– preprint –

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

Koen Lefever Gergely Székely

February 23, 2018

Abstract

For simplicity, most of the literature introduces the concept of definitional equivalence only to languages with disjoint signatures. In a recent paper, Barrett and Halvorson introduce a straightforward generalization to languages with non-disjoint signatures and they show that their generalization is not equivalent to intertranslatability in general. In this paper, we show that their generalization is not transitive and hence it is not an equivalence relation. Then we introduce the Andréka and Németi generalization as one of the many equivalent formulations for languages with disjoint signatures. We show that the Andréka–Németi generalization is the smallest equivalence relation containing the Barrett–Halvorson generalization and it is equivalent to intertranslatability even for languages with non-disjoint signatures. Finally, we investigate which definitions for definitional equivalences remain equivalent when we generalize them for theories with non-disjoint signatures.

Keywords: First-Order Logic · Definability Theory · Definitional Equivalence · Logical Translation · Logical Interpretation

1 Introduction

Definitional equivalence[1] has been studied and used by both mathematicians and philosophers of science as a possible criterion to establish the equivalence between different theories. This concept was first introduced by Montague in (Montague 1956), but there are already some traces of the idea in (Tarski et al.

[1] Definitional equivalence has also been called *logical synonymity* or *synonymy*, e.g., in (de Bouvère 1965), (Friedman and Visser 2014) and (Visser 2015).

1953). In philosophy of science, it was introduced by Glymour in (Glymour 1970), (Glymour 1977) and (Glymour 1980). Corcoran discusses in (Corcoran 1980) the history of definitional equivalence. In (Andréka et al. 2002, Section 6.3) and (Madarász 2002, Section 4.3), definitional equivalence is generalized to many-sorted definability, where even new entities can be defined and not just new relations between existing entities. (Barrett and Halvorson 2016*a*), on which the present paper is partly a commentary, and (Barrett and Halvorson 2016*b*) contain more references to examples on the use of definitional equivalence in the context of philosophy of science.

We have also recently started in (Lefever and Székely 2018) to use definitional equivalence to study the exact differences and similarities between theories which are *not* equivalent, in that case classical and relativistic kinematics. In that paper, we showed that there exists a translation of relativistic kinematics into classical kinematics, but not the other way round. We also showed that special relativity extended with a "primitive ether" is definitionally equivalent to classical kinematics. Those theories are expressed in the same language, and hence have non-disjoint signatures[2].

Barrett and Halvorson generalize in (Barrett and Halvorson 2016*a*, Definition 2) definitional equivalence from (Hodges 1993, pp. 60-61) for languages having non-disjoint vocabularies in a straightforward way. Then they show that their generalization, which we call here *definitional mergeability* to avoid ambiguity, is not equivalent to intertranslatability in general but only for theories with disjoint signatures. In this paper, we show that definitional mergeability is not an equivalence relation because it is not transitive. Then we recall Andréka and Németi's Definition 4.2 from (Andréka and Németi 2014) which is known to be equivalent to definitional mergeability for languages with disjoint signatures. Then we show that the Andréka–Németi definitional equivalence is the smallest equivalence relation containing definitional mergeablitiy and that it is equivalent to intertranslatability even for theories with languages with non-disjoint signatures. Actually, two theories are definitional equivalent iff there is a theory that is definitionally mergeable to both of them. Moreover, one of these definitional mergers can be a renaming.

Theorem 4.2 of (Andréka and Németi 2014) claims that (i) definitional equivalence, (ii) definitional mergeability, (iii) intertranslatability and (iv) model mergeability (see Definition 13 below) are equivalent in case of disjoint signatures. Here, we show that the equivalence of (i) and (iii) and that of (ii) and (iv) hold for arbitrary languages, see Theorems 8 and 7. However, since (i) and (ii) are not equivalent by Theorems 1 and 3, no other equivalence of extends to arbi-

[2]For a variant of this result in which we explicitly made the signatures disjoint, see (Lefever 2017).

trary languages. Finally, we introduce a modification of (iv) that is equivalent to (i) and (iii) for arbitrary languages, see Theorem 9.

2 Framework and definitions

Definition 1. A *signature*[3] Σ is a set of predicate symbols (relation symbols), function symbols, and constant symbols.

Definition 2. A *first-order 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.

Remark 1. For every theory T which might contain constants and functions, there is another theory T' which is formulated in a language containing only relation symbols and connected to T by all the relations investigated in this paper as candidates for definitional equivalence, see (Barrett and Halvorson 2016*a*, Proposition 2 and Theorem 1). Therefore, here we only consider languages containing only relation symbols.

Definition 3. A *sentence* is a formula without free variables.

Definition 4. A *theory* T is a set of sentences expressed in language $\mathcal{L}$.

Convention 1. We will use the notations Σ_x, Σ', etc. for the signatures, and $\mathcal{L}_x$, $\mathcal{L}'$, etc. for the languages of respective theories T_x, T', etc.

Definition 5. A *model* $\mathfrak{M} = \langle M, \langle R^{\mathfrak{M}} : R \in \Sigma \rangle \rangle$ of signature Σ consists of a non-empty underlying set[4] M, and for all relation symbols R of Σ, a relation $R^{\mathfrak{M}} \in M^n$ with the corresponding arity[5].

Definition 6. Let $\mathfrak{M}$ be a model, let M be the non-empty underlying set of $\mathfrak{M}$, let φ be a formula, let V be the set of variables and let $e : V \to M$ be an evaluation of variables, then we inductively define that e *satisfies* φ *in* $\mathfrak{M}$, in symbols

$$\mathfrak{M} \models \varphi[e],$$

as:

1. For predicate R, $\mathfrak{M} \models R(x, y, \dots, z)[e]$ holds if

$$\big(e(x), e(y), \dots, e(z)\big) \in R^{\mathfrak{M}},$$

[3]In (Andréka and Németi 2014), a *signature* is called a *vocabulary*. Since this paper is partly a comment on (Barrett and Halvorson 2016*a*), we will use their terminology, which is also being used in (Hodges 1993) and (Hodges 1997).

[4]The non-empty underlying set M is also called the *universe*, the *carrier* or the *domain* of $\mathfrak{M}$.

[5]The *arity* n is the number of variables in the relation, it is also called the *rank*, *degree*, *adicity* or *valency* of the relation. M^n denotes the Cartesian power of set M.

2. $\mathfrak{M} \models (x = y)[e]$ holds if $e(x) = e(y)$ holds,

3. $\mathfrak{M} \models \neg\varphi[e]$ holds if $\mathfrak{M} \models \varphi[e]$ does not hold,

4. $\mathfrak{M} \models (\psi \wedge \theta)[e]$ holds if both $\mathfrak{M} \models \psi[e]$ and $\mathfrak{M} \models \theta[e]$ hold,

5. $\mathfrak{M} \models \big(\exists y\psi\big)[e]$ holds if there is an element $b \in M$, such that $\mathfrak{M} \models \psi[e']$ if $e'(y) = b$ and $e'(x) = e(x)$ if $x \neq y$.

Let $\bar{x}$ be the list of all free variables of φ and let $\bar{a}$ be a list of elements of M with the same number of elements as $\bar{x}$. Then $\mathfrak{M} \models \varphi[\bar{a}]$ iff $\mathfrak{M}$ satsfies[6] φ for all (or equivalently some) evaluation e of variables for which $e(\bar{x}) = \bar{a}$, i.e., variables in $\bar{x}$ are mapped to elements of M in $\bar{a}$ in order. In case φ is a sentence, its truth does not depend on evaluation of variables. So that φ is true in $\mathfrak{M}$ is denoted by $\mathfrak{M} \models \varphi$. For theory T, $\mathfrak{M} \models T$ abbreviates that $\mathfrak{M} \models \varphi$ for all $\varphi \in T$.

Remark 2. We will use $\varphi \vee \psi$ as an abbreviation for $\neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi$ for $\neg\varphi \vee \psi$, $\varphi \leftrightarrow \psi$ for $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$ and $\forall x(\varphi)$ for $\neg\big(\exists x(\neg\varphi)\big)$.

Definition 7. $Mod(T)$ is the class of models of theory T,

$$Mod(T) \stackrel{\text{def}}{=} \{\mathfrak{M} : \mathfrak{M} \models T\}.$$

Definition 8. Two theories T_1 and T_2 are logically equivalent, in symbols

$$T_1 \equiv T_2,$$

iff[7] they have the same class of models, i.e., $Mod(T_1) = Mod(T_2)$.

Definition 9. 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 $\mathcal{L}$.

Definition 10. A *definitional extension*[8] 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 $p \in \mathcal{L}^+ \setminus \mathcal{L}$. In this paper,

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

denote that T^+ is a definitional extension of T.

[6] $\mathfrak{M} \models \varphi[\bar{a}]$ can also be read as $\varphi[\bar{a}]$ *being true in* $\mathfrak{M}$.

[7] *iff* abbreviates *if and only if*. It is denoted by $\leftrightarrow$ in the object languages (see remark 2 above) and by $\iff$ in the meta-language.

[8] We follow the definition from (Andréka and Németi 2014, Section 4.1, p.36), (Hodges 1993, p.60) and (Hodges 1997, p.53). In (Barrett and Halvorson 2016*a*, Section 3.1), the logical equivalence relation is not part of the definition.

We will use Δ_{xy} to denote the set of explicit definitions when the signature Σ_y of theory T_y is defined in terms of the signature Σ_x of theory T_x.

Definition 11. 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 $T = T_1$, $T' = T_n$, and for all $1 \leq i < n$ either $T_i \nearrow T_{i+1}$ or $T_i \nwarrow T_{i+1}$.

Remark 3. If a theory is consistent, then all theories which are definitionally equivalent to that theory are also consistent since definitions cannot make consistent theories inconsistent. Similarly, if a theory is inconsistent, then all theories which are definitionally equivalent to that theory are also inconsistent.

Definition 12. Let T_1 and T_2 be theories of languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. T_1 and T_2 are *definitionally mergeable*, in symbols

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

if there is a theory T^+ which is a common definitional extension of T_1 and T_2, i.e., $T_1 \nearrow T^+ \nwarrow T_2$.

Remark 4. From Definition 11 and Definition 12, it is immediately clear that being definitionally mergeable is a special case of being definitionally equivalent.

Lemma 1 below establishes that our Definition 12 of definitional mergeability is equivalent to the definition for definitional equivalence in (Barrett and Halvorson 2016*a*, Definition 2).

Lemma 1. Let T_1 and T_2 be two arbitrary theories. Then $T_1 \nearrow\!\nwarrow T_2$ iff there are sets of explicit definitions Δ_{12} and Δ_{21} such that $T_1 \cup \Delta_{12} \equiv T_2 \cup \Delta_{21}$.

Proof. Let $T_1 \nearrow\!\nwarrow T_2$, then there exists a T^+ such that $T_1 \nearrow T^+ \nwarrow T_2$. By the definition of definitional extension, there exist sets of explicit definitions Δ_{12} and Δ_{21} such that $T_1 \cup \Delta_{12} \equiv T^+$ and $T_2 \cup \Delta_{21} \equiv T^+$, and hence by transitivity of logical equivalence $T_1 \cup \Delta_{12} \equiv T_2 \cup \Delta_{21}$.

To prove the other direction: let T_1 and T_2 be theories such that $T_1 \cup \Delta_{12} \equiv T_2 \cup \Delta_{21}$ for some sets Δ_{12} and Δ_{21} of explicit definitions. Let $T^+ = T_1 \cup T_2 \cup \Delta_{12} \cup \Delta_{21}$. Hence $T_1 \cup \Delta_{12} \equiv T^+ \equiv T_2 \cup \Delta_{21}$ and $T_1 \nearrow T^+ \nwarrow T_2$, and therefore $T_1 \nearrow\!\nwarrow T_2$. □

Convention 2. 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

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

Definition 13. Theories T_1 and T_2 are *model mergeable*[9], 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}.

Definition 14. Let T_1 and T_2 be theories. A *translation*[10] 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., $tr\big(p(x_1, \dots, x_n)\big)$ is $\varphi_p(x_1, \dots, x_n)$.
- preserves the equality, logical connectives, and quantifiers, i.e.,
  - $tr(x_1 = x_2)$ is $x_1 = x_2$,
  - $tr(\neg\varphi)$ is $\neg tr(\varphi)$,
  - $tr(\varphi \land \psi)$ is $tr(\varphi) \land tr(\psi)$, 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 $\varphi \in \mathcal{L}_1$.

Remark 5. From (Andréka et al. 2005), we know that T being translatable into T' and T' being translatable into T is not a sufficient condition for $T \stackrel{\Delta}{\equiv} T'$.

Definition 15. Theories T_1 and T_2 are *intertranslatable*[11], in symbols

$$T_1 \rightleftarrows T_2,$$

if there are translations tr_{12} of T_1 to T_2 and tr_{21} of T_2 to T_1 such that

- $T_1 \models \forall x_1 \dots \forall x_n \big[\varphi(x_1, \dots, x_n) \leftrightarrow tr_{21}\big(tr_{12}\big(\varphi(x_1, \dots, x_n)\big)\big)\big]$
- $T_2 \models \forall x_1 \dots \forall x_n \big[\psi(x_1, \dots, x_n) \leftrightarrow tr_{21}\big(tr_{12}\big(\psi(x_1, \dots, x_n)\big)\big)\big]$

[9] We use the definition from (Andréka and Németi 2014, p. 40, item iv), which is a variant of the definition in (Henkin et al. 1971, p. 56, Remark 0.1.6).

[10] In (Andréka and Németi 2014), (Lefever 2017) and (Lefever and Székely 2018), this is called an *interpretation*, but we again follow the terminology from (Barrett and Halvorson 2016*a*) here.

[11] In (Henkin et al. 1985, p. 167, Definition 4.3.42), definitional equivalence is defined as intertranslatability.

for every formulas $\varphi(x_1, \ldots, x_n)$ and formula $\psi(x_1, \ldots, x_n)$ of languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively.

For a direct proof that intertranslatability is an equivalence relation, see e.g., (Lefever 2017, Theorem 1, p. 7). This fact also follows from Theorems 3 and 8 below.

Definition 16. The *relation defined by formula φ in $\mathfrak{M}$* is[12]:

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

Definition 17. 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

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

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

Lemma 2. Let $\mathfrak{M}$ be a model of language $\mathcal{L}_2$, let φ be a formula of language $\mathcal{L}_1$, and let $e : V \to M$ be an evaluation of variables. If $tr_{12} : \mathcal{L}_1 \to \mathcal{L}_2$ is translation of T_1 to T_2, then

$$tr^*_{12}(\mathfrak{M}) \models \varphi[e] \iff \mathfrak{M} \models tr_{12}(\varphi)[e]$$

Proof. We are going to prove Lemma 2 by induction on the complexity of φ. So let us first assume that φ is a single predicate p of language $\mathcal{L}_1$.

Let $\bar{u}$ be the e-image of the free variables of p. Then $tr^*_{12}(\mathfrak{M}) \models p[e]$ holds exactly if $tr^*_{12}(\mathfrak{M}) \models p[\bar{u}]$. By Definition 17, this holds iff

$$\Big\langle M, \big\langle \|tr_{12}(p_i)\|^{\mathfrak{M}} : p_i \in \Sigma_1 \big\rangle \Big\rangle \models p[\bar{u}]. \tag{1}$$

By Definition 16, $\|tr_{12}(p)\|^{\mathfrak{M}} = \big\{\bar{a} \in M^n : \mathfrak{M} \models tr_{12}(p)[\bar{a}]\big\}$. So (1) is equivalent to $\mathfrak{M} \models tr_{12}(p)[\bar{u}]$.

If φ is $x = y$, then we should show that

$$tr^*_{12}(\mathfrak{M}) \models (x = y)[e] \iff \mathfrak{M} \models tr_{12}(x = y)[e].$$

Since translations preserve mathematical equality by Definition 14, this is equivalent to

$$tr^*_{12}(\mathfrak{M}) \models (x = y)[e] \iff \mathfrak{M} \models (x = y)[e],$$

which holds because the underlying sets of $tr^*_{12}(\mathfrak{M})$ and $\mathfrak{M}$ are the same and both sides of the equivalence are equivalent to $e(x) = e(y)$ by Definition 6.

Let us now prove the more complex cases by induction on the complexity of formulas.

[12] $\|\varphi\|^{\mathfrak{M}}$ is basically the same as the *meaning* of formula φ in model $\mathfrak{M}$, see (Andréka et al. 2001, p. 194 Definition 34 and p. 231 Example 8).

- If φ is $\neg\psi$, then we should show that

$$tr^*_{12}(\mathfrak{M}) \models \neg\psi[e] \iff \mathfrak{M} \models tr_{12}(\neg\psi)[e].$$

Since tr_{12} is a translation, it preserves (by Definition 14) the conectives, and therefore this is equivalent to

$$tr^*_{12}(\mathfrak{M}) \models \neg\psi[e] \iff \mathfrak{M} \models \neg tr_{12}(\psi)[e],$$

which holds by Definition 6 Item 3 since we have

$$tr^*_{12}(\mathfrak{M}) \models \psi[e] \iff \mathfrak{M} \models tr_{12}(\psi)[e]$$

by induction.

- If φ is $(\psi \wedge \theta)$, then we should show that

$$tr^*_{12}(\mathfrak{M}) \models (\psi \wedge \theta)[e] \iff \mathfrak{M} \models tr_{12}(\psi \wedge \theta)[e].$$

Since tr_{12} is a translation, it preserves (by Definition 14) the conectives, and therefore $tr_{12}(\psi \wedge \theta)$ is equivalent to $tr_{12}(\psi) \wedge tr_{12}(\theta)$, and hence the above is equivalent to

$$tr^*_{12}(\mathfrak{M}) \models (\psi \wedge \theta)[e] \iff \mathfrak{M} \models \big(tr_{12}(\psi) \wedge tr_{12}(\theta)\big)[e],$$

which holds by Definition 6 Item 4 because both

$$tr^*_{12}(\mathfrak{M}) \models \psi[e] \iff \mathfrak{M} \models tr_{12}(\psi)[e]$$

and

$$tr^*_{12}(\mathfrak{M}) \models \theta[e] \iff \mathfrak{M} \models tr_{12}(\theta)[e]$$

hold by induction.

- If φ is $\exists y(\psi)$, then we should show that

$$tr^*_{12}(\mathfrak{M}) \models \big(\exists y(\psi)\big)[e] \iff \mathfrak{M} \models tr_{12}\big(\exists y(\psi)\big)[e]$$

holds. Since tr_{12} is a translation, it preserves (by Definition 14) the quantifiers, and hence this is equivalent to

$$tr^*_{12}(\mathfrak{M}) \models \big(\exists y(\psi)\big)[e] \iff \mathfrak{M} \models \big(\exists y\big(tr_{12}(\psi)\big)\big)[e].$$

By Definition 6 Item 5, both sides of he equivalence hold exactly if there exists an element $b \in M$ such that

$$tr^*_{12}(\mathfrak{M}) \models \psi[e'] \iff \mathfrak{M} \models tr_{12}(\psi)[e'],$$

where $e'(y) = b$ and $e'(x) = e(x)$ if $x \neq y$, which holds by induction because the underlying sets of $tr^*_{12}(\mathfrak{M})$ and $\mathfrak{M}$ are the same. $\square$

Corollary 1. If $tr_{12} : \mathcal{L}_1 \to \mathcal{L}_2$ is a translation of T_1 to T_2, then

$$tr_{12}^* : Mod(T_2) \to Mod(T_1),$$

that is, tr_{12}^* is a map from $Mod(T_2)$ to $Mod(T_1)$.

Proof. Let $\mathfrak{M}$ be a model of T_2 and let $\varphi \in T_1$. We should prove that $tr_{12}^*(\mathfrak{M}) \models \varphi$. By Lemma 2, we have that

$$tr_{12}^*(\mathfrak{M}) \models \varphi \iff \mathfrak{M} \models tr_{12}(\varphi).$$

Hence $tr_{12}(\varphi)$ is true in every model of T_2 as we wanted to prove. □

Remark 6. Note that while tr_{12} is a translation of T_1 to T_2, tr_{12}^* translates models the other way round from $Mod(T_2)$ to $Mod(T_1)$. For an example illustrating this for a translation from relativistic kinematics to classical kinematics, see (Lefever 2017, Chapter 7) or (Lefever and Székely 2018, Section 7).

Definition 18. 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.

Definition 19. 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.[13]

Remark 7. Note that disjoint renaming is symmetric but neither reflexive nor transitive. Also, if $T \overset{\emptyset}{\simeq} T'$, then $T \neq T'$, $T \overset{\emptyset}{\nearrow\!\nwarrow} T'$, $T \nearrow\!\nwarrow T'$, $T \overset{\Delta}{\equiv} T'$ and $T \rightleftarrows T'$.

3 Properties

Theorem 1. Definitional mergeability $\nearrow\!\nwarrow$ is not transitive. Hence it is not an equivalence relation.

The proof is based on (Barrett and Halvorson 2016*a*, Example 5). Note that the proof relies on the signatures of theories T_1 and T_2 being non-disjoint.

[13]While bijection $R^{\emptyset}_{\Sigma\Sigma'}$ is defined on signatures, it can be naturally extended to the languages using those signatures. We will use the same symbol $R^{\emptyset}_{\Sigma\Sigma'}$ for that.

Proof. 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 and T_2 are not definitionally mergeable, since they do not have a common extension as they contradict each other[14].

Let us define T_1^+ where q is defined in terms of T_1 as p and let us define T_3^+ where p is defined in terms of T_3 as q, i.e.,

$$\begin{aligned} T_1^+ &= \{\, \exists! x(x = x),\ \forall x[p(x)],\ \forall x[q(x) \leftrightarrow p(x)] \,\} \\ T_3^+ &= \{\, \exists! x(x = x),\ \forall x[q(x)],\ \forall x[p(x) \leftrightarrow q(x)] \,\}. \end{aligned}$$

Then T_1 and T_3 are definitionally mergeable because $T_1 \nearrow T_1^+$, $T_3 \nearrow T_3^+$, and $T_1^+ \equiv T_3^+$.

Let us now define T_2^+ where q is defined in terms of T_2 as $\neg p$ and let us define $T_3^\times$ where p is defined in terms of T_3 as $\neg q$, i.e.,

$$\begin{aligned} T_2^+ &= \{\, \exists! x(x = x),\ \forall x[\neg p(x)],\ \forall x[q(x) \leftrightarrow \neg p(x)] \,\} \\ T_3^\times &= \{\, \exists! x(x = x),\ \forall x[q(x)],\ \forall x[p(x) \leftrightarrow \neg q(x)] \,\}. \end{aligned}$$

Then T_2 and T_3 are definitionally mergeable because $T_2 \nearrow T_2^+$, $T_3 \nearrow T_3^\times$, and $T_2^+ \equiv T_3^\times$.

Therefore, being definitionally mergeable is not transitive and hence not an equivalence relation as $T_1 \nearrow\!\nwarrow T_3 \nearrow\!\nwarrow T_2$ but T_1 and T_2 are not definitionally mergeable. $\square$

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, i.e.,

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

Proof. Let T_1, T_2 and T_3 be theories such that $\Sigma_1 \cap \Sigma_3 = \emptyset$ and $T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_2 \overset{\emptyset}{\nearrow\!\nwarrow} T_3$.

We have from the definitions of definitional equivalence and definitional extension that there exist sets Δ_{12}, Δ_{21}, Δ_{23} and Δ_{32} of explicit definitions, such that

$$T_1 \cup \Delta_{12} \equiv T_2 \cup \Delta_{21}, \text{ i.e., } Mod(T_1 \cup \Delta_{12}) = Mod(T_2 \cup \Delta_{21}), \tag{2}$$

[14] $\exists!$ is an abbreviation for “there exists exactly one”, i.e.,

$$\exists! x\big(\varphi(x)\big) \iff \exists x\Big(\varphi(x) \wedge \neg\exists y\big(\varphi(y) \wedge x \neq y\big)\Big).$$

and

$$T_2 \cup \Delta_{23} \equiv T_3 \cup \Delta_{32}, \text{ i.e., } Mod(T_2 \cup \Delta_{23}) = Mod(T_3 \cup \Delta_{32}). \tag{3}$$

We want to prove that $T_1 \cup \Delta_{12} \cup \Delta_{23} \equiv T_3 \cup \Delta_{32} \cup \Delta_{21}$, i.e.,

$$Mod(T_1 \cup \Delta_{12} \cup \Delta_{23}) = Mod(T_3 \cup \Delta_{32} \cup \Delta_{21}).$$

If one of the theories T_1, T_2 or T_3 is inconsistent, then by Remark 3, all of them are inconsistent. In that case $T_1 \nearrow\!\!\nwarrow T_3$ is true because all statements can be proven ex falso in both theories. Let us for the rest of the proof now assume that all of them are consistent.

Let $\mathfrak{M} \in Mod(T_1 \cup \Delta_{12} \cup \Delta_{23})$. Such $\mathfrak{M}$ exists because $\Sigma_1 \cap \Sigma_2 = \emptyset$ and hence Δ_{23} cannot make consistent theory $T_1 \cup \Delta_{12}$ inconsistent.

Then $\mathfrak{M} \models T_1 \cup \Delta_{12} \cup \Delta_{23}$. Therefore $\mathfrak{M} \models T_2 \cup \Delta_{21}$ by (2) and also $\mathfrak{M} \models T_3 \cup \Delta_{32}$ because of (3) and the fact that $\mathfrak{M} \models \Delta_{23}$. Hence $\mathfrak{M} \models T_3 \cup \Delta_{32} \cup \Delta_{21}$. Consequently,

$$Mod(T_1 \cup \Delta_{12} \cup \Delta_{23}) \subseteq Mod(T_3 \cup \Delta_{32} \cup \Delta_{21}).$$

An analogous calculation shows that

$$Mod(T_1 \cup \Delta_{12} \cup \Delta_{23}) \supseteq Mod(T_3 \cup \Delta_{32} \cup \Delta_{21}).$$

So $Mod(T_1 \cup \Delta_{12} \cup \Delta_{23}) = Mod(T_3 \cup \Delta_{32} \cup \Delta_{21})$ and this is what we wanted to prove. $\square$

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

Proof. To show that definitional equivalence is an equivalence relation, we need to show that it is reflexive, symmetric and transitive:

- $\stackrel{\Delta}{\equiv}$ is reflexive because for every theory $T \nearrow T$ since the set of explicit definitions Δ can be the empty set, and hence $T \stackrel{\Delta}{\equiv} T$.
- $\stackrel{\Delta}{\equiv}$ is symmetric: if $T \stackrel{\Delta}{\equiv} T'$, then there exists a chain $T \dots T'$ of theories connected by $\equiv$, $\nearrow$ and $\nwarrow$. The reverse chain $T' \dots T$ has the same kinds of connections, and hence $T' \stackrel{\Delta}{\equiv} T$.
- $\stackrel{\Delta}{\equiv}$ is transitive: if $T_1 \stackrel{\Delta}{\equiv} T_2$ and $T_2 \stackrel{\Delta}{\equiv} T_3$, then there exists chains $T_1 \dots T_2$ and $T_2 \dots T_3$ of theories connected by $\equiv$, $\nearrow$ and $\nwarrow$. The concatenated chain $T_1 \dots T_2 \dots T_3$ has the same kinds of connections, and hence $T_1 \stackrel{\Delta}{\equiv} T_3$. $\square$

Lemma 3. If $T_1 \stackrel{\Delta}{\equiv} T_2$, then there exists a chain of definitional mergers such that

$$T_1 \nearrow\!\!\nwarrow T_a \nearrow\!\!\nwarrow \dots \nearrow\!\!\nwarrow T_z \nearrow\!\!\nwarrow T_2.$$

Proof. The finite chain of steps given by Definition 11 for definitional equivalence can be extended by adding extra extension steps $\nearrow$ or $\nwarrow$ wherever needed in the chain because definitional extension is reflexive since the set of explicit definitions Δ can be the empty set. $\square$

Lemma 4. Let T_a and T_b two theories for which $T_a \nearrow\!\!\nwarrow T_b$. Then

- if $T_b \overset{\emptyset}{\simeq} T'_b$ and $\Sigma_a \cap \Sigma'_b = \emptyset$, then $T_a \overset{\emptyset}{\nearrow\!\!\nwarrow} T'_b$,
- if $T_a \overset{\emptyset}{\simeq} T'_a$, $T_b \overset{\emptyset}{\simeq} T'_b$ and $\Sigma'_a \cap \Sigma'_b = \emptyset$, then $T'_a \overset{\emptyset}{\nearrow\!\!\nwarrow} T'_b$.

Proof. Since $T_a \nearrow\!\!\nwarrow T_b$, there are by Lemma 1 sets Δ_{ab} and Δ_{ba} of explicite definitions such that $T_a \cup \Delta_{ab} \equiv T_b \cup \Delta_{ba}$:

$$\Delta_{ab} = \{\forall \bar{x}\, [p(\bar{x}) \leftrightarrow \varphi_p(\bar{x})] : p \in \Sigma_b \text{ and } \varphi_p \in \mathcal{L}_a\}\,,$$

i.e., φ_p is the definition of predicate p from Σ_b in language $\mathcal{L}_a$.

$$\Delta_{ba} = \{\forall \bar{x}\, [q(\bar{x}) \leftrightarrow \varphi_q(\bar{x})] : q \in \Sigma_a \text{ and } \varphi_q \in \mathcal{L}_b\}\,,$$

i.e., φ_q is the definition of predicate q from Σ_a in language $\mathcal{L}_b$. We can now define $\Delta_{ab'}$ and $\Delta_{b'a}$ in the following way:

$$\Delta_{ab'} \overset{\text{def}}{=} \left\{\forall \bar{x} \left[R^{\emptyset}_{\Sigma_b\Sigma'_b}(p)(\bar{x}) \leftrightarrow \varphi_p(\bar{x})\right] : p \in \Sigma_b \text{ and } \varphi_p \in \mathcal{L}_a\right\},$$

i.e., in $\Delta_{ab'}$ the renaming $R^{\emptyset}_{\Sigma_b\Sigma'_b}(p)$ of predicate p from Σ_b is defined with the same formula φ_p as p was defined in Δ_{ab}.

$$\Delta_{b'a} \overset{\text{def}}{=} \left\{\forall \bar{x} \left[q(\bar{x}) \leftrightarrow R^{\emptyset}_{\Sigma_b\Sigma'_b}(\varphi_q)(\bar{x})\right] : q \in \Sigma_a \text{ and } \varphi_q \in \mathcal{L}_b\right\},$$

i.e., in $\Delta_{b'a}$ predicate q from Σ_a is defined with the renaming $R^{\emptyset}_{\Sigma_b\Sigma'_b}(\varphi_q)$ of the formula φ_q that was used in Δ_{ba} to define q.

Then $T_a \cup \Delta_{ab'} \equiv T'_b \cup \Delta_{b'a}$, and hence we have proven that $T_a \overset{\emptyset}{\nearrow\!\!\nwarrow} T'_b$.

Similarly, we can define $\Delta_{a'b'}$ and $\Delta_{b'a'}$ as:

$$\Delta_{a'b'} \overset{\text{def}}{=} \left\{\forall \bar{x} \left[R^{\emptyset}_{\Sigma_b\Sigma'_b}(p)(\bar{x}) \leftrightarrow R^{\emptyset}_{\Sigma_a\Sigma'_a}(\varphi_p)(\bar{x})\right] : p \in \Sigma_b \text{ and } \varphi_p \in \mathcal{L}_a\right\},$$

i.e., in $\Delta_{a'b'}$ the renaming $R^{\emptyset}_{\Sigma_b\Sigma'_b}(p)$ of predicate p from Σ_b is defined with the renaming $R^{\emptyset}_{\Sigma_b\Sigma'_b}(\varphi_p)$ of the formula φ_p that was used in Δ_{ab} to define p.

$$\Delta_{b'a'} \overset{\text{def}}{=} \left\{\forall \bar{x} \left[R^{\emptyset}_{\Sigma_a\Sigma'_a}(q)(\bar{x}) \leftrightarrow R^{\emptyset}_{\Sigma_b\Sigma'_b}(\varphi_q)(\bar{x})\right] : q \in \Sigma_a \text{ and } \varphi_q \in \mathcal{L}_b\right\},$$

i.e., in $\Delta_{b'a'}$ the renaming $R^{\emptyset}_{\Sigma_b\Sigma'_b}(q)$ of predicate q from Σ_a is defined with the renaming $R^{\emptyset}_{\Sigma_b\Sigma'_b}(\varphi_q)$ of the formula φ_q that was used in Δ_{ba} to define q.

Then $T'_a \cup \Delta_{a'b'} \equiv T'_b \cup \Delta_{b'a'}$, and hence we have proven that $T'_a \overset{\emptyset}{\nearrow\!\!\nwarrow} T'_b$. $\square$

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.,

$$T_1 \overset{\Delta}{\equiv} T_2 \iff \exists T' \big[T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_2' \text{ and } T_2' \overset{\emptyset}{\simeq} T_2 \big].$$

Proof. Let T_1 and T_2 be definitional equivalent theories. From Lemma 3, we know that there exists a finite chain of definitonal mergers

$$T_1 \nearrow\!\nwarrow T_a \nearrow\!\nwarrow \dots \nearrow\!\nwarrow T_z \nearrow\!\nwarrow T_2.$$

For all x in $\{a, \dots, z, 2\}$, let T_x' be a renaming of T_x such that $\Sigma_1 \cap \Sigma_x' = \emptyset$ and for all y in $\{a, \dots z, 2\}$, if $x \neq y$ then $\Sigma_x' \cap \Sigma_y' = \emptyset$.

By Lemma 4, $T_a', \dots, T_z', T_2'$ is another chain of mergers from T_1 to T_2

$$T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_a' \overset{\emptyset}{\nearrow\!\nwarrow} \dots T_z' \overset{\emptyset}{\nearrow\!\nwarrow} T_2' \overset{\emptyset}{\simeq} T_2,$$

where all theories in the chain have signatures which are disjoint from the signatures of all the other theories in the chain, except for T_1 and T_2 which may have signatures which are non-disjoint.

By Theorem 2, the consecutive mergers from T_1 to T_2' can be compressed into one merger. So $T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T_2' \overset{\emptyset}{\simeq} T_2$ and this is what we wanted to prove.

To show the converse direction, let us assume that T_1 and T_2 are such theories that there is a disjoint renaming theory T_2' of T_2 for which $T_1 \nearrow\!\nwarrow T_2'$. As T_2' is a disjoint renaming of T_2, we have by Remark 7 that $T_2' \overset{\emptyset}{\nearrow\!\nwarrow} T_2$. Therefore, there is a chain T^+, $T^\times$ of theories such that $T_1 \nearrow T^+ \nwarrow T_2' \nearrow T^\times \nwarrow T_2$. Hence $T_1 \overset{\Delta}{\equiv} T_2$. □

Corollary 2. Two theories are definitionally equivalent iff they can be connected by two definitional mergers:

$$T_1 \overset{\Delta}{\equiv} T_2 \iff \exists T (T_1 \overset{\emptyset}{\nearrow\!\nwarrow} T \overset{\emptyset}{\nearrow\!\nwarrow} T_2).$$

Consequently, the chain $T_1, \dots, T_n$ in Definition 11 can allways be choosed to be at most length four.

Proof. This follows immediately from Theorem 4 and Remark 7. □

Theorem 5. Definitional equivalence is the finest equivalence relation containing definitional mergeability. In fact $\overset{\Delta}{\equiv}$ is the transitive closure of relation $\nearrow\!\nwarrow$.

Proof. From Remark 4, we know that $\stackrel{\Delta}{\equiv}$ is an extension of $\nearrow\!\!\nwarrow$. To prove that $\stackrel{\Delta}{\equiv}$ is the transitive closure of $\nearrow\!\!\nwarrow$, it is enough to show that $T_1 \stackrel{\Delta}{\equiv} T_2$ holds if there is a chain $T'_1, \ldots, T'_n$ of theories such that $T_1 = T'_1$, $T_2 = T'_n$, and $T'_i \nearrow\!\!\nwarrow T'_{i+1}$ for all $1 \leq i < n$. By Theorem 4, there is a theory T' such that $T_1 \nearrow\!\!\nwarrow T' \stackrel{\emptyset}{\simeq} T_2$. By Remark 7, $T_1 \nearrow\!\!\nwarrow T' \nearrow\!\!\nwarrow T_2$ which proves our statement. □

It is known that, for languages with disjoint signatures, being definitionally mergeable and intertranslatability are equivalent, see e.g., (Barrett and Halvorson 2016*a*, Theorems 1 and 2). Now we show that, for languages with disjoint signatures, definitional equivalence also coincides with these concepts, i.e.:

Theorem 6. Let T and T' be two theories formulated in languages with disjoint signatures. Then

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

Proof. Since $T \stackrel{\emptyset}{\nearrow\!\!\nwarrow} T' \iff T \rightleftarrows T'$ is proven by (Barrett and Halvorson 2016*a*, Theorems 1 and 2), we only have to prove that $T \stackrel{\Delta}{\equiv} T' \iff T \stackrel{\emptyset}{\nearrow\!\!\nwarrow} T'$.

Let theories T and T' be definitionally equivalent theories with disjoint signatures $\Sigma \cap \Sigma' = \emptyset$. Since they are definitionally equivalent, there exists, by Theorem 4 a chain which consists of a single mergeability and a renaming step between T and T'. Since T and T' are disjoint, and since renaming by Remark 7 is also a disjoint merger, these two steps can by Theorem 2 be reduced to one step $T \stackrel{\emptyset}{\nearrow\!\!\nwarrow} T'$, and this is what we wanted to prove.

The converse direction follows straightforwardly from the definitions. □

Theorem 7. Let T_1 and T_2 be arbitrary theories, then T_1 and T_2 are mergeable iff they are model mergeable, i.e.,

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

Proof. Let T_1 and T_2 be arbitrary theories.

Let us first assume that $T_1 \nearrow\!\!\nwarrow T_2$ and prove that $Mod(T_1) \nearrow\!\!\nwarrow Mod(T_2)$. We know from Lemma 1 that there exist sets of explicit definitions Δ_{12} and Δ_{21} such that $T_1 \cup \Delta_{12} \equiv T_2 \cup \Delta_{21}$. Therefore, by Definition 8, $Mod(T_1 \cup \Delta_{12}) = Mod(T_2 \cup \Delta_{21})$. We construct map β between $Mod(T_1)$ and $Mod(T_2)$ by extending models of T_1 with the explicit definitions in Δ_{12}, which since $Mod(T_1 \cup \Delta_{12}) = Mod(T_2 \cup \Delta_{21})$ will be a model of $T_1 \cup \Delta_{12}$, and then by taking the reduct to the language of T_2. The inverse map β^{-1} can be constructed in a completely analogous manner. β is a bijection since it has an inverse defined for every model of T_2. Through this construction, 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}, and clearly the underlying set of $\mathfrak{M}$ and $\beta(\mathfrak{M})$ are the same. Hence $Mod(T_1) \nearrow\!\nwarrow Mod(T_2)$.

Let us now assume that $Mod(T_1) \nearrow\!\nwarrow Mod(T_2)$ and prove that $T_1 \nearrow\!\nwarrow T_2$. We know by Definition 13 that 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 set 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}.

Any model of both $T_1 \cup \Delta_{12}$ and $T_2 \cup \Delta_{21}$ can be obtained by listing the relations of $\mathfrak{M}$ and $\beta(\mathfrak{M})$ together over the common underlying set M. Therefore, $Mod(T_1 \cup \Delta_{12}) = Mod(T_2 \cup \Delta_{21})$, and thus by Definition 8, $T_1 \cup \Delta_{12} \equiv T_2 \cup \Delta_{21}$. Consequently, $T_1 \nearrow\!\nwarrow T_2$. □

Theorem 8. Let T_1 and T_2 be arbitrary theories. Then T_1 and T_2 are definitionally equivalent iff they are intertranslatable, i.e.,

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

Proof. Let us first assume that $T_1 \stackrel{\Delta}{\equiv} T_2$. Let T' be a disjoint renaming of T_2 to a signature which is also disjoint from the signature of T_1. By Remark 7 and the transitivity of $\stackrel{\Delta}{\equiv}$, we have $T_1 \stackrel{\Delta}{\equiv} T' \stackrel{\Delta}{\equiv} T_2$. By Theorem 6, $T_1 \rightleftarrows T' \rightleftarrows T_2$. Consequently, $T_1 \rightleftarrows T_2$ because relation $\rightleftarrows$ is transitive.

To prove the converse, let us assume that $T_1 \rightleftarrows T_2$. Let T' again be a disjoint renaming of T_2 to a signature which is also disjoint from the signature of T_1. By Remark 7 and the transitivity of $\rightleftarrows$, we have $T_1 \rightleftarrows T' \rightleftarrows T_2$. By Theorem 6, $T_1 \stackrel{\Delta}{\equiv} T' \stackrel{\Delta}{\equiv} T_2$. Consequently, $T_1 \stackrel{\Delta}{\equiv} T_2$ because relation $\stackrel{\Delta}{\equiv}$ is transitive. □

Theorem 9. Let T_1 and T_2 be arbitrary theories, then T_1 and T_2 are intertranslatable iff their models are intertranslatable, i.e.,

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

Proof. Let T_1 and T_2 be arbitrary theories. If T_1 or T_2 is inconsistent, then they are by Remark 3 both inconsistent, $Mod(T_1)$ and $Mod(T_2)$ are empty classes, and the theorem is trivially true. Let's now for the rest of the proof assume that both T_1 and T_2 are consistent theories and hence that both $Mod(T_1)$ and $Mod(T_2)$ are not empty.

Let us first assume that $T_1 \rightleftarrows T_2$ and prove that $Mod(T_1) \rightleftarrows Mod(T_2)$, i.e., that there exist $tr^*_{12} : Mod(T_2) \to Mod(T_1)$ and $tr^*_{21} : Mod(T_1) \to Mod(T_2)$ which are bijections and which are inverses of each other.

Let $\mathfrak{M}$ be a model of T_1, then

$$\mathfrak{M} \models \forall x_1 \dots \forall x_n \big[\varphi(x_1, \dots, x_n) \leftrightarrow tr_{21}\big(tr_{12}\big(\varphi(x_1, \dots, x_n)\big)\big)\big].$$

By Definition 6 and Remark 2, this is equivalent to

$$\mathfrak{M} \models \varphi[e] \iff \mathfrak{M} \models tr_{21}(tr_{12}(\varphi))[e]$$

for all evaluations $e : V \to M$.

By applying Lemma 2 twice,

$$\mathfrak{M} \models tr_{21}(tr_{12}(\varphi))[e] \iff tr^*_{21}(\mathfrak{M}) \models tr_{12}(\varphi)[e] \iff tr^*_{12}(tr^*_{21}(\mathfrak{M})) \models \varphi[e].$$

Consequently,

$$\mathfrak{M} \models \varphi[e] \iff tr^*_{12}(tr^*_{21}(\mathfrak{M})) \models \varphi[e].$$

Since M is the underlying set of both $\mathfrak{M}$ and $tr^*_{12}(tr^*_{21}(\mathfrak{M}))$, this implies that $\mathfrak{M} = tr^*_{12}(tr^*_{21}(\mathfrak{M}))$.

A completely analogous proof shows that $\mathfrak{N} = tr^*_{21}(tr^*_{12}(\mathfrak{N}))$ for all models $\mathfrak{N}$ of T_2.

Consequently, tr^*_{12} and tr^*_{21} are everywhere defined and they are inverses of each other because when we combine them we get the identity, and hence they are bijections, which is what we wanted to prove.

Let us now assume that $Mod(T_1) \rightleftarrows Mod(T_2)$ and prove that $T_1 \rightleftarrows T_2$. By Definition 18, we know that there are bijections tr^*_{12} and tr^*_{21} which are inverses of each other, and thus $\mathfrak{M} = tr^*_{12}(tr^*_{21}(\mathfrak{M}))$ for all models $\mathfrak{M}$ of T_1. Since M is the underlying set of $\mathfrak{M}$, and $tr^*_{12}(tr^*_{21}(\mathfrak{M}))$, we have that

$$\mathfrak{M} \models \varphi[e] \iff tr^*_{12}(tr^*_{21}(\mathfrak{M})) \models \varphi[e].$$

From this, by applying Lemma 2 twice, we get

$$\mathfrak{M} \models \varphi[e] \iff \mathfrak{M} \models tr_{21}(tr_{12}(\varphi))[e].$$

for all evaluations $e : V \to M$. By Definition 6 and Remark 2, this is equivalent to

$$\mathfrak{M} \models \forall x_1 \dots \forall x_n \big[\varphi(x_1, \dots, x_n) \leftrightarrow tr_{21}\big(tr_{12}\big(\varphi(x_1, \dots, x_n)\big)\big)\big].$$

A completely analogous proof shows that

$$\mathfrak{N} \models \forall x_1 \dots \forall x_n \big[\psi(x_1, \dots, x_n) \leftrightarrow tr_{12}\big(tr_{21}\big(\psi(x_1, \dots, x_n)\big)\big)\big],$$

from which follows by Definition 15 that $T_1 \rightleftarrows T_2$. $\square$

Remark 8. If we use the notations of this paper, Theorem 4.2 of (Andréka and Németi 2014) claims, without proof, that (i) definitional equivalence, (ii) definitional mergeability, (iii) intertranslatability and (iv) model mergeability are equivalent in case of disjoint signatures. In this paper, we have not only proven these statements, but we also showed which parts can be generalized to arbitrary languages and which cannot. In detail:

- item (i) is equivalent to item (iii) by Theorem 6, and we have generalized this equivalence to theories in arbitrary languages by Theorem 8,
- the equivalence of items (ii) and (iv) have been generalized to theories in arbitrary languages by Theorem 7,
- 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,
- in Definition 18, we have introduced a model theoretic counterpart of intertanslatability which, by Theorem 9, is equivalent to it even for arbitrary languages.

4 Conclusion

Since definitional mergeability is not transitive, by Theorem 1, 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. Contrary to this, the Andréka–Németi generalization of definitional equivalence is an equivalence relation, by Theorem 3. It is also equivalent to intertranslatability, by Theorem 8, and to model-intertranslatability, by Theorem 9, even for languages with non-disjoint signatures. Therefore, the Andréka–Németi generalization is more suitable to be used as the extension of definitional equivalence between theories of arbitrary languages. It is worth noting, however, that the two generalizations are really close to each-other since the Andréka–Németi generalization is the transitive closure of the Barrett-Halvorson one, see Theorem 5. Moreover, they only differ in at most one disjoint renaming, see Theorems 4 and 6, and as long as we restrict ourselves to theories which all have mutually disjoint signatures, Barrett–Halvorson's definition is transitive by Theorem 2.

Acknowledgements

The writing of the current paper was induced by questions by Marcoen Cabbolet and Sonja Smets during the public defence of (Lefever 2017). We are also grateful to Hajnal Andréka, Mohamed Khaled, Amedé Lefever, István Németi and Jean Paul Van Bendegem for enjoyable discussions and feedback while writing this paper.

References

Andréka, H., Madarász, J. X. and Németi, I. (2005), 'Mutual definability does not imply definitional equivalence, a simple example', *Mathematical Logic Quarterly* **51,6**, 591–597.

Andréka, H., Madarász, J. X., Németi, I., with contributions from: Andai, A., Sági, G., Sain, I. and Tőke, C. (2002), *On the logical structure of relativity theories*, Research report, Alfréd Rényi Institute of Mathematics, Hungar. Acad. Sci., Budapest. https://old.renyi.hu/pub/algebraic-logic/Contents.html.

Andréka, H. and Németi, I. (2014), 'Definability theory course notes'. https://old.renyi.hu/pub/algebraic-logic/DefThNotes0828.pdf.

Andréka, H., Németi, I. and Sain, I. (2001), Algebraic logic, *in* 'Handbook of Philosophical Logic Volume II', Springer Verlag, pp. 133–248.

Barrett, T. W. and Halvorson, H. (2016*a*), 'Glymour and Quine on theoretical equivalence', *Journal of Philosophical Logic* **45**(5), 467–483.

Barrett, T. W. and Halvorson, H. (2016*b*), 'Morita equivalence', *The Review of Symbolic Logic* **9**(3), 556–582.

Corcoran, J. (1980), 'On definitional equivalence and related topics', *History and Philosophy of Logic* **1**(1-2), 231–234.

de Bouvère, K. L. (1965), 'Logical synonymity', *Indagationes Mathematicae* **27**, 622–629.

Friedman, H. A. and Visser, A. (2014), 'When bi-interpretability implies synonymy'.

Glymour, C. (1970), 'Theoretical realism and theoretical equivalence', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* **1970**, 275–288.

Glymour, C. (1977), 'Symposium on space and time: The epistemology of geometry', *Noûs* **11**(3), 227–251.

Glymour, C. (1980), *Theory and Evidence*, Princeton.

Henkin, L., Monk, J. and Tarski, A. (1971), *Cylindric Algebras Part I*, North-Holland.

Henkin, L., Monk, J. and Tarski, A. (1985), *Cylindric Algebras Part II*, North-Holland.

Hodges, W. (1993), *Model Theory*, Cambridge University Press.

Hodges, W. (1997), *A Shorter Model Theory*, Cambridge University Press.

Lefever, K. (2017), Using Logical Interpretation and Definitional Equivalence to compare Classical Kinematics and Special Relativity Theory, PhD thesis, Vrije Universiteit Brussel.

Lefever, K. and Székely, G. (2018), 'Comparing classical and relativistic kinematics in first-order-logic', *Logique et Analyse* **61**(241), 57–117.

Madarász, J. X. (2002), Logic and Relativity (in the light of definability theory), PhD thesis, Eötvös Loránd Univ., Budapest.

Montague, R. (1956), Contributions to the axiomatic foundations of set theory, PhD thesis, Berkeley.

Tarski, A., Mostowski, A. and Robinson, R. (1953), *Undecidable Theories*, Elsevier.

Visser, A. (2015), 'Extension & interpretability', *Logic Group preprint series* **329**.
URL: *https://dspace.library.uu.nl/handle/1874/319941*

KOEN LEFEVER
Centre for Logic and Philosophy of Science
Vrije Universiteit Brussel
koen.lefever@vub.be
`http://homepages.vub.ac.be/~kolefeve/`

GERGELY SZÉKELY
MTA Alfréd Rényi Institute for Mathematics
szekely.gergely@renyi.mta.hu
`http://www.renyi.hu/~turms/`