Interpretations and Mathematical Logic: a Tutorial

Ali Enayat
Dept. of Philosophy, Linguistics, and Theory of Science
University of Gothenburg
Olof Wijksgatan 6
PO Box 200, SE 405 30 Gothenburg, Sweden
e-mail: ali.enayat@gu.se

1. INTRODUCTION

Interpretations are about 'seeing something as something else', an elusive yet intuitive notion that has been around since time immemorial. It was systematized and elaborated by astrologers, mystics, and theologians, and later extensively employed and adapted by artists, poets, and scientists. At first approximation, an interpretation is a structure-preserving mapping from a realm X to another realm Y, a mapping that is meant to bring hidden features of X and/or Y into view. For example, in the context of Freud's psychoanalytical theory [7], $X =$ the realm of dreams, and $Y =$ the realm of the unconscious; in Khayaam's quatrains [11], $X =$ the metaphysical realm of theologians, and $Y =$ the physical realm of human experience; and in von Neumann's foundations for Quantum Mechanics [14], $X =$ the realm of Quantum Mechanics, and $Y =$ the realm of Hilbert Space Theory.

Turning to the domain of mathematics, familiar geometrical examples of interpretations include: Khayyam's interpretation of algebraic problems in terms of geometric ones (which was the remarkably innovative element in his solution of cubic equations), Descartes' reduction of Geometry to Algebra (which moved in a direction opposite to that of Khayyam's aforementioned interpretation), the Beltrami-Poincaré interpretation of hyperbolic geometry within euclidean geometry [13]. Well-known examples in mathematical analysis include: Dedekind's interpretation of the linear continuum in terms of "cuts" of the rational line, Hamilton's interpertation of complex numbers as points in the Euclidean plane, and Cauchy's interpretation of real-valued integrals via complex-valued ones using the so-called *contour method* of integration in complex analysis.

In this expository note we will focus on interpretations between classical mathematical theories formulated in first order logic[1], for example between ZF (Zermelo-Fraenkel theory of sets) and PA (Peano's axioms for number theory), or between RCF (real-closed fields) and ACF (algebraically closed fields). This notion (relative interpretability) was introduced by Tarski and his colleagues in the landmark monograph [19], where it was developed as an effective tool for establishing powerful undecidability results. On the other hand, Tarski's paper [18] shows that a substantial fragment of Euclidean geometry known as *elementary geometry* is decidable since it is interpretable in RCF (the decidability of RCF is one of Tarski's seminal results). By now there is a large literature in mathematical logic on interpretations and their domain of interest and applicability has been extended far beyond those initially envisaged by Tarski's initial work on the subject that focused on decidability and undecidability issues. In particular, as Visser [17] has emphasized, interpretations allow us to:

- explicate intuitions of "reducible to" and "sameness" of theories and structures;
- transfer certain types of information from one theory or structure to another;
- compare theories and structures with each other; and
- import conceptual resources from one theory to another.

2. PRELIMINARIES

2.1. Definition. Suppose $\mathcal{L}$ is a list[2] of relation symbols $\mathrm{R}_1, \mathrm{R}_2, \cdots$, constant symbols $\mathrm{c}_1, \mathrm{c}_2, \cdots$, and function symbols $\mathrm{f}_1, \mathrm{f}_2, \cdots$. An $\mathcal{L}$-structure is a sequence of the form

$$\mathcal{M} := (M, R_1, R_2, \cdots, c_1, c_2, \cdots, f_1, f_2, \cdots),$$

where each R_n is a relation on M of the same arity as R_n, each f_n is a function on M of the same arity as f_n, and each $c_n \in M$. M is often referred to as the 'domain of discourse' of $\mathcal{M}$.

2.2. Examples.

(a) If $\mathcal{L}$ consists of a single binary function symbol, then $\mathcal{M}_1 := (\mathbb{R}, +)$ and $\mathcal{M}_2 := (\mathcal{P}(\mathbb{N}), \cup)$ are $\mathcal{L}$-structures.

(b) If $\mathcal{L}$ consists of two binary function symbols, one binary relation symbol, and two constant symbols, then the following are $\mathcal{L}$-structures:

[1] Here we will not discuss interpretations between different logics, e.g., between classical logic and intuitionistic logic, or between second order logic and first order logic; nor will we discuss mathematical interpretations of physical theories.

[2] The list is allowed to be empty, or to have uncountably many symbols (there is no *a priori* upper bound on the cardinality of $\mathcal{L}$).

$$\mathcal{M}_1 := (\mathbb{R}, +, \times, \leq, 0, 1)\,, \text{ and } \mathcal{M}_2 := (\mathcal{P}(\mathbb{N}), \cup, \cap, \subseteq, \varnothing, \mathbb{N}).$$

(c) If $\mathcal{L}$ consists of a single binary relation symbol, then $\mathcal{M}_1 := (\mathbb{Z}, <)$ and $\mathcal{M}_2 := (\mathcal{P}(\mathbb{N}), \subseteq)$ are $\mathcal{L}$-structures.

2.3. Definition. Given an $\mathcal{L}$-structure $\mathcal{M} = (M, f, \cdots, R, \cdots, c, \cdots)$, and some $X \subseteq M^n$ (where n is a positive integer) we say that X is $\mathcal{M}$-definable (equivalently: X is definable in $\mathcal{M}$) if there is an $\mathcal{L}$-formula $\phi(x_0, \cdots, x_{n-1})$ (using the logical symbols of first order logic, along with the symbols of $\mathcal{L}$, with precisely n free variables as indicated) such that:

$$X = \{(a_0, \cdots, a_{n-1}) \in M^n : \mathcal{M} \models \phi(a_0, \cdots, a_{n-1})\}\,.$$

2.4 Examples.

(a) The set of non-negative real numbers is definable in $(\mathbb{R}, \times)$ via the formula $\phi(x_0) := \exists y (x_0 = y^2)$.

(b) The usual ordering $\leq$ on $\mathbb{R}$ is definable in $(\mathbb{R}, +, \times)$ via

$$\psi(x, y) := \exists z\, ((x + z = y) \wedge \phi(z))\,,$$

where ϕ is as in (a) above.

(c) The usual ordering $\leq$ on $\mathbb{Z}$ is not definable in $(\mathbb{Z}, +)$. The easiest way of seeing this is to take advantage of the fact that definable subsets are invariant under automorphisms, along with the fact that that the map $f(n) = -n$ is an automorphism of $(\mathbb{Z}, +)$.

(d) Neither $\mathbb{N}$ nor $\mathbb{Q}$ is definable in $(\mathbb{R}, +, \times)$. These two highly nontrivial undefinability facts follow from putting the decidability of the first order theory of $(\mathbb{R}, +, \times)$ (due to Tarski, see [1]) together with the undecidability of the first order theories of $(\mathbb{N}, +, \times)$ (due to Gödel, see [19]) and $(\mathbb{Q}, +, \times)$ (due to Julia Robinson, see [5]).

3. INTERPRETABILITY AMONG STRUCTURES

3.1. Definition. Given two structures $\mathcal{M}$ and $\mathcal{N}$, we say that $\mathcal{M}$ is *interpretable in* $\mathcal{N}$, written $\mathcal{M} \trianglelefteq \mathcal{N}$, if the universe of discourse of $\mathcal{M}$, as well as all the relations, functions, and constants of $\mathcal{M}$, are all definable in $\mathcal{N}$.

3.2. Examples.

(a) $(\mathbb{Q}, +, \times) \trianglelefteq (\mathbb{Z},\ +,\ \times)$ since $(\mathbb{Q}, +, \times) \cong (\mathbb{Z} \times \mathbb{Z}^{\#}\ /\approx,\ \oplus,\ \otimes)$, where:

$$(a, b) \approx (c, d) \text{ iff } ad = bc \text{ and } \oplus,\ \otimes \text{ are defined as follows:}$$

$$[(a, b)] \oplus [(c, d)] = [(ad + bc,\ bd)] \text{ and } [(a, b)] \otimes [(c, d)] = [(ac,\ bd)].$$

(b) $(\mathbb{N}, +, \times) \trianglelefteq (\mathbb{Z}, +, \times)$ since by Lagrange's "four squares theorem" we have:

$$\mathbb{N} = \left\{x \in \mathbb{Z} : \exists a, b, c, d \in \mathbb{Z}\ x = a^2 + b^2 + c^2 + d^2\right\}.$$

(c) $(\mathbb{C}, +, \times) \trianglelefteq (\mathbb{R}, +, \times)$ since $(\mathbb{C}, +, \times) \cong (\mathbb{R} \times \mathbb{R}\ ,\ \oplus,\ \otimes)$, where:

$$[(a, b)] \oplus [(c, d)] = (a + b,\ c + d), \text{ and } (a, b) \otimes (c, d) = (ac - bd,\ ad + bc).$$

(d) Every structure with at least two elements can be interpreted in a lattice. This result is due to Taitslin (1962), see Theorem 5.5.2 in Hodges' text Model Theory [8].

(e) $(\mathbb{R}, +, \times) \ntrianglelefteq (\mathbb{Q}, +, \times)$. This is an immediate consequence of the fact that $\mathbb{R}$ is uncountable but $\mathbb{Q}$ is countable.

(f) $(\mathbb{R}, +, \times) \ntrianglelefteq (\mathbb{C}, +, \times)$. This is a nontrivial result.[3]

4. INTERPRETABILITY AMONG THEORIES

Suppose U and V are first order theories. An interpretation of U in V, written $\mathcal{I} : U \to V$ is given by a *translation* τ from the language of $U$to the language of V, with the requirement that V proves all the translations of sentences of U, i.e.,

$$U \vdash \varphi \Longrightarrow V \vdash \varphi^\tau.$$

The translation (for an n-dimensional interpretation) is given by a *domain formula* $\delta(x_0, \cdots, x_{n-1})$, and a mapping $P \mapsto_\tau A_P$ from the predicates of $U$to formulas of V, where an k-ary predicate P is mapped to a formula $A_P(x_0, \ldots, x_{kn-1})$. In the translation, the equality relation of U is allowed to be translated as an $2n$-ary *formula* E of V. The translation is then lifted to the full first order language in the obvious way making it commute with the propositional connectives and quantifiers, where the translated quantifiers are relativized to the domain specified by δ. See Visser's paper [15] or Friedman's paper [6] for more detail.

Each interpretation $\mathcal{I} : U \to V$gives rise to an *inner model construction* that uniformly builds a model $\mathcal{A}^\mathcal{I} \models U$ from a model of $\mathcal{A} \models V$. In other words, each interpretation

$$\mathcal{I} : U \to V$$

induces a *contravariant* functor

$$F : \mathrm{Mod}(V) \to \mathrm{Mod}(U).$$

[3]For the cognoscenti: this follows from the fact that the first order theory of the complex field is stable, but the first order theory of the real line is unstable.

4.1. Definition. Suppose U and V are first order theories. U is *interpretable* in V, written $U \trianglelefteq V$, if there is an interpretation $\mathcal{I} : U \to V$.
U and V are *mutually interpretable* when $U \trianglelefteq V$ and $V \trianglelefteq U$.

It is easy to see that if U and V are axiomatizable theories, then $U \trianglelefteq V$ implies $\mathrm{Con}(V) \Rightarrow \mathrm{Con}(U)$, but the converse may fail, e.g., consider Zermelo-Fraenkel set theory ZF and Gödel-Bernays theory of classes GB. It has long been known that $\mathrm{Con}(\mathsf{ZF}) \Rightarrow \mathrm{Con}(\mathsf{GB})$, but $\mathsf{GB} \ntrianglelefteq \mathsf{ZF}$. The first result is easy to see with a model-theoretic reasoning that shows that every model of ZF is expandable to a model of GB. The second result follows from combining three fundamental results: Gödel's incompletenss theorem, the reflection theorem for ZF, and the finite axiomatizabiliy of GB.

4.2. Examples.

(a) $\mathsf{PA} \trianglelefteq \mathsf{ZF}$ but $\mathsf{ZF} \ntrianglelefteq \mathsf{PA}$. The first result is folklore and is based on implementing arithmetic in set theory via von Neumann ordinals. The main ingredients of the second proof are Gödel's incompleteness theorem along with the fact that ZF can prove the formal consistency of PA.

(b) The theory of real closed fields RCF interprets the theory of algebraically closed fields, but not vice-versa. The former result follows from a well-known theorem of Artin and Schreier that asserts that adjoining the square root of -1 to any real closed field results in an algebraically closed field. See footnote 3 for the explanation of the latter fact.

(c) $\mathsf{PA} + \neg\mathsf{Con}(\mathsf{PA}) \trianglelefteq \mathsf{PA}$, but $\mathsf{PA} + \mathsf{Con}(\mathsf{PA}) \ntrianglelefteq \mathsf{PA}$. Both of these results are due to Feferman [4].

(d) $\mathsf{ZFC} + \mathsf{GCH} + \mathsf{AC}$ and ZF are mutally interpretable. The nontrivial direction of this fact is due to Gödel, who introduced the class of constructible sets in which, provably in ZF, the axioms of choice (AC) and generalized continuum hypothesis (GCH) hold. See Jech's text [9] for more detail.

(e) $\mathsf{ZFC} + \neg\mathsf{CH} + \neg\mathsf{AC}$ and ZF are mutually interpretable. The nontrivial direction of this fact can be established with the help of the Scott-Solovay approach to proving independence results in set theory. See Jech's text [9] for more detail.

(f) $\mathsf{Q} \trianglelefteq$ First Order Group Theory, where Q is be the theory consisting of the following axioms:

$$S(x) \neq 0, \quad (Sx = Sy) \to x = y, \quad y \neq 0 \to \exists x(Sx = y),$$
$$x + 0 = x, \quad x + Sy = S(x + y), \quad x0 = 0, \quad xSy = (xy) + x$$

This result was established in [19], and was used to show that First Order Group Theory is undecidable.

(g) AST is *mutually interpretable* with Q, where AST (Adjunctive Set Theory) is defined as follows:

$$\mathsf{AST} := \{\mathsf{Emptyset},\ \mathsf{Adjunction}\}, \text{ with:}$$

$$\mathsf{Emptyset} := \exists x\ \forall y\ y \notin x, \text{ and}$$

$$\mathsf{Adjunction} := \forall x\ \forall y\ \exists z\ \forall t\ \underbrace{(t \in z \leftrightarrow (t \in x \vee t = y))}_{z\, =\, x\, \cup\, \{y\}}.$$

This result is due to Visser [16] who refined earlier results by Tarski and others.

4.3. Definition. Suppose U and V are first order theories. U is a *retract* of V if there are interpretations $\mathcal{I}$ and $\mathcal{J}$ with:

$$\mathcal{I} : U \to V \text{ and } \mathcal{J} : V \to U,$$

and a binary U-formula F such that F is, U-verifiably, an isomorphism between id_U and $\mathcal{J} \circ \mathcal{I}$. In model-theoretic terms, this translates to the requirement that the following holds for every model $\mathcal{A} \models U$:

$$F^{\mathcal{A}} : \mathcal{A} \xrightarrow{\cong} \mathcal{A}^* := \left(\mathcal{A}^{\mathcal{I}}\right)^{\mathcal{J}}.$$

U and V are *bi-interpretable* if there are interpretations $\mathcal{I}$ and $\mathcal{J}$ with:

$$\mathcal{I} : U \to V \text{ and } \mathcal{J} : V \to U,$$

and is a binary U-formula F, and a binary V-formula G, such that F is, U-verifiably, an isomorphism between id_U and $\mathcal{J} \circ \mathcal{I}$, and G is, V-verifiably, an isomorphism between id_V and $\mathcal{I} \circ \mathcal{J}$. Note that if U and V are bi-interpretable, then given models $\mathcal{A} \models U$ and $\mathcal{B} \models V$, we have:

$$F^{\mathcal{A}} : \mathcal{A} \xrightarrow{\cong} \mathcal{A}^* := \left(\mathcal{A}^{\mathcal{I}}\right)^{\mathcal{J}} \text{ and } G^{\mathcal{A}} : \mathcal{B} \xrightarrow{\cong} \mathcal{B}^* := \left(\mathcal{B}^{\mathcal{J}}\right)^{\mathcal{I}}.$$

4.4. Examples.

(a) PA is bi-interpretable with $\mathsf{ZF}_{\mathsf{fin}}$+$\mathsf{TC}$, where

$$\mathsf{ZF}_{\mathsf{fin}} := \mathsf{ZF} \backslash \{\mathsf{Infinity}\} \cup \{\neg \mathsf{Infinity}\},$$

and TC is the axiom that states that every set has a transitive closure. This result was established by Kaye and Wong [10]. The precursor of this result is a key theorem of Ackermann, who showed that if one defines the relation E on the set $\mathbb{N}$ of natural numbers by aEb iff the a-th element of the base-2 expansion of b is 1, then we we have:

$$(\mathbb{N}, E) \cong (V_\omega, \in),$$

where V_ω is the set of hereditarily finite sets (the ω-th level of von Neumann's hierarchy of sets).

(b) PA is not bi-interpretable with $\mathsf{ZF_{fin}}$, indeed, $\mathsf{ZF_{fin}}$ is not even a retract of PA (but PA is a retract of $\mathsf{ZF_{fin}}$).This was established in the joint work of the author with Schmerl and Visser [2].

(c) $\mathsf{Z}_2+\Pi^1_\infty$-AC (Second Order Arithmetic + Choice Scheme) is bi-interpretable with $\mathsf{ZFC}\backslash\{\mathsf{Power}\}+\mathbf{V}=\mathrm{H}(\aleph_1)$, and each of these two theories is bi-interpretable with $\mathsf{KM_{fin}}$, where KM is the Kelley-Morse theory of classes, and

$$\mathsf{KM_{fin}} := \mathsf{KM}\backslash\{\mathsf{Infinity}\} \cup \{\neg\mathsf{Infinity}\}.$$

The former bi-interpretability result is detailed in section VII.3 of Simpson's monograph [12], the latter one is folklore.

(d) $\mathsf{KM}+\Pi^1_\infty$-AC is bi-intepretable with $\mathsf{ZFC}\backslash\{\mathsf{Power}\}+\exists\kappa(\kappa$ is strongly inaccessible and $\mathbf{V}=\mathrm{H}(\kappa)$. This result is folklore.

(e) Suppose U and V are deductively closed extensions of PA both of which are formulated in the language of PA; then U is a retract of V iff $V\subseteq U$. In particular, U and V are bi-interpretable iff $U=V$. This result is due to Visser [15]. The same results holds for deductively closed extensions of ZF (with a very different proof), as shown in the author's forthcoming paper [3].

References

[1] C. C. Chang and H. J. Keisler, **Model Theory**, Elsevier North Holland, Amsterdam, 1973.

[2] A. Enayat, J. Schmerl, and A. Visser, *ω-models of finite set theories*, **Set Theory, Arithmetic, and Foundations of Mathematics: Theorems, Philosophies** (edited by J. Kennedy and R. Kossak), Cambridge University Press, 2011, pp. 43-65.

[3] A. Enayat, *Set Theory and interpretations*, in preparation.

[4] S. Feferman, *Arithmetization of metamathematics in a general setting,* **Fund. Math.** 49 1960/1961, pp. 35–92.

[5] D. Flath and S. Wagon, *How to pick out the integers in the rationals: an application of number theory to logic*, **Amer. Math. Monthly** 98 (1991), no. 9, pp. 812–823.

[6] H. Friedman, *Interpretations according to Tarski*, manuscript available at:
`http://www.math.osu.edu/~friedman.8/pdf/Tarski1,052407.pdf`

[7] S. Freud, *Interpretation of Dreams,* 1899, manuscript available at:
`http://psychclassics.yorku.ca/Freud/Dreams/dreams.pdf`

[8] W. Hodges, **Model theory**, Cambridge University Press, Cambridge, 1993.

[9] T. Jech, **Set Theory**, Springer Monographs in Mathematics, Springer-Verlag, 2003.

[10] R. Kaye and T. L. Wong, *On interpretations of arithmetic and set theory*, **Notre Dame J. Formal Logic**, vol. 48 (2007), pp. 497-510.

[11] *The Rubaiyat of Omar Khayyam*, Wikipedia article available at:
`http://en.wikipedia.org/wiki/Rubaiyat_of_Omar_Khayyam`

[12] S. Simpson, **Subsystems of second order arithmetic**, Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1999.

[13] J. Stillwell, *Sources of hyperbolic geometry,* in **History of Mathematics**, vol. 10. American Mathematical Society, Providence, RI; London Mathematical Society, London, 1996.

[14] L. Van Hove, *Von Neumann's contributions to Quantum Theory,* **Bull. Amer. Math. Soc.** 64 (1958).

[15] A. Visser, *Categories of theories and interpretations,* **Logic in Tehran**, Lecture Notes in Logic, vol. 26, Association for Symbolic Logic, La Jolla, CA, 2006, pp. 284–341.

[16] ——————, *Pairs, sets and sequences in first-order theories,* **Arch. Math. Logic** 47 (2008), no. 4, pp. 299–326.

[17] ——————, *Interpretations: lecture 1*, Summer School on Set Theory and Higher Order Logic, London, 2011, available at:
`http://www.bbk.ac.uk/philosophy/our-research/ppp/summer-school`

[18] A. Tarski, *What is elementary geometry?* **The Axiomatic Method,** edited by L. Henkin, P. Suppes and A. Tarski, pp. 16–29, North-Holland Publishing Co., Amsterdam, 1959.

[19] A. Tarski, A. Mostowski, and R. M. Robinson, **Undecidable Theories**, North-Holland Publishing Co., Amsterdam 1968.