INTERPRETATIONS
Lecture I: Interpretations Introduced

Albert Visser

Department of Philosophy, Faculty of Humanities, Utrecht University

Summer School on Set Theory and Higher Order Logic
London

August 1, 2011

Examples
Why Interpretations?
Definitions

Universiteit Utrecht

Overview

Examples

Why Interpretations?

Definitions

Universiteit Utrecht

Overview

Examples

Why Interpretations?

Definitions

Overview

Examples

Why Interpretations?

Definitions

Overview

Examples

Why Interpretations?

Definitions

Descartes 1

Consider a real closed field $\mathcal{R}$. We can build a model $\mathcal{E}$ of a version of two-dimensional Euclidean Geometry employing the Cartesian plane $\mathcal{R}^2$.

Suppose we work with a points-only version of Geometry. In this case the domain of $\mathcal{E}$ will be the set of pairs (r_0, r_1) of elements of the domain of $\mathcal{R}$. E.g. the notion of collinearity will be defined by:

- $\mathrm{Coll}(r_0, r_1, s_0, s_1, t_0, t_1) :\leftrightarrow (r_0 \neq s_0 \vee r_1 \neq s_1) \wedge$
  $\exists\lambda\, ((t_0 = r_0 + \lambda(s_0 - r_0)) \wedge (t_1 = r_1 + \lambda(s_1 - r_1)))$.

We are looking at a uniform construction of *inner models*. This construction is purely syntactic. The new domain and concepts like collinearity have purely first order definitions.

This reduction gives us an *interpretation* desc of the *theory* of Euclidean Geometry EG in the *theory* RCF of real closed fields.

Descartes 2

We have:
uniform construction of inner models $\approx$ interpretation of theories.

The interpretation desc : EG $\rightarrow$ RCF is two-dimensional, unrelativised and identity preserving.

If we have a two-sorted points-and-lines version of EG, say $\mathrm{EG}^{\mathrm{pl}}$, we introduce two domains. The 'points' are again be pairs (r_0, r_1), the 'lines' are quadruples (s_0, s_1, s_2, s_3), where $s_0 \neq s_3$ or $s_1 \neq s_4$). Moreover, identity of lines is:

$$(s_0, s_1, s_2, s_3)E(t_0, t_1, t_2, t_3) :\leftrightarrow \mathrm{Coll}(s_0, s_1, s_2, s_3, t_0, t_1) \wedge \mathrm{Coll}(s_0, s_1, s_2, s_3, t_2, t_3).$$

In this case we have an interpretation, say $\mathrm{desc}^{\mathrm{pl}}$, of a two-sorted theory into a one sorted theory. The interpretation is four-dimensional, relativised and not identity preserving for the sort of lines.

Descartes 3

Clearly, $\mathrm{desc}^{\mathrm{pl}}$ is the *composition* $(\mathrm{desc} \circ K)$ of the obvious interpretation $K : \mathrm{EG}^{\mathrm{pl}} \to \mathrm{EG}$ and desc.

The interpretation desc has an *inverse* indesc that is defined as follows. Consider a model of EG, say $\mathcal{E}$. We choose two distinct points in the domain, say p and q. Our new domain becomes the points on the line $\ell := pq$. Here p interprets 0 and q interprets 1. Treating pq ad a 'unit interval' we can define addition and multiplication of points on ℓ in a purely geometric way.

We note that we did not really define a single inner model but a function for the *parameter domain* α that consists of all pairs (p, q) such that $p \neq q$, to inner models. Thus indesc is a one-dimensional, relativised, identity preserving interpretation with parameters.

Descartes 4

We can eliminate the parameters as follows. Consider two parameter pairs pq and rs in $\mathcal{E}$. We can define geometrically an isomorphism ϕ_{pqrs} from the internal real closed field $\mathrm{indesc}^{pq}(\mathcal{E})$ and the internal real closed field $\mathrm{indesc}^{rs}(\mathcal{E})$.

Now we define a new interpretation indesc^{+} as follows. The 'reals' of the interpretation are triples (p, q, r), where $p \neq q$. We define identity of triples by:

$$(p, q, r)\ E\ (p', q', r') :\leftrightarrow r' = \phi_{pqp'q'}(r).$$

We define e.g. addition by:

$$\mathrm{A}^{+}(p, q, r, p', q', r', p^\star, q^\star, r^\star) :\leftrightarrow \mathrm{A}^{p^\star q^\star}(\phi_{pqp^\star q^\star}(r), \phi_{p'q'p^\star q^\star}(r'), r^\star).$$

Here A is the adition of indesc.

Descartes 5

In what sense are desc and indesc$^+$ inverses?

If we start with $\mathcal{R}$, take the internal model $\mathcal{E} := \mathrm{desc}(\mathcal{R})$, and then take the internal model $\mathcal{R}^\star := \mathrm{indesc}(\mathcal{E})$, we have an internal model of RCF inside $\mathcal{R}$.

The model $\mathcal{R}^\star$ is *definably isomorphic* to $\mathcal{R}$. We map r in the domain of $\mathcal{R}$ to (the equivalence class of) $(0, 0, 1, 0, r, 0)$.

Consider $\mathcal{E} \mapsto \mathcal{R} \mapsto \mathcal{E}^\star$. The internal model $E^\star$ is definably isomorphic to $\mathcal{E}$. We choose distinct parameters q and r and s, where qr and qs are orthogonal and of the same length. Then, we send p to (the equivalence class of) (q, r, a, q, s, b), where a and b are the projections of p on the lines qr, resp. qs.

Note that the parameters of the isomorphism also have a parameter domain.

Descartes 6

Thus, RCF and EC are the same in a strong sense. We say that they are *bi-interpretable with parameters.*

Bi-interpretability preserves a number of good properties like finite axiomatizability and κ-categoricity. Bi-interpretability without parameters preserves, up to isomorphism, the automorphism group of a model.

This shows us that the parameters are essential in our example: $\mathbb{R}$ and $\mathbb{EG} = \text{desc}(\mathbb{R})$ have different groups of automorphisms.

There should be an appropriate adaptation to the case with parameters of the theorem of preservation the automorphism group. This should be mainly a matter of getting all the concepts right!

Neumann 1

Consider any interpretation K of a theory U in ZF. We can improve this interpretation modulo definable verifiable isomorphism to a one-dimensional interpretations, since ZF has pairing. We can improve this interpretation modulo definable verifiable isomorphism to an identity preserving interpretation using Scott's trick: we replace every equivalence class of the domain by the set of its elements of minimal rank. Note that this set needs not be *in* the equivalence class.

We interpret Peano Arithmetic PA in ZF via the interpretation neumann by taking as number domain the finite von Neumann ordinals and defining 0, successor, plus and times in the usual way.

This interpretation is *not* faithful since $\mathsf{ZF} \vdash (\mathrm{con}(\mathsf{PA}))^{\mathrm{Neumann}}$. One can show that there is another faithful interpretation of PA in ZF. W.r.t. this interpretation ZF is conservative over PA.

Neumann 2

Let PA^+ be the theory of all arithmetical consequences of ZF via Neumann. The lifted interpretation Neumann^+ is faithful.

One can show that there is an interpretation of ZF in PA^+. Thus ZF and PA^+ are *mutually interpretable*.

Neumann 3

ZF and PA^+ are not bi-interpretable via any pair of interpretations.

Proof: Suppose they were, say via $K : \mathsf{PA}^+ \to \mathsf{ZF}$ and $M : \mathsf{ZF} \to \mathsf{PA}^+$. Start with a model $\mathcal{N}$ of PA^+. We have an internal model $\mathcal{V} = M(\mathcal{N})$ of ZF in $\mathcal{N}$. We note that $\mathcal{N}$ has an initial embedding E_0 into $\mathcal{N}^+ := \mathrm{neumann}(\mathcal{V})$. We have a model $\mathcal{N}^\star = K(\mathcal{V})$ of PA^+ in $\mathcal{V}$. We have an initial embedding E_1 of $\mathcal{N}^+$ into $\mathcal{N}^\star$.

By our assumption of bi-interpretability, we have an $\mathcal{N}$-definable isomorphism F between $\mathcal{N}$ and $\mathcal{N}^\star$. We now show by induction in $\mathcal{N}$ that $E_1 \circ E_0 = F$. Thus $\mathcal{N}^+$ is definably isomorphic to $\mathcal{N}$. But inside $\mathcal{V}$ we have a truthpredicate for $\mathcal{N}^+$. Via F we can transform this into an $\mathcal{N}$-definable truthpredicate for $\mathcal{N}$. This contradicts Tarski's Theorem. $\square$

Overview

Examples

Why Interpretations?

Definitions

Universiteit Utrecht

Uses of Interpretations

We use interpretations . . .

- to explicate intuitions of sameness of theories.
  E.g., mutual interpretability, bi-interpretability.
- to transfer metamathematical information from one theory to another.
  E.g., consistency, essential undecidability.
- to compare theories w.r.t. strength.
  Reverse Mathematics.
- to import conceptual resources of one theory into another. In mathematician's terms: to increase the number of ways of 'seeing things'.
  E.g., in the proof of the Incompleteness Theorems.
- to provide a philosophical reduction of ontologies.
  As a sui-genericist, I am skeptical about this one. *A thing is what it is and not another thing.* (Bishop Joseph Butler).

Overview

Examples

Why Interpretations?

Definitions

The Basic Case: Translations

We first treat one-dimensional, one-piece, relative, non identity preserving interpretations without parameters. We work with relational signatures. Terms are treated using the term-unwinding algorithm.

A *relative translation* $\tau : \Sigma \to \Theta$ is a pair $\langle \delta, F \rangle$.

- δ is Θ-formula with one free variable v_0.
- F associates to R of Σ of arity n a Θ-formula $F(R)$ with variables among $v_0, \ldots, v_{n-1}$.

Induced extension mapping:

- $(R(y_0, \cdots, y_{n-1}))^\tau := F(R)(y_0, \cdots, y_{n-1})$;
- $(\cdot)^\tau$ commutes with propositional connectives;
- $(\forall y\, A)^\tau := \forall y\, (\delta(y) \to A^\tau)$;
- $(\exists y\, A)^\tau := \exists y\, (\delta(y) \wedge A^\tau)$.

The Basic Case: Models

Each translation $\tau : \Sigma \to \Theta$ associates to a model $\mathcal{M}$ of signature Σ an internal model $\widetilde{\tau}(\mathcal{M})$ of signature Θ, provided that $\delta_{\mathcal{M}}$ is not empty. The domain of this model is $\delta_{\mathcal{M}}$, etc.

The Basic Case: Interpretations

An interpretation K is of the form $\langle U, \tau, V \rangle$, where $\tau : \Sigma_U \to \Sigma_V$ and for all U-sentences A, we have: $U \vdash A \Rightarrow V \vdash A^\tau$.

Equivalently, we can demand that for all axioms A of U, including the ones for identity, we have: $U \vdash A \Rightarrow V \vdash A^\tau$.

An interpretation $K : V \rhd U$ is *faithful* iff, for all sentences A in the language of U, $U \vdash A$ iff $V \vdash A^\tau$.

We write for V interprets U:
$K : U \to V$, or $U \xrightarrow{K} V$, or $K : U \lhd V$, or $K : V \rhd U$.

Two interpretations $K, M : U \to V$ are the same if V proves that they are isomorphic via a definable isomorphism.
This isomorphism might contain parameters and have an associated parameter domain.

The Basic Case: Operations

We can define the identity translation id_Σ on Σ in the obvious way. Similarly for the composition $\tau \circ \sigma : \Sigma \to \Xi$ of $\sigma : \Sigma \to \Theta$ and $\tau : \Theta \to \Xi$. E.g.: $\delta_{\tau\circ\sigma}(x) :\leftrightarrow \delta_\tau(x) \wedge \delta^\tau_\sigma(x)$.

Suppose $\sigma, \tau : \Sigma \to \Theta$. Suppose A is in $\mathcal{L}_\Theta$. We define $\nu := \sigma\langle A\rangle\tau : \Sigma \to \Theta$ by:

- $\delta_\nu x :\leftrightarrow (A \wedge \delta_\sigma x) \vee (\neg A \vee \delta_\tau x)$.
- $P_\nu \vec{v} :\leftrightarrow (A \wedge P_\sigma \vec{v}) \vee (\neg A \vee P_\tau \vec{v})$.

These operations lift to operations on interpretations. Thus:

- $\mathrm{id}_U : U \to U$.
- If $K : U \to V$ and $M : V \to W$, then $M \circ K : U \to W$.
- If $K : U \to (V + A)$ and $M : U \to (V + \neg A)$, then $K\langle A\rangle M : U \to V$.

The Basic Case: Category

We can define a category of interpretations as follows. Objects are theories (e.g. one-sorted RE theories of finite signature). Arrows are interpretations modulo the designated notion of sameness.

Bi-interpretability is isomorphism in our category.

With each theory U we associate $\mathrm{MOD}(U)$, the class of models in U, modulo isomorphism. With each interpretation $K : U \to V$ we associate a mapping from models of V to models of U: by $\mathrm{MOD}(K) := \widetilde{K} := \widetilde{\tau}_K$. We have:

$$\mathrm{MOD}(M \circ K) = \mathrm{MOD}(K) \circ \mathrm{MOD}(M).$$

Thus MOD is a contravariant functor.

Much more can be said about our category, but we will not do so in this course.

The Basic Case: Reduction Relations 1

We consider three reduction relations based on interpretions/translations:

- $V \rhd U$ iff, there is a K with $K : V \rhd U$.
  This relation is *interpretability*.
- $V \rhd_{\mathsf{mod}} U$ iff, for all models $\mathcal{M}$ of V, there is an translation τ such that $\widetilde{\tau}(\mathcal{M})$ is a model of U.
  This relation is *model interpretability*.
- $V \rhd_{\mathsf{loc}} U$ iff, for all finitely axiomatized subtheories U_0 of U, $V \rhd U_0$. This relation is *local interpretability*.

Each of these reduction relations gives us a partial preorder on theories and, thus, a degree structure. We use $\equiv$, $\equiv_{\mathsf{mod}}$, $\equiv_{\mathsf{loc}}$ for the induced equivalence relations.

These relations are good for measuring strength, say in the setting of Reverse Mathematics. However, there are many properties they do *not* preserve like decidability.

The Basic Case: Reduction Relations 2

We consider theories of finite signature. Suppose A is finitely axiomatized. We have $U \rhd A$ iff $U \rhd_{\mathsf{mod}} A$.

Proof: Suppose not $U \rhd_{\mathsf{mod}} A$. Let $\tau_0, \tau_1, \ldots,$ enumerate all possible translations from the signature of A to the signature of U. Let:

- $B_i := (\exists x\, \delta_{\tau_i}(x) \wedge A^{\tau_i})$.
- $U^+ := U + \{\neg B_i \mid i \in \omega\}$.
- $\rho_k := \tau_0\langle B_0\rangle(\tau_1\langle B_1\rangle(\ldots\langle B_{k-1}\rangle\tau_k)\ldots)$.

By our assumption, U^+ cannot have a model. So it is inconsistent.

It follows that, for some k, $U \vdash \bigvee_{i \leq k} B_i$.

Ergo, $U \vdash \exists x\, \delta_{\rho_k}(x) \wedge A^{\rho_k}$. Hence, $\langle U, \rho_k, A\rangle$ is an interpretation. $\square$

The Basic Case: Reduction Relations 3

We may conclude that interpretability is included in model interpretability and model interpretability is included in local interpretability.

One can give examples that separate the three kinds of interpretability.

My separating example of model interpretability and interpretability uses some substantial theory. It would be nice to have a ‘low tech’ example. It seems that the problem becomes more difficult if we add more features to our notion of interpretability.

Adding Parameters

We can allow parameters in our interpretations. As long as we consider interpretations of finitely axiomatized theories we do not need a parameter domain. In case we want to interpret an infinitely axiomatized theory we often need a parameter domain. We can take the parameter domain as part of the data for the interpretation and not the translation.

Suppose $K = \langle U, \alpha, \tau, V \rangle$. We define: $K : V \rhd U$ iff

- $V \vdash \exists \vec{x}\, \alpha \vec{x}$,
- $V \vdash \forall \vec{x}\, (\alpha \vec{x} \to \exists y\, \delta_{\tau, \vec{x}} y)$
- For all A, if $U \vdash A$ then $V \vdash \forall \vec{x}\, (\alpha \vec{x} \to A^{\tau, \vec{x}})$.

Adding Dimensions

We can let our domains consist out of sequences of objects of fixed finite length. This does not assume that our interpreting theories have pairing: the sequences are part of the syntax.

There are hairy details concerning the choice of variables. We cannot follow the earlier strategy of translating a variable to itself. We will assume that which free variables $\vec{y}$ in an interpreting formula represent a variable x in the interpreted formula is added as extra data to the translation. E.g., suppose $f(1) = (2, 3)$ and $f(7) = (1, 5)$, then:

$$P_{\tau,f}(x_1, x_7) = F(P)(x_2, x_3, x_1, x_5).$$

Note that we have to demand that the sequences in the range of f have disjoint elements.

Alternatively we can e.g. define interpretations in the framework of Higher Order Term Rewriting, where all details of choice of variables are delegated to a version of the λ-calculus that is built in in the syntax.

Adding Sorts

We can set up the framework for many-sorted theories. We send every sort to a domain of another sort, etc. If we have sorted theories, the notion of piecewise interpretation introduced on the next slide becomes very pleasant to use.

Adding Pieces

We can let our domain consist of overlapping pieces. We consider the example of two pieces δ_0 and δ_1. We will need that at least one of the pieces is non-empty. Translations will have a sequence of pieces as extra data. E.g.,

$$P_{\tau,\{(x,1),(y,0)\}}(x,y) = F_{10}(P)(x,y).$$

Here x is supposed to be in δ_1 and y in δ_0. The clause for the universal quantifier becomes:

$$\begin{aligned}(\forall x\, Ax\vec{y})^{\tau,g} \quad :\leftrightarrow \quad & \forall x\, (\delta_0(x) \to (Ax\vec{y})^{\tau,g\cup\{(x,0)\}}) \ \wedge \\ & \forall x\, (\delta_1(x) \to (Ax\vec{y})^{\tau,g\cup\{(x,1)\}})\end{aligned}$$

We can have pieces of different dimensions and of different sorts! In case we provably have two elements in the domain of our interpreting theory pieces can be eliminated modulo sameness of interpretations by raising the dimension.