ is Σ_1-*sound* if any Σ_1-sentence provable in T is valid in $\mathbf{N}$. Σ_1-soundness is a condition weaker than (full) soundness, but stronger than consistency. One can check that with the knowledge that the formula $\mathsf{Pr}_T(x)$ is Σ_1, the proof of Gödel First Incompleteness Theorem we gave in the end of the previous section remains correct if the assumption of soundness is replaced by that of Σ_1-soundness.

Lemma 1 *Let T be Σ_1-sound and let σ be a Σ_1-sentence. Then if $T \vdash \neg\sigma$ then $\mathbf{N} \models \neg\sigma$.*

Proof Let Σ_1-sentence σ such that $T \vdash \neg\sigma$ be given. We may assume that σ is of the form $\exists v \psi(v)$ where $\psi \in \Delta_0$. From the assumption that $T \vdash \forall v \neg\psi(v)$ we have $\forall n(T \vdash \neg\psi(\overline{n}))$. Each of the sentences $\neg\psi(\overline{n})$ is bounded, hence equivalent to a Σ_1-sentence. From Σ_1-soundness we have $\forall n(\mathbf{N} \models \neg\psi(\overline{n}))$. Thus $\mathbf{N} \models \forall v \neg\psi(v)$. ∎

Theorem 2 *Let T be recursively axiomatized and Σ_1-sound theory with a language $L \supseteq L_0$. Then T is incomplete. There exist Σ_1- and Π_1-sentences independent of T.*

Proof Let A be a fixed recursively enumerable set that is *not recursive.* By the definability theorem there exists a formula $\theta(x) \in \Sigma_1$ that defines the set A, i.e. satisfies

$$\forall n(n \in A \Leftrightarrow \mathbf{N} \models \theta(\overline{n})). \tag{4}$$

Let $X = \{\, n \,;\, T \vdash \theta(\overline{n}) \,\}$ and $Y = \{\, n \,;\, T \vdash \neg\theta(\overline{n}) \,\}$. Both X and Y are recursively enumerable because T is recursively axiomatized. From Σ_1-soundness we have $X \subseteq A$. Lemma 1 says $Y \cap A = \emptyset$. So the relations between the sets X, Y and A are as shown in Fig. 1. Now recall Post's theorem: if two recursively enumerable sets are mutually complementary then each of them is recursive. Apply this theorem to A and Y; since A and Y are recursively enumerable and A is not recursive, A and Y

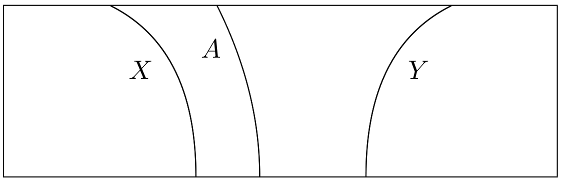


Figure 1: Provables, truths, and refutables

cannot be complementary. So there exists a number $n_0 \notin A \cup Y$. For this number n_0 we have $T \nvdash \theta(\overline{n_0})$ and $T \nvdash \neg\theta(\overline{n_0})$. The sentence $\theta(\overline{n_0})$ is an independent Σ_1-sentence and its negation $\neg\theta(\overline{n_0})$ is an independent Π_1-sentence. ∎

So we see that the Σ_1-completeness theorem is not necessary for the proof of Gödel First Incompleteness Theorem, the only knowledge we used was the definability theorem. The self-reference is also not necessary; we used the Post's theorem and the fact that recursively enumerable non-recursive sets do exist. One can, of course, argue that the proof of the existence of recursively enumerable non-recursive sets uses diagonalization, which may also be viewed as a form of self-reference.

If we add an assumption that T is an extension of Q to the assumptions of Theorem 2 then we can use the Σ_1-completeness as follows. If $n \in A$ then, by (4), $\mathbf{N} \models \theta(\overline{n})$. Since $\theta \in \Sigma_1$, it is true that $T \vdash \theta(\overline{n})$ and $n \in X$. So the additional assumption that T is an extension of Q yields an additional conclusion that $X = A$. But then the equivalence $\forall n(n \in A \Leftrightarrow T \vdash \theta(\overline{n}))$ says that the set A is *reducible* to the set $\mathrm{Thm}(T)$. The fact that a non-recursive set cannot be reducible to a recursive one is another knowledge from elementary recursion theory. So we have that $\mathrm{Thm}(T)$ is a non-recursive set, which means that the theory T is undecidable. Thus a moral we can draw from these considerations is that Σ_1-completeness is a device useful for proving undecidability, and not so necessary for proving incompleteness. Again this thesis cannot be taken too literally: the proof of Rosser's generalization of Gödel First Incompleteness Theorem, saying that even all consistent (possibly not Σ_1-sound) recursively axiomatizable extensions of Q are incomplete, would hardly go through without Σ_1-completeness.

Let me turn back to the proof of Gödel First Incompleteness Theorem given in the end of the previous section. I want to remark that the use of the condition D1 can be completely avoided there, since in the place where it was used one can also reason as follows. Assume $T \vdash \nu$. Then soundness yields $\mathbf{N} \models \nu$. The condition Def says $\mathbf{N} \models \mathsf{Pr}_T(\overline{\nu})$. So the equivalence $\nu \equiv \neg \mathsf{Pr}_T(\overline{\nu})$ is not valid in $\mathbf{N}$, which is a contradiction. With this observation we can summarize the paper as follows. The self-reference method is not the only way of proving Gödel First Incompleteness Theorem, structural proofs can provide better insight into what is going on; if one insists on proving Gödel First Incompleteness Theorem via self-reference then the usual way of reasoning is not the only possible, its special value is in the fact the it can be prolonged to a proof of Gödel *Second* Incompleteness Theorem, a topic beyond the scope of this paper.

References

[1] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic.* Springer, 1993.

[2] R. Kaye. *Models of Peano Arithmetic.* Oxford University Press, 1991.

[3] M. H. Löb. Solution of a problem of Leon Henkin. *J. Symbolic Logic*, 20:115–118, 1955.

[4] C. Smoryński. Metamathematics of Arithmetic, Chapter III: Representability and Semi-Representability. Unpublished manuscript, circa 1978.

[5] V. Švejdar. On provability logic. *Nordic J. Philosophical Logic*, 4(2):95–116, 2000.

[6] V. Švejdar. *Logika: neúplnost, složitost a nutnost (Logic: Incompleteness, Complexity, and Necessity).* Academia, Praha, 2002.

[7] A. Tarski, A. Mostowski, and R. M. Robinson. *Undecidable Theories.* North-Holland, Amsterdam, 1953.

Vítězslav Švejdar
Philosophical Faculty, Charles University
Palachovo nám. 2, 116 38, Praha 1
vitezslav.svejdar@cuni.cz

A Note on Arithmetical Completeness of Theories with Rosser Modalities

Vítězslav Švejdar[*]

Abstract We prove an arithmetical completeness theorem for the modal logic R^- of Guaspari and Solovay, which is an extension of provability logic by witness comparison modalities. Then we show an arithmetical characterization for the modal formulas A such that $\mathsf{R}^- \vdash \Box A$.

Introduction

The self-reference theorem, or better, a formulation of self-reference theorem suitable for our purposes, says that the equation $T \vdash \varphi \equiv \psi(\overline{\varphi})$, for an unknown arithmetical sentence φ, has a solution for any axiomatic theory T extending Robinson arithmetic Q and any formula $\psi(x)$ in the language of T (see [2, 1]). Recall that $\overline{\varphi}$ stands for the term $\mathrm{S}(\mathrm{S}(\,..\,\mathrm{S}(0)\,..))$ with n occurences of the symbol S, where n is the numerical code of the sentence φ—the reader may also want to consult [6] for some more details. Probably the best known application of the self-reference theorem is the proof of Gödel First Incompleteness Theorem: if an arithmetical formula $\mathsf{Pr}(x)$ is a provability predicate of a theory T and T is a

[*]This paper was supported by grant 401/01/0218 of the Grant Agency of the Czech Republic.

sound recursively axiomatized extension of Q then any solution φ of the equation $T \vdash \varphi \equiv \neg\mathsf{Pr}(\overline{\varphi})$ is independent of T. Other examples of self-referential equations are $\varphi \equiv \mathsf{Pr}(\overline{\varphi})$ and $\varphi \equiv \mathsf{Pr}(\overline{\varphi}) \to \lambda$, the latter being a *parametric equation* with parameter λ.

A self-referential equation $T \vdash \varphi \equiv \psi(\overline{\varphi})$ can be called *Gödelian* if the formula $\psi(\overline{\varphi})$ is built up from parameters and the term $\overline{\varphi}$ using logical connectives and a proof predicate $\mathsf{Pr}(x)$. Thus a solution of a Gödelian equation can be viewed as a self-referential sentence saying something about its own provability (or unprovability, or conditional provability with respect to parameters). All three equations mentioned above are Gödelian.

If T and $\mathsf{Pr}(x)$ are as above then we can choose the sentence $0 = \mathrm{S}(0)$ to represent contradiction and define Con as the sentence $\neg\mathsf{Pr}(\overline{0 = \mathrm{S}(0)})$. The sentence Con expresses, in the arithmetical language, consistency of the theory T. The proof of Gödel Second Incompleteness Theorem, saying that the sentence Con is unprovable in T, is obtained by proving that any solution φ of the equation $T \vdash \varphi \equiv \neg\mathsf{Pr}(\overline{\varphi})$ is T-equivalent to the sentence Con. This fact about the equation $T \vdash \varphi \equiv \neg\mathsf{Pr}(\overline{\varphi})$ can be widely generalized using *modal logic*. *Provability logic* GL (see [5, 4] or perhaps my [8]) is a modal propositional logic designed to model reasoning about Gödelian self-referential equations or, more generally, about arithmetical sentences built up using the provability predicate $\mathsf{Pr}(x)$. Two prominent results about self-reference obtained via modal logic say that each Gödelian equation has only (i.e. exactly) *one solution* (up to equivalence provable in T) and that the solution is *calculable*, i.e. expressible without self-reference in terms of the formula $\mathsf{Pr}(x)$ and the equation's parameters.

Provability logic can also be used to show some limitations of the Gödelian self-reference method. Namely, Gödelian self-reference is not sufficient to show the existence of a sentence φ satisfying

$$T \vdash \mathsf{Con} \to \neg\mathsf{Pr}(\overline{\varphi}) \ \&\ \neg\mathsf{Pr}(\overline{\neg\varphi}). \tag{1}$$

The sentence (1) says that (under the unavoidable assumption of consistency of the theory T) the sentence φ is independent of T. So (1) expresses *formalized independency* of φ. Note that the sentence Con satisfies only one half of (1), i.e. $T \vdash \mathsf{Con} \to \neg\mathsf{Pr}(\overline{\mathsf{Con}})$. It follows from Gödel Second Incompleteness Theorem that $\mathsf{Con} \to \neg\mathsf{Pr}(\overline{\neg\mathsf{Con}})$ is a sentence unprovable in T.

The provability predicate $\mathsf{Pr}(x)$ has the form $\exists y \mathsf{Proof}(x, y)$, where the formula $\mathsf{Proof}(x, y)$ is called *proof predicate* and can be read the number y is a proof of the number (formula) x. If $\mathsf{Proof}(x, y)$ is a proof predicate of a theory T extending Q then any solution φ of the equation

$$T \vdash \varphi \equiv \exists y(\mathsf{Proof}(\overline{\neg\varphi}, y) \,\&\, \forall v \leq y \neg \mathsf{Proof}(\overline{\varphi}, v)) \qquad (2)$$

is called *Rosser sentence* of the theory T. The equation (2) is *Rosser equation*; its solution can be described as saying there exists a proof of my negation which is smaller than any possible proof of myself. The Rosser equation is not Gödelian, and each its solution satisfies the formalized independency condition (1).

Guaspari and Solovay [3] developed a modal theory which extends provability logic by *witness comparison modalities* and which is capable of modeling the reasoning about Rosser sentences. In fact there are several modal theories in [3]: the theory R^- has a nice Kripke semantics but is not claimed to be arithmetically significant, while its extension R is complete w.r.t. arithmetical interpretations.

In this paper we follow the practice from [7] and define an arithmetical interpretation of modal logic as involving *two* axiomatic theories: one theory S is examined as to what it can say about provability in its extension T. Then we exhibit an arithmetical characterization for modal formulas A such that $\mathsf{R}^- \vdash A$, and also a (different) characterization for modal formulas A such that $\mathsf{R}^- \vdash \Box A$.

Most of the material in this paper was contained in my doctoral dissertation, then communicated to R. Solovay, but never published. Part (a) of Theorem 1 is also implicit in C. Smoryński's [4].

Theories with Rosser modalities

We consider modal propositional formulas built up from propositional atoms and the symbol $\bot$ (falsity) using logical connectives $\rightarrow$, &, $\vee$, and $\neg$, the unary modal operator $\Box$, and two binary operators $\preceq$ and $\prec$, with the restriction that $\preceq$ and $\prec$ are applicable only to formulas starting with $\Box$. So $\Box q \preceq \Box\Box\bot$ or $\Box p \preceq \Box(\Box\neg p \preceq \Box q)$ are examples of modal formulas in our modal language. The operators $\preceq$ and $\prec$ are called *witness comparison modalities*. In syntax analysis, we assume that $\rightarrow$ has higher priority than $\equiv$ but lower than & and $\vee$, and that $\preceq$ and $\prec$ have highest possible priority, higher than & and $\vee$. So e.g. $p\equiv\Box\bot\prec\Box p\&\Box\neg q$ is a shorthand for the formula $p \equiv ((\Box\bot \prec \Box p) \;\&\; \Box\neg q)$. The strings $\Box p \preceq (\Box q \preceq \Box\neg p)$, $\neg\Box\bot \preceq \Box p$, and $(\Box p \vee \Box q) \preceq \Box\neg q$ are examples of *non*formulas.

The theory R^- has the following axiom schemata A1–A6 and deduction rules R1 and R2 (see also [3] and [4]):

A1: All propositional tautologies,

A2: $\Box(A \to B) \to (\Box A \to \Box B)$,

A3: $\Box A \to \Box\Box A$,

A4: $\Box(\Box A \to A) \to \Box A$,

A5: $\Box A \preceq \Box B \to \Box(\Box A \preceq \Box B)$, $\qquad \Box A \prec \Box B \to \Box(\Box A \prec \Box B)$,

A6: $\Box A \to \Box A \preceq \Box A$,

$\Box A \preceq \Box B \to \Box A$,

$\Box A \preceq \Box B \;\&\; \Box B \preceq \Box C \to \Box A \preceq \Box C$,

$\Box A \vee \Box B \to \Box A \preceq \Box B \vee \Box B \prec \Box A$,

$\Box B \prec \Box A \equiv \Box B \preceq \Box A \;\&\; \neg(\Box A \preceq \Box B)$,

R1: $A,\ A \to B \ /\ B$,

R2: $A \ /\ \Box A$.

We read $\Box A$ as "A is necessary" or "A is provable". If convenient, we may write $\Diamond A$ as a shorthand for $\neg\Box\neg A$ and $\top$ as a shorthand for $\neg\bot$. We read $\Box A \preceq \Box B$ or $\Box A \prec \Box B$ as "A is provable by a proof which is less than or equal to (or less than, respectively) any proof of B". Note that the schemata A1–A4 and rules R1 and R2 are taken from provability logic. The five schemata A6 will be referred to as A6(i)–A6(v). These axioms essentially say that proofs are linearly preordered and that $\prec$ is a strict preorder associated with the unstrict $\preceq$. Recall that the formulation of the axiom A6(v) reflects the fact that in predicate logic without equality strict order is definable from its associated unstrict order, but not vice versa. The reader should be warned that both R^- and R (defined below) are *non-extensional* in the sense that substitution of (even necessarily) equivalent formulas into a formula may yield non-equivalent formulas. An example of reasoning in R^- is the following proof of the formula $\Box(\Box\top \prec \Box\bot)$:

If $\Box\bot$, i.e. if a contradiction is provable, then everything is provable; hence $\Box(\Box\top \prec \Box\bot)$. So assume $\neg\Box\bot$. Since $\Box\top$, by A6(iv) we have $\Box\bot \preceq \Box\top$ or $\Box\top \prec \Box\bot$. The former contradicts the assumption $\neg\Box\bot$ by A6(ii). So $\Box\top \prec \Box\bot$. From A5 we again have $\Box(\Box\top \prec \Box\bot)$.

The formulas $\Box\top$ and $\Box\bot \to \Box(\Box\top \prec \Box\bot)$, used in the proof, are derived "as usually" in modal logic.

The theory R is obtained by adding to the axioms and rules of the theory R^- the following additional inference rule:

R3: $\Box A \ / \ A$.

Recall that there is no need to add the rule R3 to provability logic, since provability logic is closed on this rule. An example of a formula provable in R is $\Box\top \prec \Box\bot$.

Now we review the *arithmetical semantics* of modal logic with Rosser modalities. Let $\mathit{Prf}(x,y)$ be an arithmetical formula. An *(arithmetical) evaluation* based on the formula Prf is any function $*$ from modal formulas to arithmetical sentences which preserves logical connectives and satisfies

$$\begin{aligned} \bot^* &= (0 = S(0)), \\ (\Box A)^* &= \exists y \mathit{Prf}(\overline{A^*}, y), \\ (\Box A \preceq \Box B)^* &= \exists y (\mathit{Prf}(\overline{A^*}, y) \mathbin{\&} \forall v{<}y \neg \mathit{Prf}(\overline{B^*}, v)), \\ (\Box A \prec \Box B)^* &= \exists y (\mathit{Prf}(\overline{A^*}, y) \mathbin{\&} \forall v{\le}y \neg \mathit{Prf}(\overline{B^*}, v)). \end{aligned} \tag{3}$$

Note that an arithmetical evaluation is determined by its values on propositional atoms. An *arithmetical interpretation* of modal logic is a pair $\langle \mathit{Prf}, S\rangle$ where S is a theory with at least the arithmetical language and $\mathit{Prf}(x,y)$ is an arithmetical formula. A modal formula A is *valid* in an arithmetical interpretation $\langle \mathit{Prf}, S\rangle$ if $S \vdash A^*$ for each evaluation $*$ based on the formula Prf. One may easily check that the set of all modal formulas valid in an intepretation $\langle \mathit{Prf}, S\rangle$ containes all instances of the schemata A1 and A6 and is closed under the modus ponens rule R1.

$\mathrm{I}\Delta_0$ is an axiomatic theory like Peano arithmetic PA but with the induction scheme restricted to Δ_0-formulas, whereas $\mathrm{I}\Delta_0$+Exp is an extension of $\mathrm{I}\Delta_0$ by the axiom Exp saying that the exponentiation function $x \mapsto 2^x$ is total. We choose $\mathrm{I}\Delta_0$+Exp as our *base theory*; in the sequel all theories are extensions of $\mathrm{I}\Delta_0$+Exp.

An arithmetical formula $Prf(x,y)$ is a *standard proof predicate* of a theory T extending $\mathrm{I}\Delta_0$+Exp if it is $\Delta_1(\mathrm{I}\Delta_0\text{+Exp})$, i.e. ($\mathrm{I}\Delta_0$+Exp)-equivalent to a Σ_1-formula and also to a Π_1-formula, and if the formula $Pr(x)$ defined as $\exists y Prf(x,y)$—which can be called *provability predicate* associated with the standard proof predicate *Prf*—satisfies the conditions

$$T \vdash \varphi \Leftrightarrow \mathbf{N} \models Pr(\overline{\varphi}), \tag{4}$$

$$\mathrm{I}\Delta_0\text{+Exp} \vdash Pr(\overline{\varphi \to \psi}) \to (Pr(\overline{\varphi}) \to Pr(\overline{\psi})), \tag{5}$$

$$\mathrm{I}\Delta_0\text{+Exp} \vdash \sigma \to Pr(\overline{\sigma}), \tag{6}$$

for all sentences φ and ψ in the language of T and every arithmetical Σ_1-sentence σ. It is straightforward to verify that if T is an extension of $\mathrm{I}\Delta_0$+Exp and *Prf* is a standard proof predicate of T then the theory R^- defined above is sound w.r.t. the arithmetical intepretation $\langle Prf, \mathrm{I}\Delta_0\text{+Exp}\rangle$. Moreover, it T is Σ_1-sound then R is sound w.r.t. the intepretation $\langle Prf, T\rangle$, see [4]. Guaspari and Solovay proved *arithmetical completeness theorem* for R: a modal formula A is provable in R if and only if it is (simultaneously) valid in all arithmetical interpretations $\langle Prf, \mathsf{PA}\rangle$ where *Prf* is a standard proof predicate of PA. Their proof remains correct if PA is replaced by any Σ_1-sound extension S of $\mathrm{I}\Delta_0$+Exp; it is however important that the formulas *Prf* run over standard proof predicates of the same theory S which stands as the second component in $\langle Prf, S\rangle$.

The choice of $\mathrm{I}\Delta_0$+Exp as a base theory is justified by the fact that it is the weakest reasonable theory capable of proving the formalized Σ_1-completeness theorem expressed by the condition (6) above. This condition is needed when verifying validity of the modal schemata A3 and A5. The reader may also ask why the proof predicates are let to run over all standard proof predicates of the theory in question; it could be of an interest to characterize all modal formulas valid in *single* interpretation say $\langle \mathsf{Proof}_\pi(x,y), \mathsf{PA}\rangle$ where $\mathsf{Proof}_\pi(x,y)$ is *the* natural proof predicate of Peano arithmetic, describing proofs in a Hilbert style predicate calculus from the axioms of PA defined "as usually". The answer is that the set of all modal formulas valid in single interpretation $\langle \mathsf{Proof}_\pi(x,y), \mathsf{PA}\rangle$ contains a lot of unwanted formulas, expressing *irrelevant facts* about ordering of proofs. For example, one of the formulas $\Box\top \prec \Box\Box\top$ and $\Box\Box\top \prec \Box\top$ is valid in the intepretation $\langle \mathsf{Proof}_\pi(x,y), \mathsf{PA}\rangle$; which it is depends on which of the (provable) sentences $\neg(0 = \mathrm{S}(0))$ and $\mathsf{Pr}_\pi(\overline{\neg(0 = \mathrm{S}(0))})$ has lower proof—a fact that may well depend on details of coding of syntax by numbers. Even the fact that each natural number is a proof of at most one formula is considered irrelevant.

A triple $\langle W, <, \Vdash \rangle$ is a *Kripke model for* R^- if $<$ is a reversely well-founded transitive relation on a non-empty set W and the truth relation $\Vdash$ satisfies (i) the usual modal condition $A \Vdash \Box A$ iff $\forall b{>}a(b \Vdash A)$, (ii) the Σ-persistency condition $A \Vdash C \Rightarrow \forall b{>}a(b \Vdash C)$ for every C of the form $\Box A \preceq \Box B$ or $\Box A \prec \Box B$, and (iii) all instances of the scheme A6. Note that there is nothing in the definition that enables us to compute the truth value of a formula $\Box A \preceq \Box B$ or $\Box A \prec \Box B$ from truth values of its components A and B. The formulas $\Box A \preceq \Box B$ and $\Box A \prec \Box B$ behave in a sense like propositional atoms; their truth values can to some extent be chosen arbitrarily, but the choice has to respect the Σ-persistency condition (i.e. the scheme A5) and the linear preorder axioms A6.

A set X of modal formula is *adequate* if it is closed under subformulas and $\preceq$ and $\prec$. For each set X there exists least adequate set $\mathrm{Ad}(X)$ containing X; if X is finite then also $\mathrm{Ad}(X)$ is finite. If A is a formula then $\mathrm{Ad}(A)$ is defined as $\mathrm{Ad}(\{A\})$. For example, if A is the formula $\Box\bot \preceq \Box\top$ then $\mathrm{Ad}(A)$ consists of the following twelve formulas: $\Box\bot \preceq \Box\top$, $\Box\top \preceq \Box\bot$, $\Box\bot \preceq \Box\bot$, $\Box\top \preceq \Box\top$, $\Box\bot \prec \Box\top$, $\Box\top \prec \Box\bot$, $\Box\bot \prec \Box\bot$, $\Box\top \prec \Box\top$, $\Box\bot$, $\Box\top$, $\bot$, $\top$. If S is an adequate set then a *Kripke pseudomodel for* S is defined as a Kripke model, but with the requirements on the truth relation $\Vdash$ restricted to S. It is known that each Kripke pseudomodel for an adequate S can be extended to a Kripke model and that the theory R^- is complete w.r.t. Kripke semantics: A is provable in R^- iff it is valid in all Kripke models for R^-. An application of these two facts is the following: choose $\Box\bot \prec \Box\top$ for the formula A, choose $W = \{r\}$, and stipulate that the formulas $\Box\bot \preceq \Box\bot$, $\Box\bot \prec \Box\top$, $\Box\bot \preceq \Box\top$, $\Box\top \preceq \Box\top$, $\Box\bot$, $\Box\top$, and $\top$ are satisfied in r while the remaining formulas from $\mathrm{Ad}(A)$ are not satisfied. The result is a Kripke pseudomodel for $\mathrm{Ad}(A)$ showing that A is not provable in R^-. As an excercise the reader may verify that the formula

$$\Box(p \equiv \Box\neg p \preceq \Box p) \mathbin{\&} \neg\Box\bot \to \neg\Box p \mathbin{\&} \neg\Box\neg p,$$

which is a modal version of Rosser incompleteness theorem, is valid in every Kripke model for R^-, and thus provable in R^-. A direct construction of a proof of this formula from the axioms of R^- is also possible.

In the sequel we always assume that if $\langle W, <, \Vdash \rangle$ is a Kripke model for R^- then the set W contains a least element w.r.t. the relation $<$, which we call a *root* of the model $\langle W, <, \Vdash \rangle$. A model $\langle W, <, \Vdash \rangle$ for R^- is *A-sound*, where A is a modal formula, if its root satisfies all formulas $\Box D \to D$, where $\Box D$ is a subformula of A. *Kripke completeness theorem for the theory* R, see again [3] or [4], says that a modal formula A is provable in R iff A is valid in all A-sound Kripke models.

The main result

In the formulation of Theorem 1 we will need the function $[A, p] \mapsto A^p$, where A is a modal formula and p a propositional atom. This function is defined by the following recursion. If A is an atom then A^p is A; if A is $B \to C$ then A^p is $B^p \to C^p$, and similarly for negation and the binary connectives including $\preceq$ and $\prec$; if A is $\Box B$ then A^p is $\Box(p \to B^p)$.

Theorem 1 *(a) If A is a modal formula in the language with Rosser modalities and p is an atom not occurring in A then the following conditions are equivalent:*

- *(i)* $\mathsf{R}^- \vdash A$,
- *(ii)* $\mathsf{R}^- \vdash A^p$,
- *(iii)* $\mathsf{R} \vdash A^p$,
- *(iv) A is simultaneously valid in all interpretations $\langle Prf, \mathrm{I}\Delta_0{+}\mathrm{Exp} \rangle$ where Prf is a standard proof predicate of a theory T extending* $\mathrm{I}\Delta_0{+}\mathrm{Exp}$.

(b) Also the following conditions are equivalent:

- *(i)* $\mathsf{R}^- \vdash \Box A$,
- *(iii)* $\mathsf{R} \vdash p \to A^p$,
- *(iv) A is simultaneously valid in all interpretations $\langle Prf, T \rangle$ where T is an extension of* $\mathrm{I}\Delta_0{+}\mathrm{Exp}$ *and Prf is a standard proof predicate of T.*

Proof (iii) $\Rightarrow$ (i). Assume $\mathsf{R}^- \nvdash A$ or $\mathsf{R}^- \nvdash \Box A$, respectively. By the Kripke completeness theorem for R^- we have a Kripke model $\langle W, <, \Vdash \rangle$ with a root r such that $r \nVdash A$ in (a) and $a \nVdash A$, where $a \neq r$, in (b).

In (a) we construct an A^p-sound Kripke pseudomodel for $\mathrm{Ad}(A^p)$ which is a counter-model to A^p, while in (b) we construct a $(p \to A^p)$-sound Kripke pseudomodel for $\mathrm{Ad}(p \to A^p)$ which is a counter-model to $p \to A^p$. We define a new truth relation $\Vdash_1$ by the following conditions. First, $r \nVdash_1 p$ and $b \Vdash_1 p$ for all $b \neq r$. If q is an atom different from p then $b \Vdash_1 q \Leftrightarrow b \Vdash q$, for all $b \in W$. If B is a formula in $\mathrm{Ad}(A)$ of the form $\Box C \preceq \Box D$ or $\Box C \prec \Box D$ then $b \Vdash_1 B^p \Leftrightarrow b \Vdash B$. Since each formula in $\mathrm{Ad}(A^p)$ or in $\mathrm{Ad}(p \to A^p)$ whose principal connective is $\preceq$ or $\prec$ equals some B^p where $B \in \mathrm{Ad}(A)$ has the form $\Box C \preceq \Box D$ or $\Box C \prec \Box D$, the relation $\Vdash_1$ is fully determined. The required instances of A5 and A6 are satisfied, hence $\langle W, <, \Vdash_1 \rangle$ is a Kripke pseudomodel. An easy induction on the complexity of B shows that $b \Vdash_1 B^p \Leftrightarrow b \Vdash B$ for each $b \in W$ and each formula $B \in \mathrm{Ad}(A)$. So in (a) we have $r \nVdash_1 A^p$ and in (b) we have $a \nVdash_1 p \to A^p$ for some $a \neq r$. It remains to show that the resulting pseudomodel $\langle W, <, \Vdash_1 \rangle$ is A^p-sound (or $(p \to A^p)$-sound in (b)). It is evident that each subformula of A^p (or of $p \to A^p$) starting with $\Box$ has the form $\Box(p \to B^p)$ for some subformula $\Box B$ of A. But then $r \Vdash_1 \Box(p \to A^p) \to (p \to A^p)$ is a consequence of $r \nVdash_1 p$.

(iv) $\Rightarrow$ (iii). Assume $\mathsf{R} \nvdash A^p$ in (a) and $\mathsf{R} \nvdash p \to A^p$ in (b). By the arithmetical completeness theorem for R there exists a proof predicate $\mathit{Prf}(x, y)$ of $\mathrm{I}\Delta_0$+Exp and an evaluation $*$ such that $\mathrm{I}\Delta_0\text{+Exp} \nvdash (A^p)^*$ in (a) and $\mathrm{I}\Delta_0\text{+Exp} \nvdash (p \to A^p)^*$ in (b). Let the formula $\mathit{Prf}_1(x, y)$ be defined as $\mathit{Prf}(\overline{p^*} \to x, y)$. The formula Prf_1 is a standard proof predicate of $\mathrm{I}\Delta_0\text{+Exp+}p^*$. Let $\sharp$ be an avaluation which coincides with $*$ on atoms and which is based on the proof predicate Prf_1. A straightforward induction on complexity of a subformula B of A shows that $B^\sharp = (B^p)^*$. In (a) we have $\mathrm{I}\Delta_0\text{+Exp} \nvdash A^\sharp$, while in (b) we have $\mathrm{I}\Delta_0\text{+Exp} \nvdash (p \to A^p)^*$, i.e. $\mathrm{I}\Delta_0\text{+Exp+}p^* \nvdash A^\sharp$. So we have found a theory $T := \mathrm{I}\Delta_0\text{+Exp+}p^*$ and its standard proof predicate Prf_1 such that A is not valid in the interpretation $\langle \mathit{Prf}_1, \mathrm{I}\Delta_0\text{+Exp} \rangle$ in (a), and A is not valid in the interpretation $\langle \mathit{Prf}_1, T \rangle$ in (b).

The implication (i) $\Rightarrow$ (iv) in (a) is the arithmetical soundness theorem for R^-. In (b) one can reason as follows. If $\mathsf{R}^- \vdash \Box A$ and Prf is a standard proof predicate of T then, by the same implication in (a), $\mathrm{I}\Delta_0\text{+Exp} \vdash \mathit{Pr}(\overline{A^*})$. The fact that $\mathbf{N}$ is a model of $\mathrm{I}\Delta_0$+Exp yields $\mathbf{N} \models \mathit{Pr}(\overline{A^*})$. By the condition (4) from the definition of a standard proof predicate we have $T \vdash A^*$.

Finally the implication (ii) $\Rightarrow$ (iii) in (a) is trivial, and (i) $\Rightarrow$ (ii) can be proved by an induction on the length of a proof of A in R^-. ∎

Theorem 1 can be paraphrased as follows. We are interested in general principles (not in irrelevant facts dependent on the choice of formalism and coding) about provability in a theory T, which are expressible in the language with witness comparison operators. The modal theory R captures the knowledge which a Σ_1-sound base theory like PA or $\mathrm{I}\Delta_0$+Exp has about provability in itself. By Theorem 1(a), the theory R^- characterizes what a base theory can say about provability in all its extensions, while the set $\{ A \,;\, R^- \vdash \Box A \}$ characterizes what all extensions T (Σ_1-sound or not) of the base theory can say about provability in themselves. An example of a formula which is provable in R but does not satisfy the conditions (b) in Theorem 1 is $\Box\top \prec \Box^n\bot$ for $n \geq 2$. Examples of formulas satisfying (b) but not (a) are $(\Box\top \prec \Box\Box\bot) \rightarrow \Box(\Box\top \prec \Box\Box\bot) \prec \Box\bot$, or $\Box^{n-1}(\Box\top \prec \Box^n\bot)$ for $n \geq 1$.

A modal theory is *invariant* if it proves A^p whenever it proves A and p is an atom not occurring in A. It follows from Theorem 1 that R^- is invariant. The importance of the notion of invariance in modal systems with arithmetical interpretations (including interpretability logic) was noticed by several people including Craig Smoryński, Albert Visser, and the author. The term "invariance" probably comes from A. Visser.

References

[1] R. Carnap. *Logische Syntax der Sprache*. Springer, 1934.

[2] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[3] D. Guaspari and R. M. Solovay. Rosser sentences. *Annals of Math. Logic*, 16:81–99, 1979.

[4] C. Smoryński. *Self-Reference and Modal Logic*. Springer, New York, 1985.

[5] R. M. Solovay. Provability interpretations of modal logic. *Israel J. Math.*, 25:287–304, 1976.

[6] V. Švejdar. On Structural Proofs of Gödel First Incompleteness Theorem. This volume, pages 115–122.

[7] V. Švejdar. Modal analysis of generalized Rosser sentences. *J. Symbolic Logic*, 48(4):986–999, 1983.

[8] V. Švejdar. On Provability Logic. *Nordic J. Philosophical Logic*, 4(2):95–116, 2000.

Vítězslav Švejdar
Philosophical Faculty, Charles University
Palachovo nám. 2, 116 38, Praha 1
vitezslav.svejdar@cuni.cz

MISCELLANEA LOGICA
TOM V

Kamila Bendová and Petr Jirků
editors

Published by Charles University in Prague
The Karolinum Press
Ovocný trh 3, 116 36 Praha 1
Prague 2003
Vice-rector editor prof. MUDr. Pavel Klener, DrSc.
Typeset by Vítězslav Švejdar using the TEX/LATEX system
Printed by The Karolinum Press
First edition
ISBN 80–246–0799–9