Two-variable logic has weak Beth definability but not strong one

Andrèka, H. and Németi, I.

August 2020

Abstract

We prove that the two-variable fragment of first-order logic has the weak Beth definability property. This makes the two-variable fragment a natural logic separating the weak and the strong Beth properties since it does not have the strong Beth definability property. The proof relies on the special property of 2-variable logic that each model is 2-equivalent to some rather homogeneous model that we dub 2-transitive.

We dedicate this paper to Harvey Friedman in respect for his work

1 Introduction

One of the many expressibility properties of first-order logic with equality FO is the Beth definability property BDP. It states that if a relation can be specified by some extra means then it can be specified explicitly without using the extra means. In more detail, if Th is a first-order logic theory on a language $\mathcal{L}$ and $\Sigma$ is another first-order logic theory on the language $\mathcal{L}$ expanded with an extra relation symbol $R$ such that in each model of Th there is at most one relation $R$ satisfying $\Sigma$, then this unique relation can be defined on the original language $\mathcal{L}$ without using the extra relation symbol, i.e., there is a formula $\varphi$ on $\mathcal{L}$ such that $\text{Th} \cup \Sigma \models R \leftrightarrow \varphi$. In this context, $\Sigma$ is called the implicit definition and $\varphi$ is called the explicit definition.

Investigating BDP for fragments of FO means showing that if all the formulas of the theory and the implicit definition belong to the fragment,
then the explicit definition, too, belongs to it. Thus, having BDP or not shows a kind of “integrity” of the fragment, and a kind of “complexity-property” of FO itself. For example, the guarded fragment GF of FO has BDP [12]. Thus, if the theory $\text{Th}$ and the implicit definition $\Sigma$ consist of guarded formulas, then the explicit definition can be chosen to be guarded, too. We note that in the strict sense, GF does not have Craig interpolation property, yet it is worth deciding when the interpolant belonging to guarded $\text{Th}$ and $\Sigma$ can be chosen to be guarded itself [13]. For work of the similar kind, see e.g., [4].

It is known that $n$-variable fragments $\text{FO}_n$ of FO do not have BDP, for all $n \geq 2$, see [1]. This means that if the theory $\text{Th}$ and the implicit definition $\Sigma$ all use only $n$ variables, the explicit definition $\varphi$ may need more than $n$ variables. That the BDP fails for $\text{FO}_2$ is kind of surprising, because $\text{FO}_2$ usually behaves “better” than $\text{FO}_n$ for $n \geq 3$. For example, $\text{FO}_2$ is decidable while $\text{FO}_n$, for $n \geq 3$ is not.

The weak Beth definability property wBDP was introduced by Harvey Friedman [9], and it is sometimes considered to be more important than the BDP. The difference between BDP and wBDP is that in wBDP an implicit definition is considered to be legitime - i.e., has to be made explicit - only if it has also the existence property, not only the uniqueness property. In mathematical practice, one almost always requires both existence and uniqueness for an implicitly defined object. It is known that $\text{FO}_n$ does not have wBDP either, whenever $n \geq 3$ ([16] for $n = 3$, [11] for $n \geq 5$, and [3] for $n \geq 3$) and it was not known whether $\text{FO}_2$ has wBDP or not.

In this paper we prove that $\text{FO}_2$ does have wBDP, this restores two-variable logic’s image that it behaves better than $n$-variable logics for $n \geq 3$. This theorem may also point to wBDP being more natural than BDP.

Since $\text{FO}_2$ has wBDP but does not have BDP, it can serve as a natural simple logic distinguishing the two definability. The difference between BDP and wBDP was not tangible so far in the sense that there was no natural example for a logic that distinguished the two properties. FO augmented with the quantifier “there exists uncountable many” $L(Q)$ was a good candidate for such a distinguishing logic, since it does not have BDP [9] and it is consistent with set theory that it has wBDP [15]. However, it is still an open problem whether wBDP can be proved for $L(Q)$ in set theory or not. We note that wBDP was intensely investigated in abstract model theory in connection with logics stronger than FO, see, e.g., [5,14]. Also, 2-variable logic and its extensions are quite popular in computer science and in modal
logic.

Our proof hinges on the fact that FO2 has a special property that FO\(n\) with \(n \geq 3\) do not have. Namely, each model of FO2 is FO2-equivalent with a rather homogeneous model in which practically nothing new can be implicitly defined because we have as many automorphisms as the FO2-types in the model allow (Thm. 1, Thm. 2). FO3 does not have this property in a very strong sense (Thm. 3).

The structure of the paper is as follows. In section 2 we define the above kind of homogeneous models that we will call 2-transitive and we prove the key property of FO2 about the abundance of 2-transitive models. In section 3 we prove that FO2 has wBDP by relying on these 2-transitive models (Thm. 4).

2 Doubly-transitive models of FO2

We use the notation of [7], if not stated otherwise. By FO we mean first-order logic with equality, but we do not allow function or constant symbols.

By FO\(n\) we mean the fragment of FO that uses only the first \(n\) variables. Strictly speaking, the fragment FO\(n\) of FO is defined by taking for all languages \(\mathcal{L}\) all the models of \(\mathcal{L}\) but restricting the set of formulas of FO to those that contain the first \(n\) variables only. Thus, relation symbols of arbitrarily high rank can be allowed in FO\(n\). For simplicity, in this paper in FO\(n\) we will allow only languages with relation symbols of rank at most \(n\). We go further, we allow relation symbols of rank \(n\) only. These are not important restrictions. Usually, we do not indicate the language, but we will always work with similar models, i.e., models having the same language, if not stated otherwise.

In FO, there is a good notion for a homogeneous model. However, for FO\(n\), the natural \(n\)-variable analogues of several equivalent properties of homogeneous models become non-equivalent, this is one of the reasons we will not call our models “homogeneous” but rather transitive ones.

Let \(n < \omega\). \(\mathfrak{M} \equiv \mathfrak{N}\) denotes that \(\mathfrak{M}\) and \(\mathfrak{N}\) are \(n\)-equivalent, i.e., the same \(n\)-variable formulas are true in them. By the \(n\)-type of \(a_1, \ldots, a_n\) in \(\mathfrak{M}\) we understand the set of FO\(n\)-formulas true for them. We say that \(\mathfrak{M}\) is \(n\)-transitive if whenever \(a_1, \ldots, a_n, b_1, \ldots, b_n\) are elements having the same \(n\)-type in it, there is an automorphism of \(\mathfrak{M}\) taking \(a_1, \ldots, a_n\) to \(b_1, \ldots, b_n\).
respectively. In concise form:

\[(\mathcal{M}, a_1, \ldots, a_n) \equiv (\mathcal{M}, b_1, \ldots, b_n) \implies (\mathcal{M}, a_1, \ldots, a_n) \cong (\mathcal{M}, b_1, \ldots, b_n).\]

Note that while \(n\)-equivalent implies \(m\)-equivalent for all \(m \leq n\), \(n\)-transitive does not imply \(m\)-transitive for any \(m\). The reason is that \(n\)-transitivity requires the existence of automorphisms only for elements of the same \(n\)-type, while it does not care for elements with the same \(m\)-type for smaller \(m\). We believe that one can prove, by using the results in [8], that all models of size \(\leq n + 1\) are \(n\)-transitive. This also shows that \(n\)-transitivity does not imply 2-transitivity, since there are models of size 4 which are not 2-transitive. In FO, the homogeneous models are abundant, but these are not 2-transitive in most cases.

2-transitivity is analogous with 2-transitivity of groups and with edge-transitivity of graphs, hence our name. The notion of 2-transitivity is rather strong: it means that the (rather weak) 2-type of each pair of points determines their (rather strong) automorphism types. Let \(R\) be a graph (an irreflexive symmetric binary relation) on a finite set \(U\) such that each point has degree different from 0 and \(|U|\). Then in \(\langle U, R \rangle\) each edge has the same 2-type. Thus \(\langle U, R \rangle\) is not 2-transitive whenever in \(R\) there are points of different degree, or sub-cycles of different size. Any \(|U|\)-cycle on \(U\) is 2-transitive. In general, successor relations in groups modulo \(k\) are typical 2-transitive relations. Let us call a model binary if all its basic relations are of rank 2. We will prove in this section that each binary model is 2-equivalent to a 2-transitive model. The idea of the proof is that we build a 2-equivalent version for any binary relation by putting together different successor relations. We note that FO2 is quite expressive on binary models.

A stronger version of the above will be proved in this section. We call a model \(\mathcal{M}\) \(n\)-homogeneous if whenever \(a_1, \ldots, a_m, b_1, \ldots, b_m, c \in M\) are such that \((\mathcal{M}, a_1, \ldots, a_m) \equiv (\mathcal{M}, b_1, \ldots, b_m)\) and \(m \leq n\), there is \(d \in M\) such that \((\mathcal{M}, a_1, \ldots, a_m, c) \equiv (\mathcal{M}, b_1, \ldots, b_m, d)\). This is a straightforward analogue of the definition of \(\alpha\)-homogeneity in [7]. Now, \(n\)-transitivity implies \(m\)-homogeneity for all \(m < n\), but not the other way round: \(n\)-transitivity does not follow from \(m\)-homogeneity for all \(m < n\).

Finally, we need the notion of \(n\)-partial isomorphism. The notion of \(n\)-partial isomorphism was defined in [3, p.259] as a natural restriction of the usual notion of partial isomorphisms between models of FO (see [7]). We recall the definition of 2-partial isomorphism in detail because we will rely
on it. The set $I$ is a 2-partial isomorphism between models $\mathcal{M}$, $\mathcal{N}$ if (i) – (iv) below hold:

(i) $I$ relates elements as well as pairs of $M$ and $N$, i.e., it is a subset of $(M \times N) \cup (M^2 \times N^2)$,

(ii) related pairs of $I$ are isomorphisms between $\mathcal{M}$ and $\mathcal{N}$ restricted to the first and second parts of the pair, respectively,

(iii) restriction property:
if $\langle (a, a'), (b, b') \rangle \in I$ then $\langle a, b \rangle \in I$ and $\langle a', b' \rangle \in I$, and

(iv) back-and-forth property:
$\forall a \in M \exists b \in N \langle a, b \rangle \in I$ and vice versa,
$\forall b \in N \exists a \in M \langle a, b \rangle \in I$, and vice versa
$\forall \langle a, b \rangle \in I \forall a' \in M \exists b' \in N \langle (a, a'), (b, b') \rangle \in I$, and vice versa
$\forall \langle a, b \rangle \in I \forall b' \in N \exists a' \in M \langle ((a, a'), (b, b')) \in I$.

Instead of 2-partial isomorphism we will simply say 2-isomorphism. It is known that if two pairs are related by a 2-isomorphism, then their 2-types are the same. (This is straightforward to show by induction.) Thus, if there is a 2-isomorphism between two models, then they satisfy the same FO2-formulas.

Being 2-isomorphic is stronger than being 2-equivalent. A corollary of the theorem below is that each binary model is 2-equivalent to a 2-transitive one, but finite binary models are 2-isomorphic to finite 2-transitive ones, see Thm.2. The stronger theorem below will be used in the proof of weak Beth definability property for 2-variable logic FO2.

**Theorem 1** A binary model $\mathcal{M}$ is 2-isomorphic to a 2-transitive model $\mathcal{N}$ if and only if it is 2-homogeneous.

**Proof.** Let $\mathcal{M}$ be any 2-homogeneous binary model. We are going to define another model $\mathcal{N}$ and a 2-isomorphism $I$ between them. $\mathcal{N}$ will be finite when $\mathcal{M}$ is so. Then we show that $\mathcal{N}$ is 2-transitive. Finally, we prove necessity of 2-homogeneity.

Some notation and terminology. The variables of FO2 will be denoted by $x, y$. Recall that by the 2-type of $a, b \in \mathcal{M}$ we understand the set of FO2-formulas $\rho(x, y)$ that are true for $a, b$ in $\mathcal{M}$. Formally,

$$\text{Type}(a, b, \mathcal{M}) = \{\rho(x, y) \in FO2 : \mathcal{M} \models \rho[a, b]\}.$$

Now we define
\[ a = \{ b \in M : \text{Type}(b, b, M) = \text{Type}(a, a, M) \}, \]

\[ [a, b] = \text{Type}(a, b, M), \]

\[ \text{Types}([a], [b]) = \{ [p, q] : p \in [a], q \in [b] \}. \]

We call the elements of \{ [a] : a \in M \} 1-Types. Note that, for convenience, we defined \([a]\) to be a subset of the model, while \([p, q]\) is a set of FO2-formulas.

The type \([p, q]\) determines \([q, p]\), we call the latter the converse type \([p, q] \rightarrow \) of \([p, q]\). Taking the converse is a bijection between \(\text{Types}([a], [b])\) and \(\text{Types}([b], [a])\). We call a type \([p, q]\) symmetric when \([p, q] = [q, p]\), otherwise we call it asymmetric. We note that there is a unique element of \(\text{Types}([a], [a])\) which contains \(x = y\), we call this the identity type on \([a]\), it is symmetric.

**Choosing groups**

We will use groups in constructing \(\mathfrak{M}\). A group is called asymmetric if its only element of order 2 is its zero element, i.e., if \(x = -x\) implies \(x = 0\) in the group. We begin choosing groups for each pair of 1-Types.

Assume that \([a]\) and \([b]\) are distinct 1-Types. Let \(G([a], [b])\) be any commutative asymmetric group on \(\text{Types}([a], [b])\). Let \(\oplus([a], [b])\) denote the group-operation of this group. We will just write \(\oplus\) in place of \(\oplus([a], [b])\), and \(\ominus\) denotes the unary as well as the binary group inverse operations. Thus, \((r \ominus s) \ominus s = r\) for any \(r, s\) in the group. We require that \((g1)\) the groups \(G([a], [b])\) and \(G([b], [a])\) be isomorphic via the bijection that sends \([p, q] \in G([a], [b])\) to \([p, q] \rightarrow\). \(\)

The choice of the groups \(G([a], [a])\) will be a bit more complicated, because \(\text{Types}([a], [a])\) contains the identity type on \([a]\) and it may contain symmetric types. The universe of \(G([a], [a])\) will be slightly different from \(\text{Types}([a], [a])\), to make up for this, we will define a function \(\tau\) mapping the group to \(\text{Types}([a], [a])\). In the following, for a while, we will write \(T\) in place of \(\text{Types}([a], [a])\) and we will write \(G\) in place of \(G([a], [a])\).

If \([a]\) is a singleton, then \(T\) consists of only one type, in this case let \(G\) be any one-element group, and let \(\tau\) be the only map between them (i.e., \(\tau(e) = [a, a]\) for the element \(e\) of the group).

Assume that \([a]\) is not a singleton. Then \(T\) contains at least one non-identity type, let’s choose one, \(t(a)\). Let \(S\) denote the set of non-identity
symmetric types in $T \setminus \{t(a), t(a)^\sim\}$. Let now $G$ be any commutative asymmetric group on 
$G([a], [a]) = (T \setminus \{t(a), t(a)^\sim\} \cup S) \cup (S \times \{1, 2\})$
such that 
(g2) the zero element $0$ of $\oplus$ is the identity type on $[a]$
and the group inverse operation $\ominus$ preserves converse of types in the following sense:
(g3) $\ominus(t) = t^\sim$ for any element of $T \setminus \{t(a), t(a)^\sim\} \cup S$, and
(g4) $\ominus(s, 1) = (s, 2)$ for any $s \in S$.
There is such a group because of the following. Take first just any asymmetric group operation $\oplus'$ on $G$. Let $H \subseteq G$ be such that exactly one of $g$ and $\ominus'g$
is in $H$ for all nonzero elements of $G$. Let $H = \{\ominus'g : g \in H\}$. Then $H, H', 0'$
is a partition of $G$. Take another similar partition of $G$ but based on the operation $\sim$ of types in place of $\ominus'$ as follows. Let $T^- = G \setminus (S \times \{1, 2\})$. Then $T^-$ consists of types. Let $K1 \subseteq T^-$ be such that exactly one of $g$
and $g^\sim$ belongs to $K1$ for each type $g$ in $T^-$. Let $K = K1 \cup S \times \{1\}, \overline{K} = \{g^\sim : g \in K1\} \cup S \times \{2\}$. Then $K, \overline{K}$
together with the identity type form a partition of $G$, too, such that $|H| = |K|$. Let $\pi$ be any bijection between $H$ and $K$, let $\beta$ be the bijection between $H$ and $\overline{H}$ that takes $g$ to
$\ominus'g$, let $\delta$ be the bijection between $K$ and $\overline{K}$ that takes $g \in K1$ to $g^\sim$ and takes $(s, 1)$ to $(s, 2)$ for $s \in S$. Extend $\pi$ to $\overline{H}$ by taking $\beta(g)$ to $\delta\pi(g)$, and extend $\pi$ to $0'$ by letting $\pi(0')$ be the identity type. Finally take $\oplus$ to be the
$\pi$-image of $\oplus'$. This $\oplus$ will do.
We define $\tau : G \rightarrow T$ as follows:
$\tau(t) = t$ for $t \in G \setminus (S \times \{1, 2\})$ and $\tau(s, j) = s$ for $s \in S, j = 1, 2$.
We have chosen our groups $G([a], [b])$ for all 1-Types $[a], [b]$.
By an $[a]$-choice we understand a function that to any 1-Type $[b]$ assigns
an element of the group $G([a], [b])$. To deal later with the selected type $t(a)$
that we left out from $G([a], [a])$, we will use a set $H(a)$ of the $[a]$-choices. We
call an $[a]$-choice $h$ non-zero when $h([a]) \neq 0$, and $\ominus h$ denotes the $[a]$-choice
that assigns to $[b]$ the group inverse $\ominus h([b])$, for any $[b]$. Now, let us choose a
subset $H(a)$ of the nonzero $[a]$-choices such that for any $[a]$-choice $h$, exactly
one of $h$ and $\ominus h$ is in $H(a)$. There is such a $H(a)$, because all our groups are
asymmetric: $h$ and $\ominus h$ are always distinct when $[a]$ is not the only 1-Type
that assigns to $[b]$ the zero of $G([a], [b])$ for all $[b]$.  

7
Definition of the model $\mathfrak{N}$

We are ready to define the model $\mathfrak{N}$. The universe $N$ of $\mathfrak{N}$ is defined as

$$N = \{([a], f) : a \in M, \ f \text{ is an } [a]\text{-choice}\}.$$ 

Let $R$ be a binary relation symbol in the language of $\mathfrak{M}$. Assume first $[a] \neq [b]$. Then we define

$$R(([a], f), ([b], g)) \in \mathfrak{N} \iff R(x, y) \in f([b]) \oplus g([a])^-.$$ 

The case when $[a] = [b]$ will be a bit more subtle because of the identity. We define

$$R(([a], f), ([a], f)) \in \mathfrak{N} \iff \ R(x, y) \text{ is in the identity type of } [a].$$ 

Assume now that $f([a]) = g([a])$ but $f \neq g$. Here is when we use that we set aside the type $t(a)$ which is not included in the group $G([a], [a])$.

If $f \ominus g \in H(a)$, then

$$R(([a], f), ([a], g)) \in \mathfrak{N} \iff R(x, y) \in t(a), \text{ and}$$

if $f \ominus g \notin H(a)$, then

$$R(([a], f), ([a], g)) \in \mathfrak{N} \iff R(x, y) \in t(a)^-.$$ 

Assume finally that $f([a]) \neq g([a])$. Define

$$R(([a], f), ([a], g)) \in \mathfrak{N} \iff R(x, y) \in \tau(f([a]) \ominus g([a])).$$

By this, the model $\mathfrak{N}$ has been defined. Clearly, $\mathfrak{N}$ is finite when $\mathfrak{M}$ is finite.

Before proving that $\mathfrak{N}$ is 2-isomorphic to $\mathfrak{M}$ and that it is 2-transitive, we prove some properties of $\mathfrak{N}$ that will be useful in these proofs.

For convenience, let $ty(([a], f), ([b], g))$ denote

(1) $f([b]) \oplus g([a])^-$ when $[a] \neq [b]$,

(2) the identity type of $[a]$ when $[a] = [b]$ and $f = g$,

(3) $t(a)$ when $[a] = [b], f([a]) = g([a]), f \neq g$ and $f \ominus g \in H(a),$

(4) $t(a)^-$ when $[a] = [b], f([a]) = g([a]), f \neq g$ and $f \ominus g \notin H(a),$

(5) $\tau(f([a]) \ominus g([a]))$ when $[a] = [b]$ and $f([a]) \neq g([a]).$
Then, by the definition of $\mathfrak{N}$, for all $p, q \in N$ we have
\[(\star1) \quad R(p, q) \text{ in } \mathfrak{N} \iff R(x, y) \in ty(p, q).\]

We will use that $ty$ respects identity and converse, thus we prove that for all $p, q \in N$ we have
\[(\star2) \quad x = y \in ty(p, q) \iff p = q \quad \text{and} \quad ty(p, q) = ty(q, p).\]

Indeed, let $p = ([a], f)$ and $q = ([b], g)$. In proving $(\star2)$, we check cases (r1)-(r5) one by one. First we check $x = y \in ty(p, q)$ iff $p = q$. We have $p = q$ in (r2), in all the other cases we have $p \neq q$. Also, we have $x = y \in ty(p, q)$ in (r2), in all the other cases $x = y \notin ty(p, q)$ because in (r1) we have $ty(p, q) \in \text{Types}([a], [b])$ with $[a] \neq [b]$, in (r3) and (r4) we have that $t(a)$ and so $t(a)^\sim$ are non-identity types, in (r5) we have that $f(a) \oplus g(a)$ is not the zero of $G([a], [a])$ by $f(a) \neq g(a)$, and so $\tau(f(a) \ominus g(a))$ is not the zero either by the definition of $\tau$, thus it is not the identity type by (g2).

Checking $ty(p, q)^\sim = ty(q, p)$ in the case (r1): $ty(p, q) = f(b) \oplus g(a)^\sim \in \text{Types}([a], [b])$ while $ty(q, p) = g(a) \oplus f(b)^\sim \in \text{Types}([b], [a])$, by definition. Let $i$ be the bijection mapping $\text{Types}([a], [b])$ to $\text{Types}([b], [a])$ defined by $i(t) = t^\sim$. Since $i$ is an isomorphism by (g1), we have that $i(f(b) \oplus g(a)^\sim) = i(f(b)) \oplus i((g(a)^\sim)) = f(b)^\sim \oplus g(a) = g(a) \oplus f(b)^\sim$ since our groups are commutative. We are done by the definition of $i$. The case of (r2) follows from the fact that the identity type on $[a]$ is symmetric. To check (r3) and (r4), assume that $[a] = [b]$, $f(a) = g(a)$ and $f \neq g$. By $f \neq g$ we have $f \ominus g \neq 0$, so exactly one of $f \ominus g$ and $g \ominus f$ is in $H(a)$, so in these cases exactly one of $ty(p, q)$ and $ty(q, p)$ is $t(a)$ and the other is then $t(a)^\sim$, so we are done. For checking (r5), assume that $[a] = [b]$ and $f(a) \neq g(a)$. Then $ty(p, q) = \tau(f(a) \ominus g(a))$ and $ty(q, p) = \tau(g(a) \ominus f(a))$. If $f(a) \ominus g(a)$ is in $G \setminus S$, then so is its group inverse $g(a) \ominus f(a)$, thus $\tau$ is the identity on both them, and group inverse coincides with converse by (g3). If $f(a) \ominus g(a)$ is $(s, j)$ for some $s \in S$ and $j \in \{1, 2\}$, then its inverse is $(s, i)$ with $i \neq j$ by (g4), thus both $ty(p, q)$ and $ty(q, p)$ equal to $s$ with $s^\sim = s$, so we are done.

We will also use a kind of union-property: Let $a, b \in M$ and let $f$ be an $[a]$-choice.
\[(\star3) \quad \text{Types}([a], [b]) = \{ty(([a], f), ([b], g)) : g \text{ is a } [b]\text{-choice}\}.\]

To prove $(\star3)$, let $a, b \in M$ and let $f$ be an $[a]$-choice and let $t \in \text{Type}([a], [b])$ be arbitrary. Assume first that $[b] \neq [a]$. Let $r = t \ominus f([b])$ and let $g
be any \([b]\)-choice such that \(g([a]) = r^\sim\). These make sense, since \(f([b]), t \in \text{Types}([a], [b]) = G([a], [b])\) and so \(r^\sim \in \text{Types}([b], [a])\). Then \(\text{ty}(([a], f), ([b], g)) = f([b] \oplus g([a]))^\sim = f([b]) \oplus r^\sim \oplus f([b]) = t\).

Assume now that \([a] = [b]\). If \(t\) is the identity type, then \(\text{ty}(([a], f), ([a], f)) = t\) by (r2). Assume that \(t\) is \(t(a)\). Then let \(g\) be any \([a]\)-choice distinct from \(f\) such that \(g(a) = f(a)\) and \(f \triangledown g \in H(a)\). We show that there is such a \(g\). Choose any \([a]\)-choice \(h\) distinct from \(f\) with \(h([a]) = f([a])\). If \(f \triangledown h \in H(a)\) then we can choose \(g\) to be \(h\) and we are done. If \(f \triangledown h \notin H(a)\) then \(h \triangledown f \in H(a)\) by the choice of \(H(a)\). Now we can choose \(g\) to be such that \(f \triangledown g = h \triangledown f\). Then \(g\) is distinct from \(f\) since \(h \triangledown f\) is nonzero and \(g([a]) = f([a])\) by \(h([a]) = f([a])\) and we are done. The case when \(t\) is \(t(a)^\sim\) is completely analogous. Assume now that \(t\) is non-identity and distinct from \(t(a), t(a)^\sim\). Let \(r\) be \((t, 1)\) if \(t\) is symmetric, and otherwise let \(r\) be \(t\). Then \(r \in G([a], [a])\) is nonzero and \(\tau(r) = t\). Let now \(g\) be any \([a]\)-choice for which \(f([a]) \oplus g([a]) = r\). Then \(g([a]) \neq f([a])\) since \(r\) is nonzero, and \(\text{ty}(([a], f), ([a], g)) = \tau(f([a]) \oplus g([a])) = t\), and (\(*3\)) has been proved.

### A 2-isomorphism between \(\mathcal{M}\) and \(\mathfrak{N}\)

We are ready to define the 2-isomorphism \(I\) between \(\mathcal{M}\) and \(\mathfrak{N}\). Let \(a, b \in M\) and \(p, q \in N\). We define \(I \subseteq (M \times N) \cup (M^2 \times N^2)\) by

\[
\langle a, p \rangle \in I \quad \text{iff} \quad p = ([a], f) \text{ for some } f,
\]

\[
\langle (a, b), (p, q) \rangle \in I \quad \text{iff} \quad [a, b] = \text{ty}(p, q).
\]

We now show that \(I\) is a 2-isomorphism between \(\mathcal{M}\) and \(\mathfrak{N}\). From the conditions defining a 2-isomorphism, \(I\) clearly satisfies (i) by its very definition.

Next we show that the restriction property (iii) holds for \(I\). Assume that \(\langle (a, b), (p, q) \rangle \in I\). This means that \([a, b] = \text{ty}(p, q)\). Assume that \(p = ([c], f)\) and \(q = ([d], g)\). By inspecting the definition of \(\text{ty}(p, q)\), we see that \(\text{ty}(p, q) \in \text{Types}([c], [d])\). Thus, \([a, b] = [r, s]\) for some \(r \in [c]\) and \(s \in [d]\). But this implies that \([a] = [r] = [c]\) and \([b] = [s] = [d]\), hence \(\langle a, p \rangle \in I\) and \(\langle b, q \rangle \in I\).

We show that \(I\) satisfies (ii). Assume that \(\langle a, p \rangle \in I\). Then \(p = ([a], f)\) for some \(f\) by the definition of \(I\), and we have to show that \(R(p, p)\) holds in \(\mathfrak{N}\) iff \(R(a, a)\) holds in \(\mathcal{M}\). This is true by (\(*1\)) and case (r2) by which \(R(p, p)\) iff \(R(x, y) \in \text{ty}(p, p) = \text{Type}(a, a, \mathcal{M})\) iff \(R(a, a)\). Assume that \(\langle (a, b), (p, q) \rangle \in I\).
Then \([a, b] = \text{ty}(p, q)\) by the definition of \(I\). We have to show that each of \(p = q, R(p, p), R(p, q), R(q, p), \) and \(R(q, q)\) holds in \(\mathfrak{N}\) iff the same holds in \(\mathfrak{M}\) for \(a, b\) in place of \(p, q\). Now, \(p = q\) iff \(x = y\) in \(\text{ty}(p, q) = [a, b] = \text{Type}(a, b, \mathfrak{M})\) iff \(a = b\), by (\(*1\)) and the definition of \([a, b]\). Similarly, \(R(p, q)\) iff \(R(x, y) \in \text{ty}(p, q) = [a, b] = \text{Type}(a, b, \mathfrak{M})\) iff \(R(a, b)\), by (\(*1\)). By (\(*2\)) we have that \(\langle (b, a), (q, p) \rangle \in I\) whenever \(\langle (a, b), (p, q) \rangle \in I\), so we also have \(R(q, p)\) iff \(R(b, a)\). Finally, \(R(p, p)\) and \(R(q, q)\) hold by the first case of (ii), since we have already shown the restriction property (iii). Thus, \(\langle (a, b), (p, q) \rangle\) indeed specifies a partial isomorphism between the suitable restricted \(\mathfrak{M}\) and \(\mathfrak{N}\).

Finally, we check the back-and-forth property (iv) for \(I\). To check the first part, notice that to any \(a \in M\) there is at least one \([a]\)-choice, because the groups \(G([a], [b])\) are all nonempty. The second part is clear, since no 1-Type is empty. To check the third and fourth parts of the back-and-forth property, assume that \(\langle a, p \rangle \in I\) with \(p = ([a], f)\). Let \(b \in M\) be arbitrary. We have to find a \([b]\)-choice \(g\) such that \([a, b] = \text{ty}(p, ([b], g))\). We can always find such a \(g\) by (\(*3\)). Let now \(q = ([b], g) \in N\) be arbitrary. We have to find \(c \in M\) such that \([a, c] = \text{ty}(p, q)\). Now, \(\text{ty}(p, q) \in \text{Types}([a], [b])\) which means that there are \(a' \in [a]\) and \(b' \in [b]\) such that \(\text{ty}(p, q) = \text{Type}(a', b', \mathfrak{M})\). By \(a' \in [a]\) we have \(\text{Type}(a', a', \mathfrak{M}) = \text{Type}(a, a, \mathfrak{M})\), so by 2-homogeneity of \(\mathfrak{M}\) there is \(c \in M\) such that \(\text{Type}(a, c, \mathfrak{M}) = \text{Type}(a', b', \mathfrak{M})\) and this shows that (iv) holds for \(I\). We have shown that \(I\) is a 2-isomorphism between \(\mathfrak{M}\) and \(\mathfrak{N}\).

\(\mathfrak{N}\) is 2-transitive

We show that \(\mathfrak{N}\) is 2-transitive. We will use automorphisms \(\alpha\) that consist of coordinated shifts. In more detail: Notice that there is a natural addition \(\oplus\) on the set of \([a]\)-choices: let \(f, g\) be \([a]\)-choices then \(f \oplus g\) is the \([a]\)-choice that assigns \(f([b]) \oplus g([b])\) to \([b]\), where \(\oplus\) is the group operation \(\oplus([a], [b])\) of \(G([a], [b])\). Let \(k\) be a function that assigns an \([a]\)-choice to any 1-Type \([a]\). For better readability, we will write \(k_a\) in place of \(k([a])\). Let \(\alpha(k)\) denote the function mapping \(N\) into \(N\) defined by

\[
\alpha(k)(([a], f)) = ([a], f \oplus k_a).
\]

Now, \(\alpha(k)\) is a permutation on \(N\) because the \(\oplus([a], [b])\) are group-operations. Let \(p = ([a], f)\) and \(q = ([b], g)\). Then \(\alpha(k)\) does not change \(\text{ty}(p, q)\) when \([a] = [b]\), i.e.,

\[
(A1) \quad \text{ty}(p, q) = \text{ty}(\alpha(k)p, \alpha(k)q) \quad \text{when} \quad [a] = [b].
\]
One can check this by noticing that the conditions in cases \((r2)-(r5)\) defining \(\tau\) are invariant by \(\alpha(k)\). Indeed, we have \(\alpha(k)p = \alpha(k)q\) whenever \(p = q\), this settles case \((r2)\). We have \((f \oplus k_a)[a] = (g \oplus k_a)[a]\) whenever \(f([a]) = g([a])\), we have \(f \oplus k_a \neq g \oplus k_a\) whenever \(f \neq g\), and \((f \oplus k_a) \ominus (g \oplus k_a) = f \ominus g\) because our groups are commutative, this settles cases \((r3)-(r5)\). To have the same for the case \([a] \neq [b]\), we have to make some restriction on \(k\). Namely,

\[
(A2) \quad \tau(y(p, q)) = \tau(\alpha(k)p, \alpha(k)q) \text{ iff } k_a([b]) = \ominus k_b([a])\sim \text{ when } [a] \neq [b].
\]

Indeed, assume \([a] \neq [b]\). Now, \(\tau(\alpha(k)p, \alpha(k)q) = (f \oplus k_a)(([b]) \oplus (g \oplus k_a)([a])\sim = f([b]) \oplus k_a([b]) \oplus (g([a]) \oplus k_b([a])\sim)\), by \((g1)\). Now, the latter equals to \(f([b]) \oplus g([a])\sim\) iff \(k_a([b]) \oplus k_b([a])\sim = 0\) since our groups are commutative, and we are done.

Notice that \((A1)\) and \((A2)\) imply that the permutation \(\alpha(k)\) is an automorphism. We are ready to show 2-transitivity of \(\mathcal{N}\).

Let \(p_i = ([a_i], f_i) \in \mathcal{N}\) for \(i = 1, 2, 3, 4\) and assume that the 2-type of \(\langle p_1, p_2 \rangle\) in \(\mathcal{N}\) is the same as that of \(\langle p_3, p_4 \rangle\). We have to show that there is an automorphism of \(\mathcal{N}\) that takes \(p_1\) to \(p_3\) and \(p_2\) to \(p_4\).

By the definition of \(I\) we have \(\langle [a_i], p_i \rangle \in I\) and so by the back-and-forth property \((iv)\) we have \(\langle ([a_i], [a'_j]), (p_i, p_j) \rangle \in I\) for some \(a'_j \in [a_j]\), for all \(i, j = 1, 2, 3, 4\). Since \(I\) is a 2-isomorphism, that implies that \(\text{Type}(a_i, a'_j, \mathcal{N}) = \text{Type}(p_i, p_j, \mathcal{N})\). Hence, our assumption that \((p_1, p_2)\) and \((p_3, p_4)\) have the same 2-type in \(\mathcal{N}\) implies that \([a_1, a'_2] = [a_3, a'_4] = \tau(p_1, p_2) = \tau(p_3, p_4)\). In particular, \([a_1] = [a_3]\) and \([a_2] = [a'_2] = [a'_3] = [a_4]\). Denote \([a] = [a_1] = [a_3]\) and \([b] = [a_2] = [a_4]\). With this notation, our assumption implies that \(\tau(p_1, p_2) = \tau(p_3, p_4) \in \text{Types}([a], [b])\).

We are ready to define the automorphism \(\alpha\) that takes \((p_1, p_2)\) to \((p_3, p_4)\). We define the “shifting constants” \(k_c\). Let \(k\) be defined by

\[
k_a = f_3 \ominus f_1, \quad k_b = f_4 \ominus f_2, \quad \text{and for all } [c] \neq [a], [b]
\]

\[
k_c([a]) = \ominus k_a([c])\sim, \quad k_c([b]) = \ominus k_b([c])\sim, \quad k_c([d]) = 0 \quad \text{for all } [d] \neq [a], [b].
\]

Clearly, \(\alpha(k)\) takes \(p_1, p_2\) to \(p_3, p_4\) respectively. For example, \(\alpha(k)(p_1) = (\langle [a_1], f_1 \oplus k_a \rangle = (\langle [a_1], f_1 \oplus f_3 \ominus f_1 \rangle = \langle [a_3], f_3 \rangle = p_3\). It remains to show that \(\alpha(k)\) is an automorphism. To this end it is enough to show that \(k\) satisfies the conditions in \((A2)\). This is stated in the definition of \(k\) explicitly for all pairs of 1-Types, except for \([a], [b]\) when \([a] \neq [b]\). Note that the condition in \((A2)\) is insensitive to the order of \([a], [b]\) because \(k_a([b]) = \ominus k_b([a])\sim\) exactly when
To prove the other direction of Theorem 1, necessity of 2-homogeneity, assume that \( k_a([a]) = \ominus k_a([b]) \). Now, \( k_a \) and \( k_b \) are determined by the goal that we want to take \( p_1, p_2 \) to \( p_3, p_4 \), but we can show that the (A2) condition hold for them by using that the pairs \((p_1, p_2), (p_3, p_4)\) have the same 2-type. We have \([a] \neq [b]\), so by \( \text{ty}(p_1, p_2) = \text{ty}(p_3, p_4) \) and (r1) we have \( f_1([b]) \oplus f_2([a]) = f_3([b]) \oplus f_4([a]) \), whence we get \( (f_3 \ominus f_1)([b]) = \ominus (f_4 - f_2)([a]) \) which is just the desired \( k_a([b]) = \ominus k_b([a]) \).

We have seen that \( \mathfrak{M} \) is 2-transitive, and this finishes the proof of one direction of Theorem 1.

**Necessity of 2-homogeneity**

To prove the other direction of Theorem 1 necessity of 2-homogeneity, assume that \( I \) is a 2-isomorphism between \( \mathfrak{M} \) and the 2-transitive \( \mathfrak{N} \). We have to show that \( \mathfrak{M} \) is 2-homogeneous. Let \( a, b, c \in M \) be such that \([a] = [b]\). By the back-and-forth property in the definition of a 2-isomorphism, there are \( a', b', c' \in N \) such that \( \langle a, a' \rangle, \langle b, b' \rangle, \langle (a, c), (a', c') \rangle \) are all in \( I \). Then \( \text{Type}(a', a', \mathfrak{N}) = \text{Type}(b', b', \mathfrak{N}) \) since \( I \) is a 2-isomorphism and \([a] = [b]\). Since \( \mathfrak{N} \) is 2-transitive, there is an automorphism of \( \mathfrak{N} \) that takes \( a' \) to \( b' \). Let \( d' \) be the image of \( c' \) under this automorphism. Then \( \text{Type}(a', c', \mathfrak{N}) = \text{Type}(b', d', \mathfrak{N}) \) since automorphisms preserve 2-types of elements. By the back-and-forth property of \( I \) again, there is \( d \in M \) such that \( \langle (b, d), (b', d') \rangle \in I \). Then \( \text{Type}(a, c, \mathfrak{M}) = \text{Type}(a', c', \mathfrak{N}) = \text{Type}(b', d', \mathfrak{N}) = \text{Type}(b, d, \mathfrak{M}) \) and we are done.

Next we state a corollary of Theorem 1.

**Theorem 2** Each binary model is 2-equivalent to a 2-transitive model. Each finite binary model is 2-isomorphic to a finite 2-transitive model.

**Proof.** First we prove

\[(S) \quad \mathfrak{M} \text{ is } \omega\text{-saturated implies that } \mathfrak{M} \text{ is 2-homogeneous.}\]

Assume that \( \mathfrak{M} \) is \( \omega \)-saturated. Let \( a, b, c \in M \) be such that \( \text{Type}(a, a, \mathfrak{M}) = \text{Type}(b, b, \mathfrak{M}) \). Let \( Y = \{ b \} \subseteq M \) and let \( \Gamma(x) = \{ \rho(x, b) \in FO2 : \rho(c, a) \text{ in } \mathfrak{M} \} \). Then \( \Gamma(x) \) is a set of formulas in the language of \( \langle \mathfrak{M}, b \rangle \). We show that it is consistent with the theory of \( \langle \mathfrak{M}, b \rangle \). Let \( \Delta \) be a finite subset of \( \Gamma(x) \), let \( \delta(y) \) denote the formula \( \exists x \land \Delta[b/y] \) that we get from \( \exists x \land \Delta \) by replacing \( b \) everywhere with \( y \). Then \( \mathfrak{M} \models \delta(a) \) by the definition of \( \Gamma(x) \), and
so $\mathcal{M} \models \delta(b)$ since $a, b$ have the same 1-Type in $\mathcal{M}$. But $\mathcal{M} \models \delta(b)$ means that $\langle \mathcal{M}, b \rangle \models \exists x \land \Delta$ that shows that $\Delta$ is consistent with the theory of $\langle \mathcal{M}, b \rangle$. Since $\Delta$ is an arbitrary finite subset of $\Gamma(x)$, this means that $\Gamma(x)$ is consistent with the theory of $\langle \mathcal{M}, b \rangle$. Since $\mathcal{M}$ is $\omega$-saturated, then there is $d \in M$ for which $\Gamma(d)$ holds. This means that $\text{Type}(a, c, \mathcal{M}) = \text{Type}(b, d, \mathcal{M})$ and we are done with showing (S).

Now, Theorem 2 follows from (S) and Theorem 1 by using that each infinite model is elementarily equivalent--hence 2-equivalent--with an $\omega$-saturated one (see [7, Lemma 5.1.4]), that each finite model is $\omega$-saturated (see [7, Prop.5.1.2]) and that the model $\mathcal{M}$ constructed in the proof of Theorem 1 is finite whenever $\mathcal{M}$ is finite.

□

Theorem 3 below serves as a contrast to Theorem 2.

**Theorem 3** There is a finite binary model that is not 3-equivalent to any 2-transitive model, not even to any model in which all elements of the same 3-type can be taken to each other by an automorphism.

**Proof.** The binary model $\mathcal{M}$ has 45 elements and 4 basic relations $S, G, R, B$. Let $(\mathbb{Z}, +)$ denote the group of non-negative numbers smaller than 5 with addition modulo 5, and let $9 = \{0, 1, \ldots, 8\}$ denote the set of non-negative integers smaller than 9. We define

$$M = 5 \times 9.$$ 

Let $s, g$ be permutations of 9 defined, in cycle form, as

$s = (012)(345)(678)$ and $g = (136)(147)(258),$ 

and let $r, b \subseteq 9 \times 9$ be defined as

$r = \{0, 3, 6\} \times \{0, 1, 2\} \cup \{1, 4, 7\} \times \{3, 4, 5\} \cup \{2, 5, 8\} \times \{6, 7, 8\}$ and $b = \{0, 4, 8\} \times \{0, 5, 7\} \cup \{1, 5, 6\} \times \{1, 3, 8\} \cup \{2, 3, 7\} \times \{2, 4, 6\}.$

Now, the basic relations of $\mathcal{M}$ are defined as

$S = \{(i, j), (i, s(j)) : i \in 5, j \in 9\},$

$G = \{(i, j), (i, g(j)) : i \in 5, j \in 9\},$

$R = \{(i, j), (i + 1, k) : i \in 5, (j, k) \in r\},$

$B = \{(i, j), (i + 2, k) : i \in 5, (j, k) \in b\}.$
We show that $\mathcal{M}$ is not 3-equivalent to any 2-transitive model. Let us call a model \textit{3,1-transitive} when any two elements of the same 3-type can be taken to each other by an automorphism. A 2-transitive model $\mathcal{N}$ is 3,1-transitive, because let $a, b \in \mathcal{N}$ have the same 3-types, then they have the same 2-types in $\mathcal{N}$, therefore there is an automorphism taking $a$ to $b$, by 2-transitivity of $\mathcal{N}$. Thus, it is enough to show that if $\mathcal{M}$ is 3-equivalent to $\mathcal{N}$ then $\mathcal{M}$ is not 3,1-transitive.

Assume that $\mathcal{M}$ is 3-equivalent to $\mathcal{N}$. For a first-order formula $\rho(x, y)$ with free variables among $x, y$ let $\rho(\mathcal{M})$ denote the relation that $\rho$ defines in $\mathcal{M}$, i.e., $\rho(\mathcal{M}) = \{(a, b) : \mathcal{M} \models \rho[a, b]\}$. Let $\mathcal{R}_{\mathcal{M}}(\mathcal{M})$ be the relation algebra of FO3-definable binary relations of $\mathcal{M}$, i.e., the universe of $\mathcal{R}_{\mathcal{M}}(\mathcal{M})$ is $\{\rho(\mathcal{M}) : \rho(x, y) \in FO3\}$, and the operations of $\mathcal{R}_{\mathcal{M}}(\mathcal{M})$ are the operations of taking union, converse and relation composition of binary relations together with the (base-sensitive) operations of taking complement in $M \times M$ and the identity constant $\{(u, u) : u \in M\}$ on $M$. Let $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ denote the similar algebra of FO3-definable binary relations of $\mathcal{N}$. We show the following:

1. $\mathcal{M}$ is 3-equivalent to $\mathcal{N}$ implies that $\mathcal{R}_{\mathcal{M}}(\mathcal{M})$ is isomorphic to $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$. Indeed, it is easy to check that the relation $\{(\rho(\mathcal{M}), \rho(\mathcal{N})) : \rho(x, y) \in FO3\}$ is an isomorphism between $\mathcal{R}_{\mathcal{M}}(\mathcal{M})$ and $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ when $\mathcal{M}$ is 3-equivalent to $\mathcal{N}$.

A \textit{base-automorphism} of $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ is a permutation $\alpha$ of $\mathcal{N}$ that leaves all elements of $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ fixed when taking $Z$ to $\{(\alpha(u), \alpha(v)) : (u, v) \in Z\}$. Now, $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ is called \textit{c-permutational} iff any element of $\mathcal{N}$ can be taken to any other by a base-automorphism.

2. $\mathcal{N}$ is 3,1-transitive implies that $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ is c-permutational. To check (2), notice first that all elements in $\mathcal{N}$ have the same 3-type. This is so because $\mathcal{M}$ is such and this property can be expressed with FO3 formulas $\{\exists x \rho(x, x) \rightarrow \forall x \rho(x, x) : \rho(x, y) \in FO3\}$. Therefore, $\mathcal{N}$ is 3,1-transitive means that each element of $\mathcal{N}$ can be taken to any other element of $\mathcal{N}$ by an automorphism of $\mathcal{N}$. Finally, an automorphism $\alpha$ of $\mathcal{N}$ is a base-automorphism of $\mathcal{R}_{\mathcal{N}}(\mathcal{N})$ and we are done.

From now on, we will use [2].

3. $\mathcal{R}_{\mathcal{M}}(\mathcal{M})$ is the algebra $\mathfrak{A}$ defined in [2 section 2]. Indeed, it can be checked that the basic relations $S, G, R, B$ of $\mathcal{M}$ coincide with the relations $s, g, r_0, b_0$ in [2 section 2]. It is stated in [2, p.375, line
that $A$ is generated by these four elements, so each element of $A$ is $\text{FO3}$-definable in $\mathfrak{M}$. In the other direction, it is a theorem of relation algebra theory that all $\text{FO3}$-definable elements of $\mathfrak{M}$ can be generated from the basic relations of $\mathfrak{M}$ with the operations of $\text{Ra}(\mathfrak{M})$, see e.g., [17, sec.3.9] or [10, Thm.3.32]. Thus, the elements of $A$ are exactly the $\text{FO3}$-definable relations of $\mathfrak{M}$ and we are done.

In the proof of [2, Theorem 1] it is proved that $A$ is not isomorphic to any c-permutational algebra, and so $\mathfrak{N}$ cannot be 3,1-permutational by (1) and (2). The proof of Theorem 3 is complete.

\section{Two-variable fragment of FO has the weak Beth definability property}

We recall the definition of when the two-variable fragment $\text{FO2}$ has the weak Beth definability property (wBDP). Let $\mathcal{L}$ be a language with relation symbols of rank 2, let $\text{Th}$ be any set of formulas of $\text{FO2}(\mathcal{L})$, the set of formulas of language $\mathcal{L}$ that contain, bound or free, only the variables $x, y$. Assume that $R$ is a binary relation symbol not occurring in $\mathcal{L}$, let $\mathcal{L}^+$ denote $\mathcal{L}$ expanded with $R$. Let $\Sigma(R)$ be a set of formulas of $\text{FO2}(\mathcal{L}^+)$. We say that $\Sigma(R)$ is a strong implicit definition of $R$ w.r.t. $\text{Th}$ when in each model $\mathfrak{M}$ of $\text{Th}$ there is exactly one relation $R$ such that $\langle \mathfrak{M}, R \rangle \models \Sigma(R)$. We say that $\Sigma(R)$ is just a (weak) implicit definition of $R$ w.r.t. $\text{Th}$ when in each model $\mathfrak{M}$ of $\text{Th}$ there is at most one relation $R$ such that $\langle \mathfrak{M}, R \rangle \models \Sigma(R)$. We say that $\Sigma(R)$ can be made explicit w.r.t. $\text{Th}$, or that $R$ has an explicit definition over $\text{Th}$ when there is a formula $\varphi \in \text{FO2}(\mathcal{L})$ such that $\text{Th} \cup \Sigma(R) \models R \leftrightarrow \varphi$. In this case, we say that $\varphi$ is an explicit definition of $R$ in $\text{Th}$. Now, $\text{FO2}$ has the weak Beth definability property means that each strong implicit definition of $\text{FO2}$ can be made explicit. We note that it is proved in [11] that there is a weak implicit definition in $\text{FO2}$ that cannot be made explicit.

\textbf{Theorem 4} \textit{$\text{FO2}$ has the weak Beth definability property.}

\textbf{Proof.} Assume that $\Sigma(R)$ is a strong implicit definition of $R$ w.r.t. $\text{Th}$. We are going to show that $\Sigma(R)$ can be made explicit, i.e., $R$ has an explicit definition over $\text{Th}$ that uses only two variables.

Take any 2-homogeneous model $\mathfrak{M}$ of $\text{Th}$, and let $\overline{\mathfrak{M}}$ be a 2-transitive model with $I$ a 2-isomorphism between $\mathfrak{M}$ and $\overline{\mathfrak{M}}$. There are such a model...
Since $Σ(R)$ is a strong implicit definition, there is $R$ which satisfies $Σ(R)$ in $\overline{M}$, i.e., $⟨\overline{M}, R⟩ \models Σ(R)$. Since $Σ(R)$ is a (weak) implicit definition of $R$ and $FO_2$ is a fragment of first-order logic $FO$, $R$ has a first-order explicit definition by the Beth definability theorem for $FO$. Our task is to show that this definition can be chosen so that it uses only two variables.

Since $R$ is $FO$-definable in $M$, it is invariant with respect to automorphisms of $M$. Since $M$ is 2-transitive, this implies that $R$ does not distinguish between pairs of the same 2-type. We now “transfer” $R$ to the model $M$ by the following definition: take any pair $(a, b) ∈ M$. We define $R$ so that this pair is related by $R$ if and only there is a pair of the same type in $M$ which is related by $R$. Formally:

$$R = \{ (a, b) ∈ M × M : ∃ c, d | R(c, d) and Type(a, b, M) = Type(c, d, \overline{M}) \}.$$  

We now show that with this definition, the 2-isomorphism $I$ between $M$ and $\overline{M}$ remains a 2-isomorphism between $⟨M, R⟩$ and $⟨\overline{M}, R⟩$. Since $I$ is a 2-isomorphism between $M$ and $\overline{M}$, it satisfies conditions (i), (iii), (iv) in the definition of a 2-isomorphism, and it also satisfies condition (ii) for atomic formulas other than $R(v, z)$. Therefore, we only have to show that if $I((a, b), (a′, b′)) ∈ I$, then $R(a, b)$ iff $\overline{R}(a′, b′)$.

Assume that $I((a, b), (a′, b′)) ∈ I$. If $R(a, b)$, then there are $c, d ∈ \overline{M}$ such that $R(c, d)$ and $Type(c, d, \overline{M}) = Type(a, b, M)$, by the definition of $R$. The 2-type of $(a′, b′)$ is also the same as that of $(a, b)$, since they are I-related by assumption. Hence the 2-type of $(c, d)$ is the same as that of $(a′, b′)$ (since they both equal the 2-type of $(a, b)$). By $R(c, d)$ we now get $\overline{R}(a′, b′)$, since we have seen that $\overline{R}$ does not distinguish elements of the same 2-type. In the other direction, assume that $\overline{R}(a′, b′)$. Since $(a, b)$ is I-related to $(a′, b′)$, their 2-types equal, hence $R(a, b)$ by the definition of $R$. By this, we have seen that $I$ is a 2-isomorphism between the expanded models $⟨M, R⟩$ and $⟨\overline{M}, R⟩$. Since the latter is a model of $Σ(R)$, we get that $⟨M, R⟩$ is a model of $Σ(R)$, too. By its definition, $R$ does not distinguish pairs of the same 2-type in $M$.

Since $Σ(R)$ is a (weak) definition over $Th$ and the 2-homogeneous $M \models Th$ was chosen arbitrarily, we get that no relation $R$ defined by $Σ(R)$ distinguishes pairs of the same 2-type in 2-homogeneous models of $Th$. Therefore, it is enough to show that this implies that $R$ is 2-definable. We say that $R$ does not cut 2-types in $M$ if for all $a, b, c, d ∈ M$ we have $R(a, b)$ iff $R(c, d)$ whenever the 2-type of $(a, b)$ is the same as that of $(c, d)$ in $M$. 

17
Lemma 1 (Cut-lemma) Assume that $\Sigma(R)$ is a weak implicit definition of $R$ in $\text{Th}$, and $\text{Th} \cup \Sigma(R)$ consists of $\text{FO}^2$ formulas. Statements (i) - (iii) below are equivalent.

(i) $\Sigma(R)$ can be made explicit in $\text{Th}$ by a $\text{FO}^2$-formula.

(ii) $R$ does not cut 2-types in any model $\mathcal{M}$ of $\text{Th}$ whenever $(\mathcal{M}, R) \models \Sigma(R)$.

(iii) $R$ does not cut 2-types in any 2-homogeneous model $\mathcal{M}$ of $\text{Th}$ whenever $(\mathcal{M}, R) \models \Sigma(R)$.

Proof. Clearly, (i) implies (ii) and (ii) implies (iii). To show that (iii) implies (i), assume that (i) does not hold, we will infer the negation of (iii).

We will refer to the negation of (i) just as “$R$ is not 2-definable”. Thus we want to show that there is a 2-homogeneous model $\mathcal{M}$ of $\text{Th}$ in which $R$ cuts a 2-type.

By a 2-partition we understand a system $\langle \pi_i : i \leq n \rangle$ of $\text{FO}^2$-formulas in the language of $\text{Th}$ such that $\text{Th} \cup \Sigma(R) \models (\forall \{ \pi_i : i \leq n \} \land \bigwedge \{ \neg (\pi_i \land \pi_j) : i < j \leq n \})$. We say that $R$ cannot cut $\pi$ when $\text{Th} \cup \Sigma(R) \models (\forall \pi, y(\pi(x, y) \rightarrow R(x, y)) \lor (\exists x, y(\pi(x, y) \rightarrow \neg R(x, y)))$. We say that $R$ cannot cut into the 2-partition $\langle \pi_i : i \leq n \rangle$ when $R$ cannot cut into any $\pi_i$ for $i \leq n$. We will use the following statement

(L1) $R$ is 2-definable iff there is a 2-partition $R$ cannot cut into.

Indeed, if $R$ is definable by the $\text{FO}^2$-formula $\rho(x, y)$, then $R$ cannot cut into $\langle \rho, \neg \rho \rangle$. In the other direction, assume that $R$ cannot cut into $\langle \pi_i : i \leq n \rangle$. Let us treat natural numbers in the von Neumann’s sense, i.e., each natural number $n$ is the set of smaller natural numbers: $n = \{ i \in \omega : i < n \}$. For $J \subseteq n$ let $\pi(J) = \bigwedge \{ \pi_j : j \in J \} \land \bigwedge \{ \neg \pi_j : j \in n \setminus J \}$. By the assumption that $R$ cannot cut into $\langle \pi_i : i < n \rangle$ we have that $R$ is a union of some $\pi_i$s in each model of $\text{Th} \cup \Sigma(R)$, i.e., $\text{Th} \cup \Sigma(R) \models (\forall R \leftrightarrow \pi(J) : J \subseteq n)$. Since $\Sigma(R)$ is an implicit definition, we have $\text{Th} \cup \Sigma(R) \cup \Sigma(R') \models R \leftrightarrow R'$ where $R'$ is a brand new binary relation symbol, so by compactness

(s) $\text{Th} \cup \Sigma_0(R) \cup \Sigma_0(R') \models R \leftrightarrow R'$ for some finite $\Sigma_0 \subseteq \Sigma$.

Let $\sigma(R) = \bigwedge \Sigma_0(R)$ for a $\Sigma_0$ satisfying (s). Assume that $(\mathcal{M}, R) \models \text{Th} \cup \Sigma(R)$. Then there is a unique $J$ such that $\mathcal{M} \models R \leftrightarrow \pi(J)$, since $\pi(J) \land \pi(K)$ is inconsistent for distinct $J$ and $K$. By $(\mathcal{M}, R) \models \Sigma(R)$ and $\Sigma_0 \subseteq \Sigma$ we
have that $\mathfrak{M} \models \sigma(\pi(J))$. Also, $\mathfrak{M} \models \sigma(\pi(K))$ is not true for $K \neq J$ by (s) since $\pi(J)$ and $\pi(K)$ define distinct relations in $\mathfrak{M}$. This shows that $\text{Th} \cup \Sigma(R) \models R \leftrightarrow \bigwedge \{\sigma(\pi(J)) \rightarrow \pi(J) : J \subseteq n\}$, and this is a 2-definition for $R$. Statement (L1) has been proved.

To prove (iii), we construct a 2-type $T$ in the language of $\text{Th}$ such that the set

\[(L2) \quad \text{Th} \cup \Sigma(R) \cup \{R(x, y), \neg R(z, v)\} \cup T(x, y) \cup T(z, v)\]

of FO-formulas is consistent. That $T$ is a 2-type means that either $\rho \in T$ or $\neg \rho \in T$ for all FO2-formulas $\rho$ in the language of $\text{Th}$. We say that a set of open FO-formulas is consistent when there is a model and an evaluation that make the set true. Equivalently, one can consider the free variables in the set to be constants, we will use this second option.

Let $\tau$ be a FO2-formula in the language of $\text{Th}$. A 2-partition of $\tau$ is a system $\langle \pi_i : i \leq n \rangle$ such that $\text{Th} \cup \Sigma(R) \models (\tau \leftrightarrow \bigvee \{\pi_i : i < n\} \land \bigwedge \{\neg(\pi_i \land \pi_j) : i < j < n\}$. Let $T$ be a set of FO2 formulas in the language of $\text{Th}$. We say that $T$ is good iff $R$ can cut into any 2-partition of $\bigwedge T_0$, for all finite subsets $T_0$ of $T$. Clearly, a directed union of good sets is a good set again, so there is a maximal good set by Zorn’s lemma. We will show that any maximal good set is a 2-type and that the set in (L2) with any good $T$ is consistent. We begin with this second statement.

Assume that $T$ is good and let $T_0 \subseteq T$ be finite. Then $R$ can cut into any 2-partition of $\bigwedge T_0$, in particular $R$ can cut into $\bigwedge T_0$. This means that there is a model of $\text{Th} \cup \Sigma(R) \cup \{R(x, y), \neg R(z, v)\} \cup T_0(x, y) \cup T_0(z, v)$. By compactness, the set in (L2) is consistent.

To show that a maximal good $T$ is a 2-type, let $T$ be any good set and let $\rho$ be any FO2 formula in the language of $\text{Th}$. We show that either $T \cup \{\rho\}$ is good or $T \cup \{\neg \rho\}$ is good. Assume that neither of $T \cup \{\rho\}$ and $T \cup \{\neg \rho\}$ is good. Then there are finite subsets $T_0, T_1$ of $T$ and 2-partitions $\pi = \langle \pi_i : i \leq n \rangle$ of $\rho \land \bigwedge T_0$ and $\delta = \langle \delta_j : j < m \rangle$ of $\neg \rho \land \bigwedge T_1$ such that $R$ cannot cut into either of these two partitions. We can now combine $\pi$ and $\delta$ to form a 2-partition $\sigma$ of $\bigwedge (T_0 \cup T_1)$ by letting the members of the partition $\sigma$ be $\pi_i \land \bigwedge T_1$ and $\delta_j \land \bigwedge T_0$ for $i < n, j < m$. Clearly, $R$ cannot cut into $\sigma$ by our assumption that $R$ cannot cut into either of $\pi$ and $\delta$; this contradicts to $T$ being good. With this, we have proved that any maximal good $T$ is a 2-type.

By the above, we now have a 2-type $T$ such that the set $\Delta = \text{Th} \cup \Sigma(R) \cup \{R(x, y), \neg R(z, v)\} \cup T(x, y) \cup T(z, v)$ is consistent. Let then $\langle \mathfrak{M}, R \rangle$
be any $\omega$-saturated model of $\Delta$. Then $\mathfrak{M}$ is also $\omega$-saturated, and so it is 2-homogeneous by statement (S) in the proof of Theorem 2. Also, $R$ cuts the 2-type $T$ in $\mathfrak{M}$ by $\langle \mathfrak{M}, R \rangle \models \Delta$. We derived the negation of (iii) from the negation of (i), and this finishes the proof of Lemma 1.

Now, Lemma 1 finishes the proof of Theorem 4. □

Acknowledgements

We thank Zalán Gyenis and Gábor Sági for enjoyable (transitive) discussions on the subject.

References

[1] Andréka, H., Comer, S. D., Madarász, J. X., Németi, I. and Sayed-Ahmed, T., Epimorphisms in cylindric algebras and definability in finite variable logic. Algebra Universalis 61,3-4 (2009), 261-282.

[2] Andréka, H., Düntsch, I. and Németi, I., A nonpermutational integral relation algebra. Michigan Math. J. 39 (1992), 371–384.

[3] Andréka, H. and Németi, I., Finite-variable logics do not have weak Beth definability property. In: The road to universal logic (Festschrift for the 50th birthday of Jean-Yves Beziau Vol II). Editors: Koslow, Arnold and Buchsbaum, Arthur., Studies in Universal Logic, Birkhouser Basel, 2015, Chapter 4, pp.125-133.

[4] Barany, V., Benedikt, M. and ten Cate, B., Some model theory of guarded negation. The Journal of Symbolic Logic 83,4 (2018), 1307-1344.

[5] Barwise, J. and Feferman, S. (eds), Model-Theoretic Logics. Springer-Verlag, 1985.

[6] Benthem, J.F.A.K., Language in action: Categories, lamgdas and dynamic logic. Elsevier, 1991.

[7] Chang, C.C. and Keisler, H.J., Model theory. North-Holland, 1973. Third Edition 1990.
[8] Comer, S.D., *Galois-theory of cylindric algebras and its applications*. Trans. Amer. Math. Soc. 286 (1984), 771-785.

[9] Friedman, H., *Beth's theorem in cardinality logics*. Israel Journal of Mathematics 14 (1973), 205-212.

[10] Hirsch, R. and Hodkinson, I., *Relation algebras by games*. North-Holland, 2002.

[11] Hodkinson, I., *Finite variable logics*. Updated version of paper appeared in Bull. Europ. Assoc. Theor. Comp. Sci. 51 (1993), 111-140. with addendum in Vol 52., [http://www.doc.ic.ac.uk/~imh/papers/fvl_revised.pdf](http://www.doc.ic.ac.uk/~imh/papers/fvl_revised.pdf) 37pp

[12] Hoogland, E., Marx, M. and Otto, M., *Beth definability for the guarded fragment*. In: Ganzinger, H., McAllester D., Voronkov A. (eds) Logic for Programming and Automated Reasoning. LPAR 1999. Lecture Notes in Compute Science vol 1705, Springer, Berlin, [https://doi.org./10.1007/3-540-48242-3_17](https://doi.org./10.1007/3-540-48242-3_17).

[13] Jung, J., and Wolter, F., *Interpolant existence in the guarded fragment*. Semantic Sholar Corpus ID:221018030, 2020.

[14] Makowsky, J.A., Shelah, S. and Stavi, J., *Delta-logics and generalized quantifiers*. Annals of Mathematical Logic 10,2 (1976), 155-192.

[15] Mekler, A.H., and Shelah, S., *Stationary logic and its friends I*. Notre Dame Journal of Formal Logic 26,2 (1985), 129-137.

[16] Sain, I. and Simon, A., *Beth properties of finite variable fragments of first order logic*. Preprint, Mathematical Institute of the Hungarian Academy of Sciences, Budapest, 1992.

[17] Tarski, A. and Givant, S.R., *A formalization of set theory without variables*. AMS Colloquium Publications Vol. 41, Providence, R.I., xxi+318pp., 1987.

Alfréd Rényi Institute of Mathematics
Budapest, Reáltanoda st. 13-15, H-1053 Hungary
andreka.hajnal@renyi.hu, nemeti.istvan@renyi.hu