On Generalization of Definitional Equivalence to Languages with Non-Disjoint Signatures

Joint research with Gergely Székely (Alfréd Rényi Institute for Mathematics).

Koen Lefever[1]

[1]Centre for Logic and Philosophy of Science
Vrije Universiteit Brussel
http://homepages.vub.ac.be/~kolefeve/

2018-03-05

Origin of this research: 2 questions during my PhD defence:

- Marcoen Cabbolet:
  - Definition of *definitional equivalence* as *intertranslatability* in my dissertation[1] is not compatible with Definition 2 and Example 5 in the paper by Barrett and Halvorson[2].
- Sonja Smets:
  - All theorems are proven syntactically, except Theorem 5 "*There is no interpretation of ClassicalKin in SpecRel*" where the proof is on models (i.e. using semantics).

[1] K. Lefever (2017), "*Using Logical Interpretation and Definitional Equivalence to Compare Classical Kinematics and Special Relativity Theory*", Vrije Universiteit Brussel.

[2] T. W. Barrett and H. Halvorson (2016), "*Glymour and Quine on Theoretical Equivalence*", Journal of Philosophical Logic, Vol. 45, Nr. 5.

Some preliminary definitions:

- A *signature* Σ is a set of predicate symbols (relation symbols)[3].
- A *model* $\mathfrak{M} = \left\langle M, \langle R^{\mathfrak{M}} : R \in \Sigma \rangle \right\rangle$ of signature Σ consists of
  - a non-empty underlying set M,
  - for all relation symbols R of Σ, a relation $R^{\mathfrak{M}} \in M^n$ with the corresponding arity.

[3] function symbols and constant symbols can be written as relation symbols.

- A *sentence* is a formula without free variables.
- A *theory* T is a set of sentences.
- Two theories T_1 and T_2 are *logically equivalent*, in symbols

$$T_1 \equiv T_2,$$

iff they have the same class of models, i.e.,

$$Mod(T_1) = Mod(T_2).$$

- A *language* $\mathcal{L}$ is a set containing a signature, as well as the terms and formulas which can be constructed from that signature using first-order logic.
- Let $\mathcal{L} \subset \mathcal{L}^+$ be two languages. An *explicit definition* of an n-ary relation symbol $p \in \mathcal{L}^+ \setminus \mathcal{L}$ in terms of $\mathcal{L}$ is a sentence of the form

$$\forall x_1 \ldots \forall x_n \big[p(x_1, \ldots, x_n) \leftrightarrow \varphi(x_1, \ldots, x_n) \big],$$

where φ is a formula of $\mathcal{L}$.

- A *definitional extension* of a theory T of language $\mathcal{L}$ to language $\mathcal{L}^+$ is a theory

$$T^+ \equiv T \cup \Delta,$$

where Δ is a set of explicit definitions in terms of language $\mathcal{L}$ for each relation symbol $p \in \mathcal{L}^+ \setminus \mathcal{L}$.

- We denote that T^+ is a definitional extension of T as

$$T \nearrow T^+ \text{ and } T^+ \nwarrow T.$$

- **Definition 2 in Barrett and Halvorson:** Let T_1 be a Σ_1-theory and T_2 be a Σ_2-theory. T_1 and T_2 are *definitionally equivalent*[4], in symbols

$$T_1 \nearrow\!\!\nwarrow T_2,$$

if there is

  - a definitional extension T_1^+ of T_1 to the signature $\Sigma_1 \cup \Sigma_1$ and
  - a definitional extension T_2^+ of T_2 to the signature $\Sigma_1 \cup \Sigma_1$

such that T_1^+ and T_2^+ are logically equivalent.

[4] To avoid confusion, we will from now on call this *definitional mergeability* in stead of *definitional equivalence*.

- A *translation* *tr* of theory T_1 to theory T_2 is a map from $\mathcal{L}_1$ to $\mathcal{L}_2$ which
  - maps every n-ary relation symbol $p \in \mathcal{L}_1$ to a corresponding formula $\varphi_p \in \mathcal{L}_2$ of n with free variables, i.e., $tr\big(p(x_1, \ldots, x_n)\big)$ is $\varphi_p(x_1, \ldots, x_n)$.
  - preserves the equality, logical connectives, and quantifiers, i.e.,
    - $tr(x_1 = x_2)$ is $x_1 = x_2$,
    - $tr(\neg\varphi)$ is $\neg tr(\varphi)$,
    - $tr(\varphi \wedge \psi)$ is $tr(\varphi) \wedge tr(\psi)$, and
    - $tr(\exists x\varphi)$ is $\exists x\big(tr(\varphi)\big)$.
  - maps consequences of T_1 into consequences of T_2, i.e., $T_1 \models \varphi$ implies $T_2 \models tr(\varphi)$ for all sentence $\varphi \in \mathcal{L}_1$.

- **Definition 5 in Barrett and Halvorson:** Let T_1 be a Σ_1-theory and T_2 a Σ_2-theory. T_1 and T_2 are *intertranslatable*[5], in symbols

$$T_1 \rightleftarrows T_2,$$

if there are translations $F : T_1 \to T_2$ and $G : T_2 \to T_1$ such that
  - $T_1 \models (\forall x_1 \dots \forall x_n)\big(\varphi(x_1, \dots, x_n) \leftrightarrow GF\varphi(x_1, \dots, x_n)\big)$
  - $T_2 \models (\forall x_1 \dots \forall x_n)\big(\psi(x_1, \dots, x_n) \leftrightarrow FG\psi(x_1, \dots, x_n)\big)$
    - for every Σ_1-formula $\varphi(x_1, \dots, x_n)$,
    - for every Σ_2-formula $\psi(x_1, \dots, x_n)$.

[5] Definition 4.3.42 in "*Cylindric Algebras Part II*" by L. Henkin, J. D. Monk, and A. Tarski (1985), defines *definitional equivalence* as *intertranslatability*.

Example 5 in Barrett and Halvorson:

- Let $\Sigma = \{p\}$ be the signature containing a unary predicate symbol p. Consider the following two Σ-theories:
  - $T_1 = \{\exists! x(x = x), \forall x\big(p(x)\big)\}$
  - $T_2 = \{\exists! x(x = x), \neg\forall x\big(p(x)\big)\}$
- T_1 and T_2 are not definitionally mergeable since they do not have a common conservative extension,
- but T_1 and T_2 are intertranslatable,
- therefore, $T_1 \nearrow\!\nwarrow T_2$ is not equivalent to $T_1 \rightleftarrows T_2$.

Theorem 1 in my dissertation: intertranslatability is an equivalence relation:

- Reflexivity: $Th \rightleftarrows Th$.
- Symmetry: $Th \rightleftarrows Th' \iff Th' \rightleftarrows Th$.
- Transitivity: $Th_1 \rightleftarrows Th_2, Th_2 \rightleftarrows Th_3 \implies Th_1 \rightleftarrows Th_3$.

Andréka–Németi[6] Definition 4.2:

- Two theories T, T' are *definitionally equivalent*, in symbols

$$T \stackrel{\Delta}{\equiv} T',$$

if there is a chain $T_1, \ldots, T_n$ of theories such that

  - $T = T_1$,
  - $T' = T_n$,
  - for all $1 \leq i < n$ either $T_i \nearrow T_{i+1}$ or $T_i \nwarrow T_{i+1}$.

[6] H. Andréka and I. Németi (2014), "*Definability Theory Course Notes*", https://old.renyi.hu/pub/algebraic-logic/DefThNotes0828.pdf.

Andréka–Németi Definition p. 40 item iv:

- Theories T_1 and T_2 are *model mergeable*, in symbols

$$Mod(T_1) \nearrow\!\nwarrow Mod(T_2),$$

iff there is a bijection β between $Mod(T_1)$ and $Mod(T_2)$ that is defined along two sets Δ_{12} and Δ_{21} of explicit definitions such that if $\mathfrak{M} \in Mod(T_1)$, then

  - the underlying sets of $\mathfrak{M}$ and $\beta(\mathfrak{M})$ are the same,
  - the relations in $\beta(\mathfrak{M})$ are the ones defined in $\mathfrak{M}$ according to Δ_{12} and vice versa, the relations in $\mathfrak{M}$ are the ones defined in $\beta(\mathfrak{M})$ according to Δ_{21}.

Andréka–Németi Theorem 4.2:

- Assume that T and T' have disjoint signatures, i.e. $\Sigma \cap \Sigma' = \emptyset$, then

$$T \stackrel{\Delta}{\equiv} T' \Leftrightarrow T \nearrow\!\nwarrow T' \Leftrightarrow T \rightleftarrows T' \Leftrightarrow Mod(T) \nearrow\!\nwarrow Mod(T').$$

On Generalization of Definitional Equivalence to Languages with Non-Disjoint Signatures
Koen Lefever and Gergely Székely (2018)

- Preprint:
  - http://philsci-archive.pitt.edu/14402/
  - https://arxiv.org/abs/1802.06844

Theorem 1:

- Definitional mergeability $\nearrow\!\nwarrow$ is not transitive. Hence it is not an equivalence relation.
  - Let p and q be unary predicate symbols. Consider the following theories T_1, T_2 and T_3:

$$
\begin{aligned}
T_1 &= \{\,\exists! x(x = x),\ \forall x[p(x)]\,\} \\
T_2 &= \{\,\exists! x(x = x),\ \forall x[\neg p(x)]\,\} \\
T_3 &= \{\,\exists! x(x = x),\ \forall x[q(x)]\,\}
\end{aligned}
$$

  - $T_1 \nearrow\!\nwarrow T_3 \nearrow\!\nwarrow T_2$
  - but T_1 and T_2 are not definitionally mergeable.
- Note that this proof depends on T_1 and T_2 having non-disjoint signatures: $\Sigma_1 = \Sigma_2 = \{p\}$.

Notation:

- If theories T_1 and T_2 are definitionally mergeable and their signatures are disjoint, i.e., $\Sigma_1 \cap \Sigma_2 = \emptyset$, we write

$$T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_2.$$

Theorem 2:

- If theories T_1, T_2 and T_3 are formulated in languages having disjoint signatures and $T_1 \nearrow\!\nwarrow T_2$ and $T_2 \nearrow\!\nwarrow T_3$, then T_1 and T_3 are also mergeable:

$$T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_2 \overset{\emptyset}{\nearrow\!\nwarrow} T_3 \text{ and } \Sigma_1 \cap \Sigma_3 = \emptyset \Longrightarrow T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_3.$$

Theorem 3: Definitional equivalence $\stackrel{\Delta}{\equiv}$ is an equivalence relation:

- Reflexivity: $T \stackrel{\Delta}{\equiv} T$.
- Symmetry: $T \stackrel{\Delta}{\equiv} T' \iff T' \stackrel{\Delta}{\equiv} T$.
- Transitivity: $T_1 \stackrel{\Delta}{\equiv} T_2, T_2 \stackrel{\Delta}{\equiv} T_3 \Longrightarrow T_1 \stackrel{\Delta}{\equiv} T_3$.

Definition:

- Theories T and T' are *disjoint renamings* of each other, in symbols

  $$T \overset{\emptyset}{\simeq} T',$$

  if their signatures Σ and Σ' are disjoint, i.e., $\Sigma \cap \Sigma' = \emptyset$, and there is a renaming bijection $R^{\emptyset}_{\Sigma\Sigma'}$ from Σ to Σ' such that the arity of the relations is preserved and that the formulas in T' are defined by renaming $R^{\emptyset}_{\Sigma\Sigma'}$ of formulas from T.

Remarks:

- Disjoint renaming is symmetric but neither reflexive nor transitive.
- If $T \overset{\emptyset}{\simeq} T'$, then

  $$T \neq T', T \overset{\emptyset}{\nearrow\!\nwarrow} T', T \nearrow\!\nwarrow T', T \overset{\Delta}{\equiv} T' \text{ and } T \rightleftarrows T'.$$

Theorem 4:

- Theories T_1 and T_2 are definitionally equivalent iff there is a theory T_2' which is the disjoint renaming of T_2 to a signature which is also disjoint from the signature of T_1 such that T_2' and T_1 are definitionally mergeable, i.e.,

$$T_1 \stackrel{\Delta}{\equiv} T_2 \iff \exists T' \big[T_1 \stackrel{\emptyset}{\rightarrowtail\!\!\leftarrowtail} T_2' \text{ and } T_2' \stackrel{\emptyset}{\simeq} T_2 \big].$$

Corollary:

- Two theories are definitionally equivalent iff they can be connected by two definitional mergers:

$$T_1 \stackrel{\Delta}{\equiv} T_2 \iff \exists T (T_1 \stackrel{\emptyset}{\rightarrowtail\!\!\leftarrowtail} T \stackrel{\emptyset}{\rightarrowtail\!\!\leftarrowtail} T_2).$$

Consequently, the chain $T_1, \ldots, T_n$ in the definition of definitional equivalence can allways be chosen to be at most length four.

Theorem 5:

- Definitional equivalence $\stackrel{\Delta}{\equiv}$ is the finest equivalence relation containing definitional mergeability $\nearrow\!\!\nwarrow$.
- Definitional equivalence $\stackrel{\Delta}{\equiv}$ is the transitive closure of definitional mergeability $\nearrow\!\!\nwarrow$.

Theorem 6:

- Let T and T' be two theories formulated in languages with disjoint signatures. Then

$$T \overset{\Delta}{\equiv} T' \iff T \overset{\emptyset}{\nearrow\!\!\nwarrow} T' \iff T \rightleftarrows T'.$$

Theorem 7:

- Let T_1 and T_2 be arbitrary theories, then T_1 and T_2 are mergeable iff they are model mergeable:

$$T_1 \nearrow\!\!\nwarrow T_2 \iff Mod(T_1) \nearrow\!\!\nwarrow Mod(T_2).$$

Theorem 8:

- Let T_1 and T_2 be arbitrary theories. Then T_1 and T_2 are definitionally equivalent iff they are intertranslatable:

$$T_1 \stackrel{\Delta}{\equiv} T_2 \iff T_1 \rightleftarrows T_2.$$

Definitions:

- The *relation defined by formula* φ *in* $\mathfrak{M}$ is:

$$\|\varphi\|^{\mathfrak{M}} \stackrel{\text{def}}{=} \left\{\bar{a} \in M^n : \mathfrak{M} \models \varphi[\bar{a}]\right\}.$$

- For all translations $tr_{12} : \mathcal{L}_1 \to \mathcal{L}_2$ of theory T_1 to theory T_2, let tr_{12}^* be defined as the map that maps model $\mathfrak{M} = \langle M, \ldots \rangle$ of T_2 to

$$tr_{12}^*(\mathfrak{M}) \stackrel{\text{def}}{=} \left\langle M, \langle \|tr_{12}(p_i)\|^{\mathfrak{M}} : p_i \in \Sigma_1 \rangle \right\rangle,$$

that is all predicates p_i of Σ_1 interpreted in model $tr_{12}^*(\mathfrak{M})$ as the relation defined by formula $tr_{12}(p_i)$.

Definition:

- Theories T_1 and T_2 are *model intertranslatable*, in symbols

$$Mod(T_1) \rightleftarrows Mod(T_2),$$

iff there are translations $tr_{12} : \mathcal{L}_1 \to \mathcal{L}_2$ of T_1 to T_2 and $tr_{21} : \mathcal{L}_2 \to \mathcal{L}_1$ of T_2 to T_1, such that $tr_{12}^* : Mod(T_2) \to Mod(T_1)$ and $tr_{21}^* : Mod(T_1) \to Mod(T_2)$ are bijections which are inverses of each other.

Theorem 9:

- Let T_1 and T_2 be arbitrary theories, then T_1 and T_2 are intertranslatable iff their models are intertranslatable:

$$T_1 \rightleftarrows T_2 \iff Mod(T_1) \rightleftarrows Mod(T_2).$$

Recalling **Andréka–Németi Theorem 4.2:**

- If T and T' have disjoint signatures, then the following are equivalent:
  - (i) $T \stackrel{\Delta}{\equiv} T'$,
  - (ii) $T \nearrow\!\nwarrow T'$,
  - (iii) $T \rightleftarrows T'$,
  - (iv) $Mod(T) \nearrow\!\nwarrow Mod(T')$.
- item (i) is equivalent to item (iii) by Theorem 6, we have generalized this to arbitrary languages by Theorem 8.

Recalling **Andréka–Németi Theorem 4.2:**

- If T and T' have disjoint signatures, then the following are equivalent:
  - (i) $T \stackrel{\Delta}{\equiv} T'$,
  - (ii) $T \nearrow\!\nwarrow T'$,
  - (iii) $T \rightleftarrows T'$,
  - (iv) $Mod(T) \nearrow\!\nwarrow Mod(T')$.
- the equivalence of items (ii) and (iv) have been generalized to theories in arbitrary languages by Theorem 7.

Recalling **Andréka–Németi Theorem 4.2:**

- If T and T' have disjoint signatures, then the following are equivalent:
  - (i) $T \stackrel{\Delta}{\equiv} T'$,
  - (ii) $T \nearrow\!\!\nwarrow T'$,
  - (iii) $T \rightleftarrows T'$,
  - (iv) $Mod(T) \nearrow\!\!\nwarrow Mod(T')$.
- items (i) and (ii) are indeed equivalent for theories with disjoint signatures by Theorem 6; however, they are not equivalent for theories with non-disjoint signatures by the counterexample in Theorem 1.

Recalling **Andréka–Németi Theorem 4.2:**

- If T and T' have disjoint signatures, then the following are equivalent:
  - (i) $T \stackrel{\Delta}{\equiv} T'$,
  - (ii) $T \nearrow\!\nwarrow T'$,
  - (iii) $T \rightleftarrows T'$,
  - (iv) $Mod(T) \nearrow\!\nwarrow Mod(T')$,
  - **(v)** $Mod(T) \rightleftarrows Mod(T')$.
- We have introduced a model theoretic counterpart of intertranslatability which, by Theorem 9, is equivalent to it even for arbitrary languages.

Conclusion:

- Since definitional mergeability is not transitive, and thus not an equivalence relation, the Barrett–Halvorson generalization is not a well-founded criterion for definitional equivalence when the signatures of theories are not disjoint.
- The Andréka–Németi generalization of definitional equivalence is an equivalence relation. It is also equivalent to intertranslatability and to model-intertranslatability, even for languages with non-disjoint signatures.
- The two generalizations are really close to each-other since the Andréka–Németi generalization is the transitive closure of the Barrett-Halvorson one. Moreover, they only differ in at most one disjoint renaming.