Defining new universes in many-sorted logic.*

by

H. Andréka, J. X. Madarász, and I. Németi

February 16, 2008

Preliminary, unfinished, draft version

Abstract

In this paper we develop definability theory in such a way that we allow to define new elements also, not only new relations on already existing elements. This is in harmony with our everyday mathematical practice, for example we define new entities when we define a geometry over a field. We will see that, in many respects, defining new elements is more harmonious in many-sorted logic than in one-sorted logic. In the first part of the paper we develop definability theory allowing to define new entities in many-sorted logic (this will amount to defining new universes i.e. new sorts), and in the second part of the paper we develop such a definability theory in one-sorted logic (where this will amount to enlarge the universe with newly defined elements). We will prove an analogon of Beth's definability theorem in this extended context, i.e. we will prove the coincidence of implicit and explicit definability, both in the many-sorted and in the one-sorted case.

1 Introduction.

We will fill this later. The first part is taken from [1], with some modifications. We filled-in proofs, but we did not eliminate parts that refer to the rest of [1]. We will do that later.

*Research supported by Hungarian National Foundation for Scientific Research grants No's T30314 and T23234, T35192.

2 Definability in many-sorted logic, defining new sorts.

Historical remark:

The theory of definability as understood in the present work is a branch of mathematical logic (and its model theory) which goes back to Tarski's pioneering work [19]. Beginning with the just quoted paper of 1934 (and its precursor from 1931), Tarski did much to help the theory of definability to become a fully developed branch of mathematical logic which is worth of studying in its own right. Of the many works illustrating Tarski's concern for the theory of definability we mention only [10, PartI], Tarski-Givant [22], Tarski-Mostowski-Robinson [23] and Tarski [19, 20], cf. also Tarski [18] and [21, Volume 1, pp. 517-548] (which first appeared in 1931 and which already addresses the theory of definability).

In passing we note that the creation of the theory of cylindric algebras can be viewed as a by-product of Tarski's interest in developing and publicizing the theory of definitions (a cylindric algebra over a model can be viewed as the collection of all relations definable in that model).

Below, we try to summarize the theory of definability (allowing definitions of new sorts) in a style tailored for the needs of the present work *and* in a spirit consistent with Tarski's original ideas and views on the subject. Here the emphasis will be on defining new sorts (which is usually not addressed in classical logic books such as e.g. Chang-Keisler [7]).

The subject matter of the present sub-section is relevant to the definability issues discussed in the literature of relativity cf. e.g. Friedman [9, pp. 62–63, 65, 378 (index)]. In Reichenbach's book "Axiomatization of the Theory of Relativity" [16] already on the first page of the Introduction (p.3) he explains the difference between explicit and implicit definitions and emphasizes their importance. (He also traces this distinction (underlying definability theory) to Hilbert's works.) In passing we note that on p.5, Reichenbach [16] also explains in considerable detail why it is desirable to start out with observational concepts first when building up our theory (like we do in Chapters 1,2) and define theoretical concepts later over observational ones using definability theory (as we do in the present chapter). For the time being we do not discuss connections between definability theory and definability issues in relativity theory explicitly, but we plan to do so in a later work.

For the physical importance of definability cf. the relevant parts of the introduction of this Chapter. Further, we note the following. If in our language we allow using certain concepts and some other concept is definable from these, then this other concept *is* available in our language even if we do not include it (explicitly). So if we allow only such concepts which are definable from observational ones, then the effect will be the same as if we allowed only observational concepts. I.e. the physical principle of Occam's razor has been respected.

$$* \quad * \quad *$$

Let $\mathfrak{M} = \langle U_0, \ldots, U_j;\ R_1, \ldots, R_l \rangle$ be a many-sorted model with universes or sorts $U_0, \ldots, U_j$, and relations $R_1, \ldots, R_l \quad (j, l \in \omega)$.[1] Since functions are special relations we do not indicate them explicitly in the present discussion. We use the semicolon ";" to separate the sorts (or universes) from the relations of $\mathfrak{M}$.

When discussing many-sorted models, we always assume that they have *finitely many sorts* only.[2] The "big universe" $Uv(\mathfrak{M})$ of the model $\mathfrak{M}$ is the union of its universes (or sorts). Formally

$$Uv \stackrel{\text{def}}{:=} Uv(\mathfrak{M}) \stackrel{\text{def}}{:=} \bigcup \{\, U_i \ : \ U_i \text{ is a universe of } \mathfrak{M} \,\}\,.^3$$

In passing we note that although the sorts $U_0, \ldots, U_j$ of $\mathfrak{M}$ need not be disjoint, the following holds. To every many-sorted model $\mathfrak{M}$ there is an isomorphic copy $\mathfrak{M}'$ of $\mathfrak{M}$ such that the sorts $U'_0, \ldots, U'_j$ of $\mathfrak{M}'$ *are* mutually disjoint (i.e. $U'_0 \cap U'_1 = \emptyset$ etc.). Therefore we *are permitted to* pretend that the sorts (i.e. universes) of $\mathfrak{M}$ are disjoint from each other whenever we would need this.

By a *reduct* of a many-sorted model $\mathfrak{M}$ we understand a model $\mathfrak{M}^-$ obtained from $\mathfrak{M}$ by omitting some of the sorts and/or some of the relations of $\mathfrak{M}$. I.e. if

$$\mathfrak{M} = \langle U_0, \ldots, U_j;\ R_1, \ldots, R_l \rangle$$

[1] The assumption that l is finite is irrelevant here in the sense that we will never make use of it (except when we state this explicitly). What we write in this section makes perfect sense if the reader replaces l with an arbitrary ordinal. As a contrast, we do use the assumption that $j \in \omega$.

[2] In some minor items there may be exceptions from this rule but then this will be clearly indicated.

[3] Although, in general, Uv is not a universe of $\mathfrak{M}$, we can *pretend* that it is a universe because there are only finitely many sorts. E.g. if we want to simulate the formula $(\exists x \in Uv)\, \psi(x)$ then we write $[\, (\exists x \in U_0)\, \psi(x) \ \vee\ (\exists x \in U_1)\, \psi(x) \ \vee\ \ldots \vee\ (\exists x \in U_j)\, \psi(x) \,]$. Then although the first formula $(\exists x \in Uv)\, \psi(x)$ does not belong to the language of $\mathfrak{M}$, the second formula "$[\, (\exists x \in U_0) \ldots]$" does belong to this language (assuming $(\exists x \in U_i)\, \psi(x)$ already belongs to the language) and the meaning of the second formula is the same as the intuitive meaning of the first one. If $(\exists x \in U_i)\, \psi(x)$ did still not belong to our many-sorted language then there is some extra routine work to do in translating this formula into our many-sorted language. This translation is explained in detail in the logic books which reduce many-sorted logic to one-sorted logic (cf. [6, 8, 13]). We also note that the quoted translation is straightforward. For more on why and how we can pretend that $Uv(\mathfrak{M})$ is a universe of $\mathfrak{M}$ we refer to the just quoted logic books.

then the reduct $\mathfrak{M}^-$ may be of the form

$$\langle U_0, \ldots, U_{j-1};\ R_1, \ldots, R_{l-1} \rangle$$

(assuming $R_1, \ldots, R_{l-1}$ do not involve the sort U_j).

A model $\mathfrak{M}^+$ is called an ***expansion*** of $\mathfrak{M}$ iff $\mathfrak{M}$ is a reduct of $\mathfrak{M}^+$. I.e. an expansion $\mathfrak{M}^+$ is obtained by adding new sorts and/or new relations to $\mathfrak{M}$. We will use the following abbreviation for denoting expansions:

$$\mathfrak{M}^+ = \langle \mathfrak{M}, U^{new};\ \bar{R}^{new} \rangle$$

where U^{new} is the new sort and $\bar{R}^{new} = \langle R_1^{new}, \ldots, R_r^{new} \rangle$ is the sequence of new relations. Of course there may be more new sorts too, then we write

$$\mathfrak{M}^+ = \langle \mathfrak{M}, U_1^{new}, \ldots, U_\varrho^{new};\ \bar{R}^{new} \rangle.$$

However, we will concentrate on the case $\varrho = 1$ (for didactical reasons). Informally the general pattern is:

"New model" = ⟨"Old model", "New sorts"; "New relations/functions"⟩.

We will ask ourselves when $\mathfrak{M}^+$ will be *(first-order logic) definable* over[4] $\mathfrak{M}$. By *definable* we will *always* (throughout this work) mean first-order logic definable. If $\langle \mathfrak{M}, U^{new};\ \bar{R}^{new} \rangle$ is definable over $\mathfrak{M}$ then we will say that the new sort U^{new} together with $\bar{R}^{new}$ are *definable* in $\mathfrak{M}$. When defining a new sort U^{new} (in an "old" model $\mathfrak{M}$) we need the new relations $\bar{R}^{new}$ too because it is $\bar{R}^{new}$ which will specify the connections between the new sort U^{new} and the old sorts of $\mathfrak{M}$.

Although we will start out with discussing definability over a single model $\mathfrak{M}$, the really important part will be when we generalize this to definability (of an expanded class K^+) *over a class* K of models (which is first-order axiomatizable).

We will discuss *two kinds* of definability in many-sorted logic: ***implicit*** definability in §2.1 and ***explicit*** definability in §2.2.[5]

Throughout model theory there is a *distinction* between symbols like Obs and objects like $Obs^{\mathfrak{M}}$ denoted by these symbols in a model $\mathfrak{M}$. This distinction between symbols and objects they denote is even more important in the theory of definitions than in other parts of logic. Therefore, in the next two items we clarify notions and notation connected to this distinction.

---

[4] "Definable over" is the same as "definable in".

[5] In passing, we note that in the *special* case of the most traditional one-sorted logic when *only* relations are defined (i.e. defining new sorts is not considered) the distinction between implicit and explicit definability is well investigated and is well understood cf. e.g. Chang-Keisler [7, p.90] or Hodges [11, pp.301-302].

CONVENTION 2.0.1 By the *vocabulary* of a model $\mathfrak{M}$ we understand the system of sort-symbols, relation symbols and function symbols interpreted by $\mathfrak{M}$. Since function symbols are special relation symbols, we will restrict our attention to sort symbols and relation symbols. Assume e.g. that $\mathfrak{M}$ is of the form

$$\mathfrak{M} = \langle U_0^{\mathfrak{M}}, \dots, U_j^{\mathfrak{M}}; R_1^{\mathfrak{M}}, \dots, R_l^{\mathfrak{M}} \rangle \ ,$$

and assume that U_i is the sort *symbol* "denoting" $U_i^{\mathfrak{M}}$ and R_i is the relation *symbol* "denoting" $R_i^{\mathfrak{M}}$. Then the vocabulary of $\mathfrak{M}$ is

$$Voc(\mathfrak{M}) \stackrel{\text{def}}{=} \langle \{U_0, \dots, U_j\}, \{R_1, \dots, R_l\} \rangle.$$

Throughout we assume that a relation symbol R' contains the extra information which we call the *rank* of R'. This can be implemented by postulating that R' is an ordered pair $R' = \langle R'_0, R'_1 \rangle$ where R'_0 is the symbol we write on paper while R'_1 is the rank of R'. E.g. in the case of the usual model $\mathfrak{N} = \langle \omega, \leq, + \rangle$ the rank of "$\leq$" is 2 while that of "$+$" is 3. If there is more than one sort, then the rank of a relation is a sequence of sort symbols. So, a vocabulary is an ordered pair

$$Voc = \langle \text{"Sort symbols"}, \text{"Relation symbols"} \rangle$$

where "Sort symbols" and "Relation symbols" are two sets as discussed above subject to the condition that the sorts occurring in the ranks of the relation symbols all occur in the set of sort symbols. Now, a model $\mathfrak{M}$ of vocabulary Voc can be regarded as a pair $\mathfrak{M} = \langle \mathfrak{M}_0, \mathfrak{M}_1 \rangle$ of functions such that

$$\mathfrak{M}_0 : \text{"Sort symbols"} \longrightarrow \text{"Universes of } \mathfrak{M}\text{"}$$

and

$$\mathfrak{M}_1 : \text{"Relation symbols"} \longrightarrow \text{"Relations of } \mathfrak{M}\text{"},$$

with the restriction that $\mathfrak{M}_1$ is "rank-preserving" in a natural sense.
E.g. if $\mathfrak{M} = \langle U_0^{\mathfrak{M}}, \dots, U_j^{\mathfrak{M}}; R_1^{\mathfrak{M}}, \dots, R_l^{\mathfrak{M}} \rangle$, then

$$\mathfrak{M}_0 : \{U_i : i \leq j\} \longrightarrow \{U_i^{\mathfrak{M}} : i \leq j\}$$

$$\mathfrak{M}_1 : \{R_i : 0 < i \leq l\} \longrightarrow \{R_i^{\mathfrak{M}} : 0 < i \leq l\}.$$

I.e. to each sort symbol in $Voc(\mathfrak{M})$, $\mathfrak{M}$ associates a universe (i.e. a set) and to each relation symbol R' in $Voc(\mathfrak{M})$, $\mathfrak{M}$ associates a relation (of rank R'_1 as indicated way above).

We call two models $\mathfrak{M}$ and $\mathfrak{N}$ *similar* if they have the same vocabulary, i.e. if $Voc(\mathfrak{M}) = Voc(\mathfrak{N})$.

Let Voc', Voc be two vocabularies. We say that Voc' is a *sub-vocabulary* of Voc if the natural conditions $Voc'_0 \subseteq Voc_0$ and $Voc'_1 \subseteq Voc_1$ hold. Assume Voc' is a sub-vocabulary of $Voc(\mathfrak{M})$ for a model $\mathfrak{M}$. Then the *reduct* $\mathfrak{M} \restriction Voc'$ of $\mathfrak{M}$ to the sub-vocabulary Voc' is defined as

$$\mathfrak{M} \restriction Voc' \stackrel{\text{def}}{=} \langle \mathfrak{M}_0 \restriction Voc'_0, \ \ \mathfrak{M}_1 \restriction Voc'_1 \rangle .$$

$\lhd$

Remark 2.0.2 (On the intuitive content of Convention 2.0.1 above) On a very intuitive informal level, one can think of a model $\mathfrak{M}$ as a *function* associating objects to symbols. E.g. $\mathfrak{M}$ associates $U_i^{\mathfrak{M}}$ to the symbol U_i and $R_i^{\mathfrak{M}}$ to R_i. It is then a matter of notational convention that we write $U_i^{\mathfrak{M}}$ for the value $\mathfrak{M}(U_i)$ and $R_i^{\mathfrak{M}}$ for $\mathfrak{M}(R_i)$. Then the domain of the function $\mathfrak{M}$ is the collection of those symbols which $\mathfrak{M}$ can interpret. Hence, the *domain* of $\mathfrak{M}$ is the same thing as its vocabulary.

If the best way (from the intuitive point of view) of thinking about a model is regarding it as a function, then why did we formalize the notion of a model as a pair of functions (instead of a single function)? The answer is that *formally* it is easier to handle models as pairs of functions, but *intuitively* we think of models as functions, we think of vocabularies as domains of these functions and we consider two models similar if they have the same domain when they are regarded as functions.[6]

$\lhd$

CONVENTION 2.0.3 Throughout, by a class K of models we understand a class of similar models, i.e. we always assume $(\forall \mathfrak{M}, \mathfrak{N} \in \mathsf{K})\ Voc(\mathfrak{M}) = Voc(\mathfrak{N})$. For any class K of similar models, $Voc(\mathsf{K}) = Voc\mathsf{K}$ denotes the vocabulary of K, that is, the vocabulary of an arbitrary element of K.

A *reduct* K^- of K is obtained from K by omitting a part of the vocabulary of K, i.e. K^- is a reduct of K iff $Voc(\mathsf{K}^-) \subseteq Voc(\mathsf{K})$ and

$$\mathsf{K}^- = \left\{ \mathfrak{M} \restriction Voc(\mathsf{K}^-) \ : \ \mathfrak{M} \in \mathsf{K} \right\}.$$

Expansion is the opposite of reduct. K^+ is an *expansion* of the class K iff K is a reduct of K^+, i.e. K^+ is an expansion of K iff $Voc(\mathsf{K}^+) \supseteq Voc(\mathsf{K})$ and

$$\mathsf{K} = \left\{ \mathfrak{M} \restriction Voc(\mathsf{K}) \ : \ \mathfrak{M} \in \mathsf{K}^+ \right\}.$$

[6] We do not claim that it is always the case that the best way of thinking about models is regarding them as functions. What we claim is that in many situations, e.g. in definability theory, this is a rather good way. In other situations it might be better to visualize a model as a set of objects equipped with some relations and functions.

Note that forming expansions or reducts of a class K is somehow *uniform* over the members of K. E.g. we forget the *same* symbols (relation symbols or sort symbols) from all models $\mathfrak{M} \in \mathsf{K}$, when taking a reduct of K.

If *Voc* is a vocabulary with $Voc \subseteq Voc(\mathsf{K})$, then we use the following abbreviation:

$$\mathsf{K} \upharpoonright Voc \stackrel{\text{def}}{=} \{\mathfrak{M} \upharpoonright Voc \ : \ \mathfrak{M} \in \mathsf{K}\}.$$

Examples: $\mathsf{FM}^- = \{\, \mathfrak{F}^{\mathfrak{M}} \ : \ \mathfrak{M} \in \mathsf{FM}\,\}$ is a reduct of our class FM of frame models. Let $\mathsf{L} = \{\, \mathbf{F} \ : \ \mathbf{F} \text{ is a field}\,\}$. Then $\{\, \langle F; +\rangle \ : \ \langle F; +, \cdot, 0, 1\rangle \in \mathsf{L}\,\}$ is a reduct of L.

Intuitively, we think of $Voc(\mathsf{K})$ as a set of symbols where each symbol contains information about its nature, i.e. about whether it is a sort symbol or a relation symbol of a certain rank. Therefore, we will write $Voc \cap Voc'$ for $\langle Voc_0 \cap Voc'_0, Voc_1 \cap Voc'_1\rangle$, similarly for $Voc \cup Voc'$, for $Voc \subseteq Voc'$ etc.

$\lhd$

CONVENTION 2.0.4 $f : A \twoheadrightarrow B$ denotes that f is a surjective function from A onto B. Further $f : A \rightarrowtail B$ denotes that f is an injective function from A into B. I.e. $\twoheadrightarrow$ denotes surjectiveness, while $\rightarrowtail$ denotes injectiveness. If we combine the two then we obtain $\rightarrowtail\!\!\!\!\twoheadrightarrow$ denoting bijectiveness. When used between german letters, i.e. structures, they denote injectiveness or surjectiveness of homomorphisms the natural way.

$\lhd$

Before getting started, we emphasize that in order to define something over a model $\mathfrak{M}$ or over a class K of models, first of all we need <u>*new symbols*</u> R_i^{new}, U_i^{new} (with i in some index set) not occurring in the language of $\mathfrak{M}$ or of K. (The new symbols may be relation symbols like R_i^{new} or sort symbols U_i^{new} or both.) What we will define then (using definability theory) will be the meanings of the new symbols in $\mathfrak{M}^+$ or K^+. Most of the time we will not talk about the new <u>*symbols*</u> like R_i^{new} because we will identify them with the new relations like $(R_i^{new})^{\mathfrak{M}^+}$ which they denote in the expansion $\mathfrak{M}^+$ of the model $\mathfrak{M}$. Our reason for identifying the "symbol" with the "object" it denotes is to simplify the discussion. However, occasionally it will be useful to remember that an expansion $\mathfrak{M}^+ = \langle \mathfrak{M}, R\rangle$ of a model $\mathfrak{M}$ involves two new things not available in $\mathfrak{M}$, namely: a relation symbol and a relation denoted by this symbol (in $\mathfrak{M}^+$).

2.1 Implicit definability in many-sorted (first-order) logic

Let $\mathfrak{M}$ be a many-sorted model. Assume, $\mathfrak{M}^+ = \langle \mathfrak{M}, U^{new};\, \bar{R}^{new} \rangle$ is an expansion of $\mathfrak{M}$. We say that $\mathfrak{M}^+$ is <u>*definable implicitly up to isomorphism*</u> over $\mathfrak{M}$ iff

for any model

$$\langle \mathfrak{M}, U';\, \bar{R}' \rangle \models \mathsf{Th}(\mathfrak{M}^+)$$

(expanding $\mathfrak{M}$) there is an isomorphism

$(\star)$
$$h : \mathfrak{M}^+ \rightarrowtail\!\!\!\twoheadrightarrow \langle \mathfrak{M}, U';\, \bar{R}' \rangle$$

such that h is the *identity* function on the sorts of $\mathfrak{M}$ (i.e. for each sort U_i of $\mathfrak{M}$ we have $h \restriction U_i = \mathrm{Id} \restriction U_i$).

$\mathfrak{M}^+$ is said to be <u>*definable implicitly without taking reducts*</u> over $\mathfrak{M}$ iff *in addition* to the above the isomorphism h mentioned above is *unique*.

We say that <u>U^{new}, $\bar{R}^{new}$ *are definable implicitly*</u> over $\mathfrak{M}$ iff $\langle \mathfrak{M}, U^{new};\, \bar{R}^{new} \rangle$ is definable implicitly without taking reducts over $\mathfrak{M}$. Informally we might say in such situations that the new sort U^{new} is definable implicitly in $\mathfrak{M}$ (but then $\bar{R}^{new}$ should be understood from the context, otherwise the definability claim is sort of under-specified).

In the above notion of definability, the *set of formulas defining* U^{new}, $\bar{R}^{new}$ *implicitly* over $\mathfrak{M}$ is $\mathsf{Th}(\mathfrak{M}^+)$. Hence, $\mathsf{Th}(\mathfrak{M}^+)$ is called an *implicit definition* of U^{new}, $\bar{R}^{new}$ over $\mathfrak{M}$ if $(\star)$ above holds and the isomorphism h is unique. Further, for any set Δ of formulas in the language of $\mathfrak{M}^+$, Δ is called an <u>*implicit definition*</u> of $U^{new}, \bar{R}^{new}$ over $\mathfrak{M}$ iff $(\star)$ above holds with Δ in place of $\mathsf{Th}(\mathfrak{M}^+)$ in such a way that h is unique.[7]

Remark 2.1.1 The reader might feel that the above notion of (implicit) definability without taking reducts (of $\mathfrak{M}^+$) is not strong enough and he might want to replace h with the identity function (requiring $U^{new} = U'$, $\bar{R}^{new} = \bar{R}'$). However, we claim that the above notion is "best possible" because (i) it is reasonable to assume that the first-order definition of $\mathfrak{M}^+$ (over $\mathfrak{M}$) is included in $\mathsf{Th}(\mathfrak{M}^+)$ and (ii) *any* isomorphic copy $\mathfrak{M}' = \langle \mathfrak{M}, U';\, \bar{R}' \rangle$ of $\mathfrak{M}^+$ will automatically validate $\mathsf{Th}(\mathfrak{M}^+)$ hence, in first-order logic we cannot define the new sort U^{new}, $\bar{R}^{new}$ more closely than up to (a

[7]The set Δ of formulas which we call an implicit definition is called a "rigidly relatively categorical" theory in Hodges [11, p.645]. If Δ is an implicit definition up to isomorphism only, then it is called a "relatively categorical" theory on p.638 of [11] (§12.5 therein).

unique) isomorphism.[8]

$\lhd$

$\mathfrak{M}^+ = \langle \mathfrak{M}, U^{new}; \bar{R}^{new} \rangle$ is said to be <u>*definable implicitly with parameters*</u> over $\mathfrak{M}$ iff there are $s \in \omega$ and $\bar{p} \in {}^sUv(\mathfrak{M})$ such that the expansion $\langle \mathfrak{M}^+, \bar{p} \rangle$ is *definable implicitly without taking reducts* over the expansion $\langle \mathfrak{M}, \bar{p} \rangle$.[9]

*   *   *

Let us turn to definability over *classes of models*. Let K be a class of models with U^{new}, $\bar{R}^{new}$ in the language of K. For $\mathfrak{M} \in \mathsf{K}$ let $\mathfrak{M}^-$ be the reduct of $\mathfrak{M}$ obtained by omitting (forgetting) U^{new}, $\bar{R}^{new}$. Let

$$\mathsf{K}^- := \left\{ \mathfrak{M}^- \ : \ \mathfrak{M} \in \mathsf{K} \right\}.$$

We ask ourselves when K is definable over K^- or equivalently (but informally) when U^{new}, $\bar{R}^{new}$ are definable over K^-. We say that the class K of models is <u>*definable implicitly without taking reducts*</u> over K^- iff there is a set $\Delta \subseteq \mathsf{Th}(\mathsf{K})$ of formulas such that condition $(\star\star)$ below holds.

$(\star\star)$ For every $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}(\Delta)$ (similar to members of K) with $\mathfrak{M}^- = \mathfrak{N}^- \in \mathsf{K}^-$, there is a unique isomorphism $h : \mathfrak{M} \rightarrowtail\!\!\!\twoheadrightarrow \mathfrak{N}$ which is the identity on the universes of $\mathfrak{M}^-$.

If the isomorphism h is not necessarily unique then we say that K is <u>*definable implicitly up to isomorphism*</u> over K^-. Informally, we say that the <u>*new sort*</u> U^{new} and $\bar{R}^{new}$ are <u>*definable implicitly*</u> over K^- iff K as understood above is definable implicitly without taking reducts over K^-. When speaking about definability of U^{new}, $\bar{R}^{new}$ over K^-, it should be clear *from context* how K is obtained from the data K^- and U^{new}, $\bar{R}^{new}$. If $(\star\star)$ holds, then Δ in $(\star\star)$ is called an <u>*implicit definition*</u> of K over K^-.

We note that in the definition of "K is implicitly definable without taking reducts over K^-" the class K of models occurs only in requiring $\Delta \subseteq \mathsf{Th}(\mathsf{K})$. Therefore, if K is implicitly definable over K^- without taking reducts, then so is $\mathbf{I}\mathsf{K}$ over $\mathbf{I}\mathsf{K}^-$, and $\mathsf{Mod}(\Delta)$ over $\mathsf{Mod}(\Delta)^-$, where Δ is any implicit definition of K over K^-.

---

[8] A possible way out of this would be if we required $\bar{R}^{new}$ to contain membership relations "$\in$" and projection functions pj_i (and then add some restrictions postulating e.g. that $\in$ and pj_i are the "real" set theoretic ones etc., cf. p.22 for the definition of the pj_i's). We will not do this because we feel that it would lead to too many complications without yielding enough benefits.

[9] We use "definable implicitly" and "implicitly definable" as synonyms. I.e. we are flexible about word order.

We leave it to the reader to generalize the above definitions to the case when we have arbitrary sequences $\bar{U}^{new}$ and $\bar{R}^{new}$ of new sorts and new relations. However, herein we restrict our attention to the case when there are finitely many new symbols (i.e. both $\bar{U}^{new}$ and $\bar{R}^{new}$ are finite sequences of sorts and relations respectively). The classical notion of definability of new relations (without new sorts) is obtained as a special case of our general notion by choosing $\bar{U}^{new} = \emptyset$, i.e. $\bar{U}^{new}$ is the empty sequence.

Let K and L be two classes of models, i.e. L is not necessarily a reduct of K. We say that $\underline{\mathsf{K}\ \textit{is definable implicitly over}\ \mathsf{L}}$ iff some expansion K^+ of K is definable implicitly without taking reducts over L. (In this case, L will be a reduct of K^+, of course.)[10] This means that statements (i) and (ii) below hold for some expansion K^+ of K:

(i) L is a reduct of K^+,

(ii) K^+ is definable implicitly over L without taking reducts. (Since here L is a reduct of K^+, our *earlier* definition of implicit definability without taking reducts on p.9 can be applied.)

We note that here we have to take seriously that our languages are finite, i.e. K^+ has only finitely many new symbols that do not occur in L.[11] In this case, we say that Δ is an $\underline{\textit{implicit definition of}\ \mathsf{K}\ \textit{over}\ \mathsf{L}}$ if Δ is an implicit definition of K^+ over L. Thus an implicit definition of K over L may contain symbols not occurring in K.

We will apply the same convention for single models too, i.e. $\underline{\mathfrak{N}\ \textit{is definable implicitly over}\ \mathfrak{M}}$ iff this holds for $\{\mathfrak{N}\}$ and $\{\mathfrak{M}\}$. We will sometime abbreviate "implicitly definable without taking reducts" by "*<u>nr-implicitly definable</u>*", where "nr" stands for "taking <u>n</u>o <u>r</u>educts". Note that $(\star\star)$ on p.9 above is a straightforward generalization of $(\star)$ on p.8. Therefore $\mathfrak{M}^+$ is definable nr-implicitly over $\mathfrak{M}$ iff the class $\{\mathfrak{M}^+\}$ is definable nr-implicitly over the class $\{\mathfrak{M}\}$.

In situations like the one involving statement $(\star\star)$ above, we also say that $\underline{U^{new}, \bar{R}^{new}\ \textit{are uniformly}}$ definable (implicitly) over K^-.[12] The set Δ of formulas is considered as a *<u>uniform (implicit) definition</u>* of U^{new}, $\bar{R}^{new}$ over K^-. We have not yet discussed *non-uniform* definability which is also called *"local"* or *"one-by-one"* definability: We will discuss this notion below Examples 2.1.6, on p.18.

[10] It would be more careful of us if we would call this new implicit definability (which permits taking reducts) weak implicit definability. This is so because when taking reducts then the uniqueness condition, cf. p.8, on isomorphisms may get lost.

[11] Cf. Examples 2.1.6 (2) on p.15.

[12] We will explain soon, beginning with item 11 of Examples 2.1.6 (p.15), what aspect of the above situation we are referring to with the adjective "uniform" here.

Although we began this sub-section with discussing definability over a single model $\mathfrak{M}$, the main emphasis in this work will be on definability over a class K of models such that $\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$ i.e. such that K is axiomatizable in first-order logic.

We note that implicit definability without taking reducts of K over K^- is strictly stronger than implicit definability up to isomorphism. This remains so even if we assume that K and K^- are first-order axiomatizable classes of models. We leave the construction of a simple counterexample to the reader, but cf. Example 2.1.6(8) way below. For the connections between the various notions of definability we refer the reader to Figure 5 on p.71.

Remark 2.1.2 (Re-formulations of the definition of nr-implicit definability) The following are intended to provide a kind of "intuitive" re-formulations implicit definability without taking reducts of a class K of models over its reduct K^- (as was defined above). Assume K^- is a reduct of the class K (i.e. K^- is of the form $\{\mathfrak{M}^- : \mathfrak{M} \in \mathsf{K}\}$).

(1) K is definable implicitly over K^- without taking reducts iff (i)–(ii) below hold.

  (i) $(\forall \mathfrak{M} \in \mathsf{K})\mathfrak{M}$ is definable nr-implicitly over its reduct $\mathfrak{M}^-$.

  (ii) There is a single set Δ of formulas such that for every $\mathfrak{M} \in \mathsf{K}$, Δ is an implicit definition of $\mathfrak{M}$ over $\mathfrak{M}^-$. In other words, not only each $\mathfrak{M}$ is nr-implicitly definable over $\mathfrak{M}^-$, but this defining can be done uniformly for the whole of K.

(2) Let $\mathsf{rd} \stackrel{\text{def}}{=} \{\langle \mathfrak{M}, \mathfrak{M}^- \rangle : \mathfrak{M} \in \mathsf{K}\}$. Then

$$\mathsf{rd} : \mathsf{K} \twoheadrightarrow \mathsf{K}^-$$

is a surjective function. Now, K is nr-inplicitly definable over K^- iff the function rd is *<u>injective up to isomorphisms</u>* – i.e. $\mathsf{rd}(\mathfrak{M}) \cong \mathsf{rd}(\mathfrak{N}) \;\Rightarrow\; \mathfrak{M} \cong \mathfrak{N}$ – and each $\mathfrak{M} \in \mathsf{K}$ is definable nr-implicitly over $\mathsf{rd}(\mathfrak{M})$ and these definitions coincide for all choices of $\mathfrak{M}$.

$\lhd$

To state the analogon of Remark 2.1.2 for “implicit definability” in place of “nr-implicit definability”, we want to spell out our usage of the word “up to isomorphism”.

Notation. When we say that a property holds up to isomorphism, we mean that the property holds modulo identifying some of the isomorphic models. In particular, let $f \subseteq \mathsf{K} \times \mathsf{L}$ be a relation between the classes K and L of models. We say that f is a *function-up-to-isomorphism* if

$$\langle \mathfrak{M}, \mathfrak{N} \rangle, \langle \mathfrak{M}, \mathfrak{N}_1 \rangle \in f \quad \Rightarrow \quad \mathfrak{N} \cong \mathfrak{N}_1 \,.$$

We say that f is *surjective-up-to-isomorphism* if $\mathbf{I} Rng(f) = \mathbf{I}\mathsf{K}$. The definitions for *injective-up-to-isomorphism* and *bijection-up-to-isomorphism* are analogous. Let $f' \subseteq \mathsf{K} \times \mathsf{L}$. We say that f' *agrees-up-to-isomorphism* with f iff f and f' induce the same functions between the isomorphism equivalence classes of K and L, i.e. between $\mathsf{K}/\cong$ and $\mathsf{L}/\cong$. Thus, e.g. f is a bijection-up-to-isomorphism if f induces a (real) bijection between $\mathsf{K}/\cong$ and $\mathsf{L}/\cong$. It is natural to work with the elements of $\mathsf{K}/\cong$ and $\mathsf{L}/\cong$. The elements of $\mathsf{K}/\cong$ are called isomorphism types in [10, Part I, p.71, lines 8–10]. Our using properties *“up to isomorphism”* is based on the practice in model theory and universal algebra of identifying isomorphic structures in some (but not in all (!)) situations. If f is a function-up-to-isomorphism, then we use the notation $f : \mathsf{K} \longrightarrow \mathsf{L}$ also. Similarly for the other notations like $f : \mathsf{K} \rightarrowtail\!\!\!\twoheadrightarrow \mathsf{L}$ and $f : \mathsf{K} \twoheadrightarrow \mathsf{L}$. $\lhd$

Remark 2.1.3 (properties of “general” definability of classes) Assume K is definable implicitly over L. Then (1)-(3) below hold, and (3) is a re-formulation of K being implicitly definable over L.

(1) K and L agree on their common vocabulary, i.e.

$$\mathsf{K} \restriction (Voc\mathsf{K} \cap Voc\mathsf{L}) = \mathsf{L} \restriction (Voc\mathsf{K} \cap Voc\mathsf{L}).$$

(2) $\mathbf{I}\mathsf{K}$ is closed under taking ultraproducts iff $\mathbf{I}\mathsf{L}$ is closed under taking ultraproducts.

(3) There is a surjective-up-to-isomorphism function $f : \mathsf{L} \twoheadrightarrow \mathsf{K}$ such that for all $\mathfrak{M} \in \mathsf{L}$, $f(\mathfrak{M})$ is implicitly definable over $\mathfrak{M}$[13]; moreover the definition of $f(\mathfrak{M})$ over $\mathfrak{M}$ is the same (set of formulas) for all choices of $\mathfrak{M}$.

[13] i.e. there is an implicit definitional expansion $\mathfrak{M}^+$ of $\mathfrak{M}$ with $f(\mathfrak{M})$ a reduct of $\mathfrak{M}^+$.

$\triangleleft$

PROPOSITION 2.1.4

(i) *Assume that* K *is axiomatizable. Then if* K^+ *is nr-implicitly definable over* K, *then* $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$*, too, is nr-implicitly definable over* K.

(ii) *If* K *is implicitly definable over* L *and* $\mathbf{I}\mathsf{L}$ *is closed under taking ultraproducts, then* $\mathbf{I}\mathsf{K}$*, too, is closed under taking ultraproducts.*

Proof. The proof is straightforward by using the definitions. Proof of (i): We only have to show that the $Voc\mathsf{K}$-reducts of the elements of $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$ are in K, i.e. $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+)) \upharpoonright Voc\mathsf{K} \subseteq \mathsf{K}$. Since K is the reduct of K^+, we have $\mathsf{K}^+ \upharpoonright Voc\mathsf{K} \subseteq \mathsf{K}$, and hence $\mathsf{K}^+ \models \mathsf{Th}(\mathsf{K})$. Then $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+)) \models \mathsf{Th}(\mathsf{K})$, and hence $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+)) \upharpoonright Voc\mathsf{K} \models \mathsf{Th}(\mathsf{K})$. Since K is axiomatizable, this means $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+)) \upharpoonright Voc\mathsf{K} \subseteq \mathsf{K}$.

Proof of (ii): K is a reduct of a definitional expansion K^+ of L. Let Δ be a definition of K^+ over L. Let $\langle \mathfrak{M}_i : i \in I\rangle$ be a system of members of $\mathbf{I}\mathsf{K}$ and let F be an ultrafilter over I. Since Δ is a definition of K over L, each $\mathfrak{M}_i$ is the reduct of an $\mathfrak{M}_i^+ \in \mathbf{I}\mathsf{K}^+$ such that $\mathfrak{M}_i^+ \upharpoonright Voc\mathsf{L} \in \mathbf{I}\mathsf{L}$. Also $\mathfrak{M}_i^+ \models \Delta$ for each $i \in I$. Then the same is true for the F-ultraproducts: $\mathfrak{M}_1 \overset{\text{def}}{=} \Pi\langle \mathfrak{M}_i : i \in I\rangle/F$ is the reduct of $\mathfrak{M}_1^+ \overset{\text{def}}{=} \Pi\langle \mathfrak{M}_i^+ : i \in I\rangle/F$, $\mathfrak{M}_1^+ \models \Delta$ and $\mathfrak{M}_1^+ \upharpoonright Voc\mathsf{L} = \Pi\langle \mathfrak{M}_i^+ \upharpoonright Voc\mathsf{L} : i \in I\rangle/F$. The latter is in $\mathbf{I}\mathsf{L}$ since it is closed under taking ultraproducts. Hence $\mathfrak{M}_1^+ \in \mathbf{I}\mathsf{K}^+$ because Δ is a definition of K^+ over L, $\mathfrak{M}_1^+ \upharpoonright Voc\mathsf{L} \in \mathbf{I}\mathsf{L}$ and $\mathfrak{M}_1^+ \models \Delta$. Hence $\mathfrak{M}_1 \in \mathbf{I}\mathsf{K}$, showing that $\mathbf{I}\mathsf{K}$ is closed under taking ultraproducts. ∎

Now we turn to giving examples.

Examples 2.1.5 (Traditional, one-sorted examples)

1. Let PA be the class of models of Peano's Arithmetic, cf. any logic book, e.g. Monk [13] or Chang-Keisler [7] for PA. The operation symbols of PA are $+, \cdot, 0, 1$. Consider the extra unary operation symbol "!" intended to denote the factorial. Let $\Delta_!$ be the set of the following two formulas

   $$!(0) = 1$$
   $$\forall x[!(x+1) = (x+1)\cdot !(x)].$$

   I.e. $\Delta_! = \{\ !(0) = 1,\ \forall x[!(x+1) = (x+1)\cdot !(x)]\ \}$. We claim that $\Delta_!$ is a (correct) implicit definition of "!" over PA. (The proof is not easy but is available in almost any logic book.) The point in the above example is that PA is an axiomatizable class and that $\Delta_!$ works over each member of PA. If we want an implicit definition over a single model instead of an axiomatizable class, that is easy:

2. Consider the model $\langle \omega, 0, \mathsf{suc}, + \rangle$.[14] Let Δ_+ be the set of the following formulas:

$$x + y = y + x$$
$$0 + x = x$$
$$x + \mathsf{suc}(y) = \mathsf{suc}(x + y).$$

Now, Δ_+ defines $+$ implicitly over the model $\langle \omega, 0, \mathsf{suc} \rangle$. However, it is important to note that over the axiomatizable hull $\mathsf{Mod}(\mathsf{Th}(\langle \omega, 0, \mathsf{suc} \rangle))$ of this model, Δ_+ is not an implicit definition[15], and moreover addition is not nr-implicitly definable in $\mathsf{Mod}(\mathsf{Th}(\langle \omega, 0, \mathsf{suc} \rangle))$.

This shows that nr-implicit definability over a single model is much weaker than nr-implicit definability over an axiomatizable class of models. (Since primarily we are interested in theories, and theories correspond to axiomatizable classes, we are more interested in definability over axiomatizable classes than over single models.)

3. Let $E = \{2 \cdot n : n \in \omega\}$ be the set of even numbers. Then E as a *unary relation* is definable nr-implicitly over the model $\langle \omega, \mathsf{suc} \rangle$.

4. Let $\mathsf{BA_0}$ be the class of Boolean algebras with "$\cap$", "$\cup$", $0, 1$ as basic operations. Now, $\{x \cap -x = 0,\ x \cup -x = 1\}$ is an implicit definition of complementation over $\mathsf{BA_0}$. This implicit definition, however, can easily be rearranged into the form of an explicit definition as follows[16]:

$$-(x) = y \quad \Leftrightarrow \quad [x \cap y = 0 \ \wedge\ x \cup y = 1].$$

5. We recommend that the reader experiment with (i) defining the Boolean partial ordering "$\leq$" over $\mathsf{BA_0}$, (ii) defining "$\cup$" over the basic operations "$\cap, -$" (and the same with the roles of "$\cup$" and "$\cap$" interchanged).

6. The model $\langle \omega\ \leq \rangle$ is implicitly definable over $\langle \omega, 0, \mathsf{suc} \rangle$, but it is not nr-implicitly definable because $\langle \omega, \leq \rangle$ is not an expansion of $\langle \omega, 0, \mathsf{suc} \rangle$. If $\mathfrak{M}^+ = \langle \mathfrak{M}; \bar{R}^{new} \rangle$, i.e. if $\mathfrak{M}^+$ does not contain new sorts, then $\mathfrak{M}^+$ is nr-implicitly definable over $\mathfrak{M}$ iff $\mathfrak{M}^+$ is implicitly definable over $\mathfrak{M}$. This is not necessarily true when $\mathfrak{M}^+$ contains new sorts, too.

$\triangleleft$

[14] Where $\mathsf{suc} : \omega \longrightarrow \omega$ is the usual successor function on ω, i.e. $\mathsf{suc}(n) = n + 1$ for all $n \in \omega$.

[15] i.e. it does not satisfy $(\star\star)$ way above

[16] We have not discussed explicit definitions yet, but they will be discussed soon (beginning with §2.2 on p.19).

Examples 2.1.6 (More advanced, many-sorted examples)

1. Let $\mathfrak{F}$ be an ordered field. Then the two-sorted model $\langle \mathfrak{F}, \mathcal{P}(F); \in \rangle$ is *not definable implicitly up to isomorphism* over $\mathfrak{F}$. Hence it is not nr-implicitly definable, either.

   Proof-idea: Assume $|F| = \omega$. Then $|\mathcal{P}(F)| > \omega$. But by the downward Löwenheim-Skolem theorem $\langle \mathfrak{F}, \mathcal{P}(F); \in \rangle$ has an elementary submodel with each sort countable.

2. Let $\bar{R}$ be any countable sequence of relations defined on the sorts $F, \mathcal{P}(F)$ in example 1 above. Then
$$\langle \mathfrak{F}, \mathcal{P}(F); \in, \bar{R} \rangle$$
   is *not definable implicitly up to isomorphism* over $\mathfrak{F}$.

   Hint: The reason remains the same as in example 1.

   This means that $\mathfrak{F}^+ \stackrel{\text{def}}{=} \langle \mathfrak{F}, \mathcal{P}(F); \in \rangle$ is *not implicitly definable* over $\mathfrak{F}$, either.

   However, there is an expansion $\mathfrak{F}^{++}$ of $\mathfrak{F}^+$ with uncountably many new relations such that $\mathfrak{F}^{++}$ is nr-implicitly definable over $\mathfrak{F}$. Indeed, let us take a new constant c_x for each element x of $F \cup \mathcal{P}(F)$. Then $\mathfrak{F}^{++} \stackrel{\text{def}}{=} \langle \mathfrak{F}, \mathcal{P}(F), \in, \langle c_x : x \in F \cup \mathcal{P}(F) \rangle \rangle$ is an nr-implicitly definable expansion of $\mathfrak{F}$. This shows the importance of allowing only finitely many relation symbols in our languages when defining implicit definability, cf. p.10.

3. Let $\mathbf{F}$ be a finite field. Then $\langle \mathbf{F}, \mathcal{P}(F); \in \rangle$ *is definable nr-implicitly* over $\mathbf{F}$. The same applies for any finite structure in place of $\mathbf{F}$.

   Notation: For any set H and cardinal κ we let $\mathcal{P}_\kappa(H)$ be the collection of those subsets of H whose cardinality is smaller than κ. In particular, $\mathcal{P}_\omega(H)$ denotes the set of finite subsets of H.

4. Let $\mathfrak{A}$ be a(n infinite) structure with universe A. Then $\langle \mathfrak{A}, \mathcal{P}_i(A); \in \rangle$ *is nr-implicitly definable* over $\mathfrak{A}$ for any $i \in \omega$.

5. Let $\mathfrak{A} = \langle \omega, \leq \rangle$ be the set of natural numbers with the usual ordering. Then the expansion $\langle \mathfrak{A}, \mathcal{P}_\omega(\omega); \in \rangle$ *is nr-implicitly definable* over $\mathfrak{A}$.

   Hint: An implicit definition is the following set of formulas:

   $\{\forall x_1 \ldots x_n \in \omega \exists y \in \mathcal{P}_\omega(\omega) y = \{x_1, \ldots, x_n\} : n \in \omega\} \cup$
   $\{\forall y \in \mathcal{P}_\omega(\omega) \exists x \in \omega \forall z \in \omega (z \in y \to z \leq x)\} \cup$

$$\{\forall y, z \in \mathcal{P}_\omega(\omega)(y = z \leftrightarrow \forall x \in \omega(x \in y \leftrightarrow x \in z))\}.$$

(In the above, $y = \{x_1, \ldots, x_n\}$ abbreviates any formula with the intended meaning.)

As a contrast, we include the following example.

6. Consider the expansion $\langle \omega, \mathcal{P}_\omega(\omega); \in \rangle$ of the "plain" structure $\langle \omega \rangle$. Then this structure (i.e. $\langle \omega, \mathcal{P}_\omega(\omega); \in \rangle$) *is not implicitly definable up to isomorphism* over $\langle \omega \rangle$.

   Hint: Take any countable elementary submodel $\mathfrak{B}$ of an ultrapower of $\langle \omega, \mathcal{P}_\omega(\omega); \in \rangle$ which contains a "nonstandard" element in $\mathcal{P}_\omega(\omega)$. Then the "$\omega$-part" of $\mathfrak{B}$ is isomorphic to $\langle \omega \rangle$, but $\mathfrak{B}$ is not isomorphic to $\langle \omega, \mathcal{P}_\omega(\omega); \in \rangle$.

7. $\langle \omega, \mathcal{P}_\omega(\omega); \mathsf{suc}, \in \rangle$ is implicitly definable over $\langle \omega, \mathsf{suc} \rangle$. We do not know whether it is nr-implicitly definable over $\langle \omega, \mathsf{suc} \rangle$ or not. (We conjecture that the answer is in the negative.)

8. $\langle \mathfrak{A}, U^{new} \rangle$ *is not implicitly definable up to isomorphism* over $\mathfrak{A}$, for any structure $\mathfrak{A}$ and infinite set U^{new}. Here U^{new} is a new sort, and there are no new relations. If $1 < |U^{new}| < \omega$, then U^{new} is implicitly definable and implicitly definable up to isomorphism, but not implicitly definable without taking reducts. If $|U^{new}| \leq 1$, then U^{new} is implicitly definable without taking reducts.

9. Let $\mathfrak{A}$ be any structure and let $\mathfrak{B}$ be any finite structure. Then $\langle \mathfrak{A}, \mathfrak{B} \rangle$ as a two-sorted structure is impicitly definable over $\mathfrak{A}$.

10. Let $\mathfrak{A}$ be a fixed structure. Consider

$$\mathsf{K} = \{ \langle \mathfrak{A}; U^{new} \rangle \ : \ |U^{new}| < \omega \} .$$

    Then K is *not nr-implicitly definable* over $\{\mathfrak{A}\}$ (not even up to isomorphism).

Understanding the examples below is *not* a prerequisite for understanding the rest of the present work. (They concern the distinction between uniform and non-uniform definability.)

11. For $k \in \omega$, let $\mathfrak{L}_k$ be the usual $k+1$ element linear ordering $\mathfrak{L}_k = \langle \{0, \ldots, k\}, < \rangle$ where "$<$" is the usual ordering of the natural numbers. Recall from set theory that $\aleph_k$ is the k'th infinite cardinal regarded as a special ordinal. E.g. $\aleph_0 = \omega$. Let

$$\mathsf{K} := \{ \langle \aleph_k, \mathfrak{L}_k \rangle \ : \ k \in \omega \}$$

where $\langle U^{new}, \bar{R}^{new}\rangle = \mathfrak{U}_k$. I.e. K^- is obtained by forgetting the $\mathfrak{U}_k$-part. Then K is *not* uniformly nr-implicitly definable over K^- although for each $\mathfrak{M} \in \mathsf{K}$, we have that $\mathfrak{M}$ is nr-implicitly definable over $\mathfrak{M}^-$, i.e. $\mathfrak{U}_k$ is nr-implicitly definable over $\langle \aleph_k \rangle$.

12. The following is a generalization of item 11 above. Let $\mathfrak{A}_0, \dots, \mathfrak{A}_k, \dots$ $(k \in \omega)$ be *any* ω-sequence of elementarily equivalent one-sorted models.[17] Let $\mathfrak{U}_k$ be as in item 11 above.

$$\mathsf{K} := \{ \langle \mathfrak{A}_k, \mathfrak{U}_k \rangle \ : \ k \in \omega \} .$$

Then K is *not uniformly nr-implicitly definable* over $\mathsf{K}^- = \{ \mathfrak{A}_k \ : \ k \in \omega \}$ while every $\mathfrak{M} \in \mathsf{K}$ *is nr-implicitly definable* over $\mathfrak{M}^-$.

Hint: The key idea can be formulated with using $\mathfrak{A}_1$, $\mathfrak{A}_2$ only. The rest of the $\mathfrak{A}_k$'s serve only as decoration. So, one starts with $\mathfrak{A}_1 \equiv_{ee} \mathfrak{A}_2$ and[18] $|U_1| \neq |U_2|$ are finite. (Where U_i is the universe of $\mathfrak{U}_i$, similarly for A_i.) It is important to note that there are no inter-sort relations permitted here i.e. sort A_i is isolated from sort U_i. Next, one uses the following property of many-sorted logic. Assume $\mathfrak{A}, \mathfrak{B}$ are two structures of *disjoint languages*. Consider the new many-sorted structure $\langle \mathfrak{A}, \mathfrak{B} \rangle$. We claim that $\mathsf{Th}(\langle \mathfrak{A}, \mathfrak{B}\rangle) = \mathsf{Th}(\mathfrak{A}) \cup \mathsf{Th}(\mathfrak{B})$. The reason for this is the fact that an atomic formula xRy belongs to a many-sorted language *only* if x and y are of the *same* sort. Hence e.g. $(\exists x \in U_0)(\exists y \in U_1)\, x \neq y$ is not a (many-sorted) formula.

The present example does not work for "implicitly definable" in place of "implicitly definable without taking reducts".

Someone might think that the reason why the above counterexample works is that all elements of K^- are elementarily equivalent. Below we show that this is *not* the case.

13. Let the language of K^- consist of countably many constant symbols $c_0, \dots, c_i, \dots$ and just one sort. Let $\mathfrak{U}_k$ $(k \in \omega)$ be as in item 11 above.

$$\begin{aligned}
\mathsf{K}^- &:= \{ \langle U, c_i \rangle_{i\in\omega} \ : \ \text{the set } \{ i \in \omega \ : \ c_i = c_0 \} \text{ is finite and} \\
&\qquad\qquad\qquad\qquad\qquad U \text{ is a set with } (\forall i \in \omega)\, c_i \in U \} . \\
\mathsf{K} &:= \{ \langle U, c_i; \mathfrak{U}_k \rangle_{i\in\omega} \ : \ k = |\{ i \in \omega \ : \ c_i = c_0 \}| \text{ and } \langle U, c_i \rangle_{i\in\omega} \in \mathsf{K}^- \} .
\end{aligned}$$

That is

[17] I.e. $(\forall k \in \omega)\, \mathsf{Th}(\mathfrak{A}_0) = \mathsf{Th}(\mathfrak{A}_k)$.

[18] Recall that $\equiv_{ee}$ denotes the binary relation of elementary equivalence defined between models.

$$\mathsf{K} \quad = \quad \big\{\, \langle \mathfrak{M};\, \mathfrak{U}_k \rangle \;:\; \mathfrak{M} \in \mathsf{K}^- \text{ and } k = |\,\{\, i \in \omega \;:\; \text{in } \mathfrak{M} \text{ we have } c_i = c_0 \,\}\,| \,\big\}\,.$$

Now, K is *not uniformly nr-implicitly definable* over K^- *while* each concrete $\mathfrak{M} \in \mathsf{K}$ *is nr-implicitly definable* over $\mathfrak{M}^-$, *further*

$$(\forall \mathfrak{M}, \mathfrak{N} \in \mathsf{K})[\,\mathfrak{M}^- \equiv_{ee} \mathfrak{N}^- \quad \Rightarrow \quad \mathfrak{M} \equiv_{ee} \mathfrak{N}\,].$$

Idea for a proof:

Assume $\Delta = \mathsf{Th}(\mathsf{K})$ defines K implicitly over K^- (up to isomorphisms). Then by using ultraproducts one can show that there is $\mathfrak{N} = \langle U, c_i;\, \mathfrak{U}_2 \rangle_{i \in \omega} \in \mathsf{Mod}(\Delta)$ such that $(\forall i > 0)(c_i \neq c_0$ holds in $\mathfrak{N})$. But clearly for $\mathfrak{M} := \langle \mathfrak{N}^-;\, \mathfrak{U}_1 \rangle$ we have $\mathfrak{N}^- = \mathfrak{M}^- \in \mathsf{K}^-$ and $\mathfrak{M} \in \mathsf{K}$ hence by $\mathfrak{M} \not\cong \mathfrak{N}$ we conclude that Δ cannot be a definition of K.

$\triangleleft$

The above three examples were designed to illustrate the difference between uniform (nr-implicit) definability and *one-by-one (nr-implicit) definability* where by the latter we understand the case when each $\mathfrak{M} \in \mathsf{K}$ is definable over its reduct $\mathfrak{M}^-$ in K^- (but these definitions might be different for different choices of $\mathfrak{M}$);in more detail: Let K be a class of models with U^{new}, $\bar{R}^{new}$ in the language of K. For $\mathfrak{M} \in \mathsf{K}$ let $\mathfrak{M}^-$ be the reduct of $\mathfrak{M}$ obtained by omitting (forgetting) U^{new}, $\bar{R}^{new}$. Let $\mathsf{K}^- := \{\, \mathfrak{M}^- \;:\; \mathfrak{M} \in \mathsf{K} \,\}$. Then we say that K is <u>*one-by-one nr-implicitly definable*</u> over K^- iff each $\mathfrak{M} \in \mathsf{K}$ is nr-implicitly definable over its reduct $\mathfrak{M}^- \in \mathsf{K}^-$. Sometimes, informally we will use instead of one-by-one definability <u>*"non-uniform"*</u> or <u>*"local"*</u> definability as synonyms. We hope that the above three examples illustrate (the generally accepted opinion) that uniform definability is a more useful concept than one-by-one definability (when considering classes K of models) and is closer to what one would intuitively understand under definability.

For completeness, we refer the interested reader to the distinction between the *"local"* and the "usual" versions of explicit definability described in Andréka-Németi-Sain [4] Definitions 55–56 (Beth definability properties) therein. We also note that most standard textbooks concentrate on uniform definability only and they do not mention what we call here one-by-one definability. We too will concentrate on uniform definability and unless otherwise specified, by <u>*definability*</u> we will always understand <u>*uniform definability*</u>.

Remark 2.1.7 A useful refinement of the notion of nr-implicit definability is *finite nr-implicit definability.* Assume K and K^- are as above statement $(\star\star)$ on p.9 (definition of nr-implicit definability). Assume, K is nr-implicitly definable over K^-.

Then K is said to be <u>*finitely nr-implicitly definable*</u> over K^- iff there is a finite set $\Delta_0 \subseteq \mathsf{Th}(\mathsf{K})$ of formulas such that Δ_0 defines K implicitly over K^-, i.e. $(\star\star)$ holds for $\Delta = \Delta_0$. In most of our concrete examples and applications we will have *finite* nr-implicit definability, but for simplicity we will write just "definability".

To illustrate the importance of finite nr-implicit definability, consider the simple model $\langle \omega, \mathsf{suc} \rangle$. There are continuum many different implicit definitions (involving one new relation symbol R) over this model while there are only countably many finite implicit definitions (and we will see that there are only countably many explicit definitions over this model). (This example cannot be generalized from a single model like $\mathfrak{M} = \langle \omega, \mathsf{suc} \rangle$ to first-order-axiomatizable classes K of models, assuming there are only finitely many sorts).[19]

$\lhd$

2.2 Explicit definability in many-sorted (first-order) logic

So far we have discussed implicit definability which is a quite general notion of definability. Below we will turn to a special kind of implicit definability which we call <u>*explicit definability*</u>. Each explicit definition can be considered as an implicit definition. The other direction is not true however, there are implicit definitions which are not explicit definitions. (I.e. there is an implicit definition Δ which in its *given form* is not an explicit definition.) In definability theory, the connection between explicit and implicit definitions is an important subject. We will return to this subject at the end of the "definability" section (§2). In particular, we will state a generalization of Beth's theorem, saying that implicit definability is equivalent with explicit definability (even in our general framework where we allow definitions of new sorts, too [besides definitions of new relations], cf. Theorem 3.3.1 and Corollary 2.5.2 on p.70.

Explicit definability will turn out to be (i) a special case of implicit definability and (ii) a strong and useful concept e.g. in the following way. Assume

[19] The reason for this is the following. In the above reasoning we heavily used the fact that every element of $\langle \omega, \mathsf{suc} \rangle$ is definable "as a constant". (Therefore infinite implicit definitions can be given by listing the elements of R and the non-elements of R.) This does not remain true in $\mathsf{Mod}(\mathsf{Th}(\langle \omega, \mathsf{suc} \rangle))$.

$\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$ and that K^+ is an expansion of K which is explicitly definable over the class K of models. Then the theories $\mathsf{Th}(\mathsf{K})$ and $\mathsf{Th}(\mathsf{K}^+)$ as well as the languages of K and K^+ will be seen to be equivalent in a rather strong sense to be explained later, see Theorems 2.3.2 and 2.3.4 on p.35.

The key ingredients of explicit definability will be introduced in items (1)–(2.2) below. Then, on p.25, they will be combined into a description of what we mean by explicit definability. The generalization from definability over single models $\mathfrak{M}$ to definability over classes K of models will be given on p.25.

<u>*Notation:*</u> Assume $\mathfrak{M}$ is a many-sorted model and that ψ is a formula in the language of $\mathfrak{M}$ such that all the free variables of ψ belong to $x_0, \ldots, x_i, \ldots$. Assume $\bar{a} \in {}^{\omega}Uv(\mathfrak{M})$ and that the sort of a_i coincides with the sort of the variable x_i, for every $i \in \omega$. Then

$$\mathfrak{M} \models \psi[\bar{a}]$$

is the standard model theoretic notation for the statement that ψ is true in $\mathfrak{M}$ under the <u>*evaluation*</u> $\bar{a}$ of its free variables cf. e.g. Monk [13], Enderton [8], Chang-Keisler [7]. Sometimes we write $\mathfrak{M} \models \psi[a_1, \ldots, a_n]$ in which case it is understood that the free variables of ψ are among $x_1, \ldots, x_n$. The latter is often indicated by writing $\psi(x_1, \ldots, x_n)$ instead of ψ. I.e. if we write $\psi(x_1, \ldots, x_n)$ in place of ψ then this means that while talking about the formula ψ we want to indicate casually that the free variables of ψ are among $x_1, \ldots, x_n$.

The following is also a standard notation from logic. Assume τ is a term. Then $\psi(x/\tau)$ denotes the formula obtained from ψ by replacing all free occurrences of x by τ. Similarly for $\psi(x_1/\tau_1, \ldots, x_n/\tau_n)$. We could say that "$(x/\tau)$" is the "operator" of substituting τ for x.

If $\psi(x)$ is a formula and y is a variable (of the same sort as x), then $\psi(y)$ denotes $\psi(x/y)$; and similarly for a sequence $\bar{x}$ of variables.

We will write "definable" for "explicitly definable" to save space. Similarly, we write "definitional expansion" for "explicit definitional expansion". In general, we will tend to omit the adjective "explicit", because our primary interest will be explicit definability.

(1) Explicit definability of relations and functions in $\mathfrak{M}$.

Let $\mathfrak{M} = \langle U_0, \ldots, U_j;\ R_1, \ldots, R_l \rangle$ be a many-sorted model with universes or sorts $U_0, \ldots, U_j$, and relations $R_1, \ldots, R_l$. Let $R^{new} \subseteq U_{i_1} \times \ldots \times U_{i_m}$ be a (new) relation, with $i_1, \ldots, i_m \in (j+1)$. Now, R^{new} is called <u>*(explicitly) definable*</u> (as a relation) over $\mathfrak{M}$ iff there is a formula $\psi(x_{i_1}, \ldots, x_{i_m})$ in the language of $\mathfrak{M}$ such that

$$R^{new} = \{ \langle a_{i_1}, \ldots, a_{i_m} \rangle \in U_{i_1} \times \ldots \times U_{i_m} \ :\ \mathfrak{M} \models \psi[a_{i_1}, \ldots, a_{i_m}] \} .$$

Such definable relations can be added to $\mathfrak{M}$ as new basic relations obtaining a(n explicit) *definitional expansion* of $\mathfrak{M}$ in the form

$$\mathfrak{M}^+ := \langle U_0, \ldots, U_j;\ R_1, \ldots, R_l, R^{new} \rangle.$$

To make $\mathfrak{M}^+$ "well defined" we have to add a *new relation symbol* to the language of $\mathfrak{M}$ denoting R^{new}. The formula $R^{new}(\bar{x}) \leftrightarrow \psi(\bar{x})$ is called an *(explicit) definition* of R^{new} (over $\mathfrak{M}$). Notice that $\Delta \stackrel{\text{def}}{=} \{R^{new}(\bar{x}) \leftrightarrow \psi(\bar{x})\}$ is also a(n implicit) definition of $\mathfrak{M}^+$ over $\mathfrak{M}$. We call Δ an *explicit definition of type (1)*. If Δ is a definition of $\mathfrak{M}'$ over $\mathfrak{M}$, then we say that *$\mathfrak{M}'$ is obtained from $\mathfrak{M}$ by step (1)*. Note that if $\mathfrak{M}'$ is defined over $\mathfrak{M}$ by Δ, then $\mathfrak{M}'$ is $\mathfrak{M}^+$ above.

(2) Explicit definability of new sorts (i.e. universes) in $\mathfrak{M}$.

Defining a new sort explicitly (in $\mathfrak{M}$) takes a bit more care than defining a new relation. This is understandable, since now we want to define (or create) a *new universe* of entities (in terms of the old universes and old relations already available in $\mathfrak{M}$) while when defining a relation we defined only a new property of *already existing* entities (or of tuples of such entities) in $\mathfrak{M}$. If we define a new relation, then this amounts to defining a new property of already existing entities. I.e. we remain on the same *ontological level*. In contrast, if we define *new entities* which "did not exist" before, then we go up to a higher ontological level.[20]

If we want to define a new sort in $\mathfrak{M}$, first of all we need a new sort-symbol, say U^{new}, which does not yet occur in the language of $\mathfrak{M}$. If there is no danger of confusion then we will *identify* a sort-symbol like U^{new} with the universe, say $(U^{new})^{\mathfrak{M}^+}$, which it denotes in a model $\mathfrak{M}^+$.

An explicit definition of a new sort, say U^{new}, describes the elements of U^{new} as being constructed from "old" elements in a systematic, "tangible" and uniform way. More concretely, first we will introduce a few (basic constructions or) basic kinds of explicit definition and then "general" explicit definitions will be obtained by iterating these basic kinds. We will refer to the just mentioned basic kinds (of explicit definition) as basic steps of explicit definitions. Our basic steps for building up explicit definitions of new sorts are described in items (2.1), (2.2) below. Our choice of basic steps might look ad-hoc at first reading, but Theorem 3.3.1 at the end of this section will say that our selected few steps (i.e. examples of explicit definitions) cover (via iteration) all cases of implicit definitions (assuming there is a

[20] In connection with defining new sorts, for completeness, we also refer e.g. to the definition of the "new" many-sorted structure A^{eq} from the "old" structure A in Hodges [11, p.151] (cf. also pp. 148, 212, 213 therein). Cf. also the definition of relative categoricity in Hodges [11] p.638 together with p.638 line 3 bottom up to p.639 line 9.

sort with more than one elements). We will return to a more careful discussion of the present issue of choosing our basic steps in Remark 2.2.4.

(2.1) The first way of defining a new sort U^{new} in $\mathfrak{M}$ explicitly.

The simplest way of defining a new sort U^{new} in a model $\mathfrak{M} = \langle U_0, \ldots, U_j;\, R_1, \ldots, R_l\rangle$ is the following. Let $R \in \{R_1, \ldots, R_l\}$ be fixed. Assume R is an r-ary relation, i.e. $R \subseteq {}^r Uv(\mathfrak{M})$. We want to postulate that U^{new} coincides with R. So the first part of our definition of U^{new} is the postulate:

$$U^{new} \overset{\text{def}}{:=} R.$$

But, if we want to expand $\mathfrak{M}$ with U^{new} as a new sort obtaining something like

$$\mathfrak{M}' := \langle U_0, \ldots, U_j, U^{new};\, R_1, \ldots, R_l\rangle$$

then we *need* some new relations or functions *connecting* the new sort U^{new} to the old ones $U_0, \ldots, U_j$. In our present case (of $U^{new} = R$) we use the <u>*projection functions*</u> $pj_i : R \longrightarrow Uv(\mathfrak{M})$ with $i < r$. Formally,

$$pj_i(\langle a_0, \ldots, a_{r-1}\rangle) \overset{\text{def}}{:=} a_i.$$

To identify the domain of pj_i we should write something like pj_i^R, but for brevity we omit the superscript R. Now, the (explicit) *definitional expansion* of $\mathfrak{M}$ obtained by the choice $U^{new} := R$ is

$$\mathfrak{M}^+ := \langle U_0, \ldots, U_j, U^{new};\, R_1, \ldots, R_l, pj_0, \ldots, pj_{r-1}\rangle = \langle \mathfrak{M}, U^{new};\, pj_i\rangle_{i<r}.$$

We note that

$$\mathfrak{M}^+ = \langle U_0, \ldots, U_j, R;\, R_1, \ldots, R_l, pj_i^R\rangle_{i<r}.$$

If x is a variable, then $(\exists\,!x)\psi(x)$ denotes the formula expressing that there is exactly one value for which ψ holds, i.e. it denotes the formula $(\exists x)(\psi(x) \wedge (\forall z)[\psi(z) \rightarrow z = x])$. Let

$$\begin{aligned}\Delta \overset{\text{def}}{=} \{&(\exists\,!u \in U^{new})(pj_1(u, x_1) \wedge \ldots \wedge pj_r(u, x_r)) \leftrightarrow R(x_1, \ldots, x_r)\ ,\\ &(\exists u \in U^{new})(pj_1(u, x_1) \wedge \ldots \wedge pj_r(u, x_r)) \rightarrow R(x_1, \ldots, x_r)\ ,\\ &(\forall u \in U^{new})(\exists\,!x_i)pj_i(u, x_i) : 1 \leq i \leq r\}\ .\end{aligned}$$

Then Δ is an implicit definition of $\mathfrak{M}^+$ over $\mathfrak{M}$. We call Δ an <u>*explicit definition of type (2.1)*</u>. If Δ is a definition of $\mathfrak{M}'$ over $\mathfrak{M}$, then we say that <u>*$\mathfrak{M}'$ is obtained from $\mathfrak{M}$ by Step (2.1)*</u>. Notice that if Δ is a definition of $\mathfrak{M}'$ over $\mathfrak{M}$, then $\mathfrak{M}'$ is isomorphic to $\mathfrak{M}^+$ above via an isomorphism which is identity on $\mathfrak{M}$.

Remark 2.2.1 This second form of $\mathfrak{M}^+$ might induce the (misleading) impression that $\mathfrak{M}^+$ contains nothing new: it consists of a rearranged version of the old parts of $\mathfrak{M}$. However, let us notice that as a first step we might define a new relation R^{new} in $\mathfrak{M}$ (in the style of item (1) above) obtaining

$$\mathfrak{M}^+ := \langle U_0, \ldots, U_j;\ R_1, \ldots R_l, R^{new} \rangle$$

and *then* we may define $U^{new} := R^{new}$ obtaining the *definitional expansion*

$$\mathfrak{M}^{++} := \langle U_0, \ldots, U_j, U^{new};\ R_1, \ldots, R^{new}, pj_i \rangle_{i<r}$$

of $\mathfrak{M}^+$. Now, we will *postulate* that a definitional expansion of a definitional expansion of $\mathfrak{M}$ is called a definitional expansion of $\mathfrak{M}$ again. Hence the above obtained $\mathfrak{M}^{++}$ is a definitional expansion of the original $\mathfrak{M}$. Using our abbreviation from p.4 we can write:

$$\langle \mathfrak{M}, U^{new};\ R^{new}, pj_i \rangle_{i<r} := \mathfrak{M}^{++}.$$

Now, *if* we do not want to have R^{new} as a relation, we can take the reduct

$$\mathfrak{M}^{++-} := \langle \mathfrak{M}, U^{new};\ pj_i \rangle_{i<r}$$

by forgetting R^{new} as a relation but not as a sort. We will call $\mathfrak{M}^{++-}$ a *generalized definitional expansion* of $\mathfrak{M}$ (cf. p.25).

$\lhd$

Example 2.2.2 Let $\mathbf{F} = \langle F, \ldots, \cdot \rangle$ be a field. We want to define the plane $F \times F$ over $\mathbf{F}$ as a new *sort* expanding $\mathbf{F}$. First we define the *relation* $R = F \times F$ by the formula $(x_0 = x_0\ \wedge\ x_1 = x_1)$. Clearly, in $\mathbf{F}$ this formula defines the relation $F \times F$. Then we expand $\mathbf{F}$ with this as a new relation obtaining

$$\mathbf{F}^+ = \langle F;\ +, \cdot\ , F \times F \rangle$$

where $F \times F$ is used as a relation interpreting the relation symbol $Rel_{F\times F}$. Now, in $\mathbf{F}^+$ we define the *new sort* $U^{new} := F \times F$ together with the projection functions as indicated above, obtaining the model

$$\mathbf{F}^{++} = \langle F, F \times F;\ +, \cdot\ , F \times F, pj_0, pj_1 \rangle$$

where $pj_i : F \times F \longrightarrow F$. Now, we take a reduct of $\mathbf{F}^{++}$ by forgetting the relation symbol $Rel_{F\times F}$, but not the sort $F \times F$. We obtain

$$\mathbf{F}^{++-} = \langle F, F \times F;\ +, \cdot\ , pj_0, pj_1 \rangle = \langle \mathbf{F}, F \times F;\ pj_0, pj_1 \rangle.$$

Clearly this model $\mathbf{F}^{++-}$ is the expansion of the field $\mathbf{F}$ with the plane $F \times F$ as a new sort as we wanted.

The above example shows that the usual expansion of $\mathbf{F}$ with the plane as a *new sort*, is indeed a definitional expansion i.e. the plane as a new sort is *(first-order) definable explicitly* in $\mathbf{F}$.

$\triangleleft$

Similarly to the above example, ${}^nF$ is first-order definable (explicitly) as a *new sort* in any frame model $\mathfrak{M}$. Later we will introduce uniform explicit definability over a class K of models. Then we will see that ${}^nF$ as a new sort is uniformly (explicitly) definable over the class of all frame models. (In defining ${}^nF$ we use $pj_i : {}^nF \longrightarrow F$, $i \in n$, the same way as we did in the case of $\mathbf{F}^{++-}$.)

(2.2) The second way of defining a new sort U^{new} in $\mathfrak{M}$ explicitly.

To define a new sort U^{new} in a model $\mathfrak{M} = \langle U_0, \ldots, U_j; R_1, \ldots, R_l \rangle$ explicitly the second way, we begin by selecting an old sort $U := U_i$ and old relation $R := R_k$ ($i \leq j$, $0 < k \leq l$) in $\mathfrak{M}$. We proceed *only if* R happens to be an equivalence relation over U (i.e. if $R \subseteq U \times U$ etc.). We define the new sort to be the quotient set of R-equivalence classes[21]

$$U^{new} := U/R.$$

Again, similarly to the case of pj_i's in item (2.1) above, we need a new relation connecting the new sort U^{new} to the old ones. Now we choose the set theoretic membership relation

$$\in \; := \; \in_{U^{new}} \; := \; \in_{U,U^{new}} \; := \{ \langle a, a/R \rangle \; : \; a \in U \}$$

acting between U and U/R. Since $\in_{U^{new}} \subseteq U_i \times U^{new}$, this relation connects the new sort U^{new} with the old one U_i. Let us notice that from the notation $\in_{U,U^{new}}$ we may omit the first index obtaining the simpler notation $\in_{U^{new}}$ or we may omit both indices obtaining $\in$. The (explicit) *definitional expansion* of $\mathfrak{M}$ obtained by the choice $U^{new} = U_i/R_k$ is defined to be the model

$$\begin{aligned}
\mathfrak{M}^+ &= \langle U_0, \ldots, U_j, U^{new}; \, R_1, \ldots R_l, \in_{U^{new}} \rangle \\
&= \langle U_0, \ldots, U_i/R_k; \, R_1, \ldots, R_l, \in \rangle \\
&= \langle \mathfrak{M}, U^{new}; \, \in_{U^{new}} \rangle \\
&= \langle \mathfrak{M}, U_i/R_k; \, \in_{U^{new}} \rangle.
\end{aligned}$$

Let

[21] $U/R \stackrel{\text{def}}{=} \{a/R : a \in U\}$ where $a/R \stackrel{\text{def}}{=} \{b \in U : \langle a, b \rangle \in R\}$. I.e. U/R is the set of all "blocks" of R, and a/R is the "block" of R a is in.

$$\Delta \stackrel{\text{def}}{=} \{(\exists u \in U^{new})(\in(x,u) \wedge \in(y,u)) \leftrightarrow R(x,y),$$
$$[\in(x,u) \wedge \in(x,v)] \to u = v \quad \}.$$

Then Δ is an implicit definition of $\mathfrak{M}^+$ over $\mathfrak{M}$. We call Δ an *explicit definition of type (2.2)*. If Δ is a definition of $\mathfrak{M}'$ over $\mathfrak{M}$, then we say that *$\mathfrak{M}'$ is obtained from $\mathfrak{M}$ by Step (2.2)*. Notice that if Δ is a definition of $\mathfrak{M}'$ over $\mathfrak{M}$, then $\mathfrak{M}'$ is isomorphic to $\mathfrak{M}^+$ above via an isomorphism which is identity on $\mathfrak{M}$.

* * *

We are ready for defining our notion of explicit definability. We call a new *sort or relation (explicitly) definable in* $\mathfrak{M}$ iff it is definable by repeated applications of the steps described in items (1), (2.1), (2.2) above.

A model $\mathfrak{N}$ is called a *definitional expansion of* $\mathfrak{M}$ iff $\mathfrak{N}$ is obtained from $\mathfrak{M}$ by repeated applications of steps (1), (2.1), (2.2) above (involving finitely many steps only). An *explicit definition of $\mathfrak{N}$ over $\mathfrak{M}$* is the union of the explicit definitions of type (1), (2.1), (2.2) involved in a sequence leading from $\mathfrak{M}$ to $\mathfrak{N}$. We call Δ an *explicit definition* if Δ is an explicit definition of some definitional expansion.

A model $\mathfrak{N}$ is called a *generalized definitional expansion* of $\mathfrak{M}$ if (i), (ii) below hold.

(i) $\mathfrak{N}$ is a reduct of a definitional expansion, say $\mathfrak{M}^+$, of $\mathfrak{M}$.

(ii) $\mathfrak{N}$ is an expansion of $\mathfrak{M}$, i.e. $\mathfrak{M}$ is a reduct of $\mathfrak{N}$.

We call $\mathfrak{N}$ *(explicitly) definable in* $\mathfrak{M}$ iff item (i) above holds. If we want to indicate that we do not take a reduct while defining say $\mathfrak{M}^+$ from $\mathfrak{M}$ explicitly (i.e. that $\mathfrak{M}^+$ is obtainable by repeatedly applying steps (1), (2.1), (2.2) to $\mathfrak{M}$) then we say that $\mathfrak{M}^+$ is *explicitly definable* in $\mathfrak{M}$ *without taking reducts*. Sometimes we write *"definitional expansion without taking reducts"* to emphasize that we mean definitional expansion and not generalized definitional expansion.

We emphasize that a precise statement claiming that U^{new} is definable as a new sort should also mention the relations and/or functions (of $\mathfrak{N}$) connecting U^{new} to the original sorts of $\mathfrak{M}$. Examples for such "connecting relations" are pj_i and $\in_{U^{new}}$ discussed above.

We note that *explicit definability with parameters* is completely analogous with implicit definability with parameters cf. p.9.

Let us turn to (explicit) definability over a *class* K of models (instead of over a single model $\mathfrak{M}$). We say that K is a(n explicit) *definitional expansion* of its reduct K^- iff K can be obtained from K^- by (a finite sequence of) repeated (uniform) appications of the steps described in items (1), (2.1), (2.2) on pp.20–25. This is equivalent to saying that there is an explicit definition which defines K over K^- (as an implicit definition). In this case we also say that K *is (explicitly) definable over (or in)* K^- *without taking reducts*. We say that K is a *generalized definitional expansion* of K^- if K is an expansion of K^- and K is a reduct of a definitinal expansion of K^-. We say that K is *(explicitly) definable in* L if K is a reduct of a definitional expansion of L.

This is completely analogous with the case of implicit definability. *Uniform (explicit) definability* and *one-by-one (explicit) definability* are obtained from the notion of (explicit) definability for single models the same way as their counterparts were obtained in the case of implicit definability, cf. pp. 10, 18.

Finally, we introduce one more notion of definability which we will call *rigid definability*. We will use this in our examples to come. About the importance of this notion see Theorem 2.3.7 on p.49.

Assume $\mathfrak{M}^+ = \langle \mathfrak{M}, \bar{U}^{new}; \bar{R}^{new} \rangle$ is an expansion of $\mathfrak{M}$ (with new sorts and relations). We say that $\mathfrak{M}^+$ is *(explicitly) rigidly definable over* $\mathfrak{M}$ if $\mathfrak{M}^+$ is definable in $\mathfrak{M}$ and the identity is the only automorphism of $\mathfrak{M}^+$ which is the identity on $\mathfrak{M}$. Informally, we will say that the new sorts and relations $\bar{U}^{new}, \bar{R}^{new}$ are rigidly definable over $\mathfrak{M}$ if $\langle \mathfrak{M}; \bar{U}^{new}, \bar{R}^{new} \rangle$ is rigidly definable over $\mathfrak{M}$.

Further, K^+ is *rigidly definable over* K iff K^+ is a generalized definitional expansion of K and each $\mathfrak{M}^+ \in \mathsf{K}^+$ is rigid(ly definable) over its K-reduct.[22]

In our opinion, rigid definability is "just as good" as definability without taking reducts. In other words, we feel that if $\bar{U}^{new}$ etc. are rigidly definable over K then $\bar{U}^{new}$ etc. are almost as well determined by K (or describable in K) as if they were definable without taking reducts. Cf. Theorem 2.3.7, Theorem 2.3.4, and Theorem 3.3.1. We note that rigid definability seems to be perhaps, our most important (or most central) version of definability.

Remark 2.2.3 (Forming disjoint union of two sorts) For didactical reasons we will refer to items (1)–(2.2) as steps (1)–(2.2) to emphasize their roles in constructing an explicit definition (for some new class K^+) in a step-by-step manner.

We could have included in this list of steps as step (2.3) the definition of a new sort as a disjoint union of two old sorts. This goes as follows:

[22]Our definition of K^+ being explicitly definable over K is strongly related to the notion of K^+ being "*coordinatisable over*" K as defined in Hodges [11, p.644], while K^+ is rigidly definable over K is strongly related to "*coordinatised over*" as defined in [11] (same page). We will return to discussing this connection in the sub-section beginning on p.69.

Assume U_k, U_m are old sorts, i.e. sorts of $\mathfrak{M}$, while U^{new} is not a sort of $\mathfrak{M}$. Then, we can define the new sort as

$$U^{new} := U_k \dot{\cup} U_m$$

with two injections

$$i_1 : U_k \rightarrowtail U^{new} \qquad \text{and} \qquad i_2 : U_m \rightarrowtail U^{new}$$

such that U^{new} is the union of $Rng(i_1)$, $Rng(i_2)$ and $Rng(i_1) \cap Rng(i_2) = \emptyset$. Here $k = m$ is permitted. But even if $k = m$, i_1 and i_2 are different. Now the expanded model is

$$\mathfrak{M}^+ := \langle \mathfrak{M}, U^{new};\ i_1, i_2 \rangle.$$

We note that such an $\mathfrak{M}^+$ is always implicitly definable over $\mathfrak{M}$, further *all the nice properties* of *explicit definitions*[23] in items (1)–(2.2) hold for this new kind of explicit definition which from now on we will consider as step (2.3) of explicit definability.

All the same, we do not include step (2.3) into the list of permitted steps of building up an explicit definition. We have two reasons for this.

(i) Step (2.3) can be reduced to (or simulated by) steps (1)–(2.2). Namely, assume $\mathfrak{M}^+$ is defined from $\mathfrak{M}$ by using step (2.3). Assume further that $\mathfrak{M}$ has a sort U_i with more than one elements (i.e. $|U_i| > 1$). Then by using steps (1)–(2.2) one can define an expansion $\mathfrak{M}^{++}$ from $\mathfrak{M}$ such that $\mathfrak{M}^+$ is a reduct of $\mathfrak{M}^{++}$.[24]

Further:

(ii) We will not need step (2.3) in the present work. I.e. in the logical analysis of relativity, explicit definitions of form (2.3) did not come up so far.

Item (i) above shows that adding step (2.3) to the permitted steps of explicit definitions would increase the collection of sorts and relations definable over $\mathfrak{M}$ *only* in the pathological case when all universes of $\mathfrak{M}$ have cardinalities ≤ 1.

Therefore while noting that step (2.3) could be included without changing the theory of explicit definability significantly, we do *not* include it. However, sometime (in some intuitive text) when we want to get "dreamy" we might refer to explicit definability as involving four steps (1)–(2.3). $\triangleleft$

[23] As an example we mention that explicitly defined symbols can be eliminated from the language, cf. sub-section 2.3 on p.34.

[24] More precisely there is a *unique* isomorphism h between $\mathfrak{M}^+$ and this reduct of $\mathfrak{M}^{++}$ such that $h \restriction \mathfrak{M}$ is the identity function.

Remark 2.2.4 One might want to develop a more systematic understanding of what explicit definitions are. For such a more systematic understanding of explicit definitions let us rearrange the basic steps into steps (1*)–(5*) below.

(1*) Definition of new relations $\bar{R}^{new}$ explicitly the classical way (as in item (1) on p.20).

(2*) Definition of new sorts as <u>*direct products*</u> of old sorts together with projection functions ($U^{new} := U_i \times U_j$ etc) (as in item (2.1) on p.22).

(3*) Definition of new sorts as <u>*disjoint unions*</u> of old sorts together with inclusion functions ($U^{new} := U_i \dot{\cup} U_j$ etc) (as in item (2.3) on p.26).

(4*) Definition of a new sort as a definable <u>*subset*</u> of an old sort together with an inclusion function. I.e.

$$U^{new} := \{\, x \in U_i \;:\; \mathfrak{M} \models \psi(x) \,\}$$

and $i_{new} : U^{new} \longrightarrow U_i$ is the usual inclusion function. The expanded model is $\mathfrak{M}^+ = \langle \mathfrak{M}, U^{new}; i_{new} \rangle$.

(5*) Definition of a new sort as a definable <u>*quotient*</u> of an old sort exactly as in item (2.2) on p.24 (i.e. $U^{new} = U_i/R$ etc).

Now, an explicit definition in the new sense is given by an arbitrary sequence (i.e. iteration) of steps (1*)–(5*) above.

If we disregard the trivial case when all sorts are singletons or empty, then explicit definitions in the new sense are equivalent with explicit definitions as introduced in §2.2. We leave checking this claim to the reader.

We would like to point out that explicit definitions as built up from steps (1*)–(5*) are not ad-hoc at all. In the category theoretic sense the formation of disjoint unions is the *dual* of the formation of direct products and the formation of sub-universes (or sub-structures) is the dual of the formation of quotients. So, we are left with two basic steps and their duals.

It is interesting to note that our steps (2*)–(5*) correspond to basic operations producing new models from old ones. (Indeed if U_i is a universe of $\mathfrak{M}$ then we can restrict $\mathfrak{M}$ to U_i and then we obtain a one-sorted reduct of $\mathfrak{M}$ with universe U_i. Hence creating new sorts from old ones is not unrelated to creating new models from old ones. All the same, we do not want to stretch this analogy too far.)

What we would like to point out here, is that steps (2*)–(5*) seem to form a natural, well balanced set of basic operations, while step (1*) has been inherited from the classical theory of definability.

Further, we note that while selecting our basic steps (e.g. steps (1*)–(5*) above) we had to be careful to keep them implicitly definable i.e. they should not lead to "explicitly definable things" which are not implicitly definable. Therefore operations like formation of powersets cf. Example 2.1.6(1) (or all finite subsets of a set cf. Example 2.1.6(6))[25] are ruled out from the beginning.

$\lhd$

CONVENTION 2.2.5 Assume K^+ is a definitional expansion of K. For $\mathfrak{M}^+ \in \mathsf{K}^+$ the reduct $\mathfrak{M}^+ \restriction Voc\mathsf{K}$ may have more than one definitional expansions in K^+. (However these expansions are isomorphic.) Therefore K may have several different definitional expansions $\mathsf{K}^\oplus$ with the same set of defining formulas say Δ which defines K^+ from K. In such cases, of course we have $\mathbf{I}\mathsf{K}^\oplus = \mathbf{I}\mathsf{K}^+$. The largest such class is called a *maximal* definitional expansion of K. Since most of the time we will be interested in classes of models *closed under isomorphisms*, sometimes, but not always, we will concentrate on maximal definitional expansions. There are important exceptions to this[26], e.g. the class of two-sorted geometries[27] is not closed under isomorphisms and despite of this we will say that it is a definitional expansion of the class of one-sorted geometries (in Tarski's sense), under some conditions of course.

$\lhd$

Remark 2.2.6 (On isomorphism-closure) In Convention 2.2.5 above, and in the definition of definitional equivalence "$\equiv_\Delta$" (p.55) way below, we are "navigating around" two different trends both present in the present work (i.e. we are trying to make the consequences of these two trends "consistent" with each other). These are the following.

Trend 1. When discussing definability over $\mathfrak{M}$ or over K, what we are really interested in is definability over $\mathbf{I}\{\mathfrak{M}\}$ or $\mathbf{I}\mathsf{K}$. More generally in the present work, most of the time, we tend to concentrate our attention to isomorphism-closed classes $\mathsf{K} = \mathbf{I}\mathsf{K}$ of models, moreover we are inclined to identify isomorphic models.

Trend 2. For purely aesthetical reasons, some of our distinguished classes of models are not quite closed under isomorphisms. E.g. in the definition of our class FM of frame models we insisted that the relation $\in$ connecting ${}^nF$ and G should be the real set theoretical membership relation.[28] This aesthetics motivated decision

[25]Seeing that $\mathcal{P}_\omega(U_i)$ leads to problems (i.e. checking Example 2.1.6(6)) is not obvious, it is not necessary to check this for understanding this work.

[26]i.e. to concentrating on maximal definitional expansions

[27]in the sense of $\langle Points, Lines; \in\rangle$, cf. p.**??**

[28]This is so if we understand the definition of FM in accordance with Convention **??** on p.**??**. (Otherwise FM can be understood in such a way that it becomes closed under isomorphisms.)

is the only reason why $\mathsf{FM} \neq \mathbf{I}\mathsf{FM}$. Similarly in our two-sorted geometries of the kind $\langle \mathit{Points}, \mathit{Lines}; \in \rangle$ we insisted that $\mathit{Lines} \subseteq \mathcal{P}(\mathit{Points})$ and "$\in$" is the real set theoretic one. This is the only reason why our two-sorted geometries are not closed under isomorphisms.

If only Trend 1 were present then we could simplify much of the presentation in this sub-section by discussing only isomorphism closed classes $\mathsf{K} = \mathbf{I}\mathsf{K}$, $\mathsf{K}^+ = \mathbf{I}\mathsf{K}^+$ etc. However, we cannot carry through this simplification because Trend 2 presents a "purely administrative" obstacle to it. We call this obstacle purely administrative because the decision behind Trend 2 is purely aesthetical (everything would go through smoothly if we worked with $\mathbf{I}\mathsf{FM}$ in place of FM). As a consequence we do the following: On the intuitive level we tend to follow the simplifications suggested by Trend 1. At the same time, on the formal level we take Trend 2 into account in order to make our results (and definitions) applicable to classes like FM or to two-sorted geometries even when we take the formal details fully into account. Therefore on the formal level, we try to make sure that our definitions make sense (and mean what they should) even when $\mathsf{K} \neq \mathbf{I}\mathsf{K}$. We suggest that the reader keep in mind the "intuitive level" (when we use only Trend 1 and replace FM with $\mathbf{I}\mathsf{FM}$ etc.) and to treat the "formal level" as secondary, because this simplifies the picture *without* loosing any of the essential ideas.

$\lhd$

We close sub-section 2.2 with some examples. More examples can be found in [1], in Chapter 6.

Example 2.2.7 (Explicit definability of the rational numbers in the ring Z of integers.)
Let $\mathbf{Z} = \langle \mathrm{Z}; 0, 1, +, \cdot \ \rangle$ be the (usual) ring of integers. We will discuss how the set Q of rationals is definable explicitly as a *new sort* in $\mathbf{Z}$.(Moreover with a little stretching of our terminology, we can say that the field $\mathbb{Q}$ of rationals is definable in $\mathbf{Z}$.) Here, the new *functions connecting* the new sort Q to the old one Z are (i) the ring-operations $+_{\mathrm{Q}}$ and $\cdot_{\mathrm{Q}}$ on the sort Q, and (ii) an injection $\mathit{repr} : \mathrm{Z} \rightarrowtail \mathrm{Q}$ representing the integers as rationals. The role of *repr* is to tell us which member of sort Z is considered to be equal with which member of the new sort Q. (Although the present "connecting-functions" do not coincide with our standard "explicit definability theoretical" ones pj_i and $\in$, we will see that they are first-order definable from the latter.)

Let us get started! We start out with $\mathbf{Z}$. First we define

$$R = \{ \langle a, b \rangle \ : \ a, b \in \mathrm{Z}, \ b \neq 0 \}$$

as a new relation, obtaining the expansion $\langle \mathbf{Z};\, R\rangle$. Then we define the new sort U to be R with projections pj_0, pj_1 *and* for simplicity we forget R as a relation (but we keep it as a sort named U). This yields the definitional expansion

$$\mathbf{Z}^+ = \langle \mathrm{Z}, U;\, 0, 1, +, \cdot\, , pj_0, pj_1\rangle = \langle \mathbf{Z}, U;\, pj_0, pj_1\rangle$$

where $pj_i : U \longrightarrow \mathrm{Z}$ are the usual. Next, we define the equivalence relation $\equiv$ on U as follows

$$\langle a, b\rangle \equiv \langle c, d\rangle \qquad \stackrel{\text{def}}{\Longleftrightarrow} \qquad a \cdot d = b \cdot c.$$

Note, that it is this point where we need the operations pj_i, namely "$\langle a, b\rangle$" is not an expression of our first-order language, but we can simulate it by using the projections as follows. We define $\equiv$ by

$$x \equiv y \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad pj_0(x) \cdot pj_1(y) = pj_1(x) \cdot pj_0(y),$$

where x, y are of sort Q. By using item (2.2) of our outline for definability, we define the *new sort* Q by $\mathrm{Q} := U/{\equiv}$ together with the usual membership relation $\in$ connecting sort U with sort Q.

Now, using the symbols $\in, pj_0, pj_1$ one can define the operations $+_{\mathrm{Q}}$, $\cdot_{\mathrm{Q}}$, *repr* as follows.

Assume $x \in \mathrm{Z}$ and $y \in \mathrm{Q}$. Then

$$repr(x) = y \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad (\exists z \in y)\,[\, pj_0(z) = x \ \wedge \ pj_1(z) = 1].$$

Assume $x, y, z \in \mathrm{Q}$. Then

$$x \cdot_{\mathrm{Q}} y = z \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad (\exists x' \in x)(\exists y' \in y)(\exists z' \in z)$$
$$[\, pj_0(x') \cdot pj_0(y') = pj_0(z') \ \wedge \ pj_1(x') \cdot pj_1(y') = pj_1(z')].$$

The rest is easy, hence we omit it.

The above shows that the structure

$$\mathbf{Z}^{++} = \langle \mathbf{Z}, \mathrm{Q};\, +_{\mathrm{Q}}, \cdot_{\mathrm{Q}}, repr\rangle$$

is definable over $\mathbf{Z}^+$ hence it is also definable over $\mathbf{Z}$.

In passing, we note that the above definitional expansion makes sense and remains first-order if instead of $\mathbf{Z}$ we start out with an arbitrary ring, say $\mathfrak{A}$.

$\lhd$

Examples 2.2.8

1. Let $\mathbf{F}$ be a field. Consider the geometric expansion

$$\mathbf{G_F} := \langle \mathbf{F}, \mathit{Points}, \mathit{Lines};\ pj_0, pj_1, \mathrm{E} \rangle$$

   of $\mathbf{F}$ where $\mathit{Points} = F \times F$ and $pj_i : F \times F \longrightarrow F$ and $\mathrm{E} \subseteq \mathit{Points} \times \mathit{Lines}$ is the incidence relation (the usual way) and $\mathit{Lines} \subseteq \mathcal{P}(\mathit{Points})$ is the set of lines in the Euclidean sense.

   Then $\mathbf{G_F}$ *is rigidly definable* over $\mathbf{F}$. See the Hint in Example 2 below.

2. To each field $\mathbf{F}$ let $\mathbf{G_F}$ be associated as in item 1 above. Then

$$\mathsf{K}^+ := \{\, \mathbf{G_F} \ :\ \mathbf{F} \text{ is a field} \,\}$$

   is rigidly definable (explicitly) over the class K of fields.[29]

   Hint: First we define $\mathit{Points} = F \times F$ (with pj_i) as a new sort. Then we define

$$R = \{\, \langle p, q \rangle \in \mathit{Points} \times \mathit{Points} \ :\ p \neq q \,\},$$

   as a new relation. Then we define the new auxiliary sort U to be R with the *new* projections $\overline{pj}_i : R \longrightarrow \mathit{Points}$ *and* we forget R as a relation (but we keep it as a sort named U). Then we define the equivalence relation $\equiv$ on U by saying

$$\langle p, q \rangle \equiv \langle r, s \rangle$$
$$\overset{\text{def}}{\Longleftrightarrow}$$
$$(p, q, r, s \text{ are collinear in the Euclidean sense}).$$

   Then we define the new sort $\mathit{Lines} := U/\equiv$ together with $\in\ \subseteq U \times \mathit{Lines}$. From these data we define our final incidence relation $\mathrm{E} := \mathrm{E}_{\mathrm{Points,Lines}}$ the usual way.[30]

$\triangleleft$

---

[29] From now on we will tend to omit "explicitly" since we agreed that definability automatically means explicit definability.

[30] I.e. $p \,\mathrm{E}\, \ell \overset{\text{def}}{\Longleftrightarrow} (\exists x \in \ell)[\,\overline{pj}_0(x), \overline{pj}_1(x), p$ are collinear as computed in $\mathbf{F}\,]$.

In the case of implicit definability we saw that uniform and one-by-one definability are wildly different. The example below is intended to demonstrate, for the case of explicit definability, the same kind of difference between uniform and one-by-one (explicit) definability. In this example we restricted ourselves to the most classical case: one sort only and the defined thing is a relation over the old sort. Besides providing explanation, this example was also designed to provide motivation for consistently sticking with the *uniform* versions of the kinds of definability we consider.

Example 2.2.9 Let $\underline{\omega} = \langle \omega;\, 0, 1, +, \cdot\, \rangle$ be the usual standard model of Arithmetic. Let us choose $R \subseteq \omega$ such that R is not explicitly definable even in higher-order logic over $\underline{\omega}$ (and even with parameters). Such an R exists.[31] Let

$$\mathsf{K} := \{\, \langle \underline{\omega};\, c, P \rangle \;:\; c \in \omega,\; P \subseteq \omega \text{ and } (c \in R \;\Rightarrow\; P = \{c\}) \text{ and } (c \notin R \;\Rightarrow\; P = \emptyset)\,\}\,.$$

Let K^- be the P-free reduct of K i.e.

$$\mathsf{K}^- := \{\, \langle \underline{\omega}, c \rangle \;:\; c \in \omega \,\}\,.$$

Claim: Each member $\mathfrak{M} = \langle \underline{\omega};\, c, P \rangle$ of K is *explicitly* definable over its P-free reduct $\mathfrak{M}^- = \langle \underline{\omega}, c \rangle$. I.e. K is one-by-one explicitly definable over its reduct K^-.

We will see that K is very far from *being uniformly explicitly* definable over K^-. (Moreover K is far from being uniformly finitely implicitly definable.)

For $n \in \omega$, we denote the constant-term $\underbrace{1 + \ldots + 1}_{n\text{-times}}$ by $\bar{n}$. Assume P is uniformly explicitly definable over K^-. Then

$$\mathsf{K} \models [\, P(x) \leftrightarrow \psi(c, x)\,],$$

for some formula $\psi(x, y)$ in the language of $\underline{\omega}$.[32] Now, for any $n \in \omega$ we have the

[31] One can choose R to be so far from being computable that R is not even in the so called Analytical Hierarchy cf. [5].

[32] This is so because $\psi(c, x)$ is in the language of K^-, which is the same as the language of $\underline{\omega}$ expanded with a constant symbol c.

following:

$$\begin{array}{lllllll}
n \in R & \Rightarrow & [\,\mathsf{K} & \models & \bar{n} = c \;\rightarrow\; P(\bar{n}) & & \text{hence} \\
 & & \;\;\mathsf{K} & \models & \bar{n} = c \;\rightarrow\; \psi(c, \bar{n}) & & \text{hence} \\
 & & \;\;\mathsf{K}^- & \models & \bar{n} = c \;\rightarrow\; \psi(c, \bar{n}) & & \text{hence} \\
 & & \;\;\mathsf{K}^- & \models & \bar{n} = c \;\rightarrow\; \psi(\bar{n}, \bar{n}) & & \text{hence}^{33} \\
 & & \;\;\mathsf{K}^- & \models & \psi(\bar{n}, \bar{n}) & & \text{hence} \\
 & & \;\;\underline{\omega} & \models & \psi(\bar{n}, \bar{n})\,]. & &
\end{array}$$

$$\begin{array}{lllllll}
n \notin R & \Rightarrow & [\,\mathsf{K} & \models & \bar{n} = c \;\rightarrow\; \neg P(\bar{n}) & & \text{moreover} \\
 & & \;\;\mathsf{K} & \models & \bar{n} = c \;\rightarrow\; P = \emptyset & & \text{hence} \\
 & & \;\;\mathsf{K}^- & \models & \bar{n} = c \;\rightarrow\; \neg\psi(c, \bar{n}) & & \text{hence} \\
 & & \;\;\mathsf{K}^- & \models & \bar{n} = c \;\rightarrow\; \neg\psi(\bar{n}, \bar{n}) & & \text{hence}^{33} \\
 & & \;\;\underline{\omega} & \models & \neg\psi(\bar{n}, \bar{n})\,]. & &
\end{array}$$

But then $\psi(x, x)$ explicitly defines $R(x)$ in $\underline{\omega}$, which is a contradiction.

We have seen that while in K^- the new relation P is one-by-one explicitly definable (in other words locally explicitly definable), P is very far from being *uniformly* explicitly definable over the same K^-.

$\triangleleft$

We hope that the above construction and proof explain why and how one-by-one definability is so much weaker than[34] uniform definability. We also hope that the above example illustrates why most authors simply identify uniform definability with definability.

2.3 Eliminability of defined concepts.

Notation 2.3.1 For a class K of (many-sorted, similar) models, $Fm(\mathsf{K})$ denotes the set of formulas of the language of K. Hence $\mathsf{Th}(\mathsf{K}) \subseteq Fm(\mathsf{K})$. Sometimes we refer to

[33] by $\mathsf{K} \not\models n \neq c$ (i.e. by $(\exists\mathfrak{M} \in \mathsf{K})\mathfrak{M} \models n = c$) and since under any evaluation of the variables (in a member of K) the value of the constant term $\bar{n}$ coincides with the element n of ω.

[34] One-by-one definability is not only weaker than uniform definability, but also it is much *less satisfactory* from the point of view of re-capturing the intuitive idea of definability. In our opinion one-by-one definability does not capture the intuitive notion of definability while uniform definability does. (All the same, one-by-one definability is useful as a mathematical *auxiliary* concept.)

$Fm(\mathsf{K})$ as the language of K.[35]

$\lhd$

THEOREM 2.3.2 (First translation theorem) *Let* K *and* K^+ *be two classes of (many-sorted) models. Assume that* K^+ *is a generalized definitional expansion of* K*. Then there is a "natural" translation mapping*

$$Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$$

having the following property (called preservation of meaning):[36]

(⋆) *Assume* $\psi(\bar{x}) \in Fm(\mathsf{K}^+)$ *is such that all its free variables (indicated as* $\bar{x}$*) belong to "old"*[37] *sorts, i.e. to sorts of* K*. Then*

$$\mathsf{K}^+ \models [\,\psi(\bar{x}) \leftrightarrow Tr(\psi)(\bar{x})\,].$$

Further, for all $\psi \in Fm(\mathsf{K}^+)$

$$\mathsf{K}^+ \models \psi \;\Leftrightarrow\; \mathsf{K} \models Tr(\psi).$$

Moreover, Tr is very simple (transparent) from the computational point of view, e.g. it is Turing-computable in linear time.

Theorem 2.3.2 follows from the stronger Theorem 2.3.4 (and its proof) to be stated soon, so we do not prove it here.

COROLLARY 2.3.3 Let K and K^+ be classes of one-sorted models such that the name of their sorts agree. Then K is definable in K^+ in the classical sense, i.e. such that we allow only step (1) in the definitions iff K is definable in K^+ in the new many-sorted sense, i.e. such that we allow the use of steps (1) - (2.2). In other words, the possiblility of defining new universes (and then forgetting them) does not create new definitional expansions among one-sorted models. $\lhd$

[35] According to our philosophy, $Fm(\mathsf{K})$ is the language, while the system of basic symbols (like relation symbols, sort symbols etc.) is the *vocabulary* of this language, cf. Convention 2.0.1 on p.5. We note this because some logic books use the word "language" for what we call the vocabulary (of a language or a model).

[36] The existence of such a translation mapping *Tr* is often called in the literature "*uniform reduction property*", cf. Hodges [11, p.640]. A result of Pillay and Shelah is that for first order axiomatizable classes implicit definability without taking reducts implies the reduction property, cf. [15]. Cf. also Lemma 12.5.1 in Hodges [11, p.641].

[37] A symbol (e.g. a sort) is called old if it is available already in K (and not only in K^+).

Before stating the stronger version of Theorem 2.3.2, let us ask ourselves in what sense Tr in Thm.2.3.2 preserves the meanings of formulas. To answer this question, let us notice that the conclusion of Theorem 2.3.2 implies (i) and (ii) below.

(i) ψ and $Tr(\psi)$ have the same free variables $\bar{x}$, and in some intuitive sense they say the same thing about these variables $\bar{x}$.

(ii) Let $\mathfrak{M} \in \mathsf{K}^+$, $\mathfrak{M}^-$ be the reduct of $\mathfrak{M}$ in K and let $\bar{a}$ be a sequence of members of $Uv(\mathfrak{M}^-)$ matching the sorts of $\bar{x}$. In other words $\bar{a}$ is an evaluation of the variables $\bar{x}$. Then

$$\mathfrak{M} \models \psi[\bar{a}] \quad \Longleftrightarrow \quad \mathfrak{M}^- \models (Tr(\psi))[\bar{a}]\,;$$

cf. the notation on p.20. Intuitively, *whatever can be said about* some "old" elements $\bar{a}$ in a model $\mathfrak{M}$ in K^+, *it can be said* (about the same elements $\bar{a}$) *already* in the "old" model $\mathfrak{M}^-$ (in K). This will be generalized to "new" elements also (i.e. to arbitrary elements), in our next theorem.

Recall that K is a reduct of K^+. In some sense (i) and (ii) above mean that the poorer class K and the richer class K^+ of models are equivalent from the point of view of expressive power of language. So, the "language + theory" of K^+ is equivalent with the "language + theory" of K in means of expression. Therefore, on some level of abstraction, we may consider the languages of K and K^+ to be the *same* except that they[38] choose different "basic vocabularies" for representing this language. (In passing we note that a stronger form of this kind of *sameness* will appear in the form of definitional equivalence $\equiv_\Delta$, cf. beginning with p.55 (and the figure on p.61).)

Generalization of Theorem 2.3.2 to permitting free variables of new sorts to occur in ψ and $Tr(\psi)$

Let us turn to discussing the restriction in Theorem 2.3.2 which says (in statement ($\star$)) that the free variables of ψ belong to the sorts of K. The theorem does admit a generalization which is without this restriction on the free variables. This will be stated in Theorem 2.3.4 below. But then two things happen discussed in items (I), (II) below.

(I) Consider the process of defining K^+ over K as a sequence of steps (as described on p.26). Assume that a relation like pj_i or $\in_U$ connecting a new sort to an old one is introduced in one step and then is *forgotten* at the last reduct step.

[38] i.e. K and K^+

Then we call the relation (e.g. pj_i) in question an _auxiliary relation_ of the definition of K^+ over K. Now, for the generalization of Theorem 2.3.2 we have in mind, we have to assume that all auxiliary relations (of the definition of K^+) remain definable in K^+. We will formulate this condition as "K^+ and K have a common (explicit) definitional expansion (without taking reducts)".

That K^+ and K have a common definitional expansion expresses that K^+ *is definable over* K *with recoverable auxiliaries* because of the following. Assume that K^{++} is a common definitional expansion of K^+ and of K. Then K^+ is a reduct of K^{++} which is a definitional expansion of K, hence K^+ is definable in K. Also, all the relations and sorts that get forgotten in the reduct-forming from K^{++} to K^+ are definable in K^+ since K^{++} is a definitional expansion of K^+.

(II) The formulation of the theorem gets somewhat complicated. Intuitively, the generalized theorem says that all new objects[39] can be represented as equivalence classes of tuples of old objects, and then (using this representation) whatever can be said about elements of $Uv(\mathfrak{M})$ in an expanded model $\mathfrak{M} \in \mathsf{K}^+$ can be already said in the reduct $\mathfrak{M}^- \in \mathsf{K}$ of $\mathfrak{M}$. This intuitive statement is intended to serve as a generalization the text below item (ii) in the discussion of the intuitive meaning of Theorem 2.3.2 (presented immediately below Theorem 2.3.2). Cf. Figure 2.

Notation: $Var(U_i)$ denotes the (infinite) set of variables of sort U_i (where U_i is treated as a sort symbol or *equivalently* U_i is the name of one of the universes of the models in K^+).

THEOREM 2.3.4 (Second translation theorem) *Assume* K *is a reduct of* K^+ *and* K *and* K^+ *have a common definitional expansion (without taking reducts). This holds e.g. whenever* K^+ *is a definitional expansion of* K. *Assume further that in* K*, every sort is nonempty and there is at least one sort which has more than one elements. Assume* $U_1^{new}, \ldots, U_k^{new}$ *are the new sorts.*[40] *Then there is a translation mapping*

$$Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$$

for which the following hold. For each U_i^{new} *there is a formula* $code_i(x, \vec{x}) \in Fm(\mathsf{K}^+)$ *such that the following 1-2 hold.*

1. $x \in \mathrm{Var}(U_i^{new})$ *and* $\vec{x}$ *is a sequence of variables of old sorts.*

[39]By objects we mean elements of some sort.
[40]i.e. they are available in K^+ but not in K.

2. *(a)–(c) below hold.*

   (a) $\mathsf{K}^+ \models \forall x \, \exists \vec{x} \; code_i(x, \vec{x})$, [41]

   (b) $\mathsf{K}^+ \models [\, code_i(x, \vec{x}) \ \wedge \ code_i(y, \vec{x}) \,] \ \rightarrow \ x = y$, *where* $y \in \mathrm{Var}(U_i^{new})$. [42]

   (c) *Our translation mapping*[43]

$$Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$$

   satisfies the following stronger[44] *property of meaning preservation. Assume* $\psi(y, \bar{z}) \in Fm(\mathsf{K}^+)$ *is such that* $y \in \mathrm{Var}(U_i^{new})$ *and* $\bar{z}$ *is (a sequence of variables) of old sorts such that the variables in* $\bar{z}$ *are* distinct *from those occurring in* $\vec{y}$*. Then*

$$\mathsf{K}^+ \ \models \ code_i(y, \vec{y}) \ \rightarrow \ [\, \psi(y, \bar{z}) \leftrightarrow (Tr(\psi))(\vec{y}, \bar{z}) \,].$$

   Intuitively, whatever is said by ψ *about* y *and* $\bar{z}$*, the same is said by the translated formula* $\mathrm{Tr}(\psi)$ *about the code* $\vec{y}$ *of* y *and* $\bar{z}$*. Cf. Figure 2. The case when* ψ *contains an arbitrary sequence, say* $\bar{y}$*, of variables of various new sorts is a straightforward generalization and is left to the reader.*

We note that the intuitive meaning of "$code_i(x, \bar{y})$" is "$\bar{y}$ codes x". Property (b) then says that "$\bar{y}$ codes only one element", property (a) says that "every new element has a code", and property (c) then tells us that "whatever can be said of a new element x in the new language, can be said of any of its codes $\bar{y}$ in the old language", cf. (II) before the statement of Theorem 2.3.4 and Figure 2.

Proof:
(I) The case of step (2.1): Assume that K^+ is obtained from K by applying step (2.1) so that we defined $U^{new} \stackrel{\text{def}}{=} R$ where R is an old r-ary relation. For simplicity we assume $r = 2$ and $R \subseteq U_0 \times U_1$ where U_0, U_1 are old sorts. Then the new symbols (in K^+) are U^{new} and pj_0, pj_1. We want to represent objects (variables) of sort U^{new} with pairs of objects of ("old") sorts. To this end, we fix an injective function

$$rep : Var(U^{new}) \rightarrowtail Var(U_0) \times Var(U_1)$$

[41] Note that here "$\forall x$" means "$\forall x \in U_i^{new}$" automatically since we know that x is of sort U_i^{new} (as a variable symbol of the language of K^+).

[42] Note that items (a), (b) mean that $code_i$ represents an unambiguous coding of elements of U_i^{new} with equivalence classes of tuples of elements of old sorts, cf. (II) preceding the statement of the theorem and the text immediately below the theorem.

[43] fixed at the beginning of the formulation of the present theorem

[44] stronger than in Theorem 2.3.2

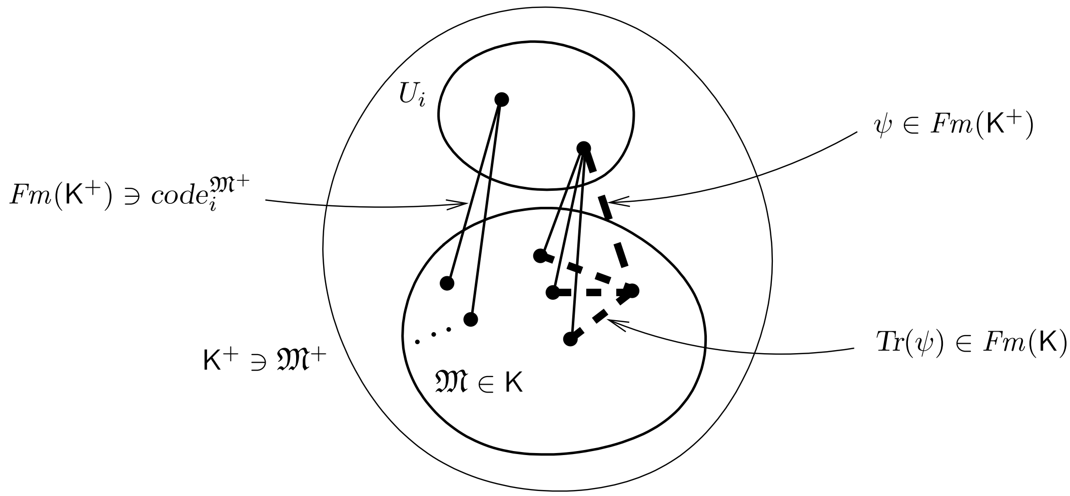

Figure 1: Illustration for the second translation theorem (Thm.2.3.4). Whatever can be said of a new element in $\mathfrak{M}^+$ can be said of its "code" in the old model $\mathfrak{M}$. (In the Figure, the codes of the new elements have lenght 1.)

such that the values $rep(x)_i$ of *rep* are all distinct.[45] For simplicity, we will denote $rep(x)_i$ by x_i.

Now, we define *Tr* by recursion as follows.

- $Tr((\exists x \in U^{new})\psi) := (\exists x_0 \in U_0, x_1 \in U_1)Tr(\psi)$; if $x \in Var(U^{new})$;
- $Tr((\exists y)\psi) := (\exists y)Tr(\psi)$; if y is a variable of old sort;
- $Tr(\neg\psi) := \neg Tr(\psi)$, $\qquad Tr(\psi \wedge \varphi) := Tr(\psi) \wedge Tr(\varphi)$;
- $Tr(x = y) := (x_0 = y_0 \ \wedge \ x_1 = y_1)$, for any $x, y \in Var(U^{new})$;
- for any other *atomic* formula ψ, $Tr(\psi)$ is obtained from ψ by replacing each occurrence of $pj_i(x)$ with x_i (i.e. with $rep(x)_i$) in ψ for *every* variable $x \in Var(U^{new})$ and $i \in 2$; i.e. $Tr(\psi) := \psi(pj_i(x)/x_i)_{x \in \mathrm{Var}(U^{new}), i<2}$.

We introduce the formula $code(x, x_0, x_1)$ (saying explicitly that the values of x_0, x_1 form really the code of the value of x) as follows:

$$code(x, x_0, x_1) \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad [\, x_0 = pj_0(x) \ \wedge \ x_1 = pj_1(x)\,].$$

[45] $rep(x) = \langle rep(x)_0, rep(x)_1 \rangle$; and $rep(x)_i = rep(y)_j$ iff $\langle x, i \rangle = \langle y, j \rangle$.

Now, it is not difficult to check that $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ is well defined, and (a)-(c) in the statement of Theorem 2.3.4 hold.

(II) The case of step (2.2): Assume that K^+ is obtained from K by applying step (2.2) so that the only new symbols (in K^+) are $U^{new} = U/R$ and $\in$, where U is an (old) sort of K, and $R(x, y) \in Fm(\mathsf{K})$ where x, y are variables of sort U.

We fix an injective function

$$rep : Var(U^{new}) \rightarrowtail Var(U)$$

and we denote $rep(x)$ by $\underline{x}$. So $\underline{x} \in Var(U)$ if $x \in Var(U^{new})$.

Now, we define Tr by recursion as follows.

- $Tr((\exists x \in U^{new})\psi) := (\exists \underline{x} \in U_0)\, Tr(\psi)$; if $x \in Var(U^{new})$;
- $Tr((\exists y)\psi) := (\exists y)\, Tr(\psi)$; if y is a variable of old sort;
- $Tr(\neg\psi) := \neg Tr(\psi), \qquad Tr(\psi \wedge \varphi) := Tr(\psi) \wedge Tr(\varphi)$;
- $Tr(x = y) := (\forall z \in U)[\in(z, \underline{x}) \leftrightarrow \in(z, \underline{y})]$, where $x, y \in Var(U^{new})$, and $z \in Var(U)$ is arbitrary;
- $Tr(\psi) \stackrel{\text{def}}{=} \psi$; for any other *atomic* formula ψ.

We introduce the formula $code(x, \underline{x})$ as follows:

$$code(x, \underline{x}) \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad \in(\underline{x}, x)\,.$$

Now, it is not difficult to check that $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ is well defined, and (a)-(c) in the statement of Theorem 2.3.4 hold.

(III) The case of explicit definability without taking reducts: If K^+ is obtained from K by step (1) then we have an obvious translation with all the good properties known from classical definability theory.[46]

By this we have covered all the steps (i.e. (1)–(2.2)) which might occur in an explicit definition. I.e. we defined *rep*, *code*, *Tr* to all three kinds of "one-step" explicit definitions represented by items (1)–(2.2).

Assume now that K^+ is explicitly defined over K without taking reducts. Now, the definition of K^+ is a finite sequence of steps with each step using one of items

[46] In the case of step (1), "*code*" is not needed because there are no new sorts involved. Hence (*if* we want to preserve uniformity of the steps) we can choose $code(x, y)$ to be $x = y$.

(1), (2.1), (2.2). Hence by the above, we have a meaning preserving translation mapping Tr_k for the k'th step for each number

$$k < n := \text{“number of steps in the definition of } \mathsf{K}^+\text{”}.$$

Besides Tr_k we also have a formula $code_k$ for each number k. Also for each Tr_k we have that (a)-(c) in the statement of Theorem 2.3.4 hold. But then we can take the composition $Tr := Tr_1 \circ Tr_2 \circ \ldots \circ Tr_n$ of these meaning preserving functions, and then the composition too will be meaning preserving if we also combine the formulas $code_1, \ldots, code_n$ into a single "big" formula $code$.

One can check that for the just defined Tr and $code$, (a)-(c) in the statement of Theorem 2.3.4 hold.

(IV) The general case: Assume now that K is a reduct of K^+ and that K^{++} is a common definitional expansion of K and K^+. By the previous case we have translation mappings $Tr_1 : Fm(\mathsf{K}^{++}) \longrightarrow Fm(\mathsf{K})$ and $Tr_2 : Fm(\mathsf{K}^{++}) \longrightarrow Fm(\mathsf{K}^+)$ together with appropriate $code_1, code_2$ which satisfy (a)-(c) in the statement of Theorem 2.3.4. Note that $Fm(\mathsf{K}^+) \subseteq Fm(\mathsf{K}^{++})$. Now we define

$$Tr \stackrel{\text{def}}{=} Tr_1 \upharpoonright Fm(\mathsf{K}^+),\ \ code(x, \vec{x}) \stackrel{\text{def}}{=} Tr_2(code_1(x, \vec{x}))$$

whenever x is a variable of new sort in the language of K^+. One can check that Tr and $code$ as defined above satisfy (a)-(c). In more detail: Assume that U_i is a new sort of K^+, i.e. U_i is not a sort of K. Then U_i is a new sort of K^{++}, therefore there is $code_i^1(x, \bar{x}) \in Fm(\mathsf{K}^{++})$ which "matches" Tr_1. We cannot use $code_i^1$ in the interpretation from K^+ to K because $code_i^1$ may not be in the language of K^+. We will use Tr_2 to translate $code_i^1$ to the language of K^+ as follows. Since K^+ is an expansion of K, all the variables in $x, \bar{x}$ have sorts which occur in K^+. Thus by the properties of Tr_2 we have

$$\mathsf{K}^{++} \models code_i^1(x, \bar{x}) \leftrightarrow Tr_2(code_i^1(x, \bar{x})).$$

Let $code_i(x, \bar{x}) \stackrel{\text{def}}{=} Tr_2(code_i^1(x, \bar{x}))$. Then $code_i(x, \bar{x}) \in Fm(\mathsf{K}^+)$ and

$$\mathsf{K}^{+} \models code_i(x, \bar{x}) \rightarrow [\psi(x, \bar{x}) \leftrightarrow Tr(\psi)(\bar{x}, \bar{z})]$$

because

$$\mathsf{K}^{++} \models code_i^1(x, \bar{x}) \rightarrow [\psi(x, \bar{z}) \leftrightarrow Tr_1(\psi)(\bar{x}, \bar{z})].$$

This finishes the proof. ∎

In our next theorem we do not have to assume that the sorts are nonempty and that there is a sort with more than one elements. The price is that the formulation becomes a little more complicated in that instead of one coding formula for each sort we will have possibly more coding formulas for each sort.

THEOREM 2.3.5 (Third translation theorem) *Assume* K *is a reduct of* K^+ *and* K *and* K^+ *have a common definitional expansion (without taking reducts). Then there is a translation mapping*

$$Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$$

for which the following hold. For each new sort U_i^{new} *there is a number* n_i *and there are formulas* $\text{code}_1^i(x, \vec{x}_1), \ldots, \text{code}_{n_i}^i(x, \vec{x}_{n_i}) \in Fm(\mathsf{K}^+)$ *such that the following hold for all* $1 \leq j \leq n_i$.

1. $x \in \text{Var}(U_i^{\text{new}})$ *and* $\vec{x}_j$ *is a sequence of variables of old sorts.*
2. *(a)–(c) below hold.*
   (a) $\mathsf{K}^+ \models \forall x \bigvee\{\, \exists \vec{x}_j \ \text{code}_j^i(x, \vec{x}_j) : 1 \leq j \leq n_i\}$, [47]
   (b) $\mathsf{K}^+ \models [\,\text{code}_j^i(x, \vec{x}_j) \ \wedge \ \text{code}_j^i(y, \vec{x}_j)\,] \ \rightarrow \ x = y$, *where* $y \in \text{Var}(U_i^{new})$.
   (c) *Our translation mapping*

   $$Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$$

   satisfies the following property of meaning preservation. Assume $\psi(y, \bar{z}) \in Fm(\mathsf{K}^+)$ *is such that* $y \in \text{Var}(U_i^{new})$ *and* $\bar{z}$ *is a sequence of variables of old sorts such that the variables in* $\bar{z}$ *are* distinct *from those occurring in the sequence* $\vec{y}$ *of variables. Then*

   $$\mathsf{K}^+ \ \models \ code_j^i(y, \vec{y}) \ \rightarrow \ [\,\psi(y, \bar{z}) \leftrightarrow (Tr(\exists \vec{y} code_j^i(y, \vec{y}) \ \wedge \ \psi))(\vec{y}, \bar{z})\,].$$

   Intuitively, whatever is said by ψ *about* y *and* $\bar{z}$*, the same is said by the translated formula* $\text{Tr}(\psi \ \wedge \ \exists \vec{y} \text{code}_j^i(y, \vec{y}))$ *about the* j*-code* $\vec{y}$ *of* y *and* $\bar{z}$*. Cf. Figure 2. The case when* ψ *contains an arbitrary sequence, say* $\bar{y}$*, of variables of various new sorts is a straightforward generalization and is left to the reader.*

When each n_i equals 1, then we get back the Second Translation Theorem.

Proof:
(I) The case of step (2.1) (direct product): Assume that K^+ is obtained from K by applying step (2.1) so that we defined $U^{new} \stackrel{\text{def}}{=} U \times V$ where U, V are old sorts. Then the new symbols (in K^+) are U^{new} (i.e. variables of sort U^{new}), and

[47] Note that here "$\forall x$" means "$\forall x \in U_i^{new}$" automatically since we know that x is of sort U_i^{new} (as a variable symbol of the language of K^+).

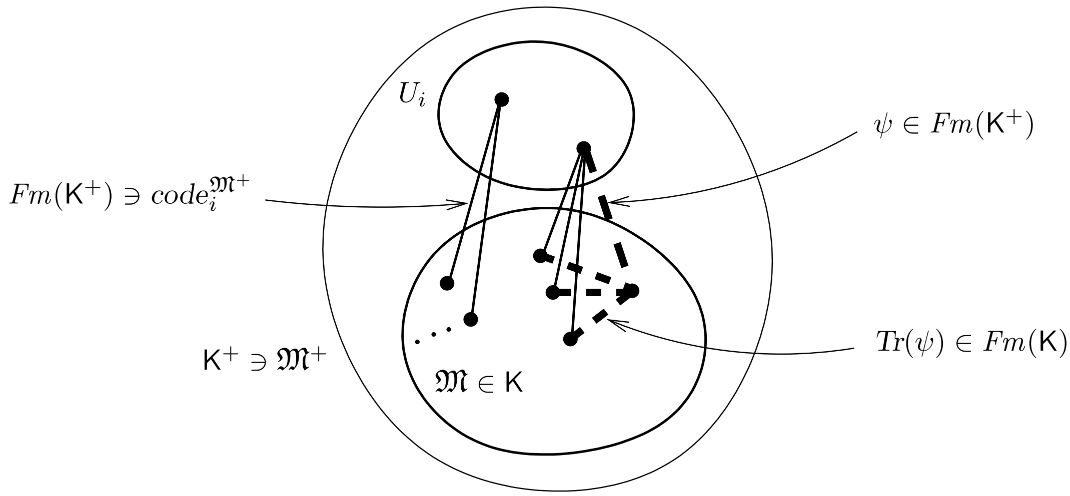

Figure 2: Illustration for the second translation theorem (Thm.2.3.4). Whatever can be said of a new element in $\mathfrak{M}^+$ can be said of its "code" in the old model $\mathfrak{M}$. (In the Figure, the codes of the new elements have lenght 1.)

pj_0, pj_1. The translation mapping $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ will be as follows. Let $\psi \in Fm(\mathsf{K}^+)$. For each $x \in Var(U^{new})$ occurring in ψ we assign variables $x_0 \in Var(U), x_1 \in Var(V)$ such that x_0, x_1 do not occur in ψ and the x_i's are all distinct from each other (i.e. x_i is the same variable as y_j iff x is y and i is j). Now we obtain $Tr(\psi)$ from ψ as follows. For all variable x of new sort occurring in ψ

- we replace $(\exists x)$ in ψ with $(\exists x_0)(\exists x_1)$
- we replace $x = y$ in ψ with $x_0 = y_0 \ \wedge\ x_1 = y_1$
- we replace $pj_i(x) = u$ in ψ with $x_i = u$, for $i = 0, 1$.

Formally, we define Tr by recursion as follows.

- $Tr((\exists x \in U^{new})\psi) := (\exists x_0 \in U)(\exists x_1 \in V)Tr(\psi)$; if $x \in Var(U^{new})$;
- $Tr((\exists y)\psi) := (\exists y)Tr(\psi)$; if y is a variable of old sort;
- $Tr(\neg\psi) := \neg Tr(\psi)$, $\qquad Tr(\psi \wedge \varphi) := Tr(\psi) \wedge Tr(\varphi)$;
- $Tr(x = y) := (x_0 = y_0 \ \wedge\ x_1 = y_1)$, where $x, y \in Var(U^{new})$;

- $Tr(pj_i(x) = u) := (x_i = u)$, where $x \in Var(U^{new})$;
- $Tr(\psi) \stackrel{\text{def}}{=} \psi$; for any other *atomic* formula ψ.

Then $Tr(\psi) \in Fm(\mathsf{K})$ because the symbols in ψ which do not occur in $Fm(\mathsf{K})$ are only those that we replaced in the above algorithm.

We introduce the formula $code(x, x_0, x_1)$ (saying explicitly that the values of x_0, x_1 form really the code of the value of x) as follows:

$$code(x, x_0, x_1) \quad \stackrel{\text{def}}{\iff} \quad [\, x_0 = pj_0(x) \ \wedge \ x_1 = pj_1(x)\,].$$

Now, it is not difficult to check that $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ is well defined, and (a)-(c) in the statement of Theorem 2.3.5 hold. The next two cases are very similar:

(II) The case of step (2.2) (subset): Assume that K^+ is obtained from K by applying step (2.2) so that the only new symbols (in K^+) are $U^{new} = S$ and in, where U is an (old) sort of K and S is an old unary relation symbol of sort U.

The translation mapping $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ is as follows. Let $\psi \in Fm(\mathsf{K}^+)$. For each $x \in Var(U^{new})$ occurring in ψ we assign a variable $\check{x} \in Var(U)$ such that $\check{x}$ does not occur in ψ and the $\check{x}$'s are all distinct from each other. We obtain $Tr(\psi)$ from ψ as follows. For all variable x of sort U^{new} occurring in ψ

- we replace $(\exists x)$ in ψ with $(\exists \check{x} \in S)$, i.e. we replace each subformula $(\exists x)\varphi$ in ψ with $(\exists \check{x})(S(\check{x}) \ \wedge \ \varphi)$
- we replace $x = y$ in ψ with $\check{x} = \check{y}$
- we replace $in(x) = u$ in ψ with $\check{x} = u$.

We define the formula $code(x, \check{x})$ as

$$code(x, \check{x}) \quad \stackrel{\text{def}}{\iff} \quad in(\check{x}) = x\,.$$

(III) The case of step (2.3) (quotient): Assume that K^+ is obtained from K by applying step (2.3) so that the only new symbols (in K^+) are $U^{new} = U/E$ and $\in$, where U is an (old) sort of K, E is an old binary relation symbol of sort U, U such that E is an equivalence relation in each model of K.

The translation mapping $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ is as follows. Let $\psi \in Fm(\mathsf{K}^+)$. For each $x \in Var(U^{new})$ occurring in ψ we assign a variable $\check{x} \in Var(U)$ such that $\check{x}$ does not occur in ψ and the $\check{x}$'s are all distinct from each other. We obtain $Tr(\psi)$ from ψ as follows. For all variable x of sort U^{new} occurring in ψ

- we replace $(\exists x)$ in ψ with $(\exists \check{x})$
- we replace $x = y$ in ψ with $E(\check{x}, \check{y})$
- we replace $\in(u, x)$ in ψ with $E(\check{x}, u)$.

We define the formula $code(x, \check{x})$ as

$$code(x, \check{x}) \quad \overset{\text{def}}{\iff} \quad \in(\check{x}, x)\,.$$

(IV) The case of step (2.4) (union): Assume that K^+ is obtained from K by applying step (2.4) so that the only new symbols (in K^+) are $U^{new} = U \dot{\cup} V$ and in_1, in_2, where U, V are (old) sorts of K.

The intuitive idea of the translation mapping is as follows. Assume $\psi \in Fm(\mathsf{K}^+)$ is given. First we assign variables x_1, x_2 of old sorts U and V respectively to variables x of new sort occurring in ψ as before. Then we rewrite ψ into a form such that each subformula contains at the beginning the information about all its free variables of new sort whether they "came from U or from V". Here, e.g. $(\exists x_1) in_1(x_1) = x$ expresses that "x came from U". Notice that we only use the translation of such formulas in the statement of Thm.2.3.5, so we have such an "assignment" for the free variables of ψ. We will call this information the "prefix" of the formula. E.g. we rewrite a subformula $(\exists y)\varphi$ of ψ where y is a variable of new sort into

$$(\exists y)[(\exists y_1) in_1(y_1, y)) \ \wedge \ \varphi] \vee (\exists y)[(\exists y_2) in_2(y_2, y)) \ \wedge \ \varphi]\,.$$

After this we proceed basically as in the previous cases, except that we replace $(\exists y)$ with $(\exists y_1)$ or with $\exists y_2$ according to the "prefix" of the formula telling us from where y came from; and similarly in the other cases: we replace $x = y$ with $x_i = y_j$ where we choose i, j according to the prefix, and we replace $in_i(z) = y$ with $z = y_i$ or with $FALSE$ according to whether the prefix tells that y came from the right place.

Below, we write out this case formally. Let X be a finite set of variables. First we define Tr^X which translates only those formulas whose all variables, free or bound, are in X. For each $y \in Var(U^{new}) \cap X$ choose $y_1 \in Var(Y) \setminus X$, $y_2 \in Var(V) \setminus X$ such that the y_i's are all distinct.

Next, for all finite $Y \subseteq Var(U^{new}) \cap X$ and $\chi : Y \longrightarrow \{1, 2\}$ we define translation functions Tr_χ by parallel recursion. But first we need to fix some notation.
If χ is as above, then we call it a "prefix".
If $y \in Var(U^{new}), i \in \{1, 2\}$ and χ is a prefix, then

$$\chi(y/i) \overset{\text{def}}{=} (\chi \setminus \{y\} \times \{1, 2\}) \cup \{\langle y, i\rangle\}\,.$$

I.e. $\chi(y/i)$ is an extension of χ which assignes i to y if χ did not assign a value to y, and otherwise $\chi(y/i)$ is the prefix we obtain from χ by changing y's value to i. For a prefix χ the formula $\gamma(\chi)$ is defined as

$$\gamma(\chi) \stackrel{\text{def}}{=} \bigwedge\{(\exists y_i) in_i(y_i) = y : \langle y, i\rangle \in \chi\}\,.$$

We will write $Tr(\chi, \psi)$ for $Tr_\chi(\psi)$. We are ready to define the Tr_χ's by recursion:

- $Tr(\chi, \exists y\psi) \stackrel{\text{def}}{=} \exists y_1 Tr(\chi(y/1), \psi) \vee \exists y_2 Tr(\chi(y/2), \psi)$, if $y \in Var(U^{new}) \cap X$.
- $Tr(\chi, \exists z\psi) \stackrel{\text{def}}{=} \exists z Tr(\chi, \psi)$, if $z \in X \setminus Var(U^{new})$,
- $Tr(\chi, \psi \wedge \varphi) \stackrel{\text{def}}{=} Tr(\chi, \psi) \wedge Tr(\chi, \varphi)$, $\quad Tr(\chi, \neg\psi) \stackrel{\text{def}}{=} \neg Tr(\chi, \psi)$,
- $Tr(\chi, y = x) \stackrel{\text{def}}{=} (y_i = x_j)$ if $\chi(y) = i, \chi(x) = j$ and $FALSE$ otherwise,
- $Tr(\chi, in_i(z) = y) \stackrel{\text{def}}{=} z = y_i$ if $\chi(y) = i$ and $FALSE$ otherwise.

By these, we have defined Tr_χ for all prefixes χ. Now we define $Tr^X(\varphi) \stackrel{\text{def}}{=} Tr(\chi, \psi)$ if φ is $\gamma(\chi) \wedge x = x \wedge \psi$ for some χ, x, ψ and $FALSE$ otherwise. (We included the subformula $x = x$ because ...) Finally,

$$Tr(\gamma(\chi) \wedge \psi) \stackrel{\text{def}}{=} Tr^X(\gamma(\chi) \wedge x = x \wedge \psi),$$

where X consists of all the variables occurring in ψ and $x \in X$. By these, we have defined our translation function $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$. We will have two coding formulas

$$code_1(x, u) \stackrel{\text{def}}{=} in_1(u) = x\ \ ,$$

$$code_2(x, v) \stackrel{\text{def}}{=} in_2(v) = x\ \ .$$

(V) The case of explicit definability without taking reducts: If K^+ is obtained from K by step (1) then we have an obvious translation with all the good properties known from classical definability theory and in this case we do not have to define coding formulas because we have no variables of new sort.

By this we have covered all the steps (i.e. (1)–(2.4)) which might occur in an explicit definition. I.e. we defined $Tr, code_1, \ldots, code_n$ to all five kinds of "one-step" explicit definitions represented by items (1)–(2.4).

Assume now that K^+ is explicitly defined over K without taking reducts. Now, the definition of K^+ is a finite sequence of steps with each step using one of items (1)-(2.4). Hence by the above, we have a meaning preserving translation mapping Tr_k for the k'th step for each number

$$k < n := \text{“number of steps in the definition of } \mathsf{K}^+\text{”}.$$

Besides Tr_k we also have formulas $code_k^1 \ldots, code_k^j$ (where j is $0, 1$ or 2) for each number k. Also for each Tr_k we have that (a)-(c) in the statement of Theorem 2.3.5 hold. But then we can take the composition $Tr := Tr_1 \circ Tr_2 \circ \ldots \circ Tr_n$ of these meaning preserving functions, and then the composition too will be meaning preserving if we also combine the formulas $code_1^i, \ldots, code_n^j$ into a single “big” formula $code$.

E.g., if $U_0^{new} = U \times V$ and $U_1^{new} = U_0^{new} \mathbin{\dot\cup} U_0^{new}$, then $code_0$ will be as in step (2.1), and we will have two coding formulas for $k = 1$, one of them being

$$code_1^1(x, z_0, z_1) := (\exists z \in U_0^{new})(in_1(z) = x \ \wedge \ pj_0(z) = z_0 \ \wedge \ pj_1(z) = z_1)\,.$$

One can check that for the just defined Tr and $code$, (a)-(c) in the statement of Theorem 2.3.5 hold.

(VI) The general case: Assume now that K is a reduct of K^+ and that K^{++} is a common definitional expansion of K and K^+. By the previous case we have translation mappings $Tr_1 : Fm(\mathsf{K}^{++}) \longrightarrow Fm(\mathsf{K})$ and $Tr_2 : Fm(\mathsf{K}^{++}) \longrightarrow Fm(\mathsf{K}^+)$ together with appropriate coding formulas $code_1^1, \ldots, code_2^n$ which satisfy (a)-(c) in the statement of Theorem 2.3.4. Note that $Fm(\mathsf{K}^+) \subseteq Fm(\mathsf{K}^{++})$. Now we define

$$Tr \stackrel{\text{def}}{=} Tr_1 \restriction Fm(\mathsf{K}^+),\ code^j(x, \vec{x}) \stackrel{\text{def}}{=} Tr_2(code_1^j(x, \vec{x}))$$

whenever x is a variable of new sort in the language of K^+ and $code_1^j$ is a coding formula for that new sort. One can check that Tr and $code^i$ as defined above satisfy (a)-(c). In more detail: Assume that U_i is a new sort of K^+, i.e. U_i is not a sort of K. Then U_i is a new sort of K^{++}, therefore there are $code_i^j(x, \bar{x}) \in Fm(\mathsf{K}^{++})$ which “match” Tr_1. We cannot use $code_i^j$ in the interpretation from K^+ to K because $code_i^j$ may not be in the language of K^+. We use Tr_2 to translate $code_i^j$ to the language of K^+ as follows. Since K^+ is an expansion of K, all the variables in $x, \bar{x}$ have sorts which occur in K^+. Thus by the properties of Tr_2 we have

$$\mathsf{K}^{++} \models code_i^j(x, \bar{x}) \leftrightarrow Tr_2(code_i^j(x, \bar{x})).$$

Let $code_i^j(x, \bar{x}) \stackrel{\text{def}}{=} Tr_2(code_i^j(x, \bar{x}))$. Then $code_i^j(x, \bar{x}) \in Fm(\mathsf{K}^+)$ and

$$\mathsf{K}^+ \models code_i^j(x, \bar{x}) \to [\psi(x, \bar{x}) \leftrightarrow Tr(\psi)(\bar{x}, \bar{z})]$$

because
$$\mathsf{K}^{++} \models code_i^j(x,\bar{x}) \to [\psi(x,\bar{z}) \leftrightarrow Tr_1(\psi)(\bar{x},\bar{z})].$$
This finishes the proof. ∎

More is true than stated in Theorem 2.3.4, namely the existence of a translation mapping as in the theorem is actually sufficient for definability, as Theorem 2.3.7 below states.

Remark 2.3.6 (In connection with Theorems 2.3.2, 2.3.4.) These theorems state that the expressive powers of two languages $Fm(\mathsf{K}^+)$ and $Fm(\mathsf{K})$ coincide. However, the proofs of these theorems prove more. Namely there exists a *computable* translation mapping *Tr* acting between the two languages. Even more than this, *Tr* preserves the logical structure of the formulas i.e. in the sense of algebraic logic, *Tr* is a "linguistic homomorphism". (Whether one is interested in this extra property of being a "linguistic homomorphism" is related to a difference between the algebraic logic approach and the abstract model theoretic approach to defining the equivalence of logics [hence, in particular, to how one approaches characterizations of logics like the celebrated Lindström theorems].)

$\triangleleft$

The following theorem says that *eliminability of new symbols is an essential feature of explicit definability*: If the new relations and sorts are arbitrary but are eliminable in the sense that there exist a mapping $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ together with "coding" formulas $code_i(x,\vec{x})$ for all new sorts U_i of K^+ which satisfy 1,2 in Theorem 2.3.4, then we can explicitly construct these new relations and sorts by using our concrete steps (1) - (2.2) (in such a way that some additional auxiliary new sorts and relations get defined in the way, but then we can forget these).

We note that both (ii) and (iii) in Theorem 2.3.7 say that K^+ is a special reduct of some definitional expansion of K. In (ii) we allow to forget relations and sorts which then can be "defined back" (i.e. K^+ is a reduct of its definitional expansion, so we forget the relations and sorts of a definitional expansion). In (iii) we allow to forget only as many relations and sorts that the remaining ones still "fix" the new sorts and relations.

If *Tr* and $code_i$ satisfy the conclusion of Theorem 2.3.4, then we say that they *interpret* K^+ in K.[48]

[48] Cf. the definition of interpretations in Hodges [11, p.212, 221]. The existence of a tuple $Tr, code_i$ interpreting K^+ in K (as in Theorem 2.3.4) is strictly stronger than the *uniform reduction property* in [11, p.640]. Actually, the existence of $Tr, code_i$ is equivalent with K^+ being coordinatised over K in the sense of [11, p.644]. This equivalence is proved in [2].

THEOREM 2.3.7 *Assume* K *is a reduct of* K^+*. Then (i) and (ii) below are equivalent and they imply (iii). If, in addition,* K^+ *is closed under taking ultraproducts, then (i)-(iii) below are equivalent.*

(i) K^+ *is interpreted in* K *by some Tr and* $code_i$*, i.e. the conclusion of Theorem 2.3.4 is true: there are Tr and* $code_i$ *satisfying 1-2 of Theorem 2.3.4.*

(ii) K^+ *and* K *have a common definitional expansion.*

(iii) K^+ *is rigidly definable over* K*.*

Proof. Proof of (i) ⇒ (ii): Assume that $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ and $code_i \in Fm(\mathsf{K}^+)$ are such that 1,2 in Theorem 2.3.4 hold. We want to show that K^+ is explicitly definable over K with recoverable auxiliaries, i.e. that K^+ and K have a common definitional expansion K^{++}. Now we set to definining K^{++}.

Let U_i be a new sort of K^+. First from the formula $code_i$ we will extract an explicit definition for U_i, cf. Figure 3.

Consider $code_i(x, \bar{x})$. Define[49]

$$\delta(\bar{x}) \stackrel{\text{def}}{=} Tr(\exists x code_i(x, \bar{x})) \quad \text{and}$$

$$\rho(\bar{x}, \bar{y}) \stackrel{\text{def}}{=} Tr(\exists x(code_i(x, \bar{x}) \wedge code_i(x, \bar{y}))).$$

Now $\delta(\bar{x}), \rho(\bar{x}, \bar{y}) \in Fm(\mathsf{K})$. Let $\bar{x} = \langle x_1, \dots, x_k \rangle$ and let the sorts of $x_1, \dots, x_k$ be $U_{j_1}, \dots, U_{j_k}$. These latter are sorts of K. Fix a model $\mathfrak{M} \in \mathsf{K}$.

First we define the relation S_i^{new} by $S_i^{new} \leftrightarrow \delta$, i.e.

$$S_i^{new} \stackrel{\text{def}}{=} \{\bar{u} \in U_{j_1} \times \ldots \times U_{j_k} : \mathfrak{M} \models \delta(\bar{u})\}.$$

Then S_i^{new} is a k-ary relation defined in $\mathfrak{M}$ by step (1). Second, from S_i^{new} we define the new sort D_i^{new} and $pj_1^i, \dots, pj_k^i$ by step (2.1):

$$D_i^{new} \stackrel{\text{def}}{=} S_i^{new} \quad \text{and}$$

$$pj_r^i \stackrel{\text{def}}{=} \{\langle \bar{u}, u_r \rangle : \bar{u} \in D_i^{new}\}, \quad \text{for } 1 \le r \le k.$$

Now we define the new binary relation R_i^{new} by step (1) as follows:

$$R_i^{new} \stackrel{\text{def}}{=} \{\langle v, w \rangle \in {}^2D_i^{new} : \mathfrak{M} \models \rho(pj_1^i(v), \dots, pj_k^i(v), pj_1^i(w), \dots, pj_k^i(w)).$$

[49] δ stands for "<u>d</u>omain of $code_i{}^{-1}$" while ρ stands for "equivalence <u>r</u>elation defined by $code_i{}^{-1}$".

Intuitively, R_i^{new} is the relation defined by ρ "projected up" to D_i^{new}. Then R_i^{new} is an equivalence relation on D_i^{new}, by the properties of $code_i, Tr$ and by the definitions of ρ, δ. We then can define, as in step (2.2), the factor-sort:

$$U_i \stackrel{\text{def}}{=} D_i^{new}/R_i^{new},$$

$$\in_i \stackrel{\text{def}}{=} \{\langle v, v/R_i^{new}\rangle : v \in D_i^{new}\}.$$

Let

$$\mathfrak{N} \stackrel{\text{def}}{=} \langle \mathfrak{M}, D_i^{new}, U_i; pj_1^i, \ldots, pj_k^i, \in_i, S_i^{new}, R_i^{new}\rangle.$$

Let $\mathfrak{M}^+ \in \mathsf{K}^+$ be any expansion of $\mathfrak{M}$. The name of the sort U_i in $\mathfrak{N}$ is the same as in $\mathfrak{M}^+$, but its "value" may be different, i.e. $U_i^{\mathfrak{N}}$ may be different from $U_i^{\mathfrak{M}^+}$. However, there is a natural bijection between these sets, as follows. Let

$$Code_i(u, \bar{x}) \stackrel{\text{def}}{=} (\exists v \in D_i^{new})(\in_i(v, u) \wedge \bigwedge_r pj_r^i(v, x_r) \wedge S_i^{new}(\bar{x})).$$

Then $Code_i(u, \bar{x})$ is in the language of $\mathfrak{N}$. By the above construction and by the properties of our translation, there is a bijection $f : U_i^{\mathfrak{N}} \longrightarrow U_i^{\mathfrak{M}^+}$ such that for all $u \in U_i^{\mathfrak{N}}$ and $\bar{a} \in {}^k Uv\mathfrak{M}$

$$\mathfrak{N} \models Code_i[u, \bar{a}] \quad \text{iff} \quad \mathfrak{M}^+ \models code_i[f(u), \bar{a}].$$

See Figure 3.

Let $\mathfrak{M}_i$ be the isomorphic copy of $\mathfrak{N}$ where we replace each element u of $U_i^{\mathfrak{N}}$ with $f(u)$. Then $\mathfrak{M}_i$ is a definitional expansion of $\mathfrak{M}$, obtained by steps (1),(2.1),(1),(2.2). Let $U_1, \ldots, U_t$ be all the new sorts of K^+ and let us do the above for all new sorts. Let $\mathfrak{M}'$ be the definitional expansion of $\mathfrak{M}$ we get by expanding $\mathfrak{M}$ with all the new sorts and relations of $\mathfrak{M}_i$, for $1 \le i \le t$. Then $\mathfrak{M}'$ contains all the new sorts of K^+, $Uv\mathfrak{M}^+ \subseteq Uv\mathfrak{M}'$, and moreover, for all $1 \le i \le t$

$$\mathfrak{M}' \models Code_i[u, \bar{a}] \quad \text{iff} \quad \mathfrak{M}^+ \models code_i[u, \bar{a}].$$

Now we set to defining the new relations of K^+ in $\mathfrak{M}'$.

Let T_j be a new relation in K^+ with arity $\langle U_1, \ldots, U_m\rangle$. Assume that, of these, $U_{i_1}, \ldots, U_{i_\ell}$ are sorts of K, while the rest, $U_{j_1}, \ldots, U_{j_s}$ are new sorts of K^+. Let $\tau \stackrel{\text{def}}{=} Tr(T_j(\bar{y}))$. Then by the properties of the translation function Tr we have that

$$\mathsf{K}^+ \models code_{j_1}(y_{j_1}, \bar{x}_1) \wedge \ldots \wedge code_{j_s}(y_{j_s}, \bar{x}_s) \rightarrow [T_j(\bar{y}) \leftrightarrow \tau(\bar{x}_1, \ldots, \bar{x}_s, y_{i_1}, \ldots, y_{i_\ell})].$$

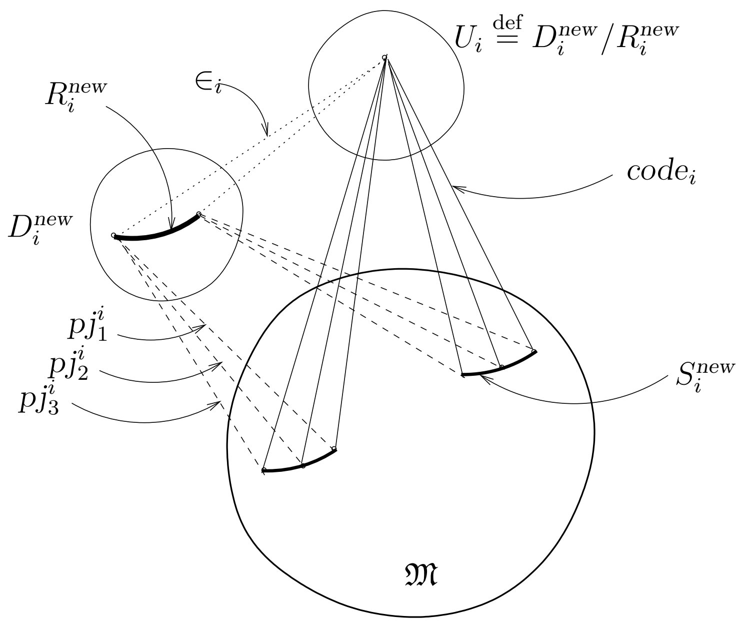

Figure 3: Illustration for the proof of Theorem 2.3.7 (i) $\Rightarrow$ (ii). From the formula $code_i$ we construct an explicit definition for U_i.

Now we define the new relation T_j in $\mathfrak{M}'$ by the formula

$$T_j(\bar{y}) \leftrightarrow \exists \bar{x}_1 \dots \bar{x}_s (\bigwedge_r Code_{j_r}(y_{j_1}, \bar{x}_r) \wedge \tau(\bar{x}_1, \dots, \bar{x}_s, y_{i_1}, \dots, y_{i_\ell})).$$

By the construction, T_j denotes the same relation in $\mathfrak{M}'$ and in $\mathfrak{M}^+$. Let now $\mathfrak{M}^{++}$ be the definitional expansion of $\mathfrak{M}'$ with all the new relations T_j, let $\mathsf{L} \stackrel{\text{def}}{=} \{\mathfrak{M}^{++} : \mathfrak{M} \in \mathsf{K}\}$ and let $\mathsf{K}^{++} \stackrel{\text{def}}{=} \{\mathfrak{N} \in \mathsf{L} : \mathfrak{N} \restriction Voc\mathsf{K}^+ \in \mathsf{K}^+\}$. Then K^{++} is a definitional expansion of K and K^+ is a reduct[50] of K^{++}. We want to show that K^{++} is a definitional expansion of K^+ also. The new (relative to K^+) sorts and relations of K^{++} are

$$S_i^{new}, D_i^{new}, pj_r^i, R_i^{new} \text{ and } \in_i$$

when U_i is a sort of K^+ which is not present in K. We define $S_i^{new}, D_i^{new}, pj_r^i$, and R_i^{new} by using steps (1),(2.1),(1) as we did in K. Since K^+ is an expansion of K, and these were all definable in K, we immediately have that the same definition will work for them in K^+, too. We then define $\in_i$ by step (1) (and not by step (2.2) since U_i is an "old" sort in K^+) as follows:

$$\in_i(v, u) \leftrightarrow \exists \bar{x}(code_i(u, \bar{x}) \wedge pj_1^i(v, x_1) \wedge \dots \wedge pj_k^i(v, x_k))).$$

By the above, (i) $\Rightarrow$ (ii) has been proved. (ii) $\Rightarrow$ (i) was proved as Theorem 2.3.4.

Proof of (i) $\Rightarrow$ (iii): Assume that K^+ is interpreted in K by some translation mapping *Tr* and formulas $code_i$. Then K^+ is definable in K, as we have seen above. By the properties of a translation mapping then K^+ is *rigidly* definable over K.

Proof of (iii) $\Rightarrow$ (i): Here we will use Beth's definability theorem for one-sorted models (i.e. for defining relations only). Assume that K^+ is rigidly explicitly definable over K. Let K^{++} be a definitional expansion of K and assume that K^+ is a reduct of K^{++}. Let $code_i, Tr$ be a translation of K^{++} to K. Since K^+ is a reduct of K^{++}, then $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$. Let U_i be a new sort of K^+. Then U_i is a new (relative to K) sort of K^{++}, therefore there is $code_i \in Fm(\mathsf{K}^{++})$ which has good properties w.r.t. *Tr*. The problem is that $code_i$ may not be in $Fm(\mathsf{K}^+)$. We will show that $code_i$ is expressible in $Fm(\mathsf{K}^+)$, i.e. it is equivalent in K^+ with a formula in the language of K^+.

For any new sort U_i of K^+ let R_i be a new relation symbol and let $\Delta(R_i)$ be the set of the following three formulas, where $\rho_i(\bar{x}, \bar{y})$ is $Tr(\exists x(code_i(x, \bar{x}) \wedge code_i(x, \bar{y}))$, as in the proof of (i) $\Rightarrow$ (ii):

[50] We introduced L into the picture only because we did not assume that K^+ is closed under isomorphism and we want K^+ be a reduct of K^{++}.

$\forall x \exists \bar{x} R_i(x, \bar{x})$

$\exists x (R_i(x, \bar{x}) \wedge R_i(x, \bar{y})) \leftrightarrow \rho_i(\bar{x}, \bar{y})$

$(R_i(x, \bar{x}) \wedge R_i(y, \bar{x})) \longrightarrow x = y$

Then $\Delta(R_i)$ is a set of formulas in the language of K expanded with one new relation symbol R_i. For any new relation T_j of K^+ let $U_{i_1}, \dots, U_{i_\ell}$ be the sorts of K, and $U_{j_1}, \dots U_{j_t}$ be the new sorts of K^+ occurring in the arity of T_j and let $\Delta(T_j)$ denote the formula

$$T_j(\bar{y}) \leftrightarrow \exists \bar{x}_1 \dots \bar{x}_t (\textstyle\bigwedge_r R_{j_1}(u_{j_r}, \bar{x}_r) \wedge Tr(T_j(\bar{y}))(\bar{x}_1, \dots, \bar{x}_t, y_{i_1}, \dots, y_{i_\ell})).$$

Let Δ be the set of all the above formulas, i.e.

$$\Delta \stackrel{\text{def}}{=} \bigcup \{\Delta(R_i) : U_i \text{ is a new sort of } \mathsf{K}^+\} \cup \{\Delta(T_j) : T_j \text{ is a new relation of } \mathsf{K}^+\}.$$

Now, Δ is a set of formulas in the language of K^+ expanded with new relation symbols R_i for all new sorts U_i. We will show that Δ is an implicit definition of $\langle R_i : U_i \text{ is a new sort in } \mathsf{K}^+ \rangle$ in K^+, in the usual sense. Indeed, let $\mathfrak{M}^+ \in \mathsf{K}^+$, $\bar{R}' \stackrel{\text{def}}{=} \langle R_i' \rangle$ and $\bar{R}'' \stackrel{\text{def}}{=} \langle R_i'' \rangle$ be systems of concrete relations in $\mathfrak{M}^+$ such that

$$\langle \mathfrak{M}^+, \bar{R}' \rangle \models \Delta \quad \text{and} \quad \langle \mathfrak{M}^+, \bar{R}'' \rangle \models \Delta.$$

Then, by using the construction of Δ, one can show that there is an isomorphism f between $\langle \mathfrak{M}^+, \bar{R}' \rangle$ and $\langle \mathfrak{M}^+, \bar{R}'' \rangle$ such that f is identity on the sorts of K, i.e.. f is identity on $\mathfrak{M} \in \mathsf{K}$, where $\mathfrak{M}$ is the reduct of $\mathfrak{M}^+$ in K. Rigidity of K^+ over K implies that then f is identity on $\mathfrak{M}^+$ also, because both Id and f are isomorphisms on $\mathfrak{M}^+$ that are identity on $\mathfrak{M}$. Since f is the identity, we get that $\bar{R}' = \bar{R}''$. Thus in each model $\mathfrak{M}^+ \in \mathsf{K}^+$ there is at most one system $\bar{R}$ of concrete relations satisfying Δ. To be able to use the Beth theorem, we need that this property holds for all $\mathfrak{M}^+$ in the axiomatizable hull $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$ of K^+ as well. By the Keisler-Shelah theorem, and by our assumption that K^+ is closed under taking ultraproducts we have that $\mathfrak{N} \in \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$ iff an ultrapower ${}^I\mathfrak{N}/F$ of $\mathfrak{N}$ is in K^+. Assume that there are two different systems of relations satisying Δ in $\mathfrak{N}$. Then the same is true in ${}^I\mathfrak{N}/F$. This contradicts our earlier argument showing that on each model $\mathfrak{M}^+ \in \mathsf{K}^+$ there is at most one system of relations satisfying Δ. Thus Δ is an implicit definition in the axiomatizable hull $\mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$ of K^+. By Beth's theorem then each of R_i is definable in the language of K^+. Let $\gamma_i(x, \bar{x}) \in Fm(\mathsf{K}^+)$ be such that $\mathsf{Th}(\mathsf{K}^+) \cup \Delta \models R_i(x, \bar{x}) \leftrightarrow \gamma_i(x, \bar{x})$. By the construction of Δ we also have that

$\mathsf{K}^{++} \models code_i(x, \bar{x}) \leftrightarrow R_i(x, \bar{x})$.

Thus, $Tr \restriction Fm(\mathsf{K}^+)$ toghether with the γ_i's is a good translation from K^+ to K. This finishes the proof of Theorem 2.3.7. ∎

Remark 2.3.8 (Discussion of Theorem 2.3.7.) **(i)** Theorem 2.3.7 is true for <u>arbitrary languages</u>, we do not need that there are only finitely many sorts or that we have only countably many symbols in the language. **(ii)** The condition that K^+ is closed under ultraproducts is needed for the direction (ii) $\Rightarrow$ (i). An example showing this is the following. Let K be the class of finite linear orderings on sort U_0. Let K^+ be the class of two-sorted models where the sorts are U_0, U_1, there is a finite linear ordering both on U_0 and on U_1 and $|U_0| = |U_1|$. Now K^+ is rigidly explicitly definable over K. But the class UpK^+ of all ultraproducts of members of K^+ is not rigid over the class UpK of all ultraproducts of members of K. (To see this, take any infinite ultraproduct of elements of K^+. Then there is a nontrivial automorphism of the linear ordering on U_1.) However, it is not difficult to see that if K and K^+ have a common definitional expansion, then UpK and UpK^+ also have a common definitional expansion, which would imply that UpK^+ is rigid over UpK. Thus K and K^+ do not have a common definitional expansion. **(iii)** Thm2.5.1 together with Thm.2.3.7 will imply that if K^+ is rigidly definable over K and K is axiomatizable, then K^+ is nr-implicitly definable over K. $\lhd$

COROLLARY 2.3.9 Assume that K^+ is an expansion of K and that K is axiomatizable. Then

$$\mathsf{K} \equiv_\Delta \mathsf{K}^+ \quad \Rightarrow \quad \mathsf{K} \equiv_\Delta \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+)).$$

Proof. By Thm.2.3.7, $\mathsf{K} \equiv_\Delta \mathsf{K}^+$ implies the existence of $Tr, code_i$ interpreting K^+ in K. Let Δ be the set of formulas expressing that $Tr, code_i$ interpret K^+ in K. Then $\mathsf{K}^+ \models \Delta \cup \mathsf{Th}(\mathsf{K})$. Moreover, let $\mathsf{K}_1 \stackrel{\text{def}}{=} \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$. Then $\mathsf{K} = \mathsf{K}_1 \restriction Voc\mathsf{K}$ by $\mathsf{K}_1 \models \mathsf{Th}(\mathsf{K})$ and $\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$, and also $Tr, code_i$ interpret K_1 in $\mathsf{K}_1 \restriction Voc\mathsf{K}$ by $\mathsf{K}_1 \models \Delta$. Thus $Tr, code_i$ interpret K_1 in K. By Thm.2.3.7 then $\mathsf{K} \equiv_\Delta \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$. ∎

2.4 Definitional equivalence of theories.

In section 2.3 we dealt with classes K and L where L was an expansion of K. In this sub-section we turn to the case when L is not necessarily an expansion of K.

Definition 2.4.1 Let K and L be two classes of models. We say that they are *definitionally equivalent*, in symbols $\mathsf{K} \equiv_\Delta \mathsf{L}$, iff they admit a common (explicit) definitional expansion M (without taking reducts).[51]

Further, $\mathfrak{M} \equiv_\Delta \mathfrak{N}$ abbreviates $\{\mathfrak{M}\} \equiv_\Delta \{\mathfrak{N}\}$. If $\mathfrak{M} \equiv_\Delta \mathfrak{N}$, then we say that $\mathfrak{M}$ and $\mathfrak{N}$ are *definitionally equivalent models*. Two *theories* Th_1, Th_2 are called *definitionally equivalent* iff $\mathsf{Mod}(Th_1) \equiv_\Delta \mathsf{Mod}(Th_2)$. $\lhd$

Cf. also in Hodges [11] under the name "*definitional equivalence*" pp. 60–61; cf. also Henkin-Monk-Tarski [10, Part I, e.g. p.56].

We will see that one can say that two definitionally equivalent theories can be regarded as being *essentially the same* theory and the difference between them is only that their *"syntactic decorations"* are different (i.e. they "choose" to represent their [essentially] common language with *different basic vocabularies*).

The same applies to classes of models K, L when $\mathsf{K} \equiv_\Delta \mathsf{L}$. As an example, choose K to be Boolean algebras with $\{\cap, -\}$ as their basic operations while choose L to be Boolean algebras with $\{\cup, -, 0, 1\}$ as basic operations. (Then $\mathsf{K} \equiv_\Delta \mathsf{L}$.) At a certain level of abstraction, K and L can be regarded as a collection of the *same* mathematical structures (namely, Boolean algebras) and the difference (between K and L) is only in the choice of their basic vocabularies (which is "$\cap, -$" in the one case while "$\cup, -, 0, 1$" in the other). Summing up: In some sense, definitionally equivalent theories $Th_1 \equiv_\Delta Th_2$ can be considered as just one theory with two *different linguistic representations*. The same applies to definitionally equivalent classes of models.

The relation $\equiv_\Delta$ defined above is symmetric and reflexive. For certain "administrative" reasons it is not transitive, but the counterexamples (to transitivity) are so artificial that we will not meet them (in this work). We could define $\equiv_\Delta^*$ to be the transitive closure of $\equiv_\Delta$ and then use $\equiv_\Delta^*$ as definitional equivalence. If this were a logic book we would do that. However, in the present work we will not need $\equiv_\Delta^*$,

[51] I.e. M is a definitional expansion (without taking reducts) of K and the same holds for L in place of K. Note that $\mathsf{Th}(\mathsf{M})$ can be regarded as an implicit definition of M over K, and the same for L in place of K.

hence we do not discuss it, and we call $\equiv_\Delta$ definitional equivalence (though it is $\equiv^*_\Delta$ which is the really satisfactory notion of definitional equivalence.)

Discussion of the definition of $\equiv_\Delta$

(1) Assume $\mathsf{K} \equiv_\Delta \mathsf{L}$. Then K and L agree on the common part of their vocabularies.[52] I.e.

$$\mathsf{K} \equiv_\Delta \mathsf{L} \qquad \Rightarrow \qquad \mathsf{K} \restriction (\mathit{Voc}\mathsf{K} \cap \mathit{Voc}\mathsf{L}) = \mathsf{L} \restriction (\mathit{Voc}\mathsf{K} \cap \mathit{Voc}\mathsf{L}).$$

(2) For any definitional expansion K^+ of K we have $\mathsf{K} \equiv_\Delta \mathsf{K}^+$. In general, if K^+ is an expansion of K and K is closed under taking ultraproducts, then $\mathsf{K}^+ \equiv_\Delta \mathsf{K}$ iff K^+ is rigidly definable over K, see Thm.2.3.7.

(3) Assume $\mathsf{K} \equiv_\Delta \mathsf{L}$. Then L and K are definable over each other. Moreover one can choose their definitions over each other to be the same. Indeed, if M is the common definitional expansion of K and L mentioned in the definition of $\equiv_\Delta$, then $\mathsf{Th}(\mathsf{M})$ is a definition of K over L as well as a definition of L over K.

(4) Definitional equivalence is stronger than mutual (explicit) definability: there exist classes K and L such that they are definable over each other, yet $\mathsf{K} \not\equiv_\Delta \mathsf{L}$ (see Examples 2.4.13, p.67). Moreover, this is so even in the one-sorted case: We can choose K and L such that both K and L have only one, common, sort. Such an example can be found in Andréka-Madarász-Németi [3].

(5) Assume $\mathsf{K} \equiv_\Delta \mathsf{L}$. Then there is a bijection-up-to-isomorphism

$$f : \mathsf{K} \rightarrowtail\!\!\!\twoheadrightarrow \mathsf{L}$$

between K and L, and there are definitions $\Delta_\mathsf{K}, \Delta_\mathsf{L}$ such that for all $\langle \mathfrak{M}, \mathfrak{N} \rangle \in f$ the following hold:

(i) $\mathfrak{M} \restriction (\mathit{Voc}\mathsf{K} \cap \mathit{Voc}\mathsf{L}) = \mathfrak{N} \restriction (\mathit{Voc}\mathsf{K} \cap \mathit{Voc}\mathsf{L})$

(ii) $\mathfrak{M}$ and $\mathfrak{N}$ have a common definitional expansion $\mathfrak{M}^+$

(iii) Δ_K defines $\mathfrak{M}^+$ over $\mathfrak{M}$ and Δ_L defines $\mathfrak{M}^+$ over $\mathfrak{N}$.

[52] As a contrast, $\mathsf{K} \equiv^*_\Delta \mathsf{L}$ does not imply this, however as we said we will not need the generality of $\equiv^*_\Delta$ in this work.

Indeed, if M is a common definitional expansion of K and L with definitions $\Delta_{\mathsf{K}}, \Delta_{\mathsf{L}}$ over K and L respectively, then we can choose f to be

$$f = \{\langle \mathfrak{M} \upharpoonright Voc\mathsf{K}, \mathfrak{M} \upharpoonright Voc\mathsf{L}\rangle : \mathfrak{M} \in \mathsf{L}\}.$$

(6) Assume $\mathsf{K} \equiv_\Delta \mathsf{L}$. Then the bijection-up-to-isomorphism $f : \mathsf{K} \rightarrowtail\!\!\!\twoheadrightarrow \mathsf{L}$ in (5) above has the following property. For all $\mathfrak{M} \in \mathsf{K}$, the automorphism group of $\mathfrak{M}$ is isomorphic to the automorphism group of $f(\mathfrak{M})$, in symbols[53]

$$\langle Aut(\mathfrak{M}), \circ\rangle \cong \langle Aut(f(\mathfrak{M})), \circ\rangle.$$

This is so because of the following. Let $\mathfrak{M}^+ \in \mathsf{M}$ be such that $\mathfrak{M}^+$ is implicitly definable without taking reducts both over $\mathfrak{M}$ and over $f(\mathfrak{M})$. Since $\mathfrak{M}^+$ is implicitly definable without taking reducts over $\mathfrak{M}$, each automorphism of $\mathfrak{M}$ extends in a unique way to an automorphism of $\mathfrak{M}^+$, and this implies that the automorphism groups of $\mathfrak{M}$ and $\mathfrak{M}^+$ are isomorphic. We get the same for $f(\mathfrak{M})$ and $\mathfrak{M}^+$ completely analogously, and this proves that the automorphism groups of $\mathfrak{M}$ and $f(\mathfrak{M})$ are isomorphic.

(7) For more on definitional equivalence, its importance, and for motivation for the way we defined and use $\equiv_\Delta$ we refer to [10, pp. 56-57, Remark 0.1.6], [11, pp. 58-61]. $\lhd$

Each of the properties in items (3) and (5) of the above discussion are equivalent with $\mathsf{K} \equiv_\Delta \mathsf{L}$. We will state this in the following theorem.

THEOREM 2.4.2 *Let K and L be two classes. Then (i)-(iii) below are equivalent.*

(i) $\mathsf{K} \equiv_\Delta \mathsf{L}$

(ii) *There is a Δ such that Δ defines K over L and Δ defines L over K.*

(iii) *For every $\mathfrak{M} \in \mathsf{K}$ there is $\mathfrak{N} \in \mathsf{L}$ and for every $\mathfrak{N} \in \mathsf{L}$ there is $\mathfrak{M} \in \mathsf{K}$ such that $\mathfrak{M}$ and $\mathfrak{N}$ have a common definitional expansion, and moreover the definitions of the expansion over $\mathfrak{M}$ and over $\mathfrak{N}$ can be chosen uniformly.*

Proof. Proof of (ii) $\Rightarrow$ (i): Assume Δ defines K over L and Δ defines L over K. Then there are K^+ and L^+ such that K^+ is a definitional expansion of K, defined by Δ, and L is a reduct of K^+ and the analogous statement for L^+. It is easy to

[53] $f(\mathfrak{M})$ exists only up to isomorphism, but we talk only about the automorphism group of $f(\mathfrak{M})$ which is defined by the isomorphism type of $\mathfrak{M}$ up to isomorphism. So this makes sense.

see that then $Voc\mathsf{K}^+ = Voc\mathsf{L}^+ = Voc\mathsf{K} \cup Voc\mathsf{L} \cup Voc\Delta$, where $Voc\Delta$ denotes the set of sort and relation symbols occurring in Δ. Let $\mathsf{M} \stackrel{\text{def}}{=} \mathsf{K}^+ \cup \mathsf{L}^+$. Then M is an expansion of both K and L. Also, $\mathsf{M} \models \Delta$ because $\mathsf{K}^+ \models \Delta$ and $\mathsf{L}^+ \models \Delta$. Thus, Δ is an nr-implicit definition of M both over K and over L.

Proof of (iii) $\Rightarrow$ (i): Let $\Delta_{\mathsf{K}}, \Delta_{\mathsf{L}}$ be the uniform definitions in (iii). Let $\mathsf{M} \stackrel{\text{def}}{=} \{\mathfrak{M} : \mathfrak{M} \restriction Voc\mathsf{K} \in \mathsf{K},\ \mathfrak{M} \restriction Voc\mathsf{L} \in \mathsf{L}$ and Δ_{K} defines $\mathfrak{M}$ over $\mathfrak{M} \restriction Voc\mathsf{K}$, Δ_{L} defines $\mathfrak{M}$ over $\mathfrak{M} \restriction Voc\mathsf{L}\}$. Then M is a class of similar models, namely $Voc\mathsf{M} = Voc\mathsf{K} \cup Voc\mathsf{L} \cup Voc\Delta_{\mathsf{K}} \cup Voc\Delta_{\mathsf{L}}$. Then both K and L are reducts of M, i.e. $\mathsf{K} = \mathsf{M} \restriction Voc\mathsf{K}$ etc., by (iii). Also, $\mathsf{M} \models \Delta_{\mathsf{K}} \cup \Delta_{\mathsf{L}}$, thus M is a definitional expansion of both K and L.

(i) $\Rightarrow$ (ii), and (i) $\Rightarrow$ (iii) were shown already in the discussion of $\equiv_\Delta$. ∎

THEOREM 2.4.3 *Let K, L be two classes of models and assume that* $\mathbf{I}\mathsf{K}$ *is closed under taking ultraproducts. Then (i) and (ii) below are equivalent.*

(i) $\mathsf{K} \equiv_\Delta \mathsf{L}$

(ii) K *and* L *have a common extension which is rigidly definable both over* K *and over* L*.*

Proof. Let M be a common rigidly definable expansion of K and L. Since $\mathbf{I}\mathsf{K}$ is closed under taking ultraproducts, then $\mathbf{I}\mathsf{M}$ is closed under taking ultraproducts, too. Hence we can apply Thm.2.3.7 to obtain common definitional expansions K^+ and L^+ of K and M and of L and M respectively. We also may assume that the new sorts and relations in K^+ and L^+ have different names, i.e. $Voc\mathsf{K}^+ \cap Voc\mathsf{L}^+ = Voc\mathsf{M}$. Then it is not difficult to see that there is a common definitional expansion M^+ of K^+ and L^+. Now, M^+ is a common definitional expansion of K and L. See Figure **??**. ∎

In connection with Theorem 2.4.3 we note that it is not difficult to see that if K^+ is nr-implicitly definable over K, then $\mathbf{I}\mathsf{K}$ is closed under ultraproducts iff $\mathbf{I}\mathsf{K}^+$ is closed under ultraproducts.

LEMMA 2.4.4 *Let K, L and K^+ be classes of models. Assume that K^+ is rigidly definable over K,* $\mathbf{I}\mathsf{L}$ *is closed under taking ultraproducts, $Voc\mathsf{K}^+ \cap Voc\mathsf{L} = Voc\mathsf{K} \cap Voc\mathsf{L}$, and $\mathsf{K} \equiv_\Delta \mathsf{L}$. Then $\mathsf{K}^+ \equiv_\Delta \mathsf{L}$.*

Proof. Assume that $\mathsf{K}, \mathsf{L}, \mathsf{K}^+$ satisfy the conditions of the lemma. Let M be a common definitional expansion of L and K and let Δ, Σ be the respective definitions of M over L and K. Since $\mathbf{I}\mathsf{L}$ is closed under taking ultraproducts and $\mathsf{K} \equiv_\Delta \mathsf{L}$, we have that $\mathbf{I}\mathsf{K}$ also is closed under taking ultraproducts, and since K^+ is definable over

K, then $\mathbf{I}\mathsf{K}^+$ is closed under taking ultraproducts. Thus by Thm.2.3.7, K and K^+ have a common definitional expansion M^+. Let Δ_1, Σ_1 be the respective definitions of M^+ over K and K^+.

We may assume that $Voc\mathsf{M}$ is disjoint from $Voc\mathsf{K}^+ \setminus Voc\mathsf{K}$ (by our assumption $Voc\mathsf{K}^+ \cap Voc\mathsf{L} = Voc\mathsf{K} \cap Voc\mathsf{L}$), and that $Voc\mathsf{M}^+ \setminus Voc\mathsf{K}^+$ is disjoint from $Voc\mathsf{M}$. Hence $Voc\mathsf{M}^+ \cap Voc\mathsf{M} = Voc\mathsf{K}$.

Let $\mathfrak{N} \in \mathsf{L}$ be arbitrary. There are $\mathfrak{M} \in \mathsf{K}$ and $\mathfrak{M}^+ \in \mathsf{M}$ such that $\mathfrak{M}^+$ is a common definitional expansion of $\mathfrak{N}$ and $\mathfrak{M}$. There are $\mathfrak{M}_1 \in \mathsf{K}^+$ and $\mathfrak{M}_1^+$ in M^+ such that $\mathfrak{M}_1^+$ is a common definitional expansion of both $\mathfrak{M}$ and $\mathfrak{M}_1$. By $Voc\mathsf{M} \cap Voc\mathsf{M}^+ = Voc\mathsf{K}$, the union of $\mathfrak{M}^+$ and $\mathfrak{M}_1^+$ is a model, $\mathfrak{M}^{++}$. Then $\mathfrak{M}^{++}$ is a common expansion of $\mathfrak{N}$ and $\mathfrak{M}_1 \in \mathsf{K}^+$. Since Δ defines $\mathfrak{M}^+$ over $\mathfrak{N}$ and Δ_1 defines $\mathfrak{M}_1^+$ over $\mathfrak{M}$, we have that $\Delta \cup \Delta_1$ defines $\mathfrak{M}^{++}$ over $\mathfrak{N}$. Similarly, $\Sigma \cup \Sigma_1$ defines $\mathfrak{M}^{++}$ over $\mathfrak{M}_1$. The proof of the other direction, $(\forall \mathfrak{M} \in \mathsf{K}^+ \exists \mathfrak{N} \in \mathsf{L}) \ldots$ is completely analogous. $\mathsf{K}^+ \equiv_\Delta \mathsf{L}$ then follows by Thm.2.4.2(iii)$\Rightarrow$(i). ∎

Remark 2.4.5 (How and why can definitionally equivalent theories [and classes of models] be regarded as identical [as a corollary of the translation theorems]?)
In addition to the text below, we also refer the reader to [10, p.56] and [11, pp.58–61] for explanations of why definitionally equivalent classes of models can be regarded as (in some sense) identical.

Let K and L be two definitionally equivalent classes of models (formally, $\mathsf{K} \equiv_\Delta \mathsf{L}$). Then, by the definition of $\equiv_\Delta$, there is a class M which is a definitional expansion (without taking reducts) of both K and L. We will argue below that this M establishes a very strong connection between K and L. (Cf. also item (5) in the discussion of the definition of $\equiv_\Delta$.) Our argument begins with the following: We can apply Theorem 2.3.2 to the pair M and K with M in place of K^+ in that theorem. The same applies to the pair M and L. By Theorem 2.3.2, then we have two translation mappings

$$\begin{tikzcd} & Fm(\mathsf{M}) \arrow[dl, "Tr_1"'] \arrow[dr, "Tr_2"] & \\ Fm(\mathsf{K}) & & Fm(\mathsf{L}) \end{tikzcd}$$

both of which preserve meaning (in the sense of Theorem 2.3.2). Both of Tr_1 and Tr_2 are surjective. Intuitively, Tr_1 identifies K with M while Tr_2 identifies M with L. Hence K *gets identified with* L. (Perhaps the best way of thinking about this is that we identify both K and L with their common expansion M. As a by-product of this we identify K and L with each other, too.)

By surjectiveness of Tr_1 and Tr_2, whatever can be said in the language $Fm(\mathsf{K})$, the same can be said in $Fm(\mathsf{M})$ and hence (using Tr_2) the same can be said in the language $Fm(\mathsf{L})$ of L. Similarly, whatever can be said in $Fm(\mathsf{L})$ the same can be said in $Fm(\mathsf{K})$, too.

Now, if we want some more detail, let $\varphi(\bar{z}) \in Fm(\mathsf{K})$ with a sequence $\bar{z}$ of variables belonging to common sorts K and L. Then there are $\varphi'(\bar{z}) \in Fm(\mathsf{M})$, $\varphi''(\bar{z}) \in Fm(\mathsf{L})$ such that $Tr_1(\varphi') = \varphi$ and $Tr_2(\varphi') = \varphi''$. I.e.

$$\varphi(\bar{z}) \xleftarrow{\quad Tr_1 \quad} \varphi'(\bar{z}) \xmapsto{\quad Tr_2 \quad} \varphi''(\bar{z}).$$

Actually, we can choose $\varphi' = \varphi$ if we want to. Using Theorem 2.3.2 we can conclude

(4) $$\mathsf{M} \models \varphi(\bar{z}) \leftrightarrow \varphi''(\bar{z}).$$

I.e. the same things can be said about the common variables $\bar{z}$ in $Fm(\mathsf{K})$ and in $Fm(\mathsf{L})$. Hence the languages of K and L have the *same* expressive power.

On the basis of (4) above and what was said before (4), we can introduce two, more direct, translation mappings

$$\begin{tikzcd} Fm(\mathsf{K}) \arrow[r, shift left, "T_2"] & Fm(\mathsf{L}) \arrow[l, shift left, "T_1"] \end{tikzcd}$$

defined as follows. In defining T_1 and T_2 we can rely on the fact that

$$Fm(\mathsf{K}) \subseteq Fm(\mathsf{M}) = Dom(Tr_1)$$

and that $Tr_1 \restriction Fm(\mathsf{K}) = \mathrm{Id} \restriction Fm(\mathsf{K})$ which is the identity function. Hence we can choose

$$\begin{aligned} T_1 &:= Tr_1 \restriction Fm(\mathsf{L}) \quad \text{and} \\ T_2 &:= Tr_2 \restriction Fm(\mathsf{K}).^{54} \end{aligned}$$

[54] In passing, we also note that Tr_1 can be regarded as *injective in the sense* that if $\psi(\bar{z}), \gamma(\bar{z}) \in Fm(\mathsf{M})$ involve free variables of K only then $[\, Tr_1(\psi) = Tr_1(\gamma) \ \Rightarrow\ \mathsf{M} \models \psi(\bar{z}) \leftrightarrow \gamma(\bar{z})]$. Similarly for Tr_2 and L.

$$\begin{tikzcd}
 & Fm(\mathsf{M}) \arrow[dl, "Tr_1"'] \arrow[dr, "Tr_2"] & \\
Fm(\mathsf{K}) \arrow[rr, shift left, "T_2"] & & Fm(\mathsf{L}) \arrow[ll, shift left, "T_1"]
\end{tikzcd}$$

Assume $\varphi \in Fm(\mathsf{M})$ involves only common free variables of K and L. Then

$$\begin{aligned}
\mathsf{M} &\models (T_2\, Tr_1\varphi) \leftrightarrow Tr_2\varphi. \\
\mathsf{M} &\models (T_1\, Tr_2\varphi) \leftrightarrow Tr_1\varphi.
\end{aligned}$$

So in this "logical sense" the above diagram commutes.

For completeness, about the above diagram we also note the following commutativity property:

$$\begin{aligned}
T_2 &\subseteq (Tr_1)^{-1} \circ Tr_2\,, \\
T_1 &\subseteq (Tr_2)^{-1} \circ Tr_1\,.
\end{aligned}$$

Here we note that $(Tr_1)^{-1} \circ Tr_2$ is a binary relation but not necessarily a function.

Using Theorem 2.3.2, and (4) way above, one can check that for all $\varphi \in Fm(\mathsf{K})$ and for all $\psi \in Fm(\mathsf{L})$, if φ and ψ use only variables of common sorts (of K and L) then:

(5)
$$\begin{aligned}
\mathsf{M} &\models \varphi(\bar z) \leftrightarrow (T_2\varphi)(\bar z), \\
\mathsf{M} &\models (T_1\psi)(\bar z) \leftrightarrow \psi(\bar z),
\end{aligned}$$
further

(6)
$$\begin{aligned}
\mathsf{K} &\models \varphi(\bar z) \leftrightarrow (T_1T_2\varphi)(\bar z), \\
\mathsf{L} &\models \psi(\bar z) \leftrightarrow (T_2T_1\psi)(\bar z).
\end{aligned}$$

These statements can be interpreted as saying that T_1 and T_2 are kind of *inverses* of each other and that they establish a kind of logical isomorphism between equivalence classes of formulas in $Fm(\mathsf{K})$ and $Fm(\mathsf{L})$ involving free variables of common sorts only. For completeness, we note that (5–6) can be generalized to formulas involving free variables of arbitrary sorts by using Theorem 2.3.4. For formulating this generalized version of (5–6) one needs to use the formulas "*code*" as they were used in Theorem 2.3.4. E.g. the first line of (6) becomes

$$\mathsf{K} \models \mathit{code}(x, \vec{x}) \rightarrow [\varphi(x, \bar{z}) \leftrightarrow (T_1 T_2 \varphi)(\vec{x}, \bar{z})],$$

where x belongs to a sort of K not in L, and $\bar{z}$ is a sequence of variables of common sorts of K and L. Here $\mathit{code}(x, \vec{x})$ is the formula we get from combining the corresponding formulas belonging to Tr_1 and Tr_2. We leave the details of generalizing (4–6) to treating free variables not in the common language to the interested reader.

(We note that the generalization of (6) above reminds us of the notion of equivalence between two categories, in the sense of category theory.)

We hope, the above shows how and to what extent we consider two definitionally equivalent classes (and theories) as being essentially identical.

$\lhd$

Weak definitional equivalence

Definition 2.4.6 Let K and L be two classes of models and let $f : \mathsf{K} \longrightarrow \mathsf{L}$ be a function. We say that f is a <u>*first-order definable meta-function*</u> iff for each $\mathfrak{M} \in \mathsf{K}$ $f(\mathfrak{M})$ is first-order definable over $\mathfrak{M}$ (in the sense of §2.2) and the definition of $f(\mathfrak{M})$ over $\mathfrak{M}$ is uniform, i.e. is the same for all choices of $\mathfrak{M} \in \mathsf{K}$.[55]

$\lhd$

A typical example for first-order definable meta-functions will be e.g. $\mathcal{G} : \mathsf{Mod}(\mathit{Th}) \longrightarrow \mathsf{Ge}(\mathit{Th})$, where $\mathcal{G} : \mathfrak{M} \mapsto \mathfrak{G}_{\mathfrak{M}}$, if Th is strong enough, cf. Thm.**??** (p.**??**). A similar example will be a kind of inverse to this function $\mathcal{M} : \mathsf{Ge}(\mathit{Th}) \longrightarrow \mathsf{Mod}(\mathit{Th})$, cf. Prop.**??** (p.**??**) and Def.**??** (p.**??**).

We note that if $f : \mathsf{K} \rightarrowtail\!\!\!\!\!\twoheadrightarrow \mathsf{L}$ is a surjective first-order definable meta-function then L is definable over K; and, more generally, if $f : \mathsf{K} \longrightarrow \mathsf{L}$ is a first-order definable meta-function then $\mathit{Rng}(f)$ is definable over K. In the other direction, if $\mathsf{L} = \mathbf{I}\mathsf{L}$ is definable over K then there is a first-order definable meta-function $f : \mathsf{K} \longrightarrow \mathsf{L}$ such that $\mathit{Rng}(f)$ is L up to isomorphism. To be able to claim this for the case when $\mathsf{L} \neq \mathbf{I}\mathsf{L}$ we make the following convention.

[55] A first-order definable <u>meta-function</u> (acting between classes of models) is a rather <u>different</u> kind of thing from an ordinary function like *factorial* : $N \longrightarrow N$ definable in a model, say in $\mathfrak{N} \in \mathsf{Mod}$(Peano's arithmetic), cf. Example 2.1.5(1) on p.13. (This is the reason why we call f a meta-function and not simply a function.)

CONVENTION 2.4.7 (Class form of the axiom of choice)
In connection with the above definition, for simplicity, throughout the present chapter we assume the *class form of the axiom of choice*. More concretely we assume that our set theoretic universe V is well orderable by the class $\mathsf{Ordinals}$ of ordinal numbers. I.e. there is a bijection

$$f : \mathsf{Ordinals} \rightarrowtail\!\!\!\twoheadrightarrow V.$$

This implies that any proper class is well orderable and therefore there exists a bijection between any two proper classes.

$\lhd$

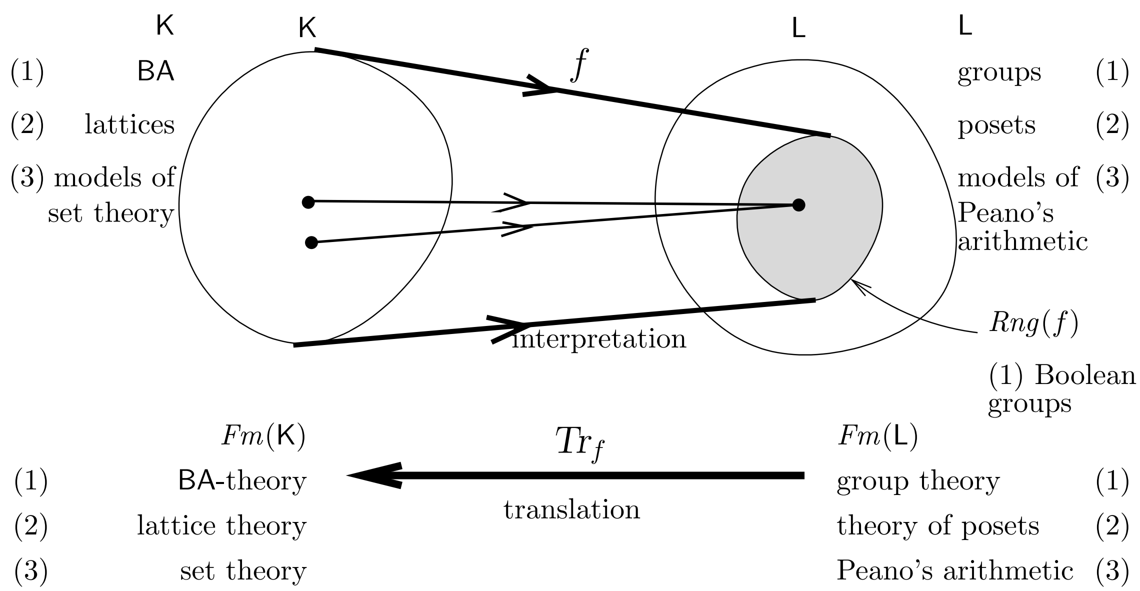

Figure 4: Examples for first-order definable meta-functions f and the induced translations between theories. For more explanation in connection with this picture cf. item (III) of Remark **??**, pp. **??**–**??**. The corresponding theories are labelled by the same numbers. E.g. BA is interpreted in "groups", "lattices" in "posets" etc.

The following proposition makes connections between the following three things: (i) "interpretations" of one theory in another, (ii) first-order definable meta-functions $f : \mathsf{K} \longrightarrow \mathsf{L}$ between classes of models, and (iii) definability of a class $Rng(f)$ over another class K, see Fig.4. In this context the function Tr_f (in the proposition) below is what we call an interpretation (or translation).[56] In particular the proposition says that any first-order definable meta-function $f : \mathsf{K} \longrightarrow \mathsf{L}$ induces a natural syntactical translation mapping from the language $Fm(\mathsf{L})$ of L to that of K. Moreover, this translation is meaning preserving w.r.t. the semantical function f.[57]

PROPOSITION 2.4.8 *Assume* $f : \mathsf{K} \longrightarrow \mathsf{L}$ *is a first-order definable meta-function. Then there is a "natural" translation mapping*

$$Tr_f : Fm(\mathsf{L}) \longrightarrow Fm(\mathsf{K})$$

such that for every $\varphi(\bar{x}) \in Fm(\mathsf{L})$ *with all free variables belonging to common sorts of* K *and* L [58], $\mathfrak{A} \in \mathsf{K}$ *and evaluation* $\bar{a}$ *of* $\bar{x}$ *in the common sorts (i.e. universes) of* $\mathfrak{A}$ *and* $f(\mathfrak{A})$ *the following holds.*[59]

$$f(\mathfrak{A}) \models \varphi[\bar{a}] \quad \Leftrightarrow \quad \mathfrak{A} \models Tr_f(\varphi)[\bar{a}].$$

Cf. Fig.4.

Proof: The proposition follows easily by Thm.2.3.2 (first translation theorem) on p.35. In more detail: Assume $f : \mathsf{K} \longrightarrow \mathsf{L}$ is a first-order definable meta-function. Then there is an expansion K^+ of $Rng(f)$ such that K^+ is definable over K without taking reducts. Then, by Thm.2.3.2, there is a translation mapping $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ such that $(\star)$ in Thm.2.3.2 holds. Let $Tr_f := Tr \restriction Fm(\mathsf{L})$. One can check that Tr_f has the desired properties. ∎

The following is a weaker form of definitional equivalence. We will use it e.g. in Thm.**??** (p.**??**).

Esetleg azt hogy K és L definicióssan ekvivalens úgy definiálni, hogy $\mathbf{I}\mathsf{K}$ és $\mathbf{I}\mathsf{L}$ definicióssan ekvivalens.

[56] In the one-sorted case an interpretation $Tr : Fm(\mathsf{L}) \longrightarrow Fm(\mathsf{K})$ is the same thing as a cylindric algebraic homomorphism between the cylindric algebras of formulas $Fm(\mathsf{L})$ and $Fm(\mathsf{K})$. I.e. if we endow $Fm(\mathsf{L})$ with the cylindric algebraic structure (of first-order formulas) and do the same with $Fm(\mathsf{K})$ then the homomorphisms between the two algebras of formulas are typical examples of interpretations.

[57] Translation functions of the type $Tr : Fm(\mathsf{L}) \longrightarrow Fm(\mathsf{K})$ play an important role in the present work. They have two important features: (i) they are meaning preserving, and (ii) they *respect the logical structure* of the languages involved, e.g. $Tr(\neg\varphi) = \neg Tr(\varphi)$ and analogously for the remaining parts of our logic. (We do not discuss property (ii) explicitly, but since it is important we mention that it is discussed in the algebraic logic works e.g. in Andréka et al. [4].) In other words (ii) could be interpreted as saying that our translation mappings are grammatical, i.e. they respect the grammar of the languages involved. Cf. Remark 2.3.6 on p.48.

[58] i.e. to $Voc_0\mathsf{K} \cap Voc_0\mathsf{L}$

[59] We note that the formulas φ and $Tr_f(\varphi)$ have the same free variables (therefore the statement below makes sense).

Definition 2.4.9 (Weak definitional equivalence)
Let K, L be two classes of models. K and L are called *weakly definitionally equivalent*, in symbols

$$\mathsf{K} \equiv^w_\Delta \mathsf{L},$$

iff there are first-order definable meta-functions

$$f : \mathsf{K} \longrightarrow \mathsf{L} \quad \text{and} \quad g : \mathsf{L} \longrightarrow \mathsf{K}$$

such that for any $\mathfrak{M} \in \mathsf{K}$ and $\mathfrak{G} \in \mathsf{L}$, (i) and (ii) below hold.

(i) $(f \circ g)(\mathfrak{M}) \cong \mathfrak{M}$ and $(g \circ f)(\mathfrak{G}) \cong \mathfrak{G}$, and

(ii) moreover there is an isomorphism between the two structures $\mathfrak{M}$ and $(f \circ g)(\mathfrak{M})$ which is the identity map on the reduct $\mathfrak{M} \restriction (Voc\mathsf{K} \cap Voc\mathsf{L})$ [60] of $\mathfrak{M}$. Similarly for structures $\mathfrak{G}$ and $(g \circ f)(\mathfrak{G})$.

$\lhd$

Intuitively, K and L are weakly definitionally equivalent iff they are definable over each other and the first-order definable meta-functions induced by these definitions are inverses of each other up to isomorphism.

PROPOSITION 2.4.10 *Assume K, L are two classes of models. Then*

$$\mathsf{K} \equiv_\Delta \mathsf{L} \quad \Rightarrow \quad \mathsf{K} \equiv^w_\Delta \mathsf{L},$$

i.e. if K and L are definitionally equivalent then they are also weakly definitionally equivalent.

We omit the **proof**. ∎

In connection with the above proposition we note that the other direction does not hold in general, i.e.

$$\mathsf{K} \equiv^w_\Delta \mathsf{L} \quad \not\Rightarrow \quad \mathsf{K} \equiv_\Delta \mathsf{L}.$$

This (i.e. $\not\Rightarrow$) is so even if we assume that K and L are both axiomatizable, cf. Examples 2.4.13 (p.67) and Thm.**??** (p.**??**).

Examples come at the end of this section.

[60] $Voc\mathsf{K} \cap Voc\mathsf{L}$ is the common part of the vocabularies of K and L.

Remark 2.4.11 Assume that $f : \mathsf{K} \longrightarrow \mathsf{L}$ and $g : \mathsf{L} \longrightarrow \mathsf{K}$ are first-order definable meta-functions as in Def.2.4.9. Then $Rng(f)$ is L up to isomorphism and $Rng(g)$ is K up to isomorphism. Moreover, for every $\mathfrak{A} \in \mathsf{L}$ there is $\mathfrak{A}' \in Rng(f)$ such that there is an isomorphism between the structures $\mathfrak{A}$ and $\mathfrak{A}'$ which is the identity map on the reduct $\mathfrak{A} \restriction (Voc\mathsf{K} \cap Voc\mathsf{L})$ of $\mathfrak{A}$; and the analogous statement holds for every $\mathfrak{B} \in \mathsf{K}$.

$\triangleleft$

The following proposition says that if $\mathsf{K} \equiv^w_\Delta \mathsf{L}$ then the language $Fm(\mathsf{K})$ of K can be translated into the language $Fm(\mathsf{L})$ of L in a meaning preserving way and vice-versa; more precisely these translations work well for the sentences[61] only or more generally for those formulas which contain only such free variables that range over the common sorts of K and L. Moreover these translation mappings are inverses of each other (up to logical equivalence "$\leftrightarrow$"). We note that if in addition we have $\equiv_\Delta$ in place of $\equiv^w_\Delta$ [62] then this nice, meaning preserving translation mapping extends to all formulas, cf. the end of Remark 7 on p.62.

PROPOSITION 2.4.12 *Assume* $\mathsf{K} \equiv^w_\Delta \mathsf{L}$. *Then there are "natural" translation mappings*

$$T_f : Fm(\mathsf{L}) \longrightarrow Fm(\mathsf{K}) \quad \textit{and} \quad T_g : Fm(\mathsf{K}) \longrightarrow Fm(\mathsf{L})$$

such that for every $\varphi(\bar{x}) \in Fm(\mathsf{L})$, $\psi(\bar{y}) \in Fm(\mathsf{K})$ *with all their free variables belonging to common sorts of* K *and* L, $\mathfrak{A} \in \mathsf{L}$ *and* $\mathfrak{B} \in \mathsf{K}$, *and evaluations* $\bar{a}, \bar{b}$ *of the variables* $\bar{x}, \bar{y}$, *respectively, (i)–(iv) below hold, where* f *and* g *are as in Def.2.4.9.*

(i) $f(\mathfrak{B}) \models \varphi[\bar{a}] \;\Leftrightarrow\; \mathfrak{B} \models T_f(\varphi)[\bar{a}]$ *and* $g(\mathfrak{A}) \models \psi[\bar{b}] \;\Leftrightarrow\; \mathfrak{A} \models T_g(\psi)[\bar{b}]$.

(ii) $\mathfrak{A} \models \varphi[\bar{a}] \;\Leftrightarrow\; g(\mathfrak{A}) \models T_f(\varphi)[\bar{a}]$ *and* $\mathfrak{B} \models \psi[\bar{b}] \;\Leftrightarrow\; f(\mathfrak{B}) \models T_g(\psi)[\bar{b}]$.

(iii) $\mathfrak{A} \models \varphi(\bar{x}) \leftrightarrow (T_f \circ T_g)(\varphi)(\bar{x})$ *and* $\mathfrak{B} \models \psi(\bar{y}) \leftrightarrow (T_g \circ T_f)(\psi)(\bar{y})$.

(iv) $\mathsf{L} \models \varphi \;\Leftrightarrow\; \mathsf{K} \models T_f(\varphi)$ *and* $\mathsf{K} \models \psi \;\Leftrightarrow\; \mathsf{L} \models T_g(\psi)$.

Proof: Item (i) of the proposition follows by Prop.2.4.8 above. Items (ii)–(iv) follow by item (i) and Remark 2.4.11. ∎

[61] *Sentence* means closed formula, i.e. formula without free variables.
[62] i.e. $\mathsf{K} \equiv_\Delta \mathsf{L}$

Examples 2.4.13 In all three examples below we state $\mathsf{K} \not\equiv_\Delta \mathsf{L}$ for some classes K, L. In all three examples we can use item (6) on p.57 to prove $\mathsf{K} \not\equiv_\Delta \mathsf{L}$.

1. Let K be the class of two-element algebras without operations. I.e.

$$\mathsf{K} = \{ A \,:\, |A| = 2 \}.$$

   Let L be the class of two-element ordered sets. Important: The sort symbol of K and the sort symbol of L are different. Then

$$\mathsf{K} \equiv_\Delta^w \mathsf{L}, \quad \text{but} \quad \mathsf{K} \not\equiv_\Delta \mathsf{L}.$$

2. Let K_2 be the same as K was in item 1. above. Let K_3 be the class of three element algebras without operations. Let the sort symbols of K_2 and K_3 be different. Then

$$\mathsf{K}_2 \equiv_\Delta^w \mathsf{K}_3, \quad \text{but} \quad \mathsf{K}_2 \not\equiv_\Delta \mathsf{K}_3.$$

3. More sophisticated example, <u>*affine structures*</u>: Let AB be the class of Abelian (i.e. commutative) groups.

   Assume $\mathfrak{G} = \langle G; +, -, 0 \rangle \in \mathsf{AB}$.

   We define the affine relation R_+ on G as follows.

$$R_+(a,b,c,d,e,f) \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad (a-b)+(c-d)=(e-f).$$

   The <u>*affine structure*</u> associated to the group $\mathfrak{G}$ is

$$\mathfrak{A}_{\mathfrak{G}} := \langle G;\, R_+ \rangle.$$

   The class of affine structures is

$$\mathsf{Af} := \{ \mathfrak{A}_{\mathfrak{G}} \,:\, \mathfrak{G} \in \mathsf{AB} \}.$$

   Let the sort symbols of AB and Af be different. Claim:

$$\mathsf{AB} \equiv_\Delta^w \mathsf{Af}, \quad \text{but} \quad \mathsf{AB} \not\equiv_\Delta \mathsf{Af}.$$

   Hint: Definability of Af over AB is trivial. Definability of AB over Af: Let $\langle G;\, R_+ \rangle \in \mathsf{Af}$. We define a new relation eq as follows.

$$\langle a,b \rangle \; eq \; \langle c,d \rangle \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad R_+(a,b,a,a,c,d).$$

Let us notice that eq is an equivalence relation on $G \times G$. Now, let

$$A := G \times G/eq$$

be a new sort. Further

$$\langle a, b\rangle/eq + \langle c, d\rangle/eq = \langle e, f\rangle/eq \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad R_+(a, b, c, d, e, f).$$

Now, defining the rest of the Abelian group $\langle A, +, \ldots\rangle$ over the affine structure $\langle G;\, R_+\rangle$ is left to the reader.

The proof of $\not\equiv_\Delta$ is based on looking at the large number of automorphisms of the affine structure $\langle G;\, R_+\rangle$. We omit the details. (The idea is similar to that of example 1.)

$\lhd$

Remark 2.4.14 (Making $\equiv^w_\Delta$ strong by using parameters)
Consider the applications of $\equiv^w_\Delta$ in items (i), (ii) below.

(i) In Thm.**??** (p.**??**) it is stated that

$$(\text{Fields}) \ \equiv^w_\Delta \ (\textbf{pag}\text{-geometries}).$$

Theorems **??**, **??** are analogous.

(ii) $\mathsf{Mod}(Th) \ \equiv^w_\Delta \ \mathsf{Mog}(TH)$ for certain choices of Th, TH, where the class $\mathsf{Mog}(TH)$ of geometries is defined on p.**??**. We note that this is not proved or even stated in the present work, but elaborating this can be considered as a useful research exercise for the reader.

Now, if in the context (or background) of items (i), (ii) above we replace the notion of definability with <u>parametric definability</u> using finitely many parameters only (in the usual sense cf. p.**??** and p.9, immediately below Remark 2.1.1, or e.g. Hodges [11, pp. 27–28])[63] then we will obtain that the classes in question e.g. $\mathsf{Mod}(Th)$ and $\mathsf{Mog}(TH)$ become definitionally equivalent in this weaker parametric sense. (I.e. they have a single common parametrically definable definitional expansion etc.) More concretely we could add $(n+1)$-many new constants to **pag** geometries such that

$$(\text{Fields}) \ \equiv_\Delta \ (\textbf{pag}\text{-geometries} \ + \ \text{these constants}).$$

[63] Parametric definability is a slightly weaker notion than definability.

Completely analogous improved versions of Theorems **??**, **??** (pp. **??**, **??**) are also true.

Also we could add $n+1$ new constants to $\mathsf{Mog}(TH)$ and a constant (a distinguished observer) to $\mathsf{Mod}(Th)$ yielding

$$(\mathsf{Mod}(Th)\ +\ \text{new constant})\ \equiv_\Delta\ (\mathsf{Mog}(TH)\ +\ \text{new constants}),$$

for certain choices of Th and TH. This works even if we assume **Ax(eqtime)** $\in Th$ (cf. Conjecture **??** on p.**??**).
It is these new auxiliary constants which are called parameters in the theory of parametric definability.

We leave elaborating the details of this parametric direction to the interested reader.

$\lhd$

2.5 An extension of Beth's theorem. Connections between the various definability notions.

Many-sorted definability theory with new sorts (i.e. the notion of implicit and explicit definition) is a generalization of one-sorted definability theory (without new elements) discussed in traditional logic books. This observation leads to several natural questions which we discuss here only tangentially. One of these is the question whether Beth's theorem (about the equivalence of the two notions of definability) generalizes to our present case.

THEOREM 2.5.1 *Assume* $\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$ *is a reduct of* K^+ *such that* K^+ *has only finitely many new sorts. Assume that the language of* K^+ *is countable, and that* K *has a sort with more than one element.*[64] *Then (i) and (ii) below are equivalent.*

(i) K^+ *is implicitly definable over* K *without taking reducts.*

(ii) K^+ *is a definitionally equivalent expansion of* K.

[64] I.e. if $U_1,\dots,U_n$ are the sorts of K, then $\mathsf{K} \models |U_1| > 1 \vee \ldots \vee |U_n| > 1$.

The **proof** uses Gaifman's theorem (cf. Hodges [11, Thm.12.5.8, p.645]), which is about one-sorted structures, together with ideas from Pillay & Shelah [15], and can be found in Andréka-Madarász-Németi [2]. ∎

COROLLARY 2.5.2 (Beth's theorem generalized to defining new sorts) Assume $\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$ is a reduct of K^+ such that K^+ has only finitely many new sorts. Assume that the language of K^+ is countable, and that K has a sort with more than one element. Then (i) and (ii) below are equivalent.

(i) K^+ is implicitly definable over K.

(ii) K^+ is explicitly definable over K.

QUESTION 2.5.3 *Can Theorem 2.5.1 and Corollary 2.5.2 above be generalized for the case when infinitely many new sorts U_i^{new} $(i \in I)$ are allowed? (First one has to generalize the definition of explicit definability. This can be done easily, e.g. we may allow iteration of steps (1), (2.1), (2.2) along an infinite ordinal, taking "unions" of ascending chains of expansions in the limit steps.)* $\lhd$

On the above question: If there are only finitely many old sorts (i.e. in K), then the answer is affirmative. The question is interesting when there are infinitely many old sorts as well as infinitely many new ones.

Connections between the various notions of definability

Figure 5 below shows the connections between the various notions introduced in this sub-section. It also indicates the above outlined connections with some notions used in the literature (relative categoricity, coordinatisability). The connections indicated are fairly easy to show, except for the following proposition (and, of course where Theorem 2.5.1 and Corollary 2.5.2 are indicated).

PROPOSITION 2.5.4 (Hodges [11]) *Assume the hypotheses of Theorem 2.5.1 (which are the same as the hypotheses used in Figure 5). Then "K^+ is implicitly definable over K up to isomorphism" does not imply "K^+ is implicitly definable over K".*

Proof. A 6-element counterexample proving this is given in Hodges [11, Example 2 on p.625]. There two structures are defined, **A** and **B**, with **A** a reduct of **B**. **B** is implicitly definable up to isomorphism over **A** (this follows from the fact that **B**

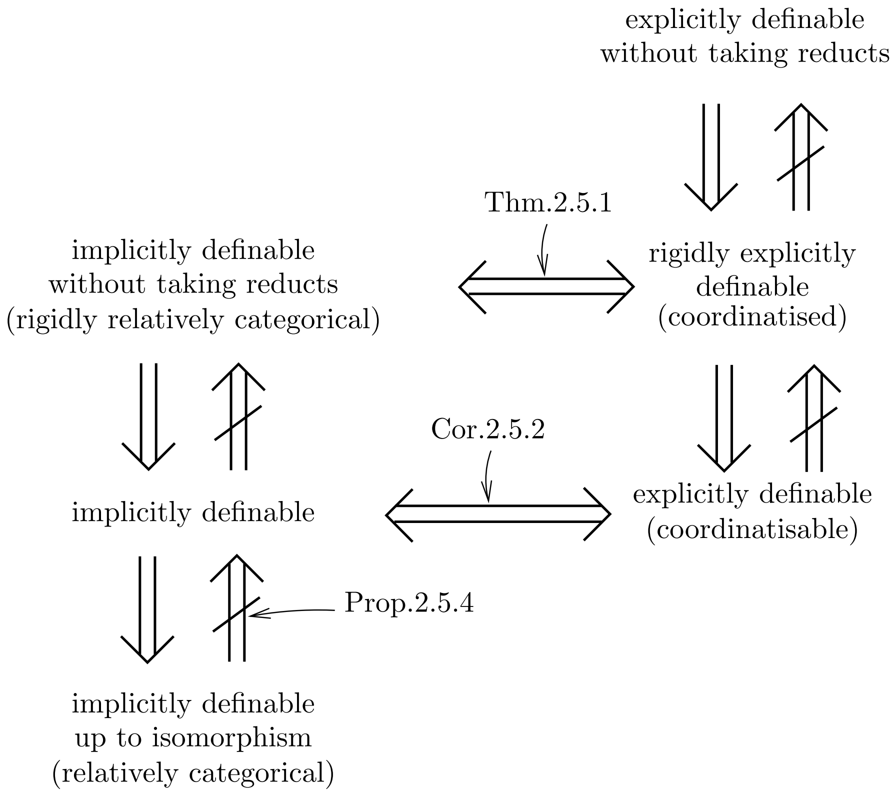

Figure 5: Connections between the various notions of definability. We assume that $\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$ is a reduct of K^+ such that K^+ has only finitely many new sorts. We also assume that the language of K^+ is countable, and that K has a sort with more than one element. On the figure we write "implicitly definable without taking reducts" for "K^+ is implicitly definable over K without taking reducts", and similarly for the other notions. For the implication "implicitly definable" to "implicitly definable up to isomorphisms" we need the extra assumption $\mathsf{K}^+ = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$.

is finite). At the same time, $\mathbf{B}$ is not definable implicitly over $\mathbf{A}$, because $\mathbf{A}$ has an automorphism α of order 2 (i.e. $\alpha \circ \alpha = Id_A$) which cannot be extended to an automorphism β of $\mathbf{B}$ of order 2. Indeed, if $\mathbf{B}$ was implicitly definable over $\mathbf{A}$, then an expansion $\mathbf{B}^+$ of $\mathbf{B}$ would be implicitly definable over $\mathbf{A}$ without taking reducts. Hence the automorphism α would extend to an automorphism β of $\mathbf{B}^+$. Since the identity of $\mathbf{A}$ extends to a unique automorphism of $\mathbf{B}^+$, then $\beta \circ \beta = Id_{B^+}$ should hold. But then $\beta \restriction B$ would be an automorphism of $\mathbf{B}$ of order 2 and extending α. (Cf. Thm.12.5.7 in [11, p.644].) Since $\mathbf{A}$ and $\mathbf{B}$ are finite structures, we can take $\mathsf{K} = \mathbf{I}\{\mathbf{A}\}$ and $\mathsf{K}^+ = \mathbf{I}\{\mathbf{B}\}$, and then the hypotheses of Proposition 2.5.4 hold for K and K^+. This finishes the proof. ∎

3 Definability in one-sorted first-order logic allowing to enlarge the universes of models

We use the notation and definitions in Hodges [11]. Throughout we assume that L and L^+ are first-order languages without function symbols (but possibly with constants) such that $L \subseteq L^+$. One of the symbols in L^+ is a unary relation symbol P. T usually denotes a theory in the language L^+, we do not assume that T is complete. We usually denote models of T with $\mathfrak{B}$. If $\mathfrak{B}$ is a model of T, then $\mathfrak{B}_P$ denotes the reduct of $\mathfrak{B}$ to the language L restricted to the interpretation of P in $\mathfrak{B}$, i.e. the universe of $\mathfrak{B}_P$ is $P^{\mathfrak{B}}$ and the interpretations of the symbols of L in $\mathfrak{B}_P$ are those in $\mathfrak{B}$ restricted to $P^{\mathfrak{B}}$. For $\psi \in L^+$, ψ^P denotes the formula ψ relativised to P, further, the notation ψ^ϕ, is meaningful for any formula ϕ. $\varphi^{\mathfrak{A}} = \{\bar{a} : \mathfrak{A} \models \varphi(\bar{a})\}$ denotes the relation defined by φ in $\mathfrak{A}$. Thus $\varphi^{\mathfrak{A}}$ is an n-ary relation on A if φ has n free variables. We also use some notation borrowed from [1, §6.2], e.g. $\mathfrak{B} \restriction L$, $L(\mathfrak{A})$, $Voc(L)$, $Voc(\mathfrak{A})$.

3.1 Explicitly defining a set disjoint from the universe

Let $\mathfrak{A}$ be a model. We say that a formula $\varphi(\bar{x}, \bar{y})$ *defines an equivalence relation in $\mathfrak{A}$* if $\mathfrak{A} \models [(\varphi(\bar{x}, \bar{y}) \wedge \varphi(\bar{z}, \bar{y})) \rightarrow (\varphi(\bar{x}, \bar{x}) \wedge \varphi(\bar{y}, \bar{x}) \wedge \varphi(\bar{x}, \bar{z}))]$. (We assume, of course, that the length of the seqeunces $\bar{x}$ and $\bar{y}$ of variables is the same.) This is the same as saying that $\varphi^{\mathfrak{A}}$ is an equivalence relation on $\{\bar{a} : \langle \bar{a}, \bar{a}\rangle \in \varphi^{\mathfrak{A}}\}$. In the following we will write $\varphi(\bar{a}, \bar{a})$ in place of $\langle \bar{a}, \bar{a}\rangle \in \varphi^{\mathfrak{A}}$.

If R is an equivalence relation on U, then $U/R \stackrel{\text{def}}{=} \{u/R : u \in U\}$ where $u/R \stackrel{\text{def}}{=} \{v \in U : \langle u, v\rangle \in R\}$.

Definition 3.1.1 (definable extension of a model) Let $\mathfrak{A}$ be a model and assume that $\varphi(\bar{x}, \bar{y}) \in L(\mathfrak{A})$ defines an equivalence relation in $\mathfrak{A}$. For simplicity, assume first $A \cap \mathcal{P}({}^nA) = \emptyset$ where n is the length of the sequence $\bar{x}$ of variables.

We define a new model $\mathfrak{A}(\varphi)$ as follows. Let

$$U \stackrel{\text{def}}{=} \{\bar{a} : \varphi(\bar{a}, \bar{a})\}/\varphi^{\mathfrak{A}}, \qquad \text{and}$$

$$\varepsilon \stackrel{\text{def}}{=} \{\langle a_1, \ldots, a_n, \bar{a}/\varphi^{\mathfrak{A}}\rangle : \varphi(\bar{a}, \bar{a})\}.$$

The universe of $\mathfrak{A}(\varphi)$ will be $A \cup U$, the language of $\mathfrak{A}(\varphi)$ will be that of $\mathfrak{A}$ expanded with a new $n+1$-ary relation symbol, say E, which will denote ε in $\mathfrak{A}(\varphi)$, and the old relation symbols will denote in $\mathfrak{A}(\varphi)$ the same what they denoted in $\mathfrak{A}$. Thus

$$\mathfrak{A}(\varphi) \stackrel{\text{def}}{=} \langle A \cup U, \varepsilon, R^{\mathfrak{A}}\rangle_{R \in Voc(\mathfrak{A})}.$$

Intuitively, $\mathfrak{A}(\varphi)$ is the model $\mathfrak{A}$ such that we enlarged the universe with U and we added the relation ε to fix the connection between U and A. Note that U is disjoint from A by our assumption.

We call $\mathfrak{A}(\varphi)$ <u>*the extension of* $\mathfrak{A}$ *with new elements defined by* φ</u>.

We want to use this notion up to isomorphism. Therefore, instead of requiring $A \cap \mathcal{P}({}^nA) = \emptyset$, we define the universe of $\mathfrak{A}(\varphi)$ to be the *disjoint* union $A \mathbin{\dot\cup} U$, which we define to be $A \cup (U \smallsetminus A) \cup (U \cap A) \times \{A\}$. We then modify ε accordingly.

We say that $\mathfrak{B}$ is <u>*a definable extension of* $\mathfrak{A}$</u> (with the new elements defined by φ) if $\mathfrak{B}$ is isomorphic to $\mathfrak{A}(\varphi)$ via an isomorphism that is the identity on A.

$\lhd$

The relation ε is basically the "element-of" relation "$\{\langle \bar{a}, \bar{a}/\varphi^{\mathfrak{A}}\rangle : \varphi(\bar{a}, \bar{a})\}$". ε is the relation that tells us the connection between U and A. Looking at ε in another way, $\varepsilon(\bar{x}, z)$ means that $z = \bar{x}/\varphi^{\mathfrak{A}}$. Thus ε "gives names" to the elements of U. So in some sense ε is the "definition" of U.

Recall that we say that $\mathfrak{B}$ is an <u>*extension*</u> of $\mathfrak{A}$ if $\mathfrak{A}$ is a submodel of $\mathfrak{B}$, and we say that $\mathfrak{B}$ is an <u>*expansion*</u> of $\mathfrak{A}$ if $\mathfrak{A}$ is a reduct of $\mathfrak{B}$. Thus we get an extension by adding new elements to $\mathfrak{A}$, while we get an expansion by adding new relations to $\mathfrak{A}$. Thus, strictly speaking, $\mathfrak{A}(\varphi)$ is an expansion of an extension of $\mathfrak{A}$.

Examples If we choose φ to be $x = x \wedge y = y$, then $U = \{\{\langle a, b\rangle\} : a, b \in A\}$ and $\varepsilon = \{\langle a, b, \{\langle a, b\rangle\}\rangle : a, b \in A\}$. Thus this case is adding $A \times A$ together with the projection functions to $\mathfrak{A}$, basically. If we choose $n = 1$, then $U = \{\{a\} : a \in \varphi^{\mathfrak{A}}\}$, and $\varepsilon = \{\langle a, \{a\}\rangle : a \in \varphi^{\mathfrak{A}}\}$. So in this case defining U is essentially adding a disjoint copy of a definable subset of $\mathfrak{A}$ to A. To see the use of factoring, choose e.g. $\varphi(x_1, x_2, y_1, y_2)$ to be $x_1 = y_1 \wedge (x_1 = x_2 \leftrightarrow y_1 = y_2)$. Then the blocks of $\varphi^{\mathfrak{A}}$ are $\{\langle a, a\rangle\}$ and $\{\langle a, b\rangle : b \in A, b \neq a\} = \{a\} \times (A \smallsetminus \{a\})$ for $a \in A$. Thus the new model we get by using φ is adding two copies of A to $\mathfrak{A}$, see Figure **??**.

Figure example-fig.

In fact, by using the above features together, we can define an arbitrarily large finite set U in any model $\mathfrak{A}$ which has at least two elements. The idea is the following. Assume we want to add four new elements to $\mathfrak{A}$. Let $\varphi(x_1, x_2, x_3, y_1, y_2, y_3)$ be the following formula:

$$\bigwedge\{x_i = x_j \leftrightarrow y_i = y_j : 1 \leq i, j \leq 3\} \wedge \bigvee\{x_i = x_j : 1 \leq i, j \leq 3\}.$$

Then $\varphi^{\mathfrak{A}}$ is an equivalence relation on ${}^3A$ whenever $|A| \geq 2$ with the following four equivalence classes: $u_1 \stackrel{\text{def}}{=} \{\langle a, a, a\rangle : a \in A\}$, $u_2 \stackrel{\text{def}}{=} \{\langle a, a, b\rangle : a, b \in A, a \neq b\}$, $u_3 \stackrel{\text{def}}{=} \{\langle a, b, a\rangle : a, b \in A, a \neq b\}$, $u_4 \stackrel{\text{def}}{=} \{\langle b, a, a\rangle : a, b \in A, a \neq b\}$. Then $\varepsilon(a, a, a, u)$ implies $u = u_1$, $\varepsilon(a, a, b, u) \wedge a \neq b$ implies $u = u_2$ etc. See Figure **??**. The formula ε tells us the way the elements of U are coded up in $\mathfrak{A}$.

Figure coding-fig

We say that a relation $R \subseteq {}^nA$ is *definable* in $\mathfrak{A}$ if there is a formula $\varphi \in L(\mathfrak{A})$ such that $R = \varphi^{\mathfrak{A}}$.

Definition 3.1.2 (explicit definability)

(i) We say that $\mathfrak{B}$ is <u>*explicitly definable over P*</u> if there is a formula $\varepsilon \in L^+$ such that $(\mathfrak{B} \restriction L, \varepsilon^{\mathfrak{B}})$ is a definable extension of $\mathfrak{B}_P$, and every new relation of $\mathfrak{B}$ is definable in $(\mathfrak{B} \restriction L, \varepsilon^{\mathfrak{B}})$.

(ii) Assume that $\mathfrak{B}$ is explicitly definable over P. Let $\varphi \in L$, $\varepsilon \in L^+$ be such that $\mathfrak{B}^e \stackrel{\text{def}}{=} (\mathfrak{B} \restriction L, \varepsilon^{\mathfrak{B}})$ is isomorphic to $\mathfrak{B}_P(\varphi)$ via an isomorphism which is identity on $P^{\mathfrak{B}}$. Further, for every new relation symbol S of $\mathfrak{B}$ let $\varphi_S \in L^E$ be such that φ_S defines $S^{\mathfrak{B}}$ in $\mathfrak{B}^e$. We say that an <u>*explicit definition*</u> of $\mathfrak{B}$ over P is the set of the following formulas:

$\exists\bar{x}\varepsilon(\bar{x}, y) \leftrightarrow \neg P(y)$

$(\varepsilon(\bar{x}, y) \wedge \varepsilon(\bar{x}, z)) \rightarrow y = z$

$\exists y(\varepsilon(\bar{x}, y) \wedge \varepsilon(\bar{z}, y)) \leftrightarrow \varphi^P(\bar{x}, \bar{z})$

$\forall\bar{x}\forall z(E(\bar{x}, z) \leftrightarrow \varepsilon(\bar{x}, z)) \rightarrow [S(y_1, \ldots, y_m) \leftrightarrow \varphi_S(y_1, \ldots, y_m)]$, for all new relation symbol S.

(iii) We say that <u>*T is explicitly definable over P*</u> if every model $\mathfrak{B}$ of T is explicitly definable over P, and with a uniform explicit definition. I.e. T is explicitly definable over P if there is an explicit definition which defines $\mathfrak{B}$ over $\mathfrak{B}_P$ for every model $\mathfrak{B}$ of T.

$\triangleleft$

Definition 3.1.3 (Explicit definability in P)

(i) We say that $\mathfrak{B}$ is <u>*explicitly definable*</u> in $\mathfrak{B}_P$ if (1)-(2) below hold.

(1) There are formulas $\varphi \in L$ and $\varepsilon \in L^+$ such that the following formulas are valid in $\mathfrak{B}$:

$\exists\bar{x}\varepsilon(\bar{x}, y) \leftrightarrow \neg P(y)$

$(\varepsilon(\bar{x}, y) \wedge \varepsilon(\bar{x}, z)) \rightarrow y = z$

$\exists y(\varepsilon(\bar{x}, y) \wedge \varepsilon(\bar{z}, y)) \leftrightarrow \varphi^P(\bar{x}, \bar{z})$

Intuitively, ε is an interpretation of the new elements as "sets of n-tuples" of old elements. See Figure **??**.

(2) Every relation S in L^+ is defined by old relation symbols with the help of ε, i.e., let S be an n-ary relation symbol in L^+, and let $i_1, \ldots, i_k < n$, let $H = \{i_1, \ldots, i_k\}$, $n \setminus H = \{j_1, \ldots, j_\ell\}$. Then we require that there is a formula $\varphi_S \in L$ such that $\mathfrak{B}$ models the following formula:

$$S(\bar{x}) \wedge \bigwedge_{i \in H} P(x_i) \wedge \bigwedge_{j \notin H} \neg P(x_j) \leftrightarrow$$
$$\exists \bar{y}_1, \ldots, \bar{y}_\ell(\varepsilon(\bar{y}_1, x_{j_1}) \wedge \ldots \wedge \varepsilon(\bar{y}_\ell, x_{j_\ell}) \wedge \varphi_S^P(x_{i_1}, \ldots, x_{i_k}, \bar{y}_1, \ldots, \bar{y}_\ell).$$

We say that the collection Δ of displayed formulas in (1) and (2) above is an *explicit definition* of $\mathfrak{B}$ in $\mathfrak{B}_P$.

(ii) We say that *T is explicitly definable in P* if every model $\mathfrak{B}$ of T is explicitly definable in $\mathfrak{B}_P$ such that there is a *uniform* explicit definition for all models of T.

$\lhd$

We note that the above notion of explicit definability is a very concrete one. If we choose φ such that $\varphi^{\mathfrak{A}} = \emptyset$ (this corresponds to the case that we did not define new elements, or that $P = B$), then we get back usual explicit definability. If $\mathfrak{B}$ is explicitly definable in $\mathfrak{B}_P$ with φ and ε as in Definition 3.1.3 above, then $B \setminus B_P$ is explicitly definable in $\mathfrak{B}_P$ by φ in the sense of the definition given in the Motivation part, up to isomorphism, and ε is the same as the one in the Motivation part. Thus ε "gives names" to the new elements of the universe.

We will show, in Theorem 3.2.3 below, that this explicit definability is the same as "coordinatised" in Hodges [11]. But first we define some equivalent forms of this notion.

3.2 Equivalent forms of explicit definability.

Recall from Hodges [11, p.640] that T is said to have the *(uniform) reduction property* over P if to any formula $\psi(\bar{x}) \in L^+$ there is a formula $\varphi(\bar{x}) \in L$ such that $T \models \forall \bar{x}(P(x_1) \wedge \ldots \wedge P(x_n) \to [\psi(\bar{x}) \leftrightarrow \varphi^P(\bar{x})])$.

Definition 3.2.1 **(T has the total reduction property over P)**

We say that <u>T *has the total (uniform) reduction property over* P</u> if there is a formula $\varepsilon \in L^+$ such that $T \models \forall y(\neg P(y) \to \exists \bar{x} \varepsilon(\bar{x}, y))$ and for every $\psi(\bar{x}) \in L^+$ and for any $H \subseteq \{x_1, \ldots, x_n\}$, if $H = \{x_{i_1}, \ldots, x_{i_k}\}$ and $\{x_1, \ldots, x_n\} \setminus H = \{x_{j_1}, \ldots, x_{j_\ell}\}$, then there is a $\varphi(\bar{x}) \in L$ such that

$$T \models (P(x_{i_1}) \wedge \ldots \wedge P(x_{i_k}) \wedge \neg P(x_{j_1}) \wedge \ldots \wedge \neg P(x_{j_\ell}) \wedge$$
$$\varepsilon(\bar{y}_1, x_{j_1}) \wedge \ldots \wedge \varepsilon(\bar{y}_\ell, x_{j_\ell})) \longrightarrow [\psi(\bar{x}) \leftrightarrow \varphi^P(x_{i_1}, \ldots, x_{i_k}, \bar{y}_1, \ldots, \bar{y}_\ell)].$$

In this case we say that φ is a <u>*reduction*</u> of ψ w.r.t. H. $\lhd$

We note that "total reduction property" implies "reduction property" as defined in Hodges [11]. Also it is easy to see that if T has the total reduction property, then types over P are definable for T, and are isolated. (For the notions of types are definable over T and types are isolated see Hodges [11].)

Recall from Hodges [11, p.212] that an *interpretation of* $\mathfrak{B}$ *in* $\mathfrak{A}$ is a tuple $\langle \delta, f, \varphi_S : S \in At^+ \rangle$ where At^+ denotes the set of atomic formulas of the language of $\mathfrak{B}$, δ and φ_S are formulas in the language of $\mathfrak{A}$, f maps $\delta^{\mathfrak{A}}$ onto B such that for all atomic formula $S(x_1, \ldots, x_m)$ of $\mathfrak{B}$

$$\mathfrak{B} \models S(f(\bar{a}_1), \ldots, f(\bar{a}_m)) \quad \Leftrightarrow \quad \mathfrak{A} \models \varphi_S(\bar{a}_1, \ldots, \bar{a}_m).$$

Here f is called the coordinate map.

Definition 3.2.2 (**T is interpretable in P**)

(i) We say that <u>$\mathfrak{B}$ *is interpretable in* P</u> if there is an interpretation of $\mathfrak{B}$ in $\mathfrak{B}_P$ such that the coordinate map of the interpretation is definable in $\mathfrak{B}$ and moreover, the restriction of the coordinate map to $\mathfrak{B}_P$ is definable in $\mathfrak{B}_P$. In more detail, this means that there is an interpretation $\langle \delta(x_1, \ldots, x_n), f, \varphi_S : S \in At^+ \rangle$ of $\mathfrak{B}$ in $\mathfrak{A} \stackrel{\text{def}}{=} \mathfrak{B}_P$ such that $f : \delta^{\mathfrak{A}} \longrightarrow B$ is definable by a formula $\pi(x_1, \ldots, x_n, z) \in L^+$ and $f \cap {}^{n+1}A$ is definable by a formula $\rho(x_1, \ldots, x_n, z) \in L$. This means that $\pi \in L^+$ and $\delta, \rho, \varphi_S \in L$ are such that the following formulas are valid in $\mathfrak{B}$:

$\delta^P(\bar{x}) \leftrightarrow \exists z \pi(\bar{x}, z)$	(the domain of $\pi^{\mathfrak{B}}$ is $\delta^{\mathfrak{A}}$)
$\pi(\bar{x}, z) \wedge \pi(\bar{x}, w) \to z = w$	($\pi^{\mathfrak{B}}$ is a function)
$\forall z \exists \bar{x} \pi(\bar{x}, z)$	($\pi^{\mathfrak{B}}$ is onto B)
$\rho^P(\bar{x}, z) \leftrightarrow (\pi(\bar{x}, z) \wedge P(z))$	(ρ defines the restriction of $\pi^{\mathfrak{B}}$ to A)

$$\exists y_1 \ldots y_m (S(y_1, \ldots, y_m) \wedge \bigwedge_i \pi(\bar{x}_i, y_i)) \leftrightarrow \varphi_S^P(\bar{x}_1, \ldots, \bar{x}_m).$$

(ii) We say that <u>*T is interpretable in P*</u> if every model $\mathfrak{B}$ of T is interpretable in P and these interpretations are the same. I.e. T is interpretable in P if there are $\delta, \pi, \rho, \varphi_S$ as above such that the set of displayed formulas is valid in T. In this case we say that $\langle \delta, \pi, \rho, \varphi_S : S \in At^+ \rangle$ is an <u>*interpretation*</u> of T in P.

$\lhd$

Recall from Hodges [11, p.627,644], or from Hodges, Hodkinson, Macpherson [12], that $\mathfrak{B}$ is said to be coordinatised over P if $Th(\mathfrak{B})$ has the reduction property and every element of $B \setminus B_P$ is in the definable closure of B_P, i.e. for all $b \in B \setminus B_P$ there is a formula $\psi(\bar{x}, z) \in L^+$ such that $\mathfrak{B} \models \exists \bar{x}(P(x_1) \wedge \ldots \wedge P(x_n) \wedge \psi(\bar{x}, b) \wedge \forall z(\psi(\bar{x}, z) \rightarrow z = b))$. A theory T is said to be coordinatised over P if T has the uniform reduction property and every model $\mathfrak{B}$ of T is coordinatised over P.

The next theorem says that our notion of explicit definability coincides with the notion of "coordinatised" which further coincides with our natural strengthenings of the notions "interpretable" and "reduced".

THEOREM 3.2.3 *The following statements (i)-(iii) are equivalent and they are implied by (iv). If we assume that $T \models |P| \geq 2$, i.e. $T \models \exists y, z(P(y) \wedge P(z) \wedge y \neq z)$, then all four statements (i)-(iv) are equivalent.*

(i) *T is explicitly definable in P.*

(ii) *T is coordinatised over P.*

(iii) *T has the total reduction property.*

(iv) *T is interpretable in P.*

Proof.

<u>Proof of (ii) $\Rightarrow$ (i):</u> Assume that T is coordinatised over P. Then in each model of T each element not in P is "generated" by some formula. (We say that x is generated by γ if there is $\bar{y} \in {}^nP$ such that $\gamma(\bar{y}, x)$ and $\forall z(\gamma(\bar{y}, z) \rightarrow z = x)$.) First we show that there is a finite set Γ of formulas such that in each model of T each element not in P is "generated" by one of the formulas in Γ, i.e.:[65]

[65]In this proof we often write parts of formulas informally, e.g. we write $x \notin P$ in place $\neg P(x)$ and we write $\bar{y} \in {}^nP$ in place of $P(y_1) \wedge \ldots \wedge P(y_n)$ where $\bar{y} = \langle y_1, \ldots, y_n \rangle$.

(1) $T \models \forall x \notin P \exists \bar{y} \in {}^nP(\bigvee\{\gamma(\bar{y}, x) \wedge \forall z(\gamma(\bar{y}, z) \to z = x) : \gamma \in \Gamma\})$.

Indeed, assume that there is no such Γ. Then for all finite $\Gamma \subseteq L^+$ there are $\mathfrak{B}_\Gamma \models T$ and $b_\Gamma \in B_\Gamma$ such that

$$\mathfrak{B}_\Gamma \models \forall \bar{y} \in {}^nP(\bigwedge\{\neg\gamma(\bar{y}, b_\Gamma) \vee (\exists x \neq b_\Gamma)\gamma(\bar{y}, x) : \gamma \in \Gamma\}).$$

Let I be the set of all finite subsets of L^+, let F be an ultrafilter on I such that $\{\Gamma \in I : \varphi \in \Gamma\} \in F$ for all $\varphi \in L^+$ and let $\mathfrak{B} \stackrel{\text{def}}{=} P\mathfrak{B}_\Gamma/F$, $b \stackrel{\text{def}}{=} Pb_\Gamma/F$.Then $\mathfrak{B} \models T$ and $b \in B$. Let $\varphi \in L^+$ and $X = \{\Gamma \in I : \varphi \in \Gamma\} \in F$. Then for all $\Gamma \in X$

$$\mathfrak{B}_\Gamma \models \forall \bar{y} \in {}^nP(\neg\varphi(\bar{y}, b_\Gamma) \vee (\exists x \neq b_\Gamma)\gamma(\bar{y}, x),$$

thus this same formula is true in $\mathfrak{B}$ for b in place of b_Γ. This contradicts $\mathfrak{B} \models T$, $b \notin P^{\mathfrak{B}}$ and that b is in the definable closure of $P^{\mathfrak{B}}$. By this, (1) has been proved.

Let Γ be a finite set of formulas that satisfies (1). Next we show that this Γ can be coded into one formula ε that generates each element not in P. Let $\Gamma \stackrel{\text{def}}{=} \{\gamma_1, \ldots, \gamma_k\}$. We may assume that the free variables of each element of Γ are $x_1, \ldots, x_n$. Let $\bar{x}_1, \ldots, \bar{x}_k$ be sequences of variables of length n such that the sequence $\bar{x}_1 \ldots \bar{x}_k$ is repetition-free. Define for all $1 \leq i \leq k$

$$\delta_i(\bar{x}, y) \stackrel{\text{def}}{=} \gamma_i(\bar{x}, y) \wedge \forall z(\gamma_i(\bar{x}, z) \to z = y),$$

$$\varepsilon_i(\bar{x}_1, \ldots, \bar{x}_k, y) \stackrel{\text{def}}{=} \neg\delta_1(\bar{x}_1, y) \wedge \ldots \wedge \neg\delta_{i-1}(\bar{x}_{i-1}, y) \wedge \delta_i(\bar{x}_i, y) \wedge \bar{x}_i = \ldots = \bar{x}_k,$$

$$\varepsilon(\bar{x}_1, \ldots, \bar{x}_k, y) \stackrel{\text{def}}{=} \bigvee\{\varepsilon_i(\bar{x}_1, \ldots, \bar{x}_k, y) : 1 \leq i \leq k\} \wedge \bar{x}_1, \ldots, \bar{x}_k \in {}^nP \wedge \neg P(y).$$

Now we show that ε as defined above satisfies the first two formulas in Definition 3.1.3(i)(1). Let $\bar{x}$ denote the sequence $\bar{x}_1, \ldots, \bar{x}_k$ of variables.

Proof of $\varepsilon(\bar{x}, y) \wedge \varepsilon(\bar{x}, z) \to y = z$: If the antecedent of the formula holds, then there are i, j such that $\varepsilon_i(\bar{x}, y) \wedge \varepsilon_j(\bar{x}, y)$. If $i = j$ then $\delta_i(\bar{x}_i, y) \wedge \delta_i(\bar{x}_i, z)$ holds, which implies $y = z$. Assume now $i \neq j$, say $i < j$. Then $\varepsilon_i(\bar{x}, y)$ implies $\delta_i(\bar{x}_i, y) \wedge \bar{x}_i = \bar{x}_j$ and $\varepsilon_j(\bar{x}, y)$ implies $\neg\delta_i(\bar{x}_j, y)$, a contradiction.

Proof of $\exists \bar{x}\varepsilon(\bar{x}, y) \leftrightarrow \neg P(y)$: The proof of $\exists \bar{x}(\varepsilon(\bar{x}, y) \to \neg P(y)$ follows immediately from the definition of ε. To show $\neg P(y) \to \exists \bar{x}\varepsilon(\bar{x}, y)$, let $y \notin P$. By (1) then there are i and $\bar{x}_i \in {}^nP$ such that $\gamma_i(\bar{x}_i, y) \wedge \forall(\gamma_i(\bar{x}_i, z) \to z - y)$, i.e. $\delta_i(\bar{x}_i, y)$. Let i be the smallest such. Then $\varepsilon(\bar{x}_i, \ldots, \bar{x}_i, y)$ holds, and so $\varepsilon(\bar{x}_i, \ldots, \bar{x}_i, y)$ also holds.

We have shown that ε satisfies the first two formulas required in the definition of explicit definability. Since T is coordinatised over P, T satisfies the reduction property. By the reduction property there is $\varphi \in L$ such that $T \models \forall \bar{x}, \bar{z} \in {}^m P[\exists y(\varepsilon(\bar{x}, y) \wedge \varepsilon(\bar{z}, y)) \leftrightarrow \varphi(\bar{x}, \bar{z})]$. Thus the third formula in Def. 3.1.3(i)(1) also holds.

To show that (i)(2) of Def.3.1.3 is satisfied, let S, $H = \{i_1, \ldots, i_k\}$ and $n \setminus H = \{j_1, \ldots, j_\ell\}$ be as in (i)(2) of Def.3.1.3. For simplifying the notation, let us assume that $\bar{x} = \langle x_i, x_j \rangle$ and $H = \{i\}$. By the reduction property there is $\varphi_S \in L$ such that $T \models (\exists x_j)[S(x_i, x_j) \wedge P(x_i) \wedge \neg P(x_j) \wedge \varepsilon(\bar{y}_j, x_j)] \leftrightarrow \varphi_S^P(x_i, \bar{y}_j)$. By the properties of ε we then have that

$S(x_i, x_j) \wedge P(x_i) \wedge \neg P(x_j)$ is equivalent to

$(\exists \bar{y}_j)[S(x_i, x_j) \wedge P(x_i) \wedge \neg P(x_j) \wedge \varepsilon(\bar{y}_j, x_j)]$, which further is equivalent to

$(\exists \bar{y}_j)[\varepsilon(\bar{y}_j, x_j) \wedge (\exists x_j)(S(x_i, x_j) \wedge P(x_i) \wedge \neg P(x_j) \wedge \varepsilon(\bar{y}_j, x_j)]$, which is equivalent to

$(\exists \bar{y}_j)[\varepsilon(\bar{y}_j, x_j) \wedge \varphi_S^P(x_i, \bar{y}_j)]$.

By this, (ii) $\Rightarrow$ (i) has been proved.

Proof of (i) $\Rightarrow$ (iii): The proof goes by induction. Assume that T is explicitly definable over P, with formulas $\varepsilon, \delta, \varphi_S$ as in Def.3.1.3. Then the following are not difficult to check:

For a reduction of $S(\bar{x})$ we can take $\exists \bar{w}_1 \ldots \bar{w}_\ell(\varphi(\bar{w}_1, \bar{y}_1) \wedge \ldots \wedge \varphi(\bar{w}_\ell, \bar{y}_\ell) \wedge \varphi_S(x_{i_1} \ldots x_{i_k}, \bar{w}_1 \ldots \bar{w}_\ell)$, see Def.3.2.1. This is so because $\exists z(\varepsilon(\bar{y}, z) \wedge \varepsilon(\bar{w}, z)) \leftrightarrow \varphi(\bar{w}, \bar{y})$.

For a reduction of $x = z \wedge \neg P(x)$ we can take $\varphi(\bar{y}, \bar{w})$, and for a reduction of $x = z \wedge P(x)$ we can take $x = z$.

By the definition of a reduction, it is immediate that if φ, ψ are reductions of η and δ respectively, then $\varphi \wedge \psi$ and $\neg\varphi$ are reductions of $\eta \wedge \delta$ and $\neg\eta$ respectively.

Assume that $\psi(x, \bar{x})$ has a reduction. To give a reduction for $\exists x \psi(x, \bar{x})$ we will use that the latter formula is equivalent to $\exists x(P(x) \wedge \psi(x, \bar{x}) \vee \exists x(\neg P(x) \wedge \psi(x, \bar{x}))$. For notational convenience, let $\alpha(\bar{x}\bar{y})$ denote $P(x_{i_1}) \wedge \ldots \wedge P(x_{i_n}) \wedge \neg P(x_{j_1}) \wedge \ldots \wedge \neg P(x_{j_\ell}) \wedge \varepsilon(\bar{y}_1, x_{i_1}) \wedge \ldots \wedge \varepsilon(\bar{y}_\ell, x_{j_\ell})$. By our induction hypothesis, there are $\psi_1, \psi_2 \in L$ such that

$\alpha(\bar{x}, \bar{y}) \wedge P(x) \rightarrow [\psi(x, \bar{x}) \leftrightarrow \psi_1(x, \bar{x}, \bar{y})]$ and

$$\alpha(x,\bar{y}) \wedge \neg P(x) \wedge \varepsilon(\bar{w},x) \rightarrow [\psi(x,\bar{x}) \leftrightarrow \psi_2(\bar{w}\bar{x}\bar{y})].$$

Using the above and properties of ε we get

$$\alpha(\bar{x}\bar{y}) \rightarrow [(\exists x)\psi(x,\bar{x}) \leftrightarrow \exists x(P(x) \wedge \psi_1(x,\bar{x},\bar{y})) \vee \exists \bar{w}(\psi_2(\bar{w},\bar{x},\bar{w}))].$$

Thus, for a reduction of $\exists x \psi(x,\bar{x})$ we can take $\exists x \psi_1(x,\bar{x},\bar{y}) \vee \exists \bar{w} \psi_2(\bar{w},\bar{x},\bar{y})$.

By the above, (i) $\Rightarrow$ (iii) has been proved.

<u>Proof of (iii) $\Rightarrow$ (ii):</u> Assume that T has the total reduction property over P with $\varepsilon \in L^+$. Then T has the reduction property. We are going to show that every element is in the definable closure of P. Let $\eta \stackrel{\text{def}}{=} \varepsilon(\bar{x},y) \wedge \neg P(y)$.

Let $\varphi^=$ be the reduction of the formula $y = y$. Then $\neg P(y) \wedge \varepsilon(\bar{x},y) \rightarrow [y = y \leftrightarrow \varphi^{=P}(\bar{x})]$, i.e. $\eta(\bar{x},y) \rightarrow (P(x_1) \wedge \ldots \wedge P(x_n))$.

Let $\varphi^{\neq}$ be the reduction of the formula $y \neq z$. Then $\neg P(y) \wedge \neg P(z) \wedge \varepsilon(\bar{x},y) \wedge \varepsilon(\bar{w},z) \rightarrow [y \neq z \leftrightarrow \varphi^{\neq P}(\bar{x},\bar{w})]$. Then $\neg P(y) \wedge \varepsilon(\bar{x},y) \rightarrow [y \neq y \leftrightarrow \varphi^{\neq P}(\bar{x},\bar{x})]$, i.e. $\eta(\bar{x},y) \rightarrow \neg\varphi^{\neq P}(\bar{x},\bar{x})$.

Now, $\eta(\bar{x},y) \wedge \eta(\bar{x},z) \rightarrow [y \neq z \leftrightarrow \varphi^{\neq P}(\bar{x},\bar{x})]$, so $\eta(\bar{x},y) \wedge \eta(\bar{x},z) \rightarrow y = z$. Thus $\eta(\bar{x},y)$ defines a function whose domain is in P. By $T \models \forall y(\neg P(y) \rightarrow \exists \bar{x} \varepsilon(\bar{x},y))$ then we obtain that every $y \notin P$ is defined over P, namely uniformly by the formula $\eta(\bar{x},y)$. (iii) $\Rightarrow$ (ii) has been proved.

<u>Proof of (iii) $\Rightarrow$ (iv):</u> Here we need $T \models |P| \geq 2$. Assume that T has the total reduction property with $\varepsilon(x_1,\ldots,x_n) \in L^+$. Let $\bar{x} \stackrel{\text{def}}{=} \langle x_1,\ldots,x_n \rangle$. Define

$$\pi(\bar{x},y,z) \stackrel{\text{def}}{=} P(x_1,) \wedge \ldots \wedge P(x_n) \wedge P(y) \wedge (x_1 = y \rightarrow \varepsilon(\bar{x},z)) \wedge (x_1 \neq y \rightarrow z = y).$$

It is not difficult to check that π defines a surjective function ${}^{(n+1)}P^{\mathfrak{B}} \longrightarrow B$ in any model $\mathfrak{B}$ of T. To check that π defines a function, we use the part of the previous proof of (iii) $\Rightarrow$ (i) which shows that $\varepsilon(\bar{x},z) \wedge \neg P(z)$ defines a function. We use the condition $|P| \geq 2$ when showing that this function is surjective.

Let $\delta(\bar{x},y)$ and $\rho(\bar{x},y,z)$ be reductions of the formulas $\exists z \pi(\bar{x},y,z)$ and $\pi(\bar{x},y,z) \wedge P(z)$ respectively. For $S(y_1,\ldots,y_m) \in At^+$ let φ_S be a reduction of the formula $\exists y_1 \ldots y_m(S(y_1,\ldots,y_)) \wedge \pi(\bar{x}_1,y_1) \wedge \ldots \wedge \pi(\bar{x}_m,y_m)$ where $\bar{x}_i$ are sequences of distinct variables of length $n+1$.

It is not difficult to check that $\langle \delta,\pi,\rho,\varphi_S : S \in At^+ \rangle$ as defined above is an interpretation of T in P.

Proof of (iv) $\Rightarrow$ (i): Assume that T is interpretable in P, let $\langle \delta, \pi, \rho, \varphi_S : S \in At^+ \rangle$ be an interpretation of T in P. Let $\varphi_P \in L$ and $\varphi_= \in L$ be such that $\varphi_P^P(\bar{x}) \leftrightarrow (\exists y(\pi(\bar{x}, y) \wedge P(y))$ and $\varphi_=^P(\bar{x}, \bar{z}) \leftrightarrow (\exists y(\pi(\bar{x}, y) \wedge \pi(\bar{z}, y))$ are valid in T. Such φ_P and $\varphi_=$ exist because of our assumption on T. Let us define

$$\varepsilon(\bar{x}, y) \stackrel{\text{def}}{=} \pi(\bar{x}, y) \wedge \neg\varphi_P^P(\bar{x}),$$

$$\varphi(\bar{x}, \bar{z}) \stackrel{\text{def}}{=} \varphi_=^P(\bar{x}, \bar{z}).$$

Let now S and H be as in (i)(2) of Definition 3.1.3. Define

$$\psi_S \stackrel{\text{def}}{=} \exists \bar{w}_1 \dots \bar{w}_k (\textstyle\bigwedge_i \rho(\bar{w}_i, x_i) \wedge \varphi_S(\bar{w}_1, \dots, \bar{w}_k, \bar{y}_1, \dots, \bar{y}_k)).$$

It is not difficult to check that $\varepsilon, \varphi, \psi_S$ as defined above gives an explicit definition of T in P.

Theorem 3.2.3 has been proved. ∎

3.3 Equivalence of implicit and explicit definability.

Recall from Hodges [11, p.645] that a theory T is called *rigidly relatively categorical* over P if whenever $\mathfrak{B}$ and $\mathfrak{B}'$ are models of T such that $\mathfrak{B}_P = \mathfrak{B}'_P$, then there is a unique isomorphism $f : \mathfrak{B} \longrightarrow \mathfrak{B}'$ which is the identity on $\mathfrak{B}_P$. In the literature, usually "relatively categorical" is considered as the notion corresponding to implicit definability (cf. e.g. [11]). In [1] we call a theory T implicitly definable over P if T is rigidly relatively categorical over P.

We show that the above notions all coincide with "rigidly relatively categorical" if the language is countable. This is proved for complete theories T under assuming that the P-part of the models of T are infinite as Gaifman's theorem in Hodges [11, Thm.12.5.8]. We give a proof here for the general case that we will need (i.e. we do not assume that T is complete and that it has infinite models).

THEOREM 3.3.1 (Gaifman's theorem, cf. [11, Thm.12.5.8, p.645]) *Assume that L is countable. Then the following are equivalent.*

(i) *T is rigidly relatively categorical over P.*

(ii) *T is coordinatised over P.*

To prove Theorem 3.3.1, we will use the following theorems.

PROPOSITION 3.3.2 *If T is explicitly definable in P, then T is rigidly relatively categorical over P.*

The proof of Prop.3.3.2 is straightforward by using the definition of T being explicitly definable in P. We omit it.

The theorem below is from Pillay and Shelah [15] and the proof is basically from Hodges [11, Lemma 12.5.1, p.641]. In both cases the theorem was stated for complete theories T only. (The difference in the proof is that we eliminated the use of completeness of T and used ultraproducts instead of compactness and ω-homogeneous elementary extensions).

THEOREM 3.3.3 (Pillay and Shelah [15]) *Condition (i) implies (ii) below.*

(i) *For any model $\mathfrak{B}$ of T, every automorphism of $\mathfrak{B}_P$ extends to an automorphism of $\mathfrak{B}$.*

(ii) *T has the uniform reduction property.*

Proof. If $\bar{x}$ is a sequence of variables, then $L^{(\bar{x})}$ denotes the set of those formulas of L in which the free variables are among the members of $\bar{x}$.

Suppose that (ii) does not hold, let $\varphi(\bar{x}) \in L^+$ be such that for all $\psi \in L^{(\bar{x})}$

(1) $T \not\models \forall \bar{x} \in P(\varphi(\bar{x}) \leftrightarrow \psi^P(\bar{x}))$.

For any finite $\Gamma \subseteq L^{(\bar{x})}$ define

$$\Phi(\Gamma) \stackrel{\text{def}}{=} \forall \bar{x}\bar{y} \in P\Big(\bigwedge \{\gamma^P(\bar{x}) \leftrightarrow \gamma^P(\bar{y}) : \gamma \in \Gamma\} \leftrightarrow (\varphi(\bar{x}) \leftrightarrow \varphi(\bar{y}))\Big).$$

By (1) we have that $T \not\models \Phi(\Gamma)$ for all finite $\Gamma \subseteq L^{(\bar{x})}$. (Otherwise, it would be easy to put together a formula ψ from Γ that would be a "uniform reduction" of φ.) $T \not\models \Phi(\Gamma)$ means that there are $\mathfrak{B}(\Gamma) \models T$ and $\bar{x}^\Gamma, \bar{y}^\Gamma \in {}^nB(\Gamma)$ such that

(2) $\mathfrak{B}(\Gamma) \models \gamma^P(\bar{x}^\Gamma) \leftrightarrow \gamma^P(\bar{y}^\Gamma)$ for all $\gamma \in \Gamma$ and $\mathfrak{B}(\Gamma) \models \varphi(\bar{x}^\Gamma) \wedge \neg\varphi(\bar{y}^\Gamma)$.

We will put together from these $\mathfrak{B}(\Gamma), \bar{x}^\Gamma, \bar{y}^\Gamma$'s a model $\mathfrak{B} \models T$ together with $\bar{x}, \bar{y} \in {}^nB$ such that

(3) $\mathfrak{B} \models \gamma^P(\bar{x}) \leftrightarrow \gamma^P(\bar{y})$ for all $\gamma \in L^{(\bar{x})}$ and $\mathfrak{B} \models \varphi(\bar{x}) \wedge \neg\varphi(\bar{y})$.

Indeed, let I be the set of all finite subsets of $L^{(\bar{x})}$. Let F be an ultrafilter on I such that $\{\Gamma \in I : \psi \in \Gamma\} \in F$ for all $\psi \in L^{(\bar{x})}$. Then $\mathfrak{B} \stackrel{\text{def}}{=} P_{\Gamma\in I}\mathfrak{B}(\Gamma)/F$, $\bar{x} \stackrel{\text{def}}{=} P_{\Gamma\in I}\bar{x}^{\Gamma}/F$, and $\bar{y} \stackrel{\text{def}}{=} P_{\Gamma\in I}\bar{y}^{\Gamma}/F$ will satisfy (3) by the Los-lemma.

Let L' be the language we get from L by adding n new constants $c_1, \ldots, c_n$. Then by (3) we have that the model $(\mathfrak{B}_P, \bar{x})$ is elementarily equivalent with the model $(\mathfrak{B}_P, \bar{y})$. By the Keisler-Shelah theorem, then there is an ultrafilter G over some set J such that

(4) ${}^J(\mathfrak{B}_P, \bar{x})/G \cong {}^J(\mathfrak{B}_P, \bar{y})/G$.

Let $\mathfrak{B}' \stackrel{\text{def}}{=} {}^J\mathfrak{B}/G$, $\bar{x}' \stackrel{\text{def}}{=} {}^J\bar{x}/G$, and $\bar{y}' \stackrel{\text{def}}{=} {}^J\bar{y}/G$. Then

(5) $\mathfrak{B}'_P = {}^J\mathfrak{B}_P/G$ and $\mathfrak{B}' \models \varphi(\bar{x}') \wedge \neg\varphi(\bar{y}')$,

by the Los-lemma. By (4) we have an automorphism of $\mathfrak{B}'_P$ that takes $\bar{x}'$ to $\bar{y}'$. This automorphism cannot be extended to an automorphism of $\mathfrak{B}'$ by (5). ∎

We note that if T is relatively categorical over P, then T satisfies (i) of Theorem 3.3.3. Thus any relatively categorical theory has the uniform reduction property.

Next we state (an analogon of) Gaifman's theorem for finite models.

THEOREM 3.3.4 *Let $\mathfrak{B}$ be finite. Then (a)-(e) below are equivalent.*

(a) $\mathsf{Th}(\mathfrak{B})$ *is rigidly relatively categorical over P.*

(b) *Every automorphism of $\mathfrak{B}_P$ extends to an automorphism of $\mathfrak{B}$ in a unique way.*

(c) $\mathfrak{B}$ *is coordinatised over P.*

(d) $\mathfrak{B}$ *is interpretable in $\mathfrak{B}_P$.*

(e) $\mathfrak{B}$ *is explicitly definable in $\mathfrak{B}_P$.*

Proof: We will only sketch the proofs, by referring to relevant parts of Hodges [11]. To prove Thm.3.3.1 we will only need (a) ⇒ (c). For the proof of (a) ⇒ (b) see e.g. p.638 in [11]. Let $\mathfrak{A} \stackrel{\text{def}}{=} \mathfrak{B}_P$. Proof of (b) ⇒ (c): Let $\mathfrak{B}^A$ denote the model we get from $\mathfrak{B}$ by making the elements of $\mathfrak{A}$ constants. Then, by our assumption (b), $\mathfrak{B}^A$ is rigid, i.e. it has no nontrivial automorphism. Therefore every element of $\mathfrak{B}^A$ is definable in the first-order language of $\mathfrak{B}^A$. This means that every element of B is first-order definable over $\mathfrak{A}$, using the notation of [11]: B is in the closure of $\mathfrak{A}$. The reduction property holds by Theorem 3.3.3. (Here is a short elementary proof for the finite case: Let $\varphi \in L(\mathfrak{B}^A)$ be an arbitrary formula with n free variables. Let $\equiv_n$ denote the equivalence relation on ${}^nA$ defined by $\bar{a} \equiv_n \bar{b}$ iff $\bar{a}$ and $\bar{b}$ cannot be distinguished by formulas in $\mathfrak{A}$. Each equivalence block of this equivalence relation $\equiv_n$ is definable in $\mathfrak{A}$. (Why?) If $\varphi \in L(\mathfrak{B})$ "cuts" no block of $\equiv_n$, then φ is reducible to Ł$(\mathfrak{A})$. Assume that φ cuts a block of $\equiv_n$, i.e. $\bar{a} \equiv_n \bar{b}$ but $\varphi(\bar{a})$ while $\neg\varphi(\bar{b})$ for some $\bar{a}, \bar{b}$. Then there is an automorphism of $\mathfrak{A}$ which takes $\bar{a}$ to $\bar{b}$. This automorphism cannot extend to $\mathfrak{B}$.) Proof of (c) $\Rightarrow$ (d): Since $\mathfrak{B}$ is finite, $\mathfrak{B}$ is *finitely* coordinatised over $\mathfrak{A}$, and then see p.647 in [11] (cf. also Exercises 13, 14 in [11]). (c) $\Leftrightarrow$ (e) is stated in Thm.3.2.3. The proofs of (e) $\Rightarrow$ (a) and (d) $\Rightarrow$ (a) are straightforward by using the definitions. This finishes the proof of Theorem 3.3.4. ∎

We are ready to prove Theorem 3.3.1.

Proof of Theorem 3.3.1: Assume that T is coordinatised over P. Then T is explicitly definable in P by Theorem 3.2.3, and then T is rigidly relatively categorical over P by Proposition 3.3.2. This proves (ii) $\Rightarrow$ (i).

To prove (i) $\Rightarrow$ (ii), assume that T is rigidly relatively categorical over P, we want to prove that T is coordinatised over P. Now T has the uniform reduction property by Theorem 3.3.3 and the remark following it. Thus it is enough to prove that for every model $\mathfrak{B}$ of T, $\mathfrak{B}$ is coordinatised over $\mathfrak{B}_P$. If T is categorical and B_P is infinite, then this holds by Theorem 12.5.8 (p.645) in Hodges [11]. (Here is where we have to use that the language of T is countable.) Assume that T is categorical and B_P is finite. Then $\mathfrak{B}$ is finite because T is relatively categorical: if $\mathfrak{B}$ would be infinite, then $\mathfrak{B}_P$ would have two extensions in T of different cardinalities, which contradicts our assumption (i). Thus (ii) holds if T is categorical. Assume now that T is arbitrary and let $\mathfrak{B}$ be a model of T. Then the theory $\mathsf{Th}(\mathfrak{B})$ is complete, and it satisfies (i) because $\mathsf{Th}(\mathfrak{B}) \supseteq T$. Thus $\mathfrak{B}$ is coordinatised over $\mathfrak{B}_P$ since we have already seen that Thm.3.3.1 holds for complete theories. This finishes the proof of Theorem 3.3.1. ∎

3.4 Connections between the one-sorted and the many-sorted notions

For investigations related to definability of new sorts as discussed in the present section (§2 herein) we refer to Hodges [11] Chapter 12, and within that chapter to §12.3 (pp.624-632), §12.5 (pp.638-652). E.g. p.638 last 3 lines – p.639 line 9 discusses generalizability of Beth's theorem, and similarly for p.645 line 6, p.649 lines 5-6. (We would also like to point out Exercises 13, 14 on p.649 of [11].) We also refer to Myers [14], Hodges-Hodkinson-Macpherson [12], Pillay-Shelah [15], Shelah [17]. In passing we note that our subject matter (i.e. definability of new sorts) is related to the directions in recent (one-sorted) model theory called "relative categoricity" or "categoricity over a predicate", and "theory of stability over a predicate".

Below we outline some connections between our notions and the ones used in a substantial part of the above quoted (one-sorted) literature. We will systematically refer to Hodges [11].

Assume $\mathsf{K}^+ = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$ and that K has finitely many sorts $U_0, \ldots, U_k$. Let $P = U_0 \cup \ldots \cup U_k$ be the union of these sorts regarded as a unary predicate. Then:

(1) "K^+ is implicitly definable up to isomorphism over K" is equivalent with "$\mathsf{Th}(\mathsf{K}^+)$ is relatively categorical over P".

(2) "K^+ is implicitly definable without taking reducts over K" is equivalent with "$\mathsf{Th}(\mathsf{K}^+)$ is rigidly relatively categorical over P".

(3) "K^+ is explicitly definable over K" is equivalent with "$\mathsf{Th}(\mathsf{K}^+)$ is coordinatisable over P".

(4) "K^+ is a definitionally equivalent expansion of K" is equivalent with "$\mathsf{Th}(\mathsf{K}^+)$ is coordinatised over P".

In items (1)-(4) above, on the left hand side we have many-sorted notions, while on the right-hand side we have one-sorted notions (like relative categoricity). So it needs some explanation what we mean by claiming their equivalence. The answer is the following: First we translate our many-sorted notions to one-sorted ones (by treating the sorts as unary predicates of one-sorted logic) the usual, natural way, and then we claim that the so translated version of our many-sorted notion is equivalent with the other one-sorted notion quoted from Hodges [11]. E.g., the so elaborated version of item (1) looks like the following. "The one-sorted translation of (K^+ is implicitly definable up to isomorphism over K)" is equivalent with "(the one-sorted version of $\mathsf{Th}(\mathsf{K}^+)$) is relatively categorical over P". The point here is that relative categoricity is defined only for one-sorted logic in Hodges [11]. Therefore, to use it as a possible equivalent of (our many-sorted) "implicit definability up to isomorphism", first we have to translate everything to one-sorted logic, and then make the comparison. Indeed, items (1)-(4) are understood this way.

To do the above in a precise way, first we define the one-sorted version of a many-sorted expansion. This is a slight modification of the one used in the literature, and the difference is that we use an extra unary relation P to denote the universe of the reduct.

Definition 3.4.1 (One-sorted version of a many-sorted expansion) Let K^+ be an expansion of K. We define the <u>*one-sorted version* $\overline{\mathsf{K}^+}$ *of* K^+</u> as follows. The vocabulary *Voc* of $\overline{\mathsf{K}^+}$ consists of the relation symbols of K^+ together with unary relation symbols U_i for each sort U_i of K^+, and another unary relation symbol P. (We assume that P is a new symbol not occurring in $Voc\mathsf{K}^+$.) For any many-sorted model $\mathfrak{M}$ of vocabulary $Voc\mathsf{K}^+$ we define the one-sorted model $\overline{\mathfrak{M}}$ of vocabulary *Voc* as follows.

- The universe of $\overline{\mathfrak{M}}$ is the union of all the sorts of $\mathfrak{M}$.
- The interpretations of the relation symbols of $\mathfrak{M}$ is the same in $\overline{\mathfrak{M}}$ as in $\mathfrak{M}$.
- The interpretation of the unary relation symbol U_i in $\overline{\mathfrak{M}}$ is the sort U_i of $\mathfrak{M}$.
- The interpretation of P in $\overline{\mathfrak{M}}$ is the union of the old sorts, i.e. of the sorts of K in $\mathfrak{M}$.

Finally, we define $\overline{\mathsf{K}^+} \stackrel{\text{def}}{=} \{\overline{\mathfrak{M}} : \mathfrak{M} \in \mathsf{K}^+\}$. $\lhd$

Remark 3.4.2 (Converting formulas) It is not difficult to see that there are meaning-preserving translations between $Fm(\mathsf{K}^+)$ and $Fm(\overline{\mathsf{K}^+})$ in both ways.

(1) For all $\varphi(\bar{x}) \in Fm(\mathsf{K}^+)$ there is a $\overline{\varphi}(\bar{x}) \in Fm(\overline{\mathsf{K}^+})$ such that for all $\mathfrak{M} \in \mathsf{K}^+$ and sequence $\bar{a}$ of elements of $\mathfrak{M}$ of suitable sorts we have

$$\mathfrak{M} \models \varphi[\bar{a}] \qquad \text{iff} \qquad \overline{\mathfrak{M}} \models \overline{\varphi}[\bar{a}].$$

(2) For all $\psi(\bar{x}) \in Fm(\overline{\mathsf{K}^+})$ and sequence $\bar{U}$ of sorts (of the same length as $\bar{x}$) there is $\varphi(\bar{y}) \in Fm(\mathsf{K}^+)$ such that for all $\mathfrak{M} \in \mathsf{K}^+$ and sequence $\bar{a}$ of elements of $\mathfrak{M}$ of sorts $\bar{U}$ we have that

$$\mathfrak{M} \models \varphi[\bar{a}] \qquad \text{iff} \qquad \overline{\mathfrak{M}} \models \psi[\bar{a}].$$

The idea of the translation in (1) is to replace $\forall x\varphi$ with $(\forall x)[U_i(x) \to \varphi]$ when x is a variable of sort U_i. This translation can be found in almost all logic books, see e.g. Monk [13], Enderton [8], or Barwise-Feferman [6]. For the translation in (2) we use that we have only finitely many sorts, and the main idea is the following: we replace the free variables with variables of the given sorts, we replace the universal quantifiers so that we replace first the outmost ones and then we go inward and we replace $\forall x\varphi(x)$ with $\forall x_1\varphi(x_1) \wedge \ldots \wedge \forall x_n\varphi(x_n)$ where $x_1, \ldots, x_n$ are variables of sorts $U_1, \ldots, U_n$ and $U_1, \ldots, U_n$ are all the sorts, and finally we replace all occurrences of $P(x)$ with $U_1(x) \vee \ldots \vee U_k(x)$ where $U_1, \ldots, U_k$ are all the old sorts, i.e. sorts of K. We omit the details. $\lhd$

Definition 3.4.3 For $\varphi(\bar{x}) \in Fm(\mathsf{K}^+)$ we denote by

$$\overline{\varphi}(\bar{x}) \in Fm(\overline{\mathsf{K}^+})$$

the formula described in (1) of Remark 3.4.2. We call the formula $\varphi(\bar{y}) \in Fm(\mathsf{K}^+)$ assigned to $\psi(\bar{x}) \in Fm(\overline{\mathsf{K}^+})$ and the sequence of sorts $\bar{U}$ <u>*the many-sorted version of* $\psi(\bar{x})$ *with sequence* $\bar{U}$ *of sorts*</u>. $\lhd$

THEOREM 3.4.4 *Assume that* K^+ *is an axiomatizable expansion of* K*, and let* $\overline{\mathsf{K}^+}$ *denote the one-sorted version of* K^+ *as defined in Def.3.4.1. Then (i) below implies (ii). If in addition* $\mathsf{K}^+ \models$ *"one of the old sorts has more than one elements" than (i) and (ii) are equivalent.*

(i) K^+ *is interpreted in* K *by some* Tr *and* $code_i$ *(as in Thm.2.3.7(i)).*

(ii) $\mathsf{Th}(\overline{\mathsf{K}^+})$ *has the total reduction property over* P*.*

Proof. Assume first (i), and let $Tr, code_i$ be the function and formulas interpreting K^+ in K. We may assume that the lengths of sequences $\bar{x}_i$ in $code_i(\bar{x}_i, x)$ are the same for all new sorts U_i of K^+. Let $\varepsilon(\bar{x}, y) \stackrel{\text{def}}{=} \bigvee\{\overline{code_i}(\bar{x}, y) : U_i \text{ is a new sort of } \mathsf{K}^+\}$, where $\overline{code_i}(\bar{x}, y) \in Fm(\overline{\mathsf{K}^+})$ is the one-sorted version of the formula $code_i(\bar{x}_i, y)$ such that we use the same variables $\bar{x}, y$ for all i. Let $\psi(\bar{x}), H, \{x_{i_1}, \ldots, x_{i_k}\}$ and $\{x_{j_1}, \ldots, x_{j_\ell}\}$ be as in the definition of the total reduction property. For a sequence $\bar{U} = \langle U_1, \ldots, U_\ell\rangle$ of new sorts of K^+ let $\hat{\psi}(\bar{U}, \bar{y}) \in Fm(\mathsf{K}^+)$ be the many-sorted version of $\psi(\bar{x})$ and $\bar{U}$ as in (2) of Remark 3.4.2, and let φ be

$$\bigwedge\{U_1(x_{j_1}) \wedge \ldots \wedge U_\ell(x_{j_\ell}) \to Tr(\hat{\psi}(\bar{U}, \bar{y})) : \bar{U} \text{ is a sequence of new sorts of } \mathsf{K}^+\}.$$

Then the one-sorted version $\overline{\varphi}$ of φ (as in (1) of Remark 3.4.2) will be a reduction of ψ w.r.t. H.

To prove the other direction, assume now (ii), and let $\varepsilon(\bar{x}, y) \in Fm(\overline{\mathsf{K}^+})$ be the "coding formula", where $\bar{x}$ is a sequence of variables of lengh k, say. Let $T \stackrel{\text{def}}{=} \mathsf{Th}(\overline{\mathsf{K}^+})$. The problem in translating $\varepsilon(\bar{x}, y)$ to many-sorted formulas $code_i$ is that there may be no old sorts $U_{i_1}, \ldots, U_{i_k}$ such that $T \models \varepsilon(\bar{x}, y) \wedge U_i(y) \to U_{i_1}(x_1) \wedge \ldots \wedge U_{i_k}(x_k)$. We will overcome this problem with increasing the length of $\bar{x}$ and using some coding. In doing this we will rely on the fact that there are only finite many old sorts, each of them nonempty, and one of them has more than one elements.

Let $U_1, \ldots, U_n$ be the old sorts, i.e. sorts of K, and assume that $T \models \bigvee\{|U_i| > 1 : 1 \le i \le n\}$. Assume $\bar{x} = \langle x_1, \ldots, x_k\rangle$ and let $\bar{U} = \langle U_{i_1}, \ldots, U_{i_k}\rangle$ be a sequence of old sorts. Then $\bar{U}(\bar{x})$ denotes the formula $U_{i_1}(x_1) \wedge \ldots \wedge U_{i_k}(x_k)$. Let $\mathcal{U}$ denote the set of all k-sequences of old sorts.

As a first step, we show that we may assume that $\varepsilon(\bar{x}, y)$ has the following coherence property:

(*) If $\varepsilon(\bar{x}, y)$ and $\bar{U}(\bar{x})$, then for all $\bar{z}$ such that $\varepsilon(\bar{z}, y)$ we have $\bar{U}(\bar{z})$.

To show this, let us make an (arbitrary) linear ordering $\le$ on the k-tuples of old sorts, i.e. on $\mathcal{U}$. The idea coded in the formula below is that for each y we select the smallest $\bar{U}$ such that $\varepsilon(\bar{x}, y) \wedge \bar{U}(\bar{x})$, and then we keep in relation only those $\bar{z}$'s for which $\varepsilon(\bar{z}, y) \wedge \bar{U}(\bar{z})$. Indeed, let $\varepsilon'(\bar{x}, y)$ denote the formula

$$\varepsilon(\bar{x}, y) \wedge \bigwedge\{\forall \bar{z}[(\varepsilon(\bar{z}, y) \wedge \bar{U}(\bar{x})) \to \bigvee\{\bar{U}'(\bar{z}) : \bar{U}' > \bar{U}\}] : \bar{U} \in \mathcal{U}\}.$$

Then $\varepsilon'(\bar{x}, y)$ is also a good coding formula, and $\varepsilon'(\bar{z}, y)$ has property (*). So from now we assume that ε itself has property (*).

To show the idea, assume first $k = 1$, i.e. $\bar{x} = \langle x_1\rangle$. Let $\bar{w} = \langle w_1^1, w_2^1, w_1^2, w_2^2, \ldots, w_1^n, w_2^n\rangle$ be a sequence of distinct variables. The coding will be that $code_i(\bar{w}, y)$ will be the many-sorted version of the formula "there is exactly one ℓ such that $w_1^\ell \ne w_2^\ell$ and for this ℓ we have $\varepsilon(w_1^\ell, y)$" and we assigne the sorts to the free variables as $U_i(y)$ and $U_j(w_1^j) \wedge U_j(w_2^j)$ for all $1 \le j \le n$. Formally let γ denote the formula

$$U_i(y) \wedge U_1(w_1^1) \wedge U_1(w_2^1) \wedge \ldots \wedge U_n(w_1^n) \wedge U_n(w_2^n) \wedge$$

$$\bigvee\{w_1^1 = w_2^1 \wedge \ldots \wedge w_1^{\ell-1} = w_2^{\ell-1} \wedge w_1^\ell \ne w_2^\ell \wedge w_1^{\ell+1} = w_2^{\ell+1} \wedge \ldots \wedge w_1^n = w_2^n \wedge \varepsilon(w_1^\ell, y) : 1 \le \ell \le n\}.$$

We then define $code_i(\bar{w}, y)$ to be the many-sorted version of γ. (Note that this is defined since we assigned sorts to each free variables of γ.)

It is straightforward to adapt the above idea to the case when $\bar{x} = \langle x_1, \ldots, x_k \rangle$ (with k arbitrary). We will do that now.

Let us define a linear ordering on the set $\mathcal{U}$ of k-tuples of old sorts, and let $\bar{U}_1 < \bar{U}_1 < \ldots < \bar{U}_N$ be this ordering (where $\mathcal{U} = \{\bar{U}_1, \ldots, \bar{U}_N\}$). Let

$$\bar{w} = \bar{w}_1^1 \bar{w}_2^1 \ldots \bar{w}_1^N \bar{w}_2^N$$

be a sequence of $k \cdot N \cdot 2$ distinct variables where $\bar{w}_i^j$ are sequences of variables, of lenght k. For $1 \leq j \leq N$ let $\bar{w}_1^j = \bar{w}_2^j$ denote the formula $\bigwedge\{w_{1\ell} = w_{2\ell} : 1 \leq \ell \leq k\}$ where $\bar{w}_i^j = \langle w_{1i}, \ldots, w_{ki} \rangle$ $(i = 1, 2)$. Let γ denote the formula

$$U_i(y) \wedge \bar{U}_1(\bar{w}_1^1) \wedge \bar{U}_1(\bar{w}_2^1) \wedge \ldots \wedge \bar{U}_N(\bar{w}_1^N) \wedge \bar{U}_N(\bar{w}_2^N) \wedge$$

$$\bigvee\{\bar{w}_1^1 = \bar{w}_2^1 \wedge \ldots \wedge \bar{w}_1^{\ell-1} = \bar{w}_2^{\ell-1} \wedge \neg \bar{w}_1^\ell = \bar{w}_2^\ell \wedge \bar{w}_1^{\ell+1} = \bar{w}_2^{\ell+1} \wedge \ldots \wedge \bar{w}_1^N = \bar{w}_2^N \wedge \varepsilon(\bar{w}_1^\ell, y) : 1 \leq \ell \leq N\}.$$

Then $code_i(\bar{w}, y)$ is the many-sorted version of γ.

The next claim states that $code_i(\bar{w}, y)$ really code $\varepsilon(\bar{x}, y)$.

Claim 3.4.5 Let $\mathfrak{M} \in \mathsf{K}^+$. Let U_i be a new sort of K^+, let $b \in U_i^{\mathfrak{M}}$ and let $\bar{U}_\ell$ be such that there is $\bar{e}$ with $\bar{U}_\ell(\bar{e})$ and $\varepsilon(\bar{e}, b)$ in $\mathfrak{M}$. Then

$$\{\bar{a} : \overline{\mathfrak{M}} \models \varepsilon(\bar{a}, b)\} = \{\bar{a} : \mathfrak{M} \models code_i(\bar{w}, b) \text{ and } \bar{a} = \bar{w}_1^\ell\}.$$

$\lhd$

We omit the proof of the claim. Thus, $code_i$ satisfy the requirements in the second translation theorem ((a), (b) in Theorem 2.3.4).

The rest of the proof presents no difficulty. We have to give the translation function $Tr : Fm(\mathsf{K}^+) \longrightarrow Fm(\mathsf{K})$ which has the desired properties w.r.t. the formulas $code_i$ (U_i is a new sort of K^+). We show this on a simple case. Let $\psi(y, \bar{z}) \in Fm(\mathsf{K}^+)$, $y \in Var(U_i)$ and $\bar{z}$ variables of old sorts, and let $\overline{\psi}(y, \bar{z}) \in Fm(\overline{\mathsf{K}^+})$ be its one-sorted version. Let φ be the reduction of $\overline{\psi}$ such that

$$\overline{\mathsf{K}^+} \models P(\bar{z}) \wedge \neg P(y) \wedge \varepsilon(\bar{z}, y) \longrightarrow [\overline{\psi}(y, \bar{z}) \leftrightarrow \varphi^P(\bar{z}, \bar{z})].$$

Let $\tau(\bar{w}, \bar{z})$ be the formula

$$\exists y(\gamma(\bar{w}, y) \wedge \exists \bar{x} \varepsilon(\bar{x}, y) \wedge \bar{U}^\ell(\bar{x}) \longrightarrow \varphi(\bar{s}_1^\ell, \bar{z}),$$

and let $Tr(\psi(y, \bar{x}))$ be the many-sorted version of $\tau(\bar{w}, \bar{z})$ (where $\bar{z}$ have the sorts as in the many-sorted formula $\psi(y, \bar{z})$ and $\bar{w}$ have the sorts as indicated in $code_i(\bar{w}, y)$). This finishes the proof of Theorem 3.4.4. ∎

Remark 3.4.6 The condition that at least one old sort has more than one element is necessary in Theorem 3.4.4 above. To see this, let $\mathsf{K}^+ = \mathbf{I}\{\mathfrak{M}\}$ where $\mathfrak{M}$ has three sorts, U_1, U_2, U_3 where U_1, U_2 are old sorts, they are singletons in $\mathfrak{M}$, and U_3 has two elements a, b and a binary relation $R = \{\langle a, b\rangle\}$ on U_3, see the Figure. Then $\varepsilon(x, y) \stackrel{\text{def}}{=} (U_1(x) \wedge \exists z R(x, z)) \vee (U_2(x) \wedge \exists z R(z, x))$ is a good coding formula in $\overline{\mathfrak{M}}$ showing that $\overline{\mathfrak{M}}$ has the total reduction property over P. On the other hand, let $\mathfrak{M}^-$ be the reduct of $\mathfrak{M}$ to the sorts U_1 and U_2, and $\mathsf{K} = \mathbf{I}\{\mathfrak{M}^-\}$. Then $\mathfrak{M}$ is not rigid over $\mathfrak{M}^-$, hence K^+ is not rigidly definable over K but K^+ is axiomatizable. Hence K^+ is not interpreted by some $Tr, code_i$ in K, by Theorem 2.3.7. $\lhd$

THEOREM 3.4.7 *Assume that K^+ is an axiomatizable expansion of K, and let $\overline{\mathsf{K}^+}$ be the one-sorted version of K^+ as defined in Def.3.4.1. Then (i) and (ii) below are equivalent.*

(i) K^+ *is nr-implicitly definable over* K.

(ii) $\mathsf{Th}(\overline{\mathsf{K}^+})$ *is rigidly relatively categorical over P.*

Proof. (ii) $\Rightarrow$ (i) is straightforward by using Def.3.4.1. Proof of (i) $\Rightarrow$ (ii): Let $\overline{\mathfrak{M}}, \overline{\mathfrak{N}} \models \mathsf{Th}(\overline{\mathsf{K}^+})$ be such that $\overline{\mathfrak{M}}_P = \overline{\mathfrak{N}}_P$. Let $\mathfrak{M}, \mathfrak{N}$ be such that $\overline{\mathfrak{M}}$ and $\overline{\mathfrak{N}}$ are the one-sorted versions of $\mathfrak{M}$ and $\mathfrak{N}$ respectively. We want to show that $\mathfrak{M}, \mathfrak{N} \in \mathsf{K}^+$. In order to show $\mathfrak{M} \in \mathsf{K}^+$, it is enough to show $\mathfrak{M} \models \mathsf{Th}(\mathsf{K}^+)$ since K^+ is axiomatizable. Let ψ be such that $\mathsf{K}^+ \models \psi$. Then the one-sorted version $\overline{\psi}$ of ψ is valid in $\overline{\mathsf{K}^+}$ by Remark 3.4.2, so $\overline{\mathfrak{M}} \models \overline{\psi}$. But then $\mathfrak{M} \models \psi$ by Remark 3.4.2. ∎

We are ready to prove the equivalence of implicit and explicit definability in many-sorted logic.

Proof of Theorem 2.5.1. By Cor.2.3.9, Prop.2.1.4 and $\mathsf{K} = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}))$ we may assume that K^+ is axiomatizable, too, i.e. that $\mathsf{K}^+ = \mathsf{Mod}(\mathsf{Th}(\mathsf{K}^+))$. So, assume that K^+ is axiomatizable and K is a reduct of K^+. Then K^+ is a definitionally equivalent expansion of K iff K^+ is interpreted in K by some Tr and $code_i$, by Thm.2.3.7. Since K^+ is axiomatizable, the latter holds iff $\mathsf{Th}(\overline{\mathsf{K}^+})$ has the total reduction property, by Thm.3.4.4. By Thm.3.2.3, the latter holds iff $\mathsf{Th}(\overline{\mathsf{K}^+})$ is coordinatised over P and by Thm.3.3.1 the latter holds iff $\mathsf{Th}(\overline{\mathsf{K}^+})$ is rigidly relatively categorical over P. Finally, by Thm.3.4.7 the latter holds iff K^+ is nr-implicitly definable over K. ∎

References

[1] H. Andréka, J.X. Madarász, and I. Németi. The logical structure of relativity theories., 2001. (With contributions from A. Andai, G. Sági, I. Sain and Cs. Tőke).

[2] H. Andréka, J.X. Madarász, and I. Németi. Defining new universes in many-sorted logic, January 2001. A. Rényi Institute of Mathematics, Budapest.

[3] H. Andréka, J.X. Madarász, and I. Németi. A note on definability and definitional equivalence in model theory, June 2000. A. Rényi Institute of Mathematics, Budapest.

[4] H. Andréka, I. Németi, and I. Sain. Algebraic logic. In *Handbook of Philosophical Logic Vol. II, second edition*, page To appear. D. Gabbay and F. Guenthner (eds), D. Reidel Publ. Co., 1997.

[5] J. Barwise. *Handbook of Mathematical Logic*. North-Holland, Amsterdam, 1977.

[6] J. Barwise and S. Feferman, editors. *Model-Theoretic Logics*. Springer–Verlag, Berlin, 1985.

[7] C.C. Chang and H.J. Keisler. *Model Theory*. North–Holland, Amsterdam, 1973, 1990.

[8] H.B. Enderton. *Introduction to Mathematical Logic*. North–Holland, Amsterdam, 1972.

[9] M. Friedman. *Foundations of Space-Time Theories. Relativistic Physics and Philosophy of Science.* Princeton University Press, 1983.

[10] L. Henkin, J.D. Monk, and A. Tarski. *Cylindric Algebras Parts I and II.* North–Holland, Amsterdam, 1971 and 1985.

[11] W. Hodges. *Model Theory*. Cambridge University Press, 1993.

[12] W.A. Hodges, I. Hodkinson, and H.D. Macpherson. Omega-categoricity, relative categoricity and coordinatisation. *Annals of Pure and Applied Logic*, 46(4):169–199, 1990.

[13] J.D. Monk. *Mathematical Logic.* Springer Verlag, 1976.

[14] D. Myers. An interpretive isomorphism between binary and ternary relations. In *Structures in logic and computer science*, 1997. ed. Jan Mycielski, Grzegorz Rozenberg, and Arto Salomaa, Lecture Notes in Computer Science 1261, Springer, Berlin 1997, pp.84-105.

[15] A. Pillay and S. Shelah. Classification theory over a predicate I. *Notre Dame Journal of Formal Logic*, 26(4):361–376, 1985.

[16] H. Reichenbach. *Axiomatization of the theory of relativity.* University of California Press, Berkeley, 1969. Translated by M. Reichenbach. Original German edition published in 1924.

[17] S. Shelah. *Around classification theory of models.* Lecture Notes in Mathematics 1182, Springer-Verlag, Berlin, 1986.

[18] A. Tarski. A problem concerning the notion of definability. In [21], Vol. 3, pp. 163-170.

[19] A. Tarski. Some methodological investigations on the definability of concepts. (in German), in [21], Voll 1, pp.637-639. The Polish version contains more detail: in Revue Philosophique, Vol.37 (1934).

[20] A. Tarski. Fundamentale begriffe der methodologie der deduktiven wissenchaften, i. *Monatsh. Math. Phys.*, 37:361–404, 1930. Also in "Logic, Semantics, Metamathematics. Papers from 1923 to 1938. Clarendon Press, Oxford, 1956. pp.60–109.

[21] A. Tarski. *Collected Papers, Volumes 1–4.* Birkhäuser (Basel–Boston–Stuttgart), 1986. Editors: Steven R. Givant and Ralph N. McKenzie.

[22] A. Tarski and S. Givant. *A Formalization of Set Theory Without Variables*, volume 41 of *AMS Colloquium Publications.* Providence, Rhode Island, 1987.

[23] A. Tarski, A. Mostowski and R.M. Robinson. *Undecidable theories.* North-Holland, Amsterdam, 1953. xii+98pp.