The Review of Symbolic Logic
Volume 17, Number 4, December 2024

TESTING DEFINITIONAL EQUIVALENCE OF THEORIES VIA AUTOMORPHISM GROUPS

HAJNAL ANDRÉKA
Alfréd Rényi Institute of Mathematics
JUDIT MADARÁSZ
Alfréd Rényi Institute of Mathematics
ISTVÁN NÉMETI
Alfréd Rényi Institute of Mathematics
and
GERGELY SZÉKELY
Alfréd Rényi Institute of Mathematics
and
Department of Natural Science, University of Public Service

Abstract. Two first-order logic theories are definitionally equivalent if and only if there is a bijection between their model classes that preserves isomorphisms and ultraproducts (Theorem 2). This is a variant of a prior theorem of van Benthem and Pearce. In Example 2, uncountably many pairs of definitionally inequivalent theories are given such that their model categories are concretely isomorphic via bijections that preserve ultraproducts in the model categories up to isomorphism. Based on these results, we settle several conjectures of Barrett, Glymour and Halvorson.

§1. Introduction.

1.1. Classical definitional equivalence. The subject of the present paper is the notion of (classical) definitional equivalence of first-order logic theories. There are various definitions of this notion scattered in the literature. Most of these define the notion for theories with disjoint languages only. We use the version defined in [22, definition 11], which does not require the languages to be disjoint. According to this definition, definitional equivalence of theories is the symmetric and transitive closure of the relation "definitional extension." This notion of definitional equivalence is shown to be the same as the more prevailing ones for disjoint languages. For example, it coincides with inter-translatability [22, theorem 8] and "having a joint definitional extension" [22, theorem 4]. We believe that making the vocabularies of theories

Received: November 27, 2022.
2020 *Mathematics Subject Classification*: Primary 03A10, 03B10, 03C20, 03C40, 08A35, Secondary 18Cxx.
Key words and phrases: definitional equivalence of theories, first-order logic, automorphism group, ultraproduct, categorical equivalence of theories, many-dimensional definitional equivalence, Morita equivalence, category of models.

doi:10.1017/S1755020323000242

disjoint is a superfluous administrative task. Besides, making vocabularies disjoint masks important intuitive features in many cases. This would be the case in the present paper, too, e.g., in Example 2 and Theorem 3.

Definitional equivalence is also defined by means of a bijection between two model classes in [17, p. 56]. According to this definition, two theories are definitionally equivalent when there is a bijection between their model classes such that connected models are definitionally equivalent via the same definitions. This property is called "model mergeability" in [22, definition 13] and is proved to coincide with definitional equivalence as used in this paper [22, theorem 7]. One of the advantages of model mergeability is that it is a kind of language-free in so far that it is insensitive to whether the signatures of the two theories overlap or not. Model mergeability is a mix of semantic and syntactic features.

A purely semantic characterization of definitional equivalence is given in [12], as follows. Two theories on disjoint languages are definitionally equivalent if and only if there is a third theory on the union of their languages such that both reduct-formation functions, from the model class of the third theory to the model classes of the two theories, respectively, are bijections. For variants of this characterization, see [7, corollary 2] and [23, claim 4]. This semantic characterization is in terms of the concrete reduct-formation functions between model classes. Theorem 2 in the present paper is a similar characterization for definitional equivalence: two theories are definitionally equivalent if and only if there is a bijection between their model classes that preserves universes, isomorphisms and ultraproducts. This is a purely semantic characterization of definitional equivalence similar to the one in [17] and different from the one in [12]. The difference is that no third theory is used and arbitrary function is used in place of the concrete reduct-formation one. The idea of using functions that preserve isomorphisms and ultraproducts already occurs in [27] where relative interpretability between first-order theories is characterized in place of definitional equivalence. For more on this, see Remark 5.

1.2. Philosophy of science. Definability theory is used quite extensively in recent philosophy of science papers (see, for example, [9, 13, 15, 19, 29]). In philosophy of science, just as in mathematical logic, several notions of equivalence are used for comparing theories. One is many-sorted definitional equivalence [4, 16, 24], which is also called many-dimensional definitional equivalence [18, 28] or Morita-equivalence [8, 15]. Many-sorted definitional equivalence allows one to re-define the universes of models in a theory; therefore, it is rather important. To distinguish definitional equivalence from many-sorted one, we sometimes call it classical definitional equivalence. Another version of equivalence of theories is bi-interpretability (see [18, 28]). Categorical equivalence of theories [8, 29] is perhaps the weakest among the equivalences used for comparing theories.

It is shown in [8] that classical definitional equivalence, many-sorted definitional equivalence and categorical equivalence of theories are strictly weaker in this order.[1] Example 2 in this paper contains pairs of theories on finite signatures that are categorically equivalent but not many-sorted definitionally equivalent (nor bi-interpretable). With this, we answer Barrett and Halvorson's questions [8, question 6.1] and [6, question 1, p. 77] concerning the importance of infinite signature in their

[1] It is not clear to us how bi-interpretability fits into this sequence.

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

counterexample. In this context, it is natural to ask how much weaker categorical equivalence is than many-sorted definitional equivalence. Theorem 2 and especially its corollaries Theorem 4 and Corollary 3 in the present paper provide a property $\mathfrak{P}$ of functors such that a functor establishing the categorical equivalence satisfies $\mathfrak{P}$ if and only if the theories are classically definitionally equivalent. This property is that the functor is concrete and preserves ultraproducts. This is an answer to [7, the question below corollary 2], [6, question 2] and [29, note 23].

The investigations in the present paper are also relevant to the so-called syntax–semantics debate in philosophy of science. The issue here is, roughly, whether it is better to consider theories occurring in science as collections of linguistical objects (e.g., sentences of a given language), or as collections of structural objects of some kind. For a summary of the debate, see [20, 23]. In this context, the need for a semantic characterization of definitional equivalence was raised in [14]. Glymour [13] pointed out that de Bouvère [12] contains such a characterization. Theorem 2 in the present paper is another such semantic characterization. An advantage of Theorem 2 is that it gives intuition about what properties of theories are preserved by definitional equivalence. Namely, by Theorem 2, a property of a theory is preserved when it can be expressed in terms of universes, isomorphisms and ultraproducts of models. Glymour [13, p. 296] conjectures that each of the following four properties is preserved by classical definitional equivalence: having a one-element model, the model class being closed under substructures, the model class being closed under unions of chains, and having an equational axiomatization. Of these, the first property is clearly preserved by definitional equivalence because it is expressed by using the universes of the models. We show, after Theorem 3, that neither one of the remaining three properties is preserved by classical definitional equivalence.

Halvorson [14, section 7] proposes the programme to investigate what structure a model class naturally has and Glymour [13, p. 297] appreciates this programme. This programme involves to endow the model class of a theory in such a way that from this structure on the model class, the theory can be recovered up to definitional equivalence. For propositional logic, Stone duality provides such a structure in the form of the Stone topology on the model class. Stone duality has been generalized to first-order logic by several authors, e.g., Makkai [26] and Awodey and Forssell [5]. Halvorson points out the relevance of Stone duality for his programme and he mentions [5, 26]. Now, from the model-structures proposed in these two papers, the first-order theory can be recovered only up to the weaker many-sorted definitional equivalence. Theorem 2 in the present paper suggests a structure on the model classes, we call this concrete ultracategory, from which a theory can be recovered up to classical definitional equivalence (and not only up to many-sorted definitional equivalence). See Remark 7. We do not know of any other structure proposed in the literature on the model classes from which a theory can be recovered up to classical definitional equivalence.

Example 2 points to an interesting difference between structural and language-based equivalences of theories. Namely, Example 2 contains pairs of theories which are not equivalent with respect to any finitely linguistic-based equivalence (see the proof of Lemma 1), yet there is a bijection between their model classes that preserves isomorphisms and ultraproducts up to isomorphism. If such a bijection preserves ultraproducts not only up to isomorphism, then it establishes definitional equivalence according to Theorem 2. This shows that preserving ultraproducts only up

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

to isomorphism, which structural properties usually do, is not enough for establishing classical definitional equivalence.

1.3. On the approach taken in the present paper. It is known that definability and automorphisms are intimately connected. Though it is not true that a relation is definable in a model if and only if all automorphisms of the model preserve the relation, something close is true: a relation is definable if and only if all automorphisms of all ultrapowers preserve (the corresponding ultrapower of) the relation (see [2, lemma 6.7.5]). This theorem has proved to be quite useful so far for establishing definability and non-definability of relations.

This paper can be viewed as a search for a similar complete method for establishing definitional equivalence and inequivalence of theories. Section 2 contains two examples. The warm-up Example 1 shows that having the same classes of automorphism groups does not entail definitional equivalence. It also motivates the notion of spectrum of concrete automorphism groups. Example 2 shows that having the same spectrum of concrete automorphism groups still does not entail definitional equivalence. It also shows the importance of preserving ultraproducts. Section 3 contains a purely semantic characterization of definitional equivalence (Theorem 2), which is also a complete method for establishing definitional equivalence by using concrete automorphism groups and ultraproducts. We then show how to use this method for establishing definitional inequivalence of two theories from Example 2 (Theorem 3). Finally, we make connections with related recent philosophy of science papers.

If not stated otherwise, we use the notation of Chang and Keisler [11].

§2. Testing with automorphism groups. We are in first-order logic. Two theories T_1 and T_2 are said to be *definitionally equivalent* when there are copies of these theories with disjoint languages which have a joint definitional extension. A copy of a theory T is a theory T' which is obtained from T by renaming some elements of the vocabulary. A definitional extension of a theory is the theory where some defined relations are added to the language. For discussion of this definition of definitional equivalence of theories, see the introduction and [22, definitions 10 and 19 and theorem 4]. Two theories are said to be *definitionally inequivalent* when they are not definitionally equivalent. When T is a theory, $\mathsf{Mod}(\mathsf{T})$ denotes the class of its models, and when K is a class of similar models, $\mathsf{Th}(\mathsf{K})$ denotes its theory, i.e., the set of formulas valid in it. When $\mathfrak{M}$ is a model, $\mathsf{Aut}(\mathfrak{M})$ denotes its concrete automorphism group, i.e., the universe of $\mathsf{Aut}(\mathfrak{M})$ is the set of all *automorphisms* of $\mathfrak{M}$ (i.e., permutations of the universe of $\mathfrak{M}$ which leave all relations of $\mathfrak{M}$ unchanged as sets) and the sole operation of $\mathsf{Aut}(\mathfrak{M})$ is the operation of composition.

$$\mathsf{Aut}(\mathsf{T}) = \{\mathsf{Aut}(\mathfrak{M}) : \mathfrak{M} \in \mathsf{Mod}(\mathsf{T})\}.$$

We begin with two examples. The first example serves to show that searching for automorphism groups occurring in one but not the other of the theories is not a complete method for showing failure of definitional equivalence.

EXAMPLE 1 (Definitionally inequivalent theories with the same automorphism groups). *We present theories* T_1 *and* T_2 *such that* $\mathsf{Aut}(\mathsf{T}_1) = \mathsf{Aut}(\mathsf{T}_2)$ *and* T_1 *is not definitionally equivalent to* T_2. *The two theories have the same language*; *this language contains two binary relation symbols* S, R. *The first theory,* T_1, *states that at most one*

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

of S *and* R *can be non-empty. The second theory,* T_2*, states in addition that when* R *is non-empty, it is asymmetric*:

$$\mathsf{T}_1 = \{\forall xy\neg\mathsf{S}(xy) \vee \forall xy\neg\mathsf{R}(xy)\}, \quad \mathsf{T}_2 = \mathsf{T}_1 \cup \{\forall xy(\mathsf{R}(xy) \rightarrow \neg\mathsf{R}(yx))\}.$$

The two theories have the same automorphism groups because of the following. Let G *denote the class of automorphism groups of all models with one binary relation, i.e.,* $\mathsf{G} = \{\mathsf{Aut}(\langle M, S\rangle) : S \subseteq M \times M\}$*. Clearly,* $\mathsf{Aut}(\mathsf{T}_1) = \mathsf{Aut}(\mathsf{T}_2) = \mathsf{G}$ *because in any model of* T_1 *or* T_2 *there is at most one nonempty relation and the empty relation does not affect the automorphism group, so* $\mathsf{Aut}(\mathsf{T}_1) \cup \mathsf{Aut}(\mathsf{T}_2) \subseteq \mathsf{G}$*. The other containment follows from the fact that neither of the theories make any restriction on* S*.*

To show that T_1 *and* T_2 *are not definitionally equivalent, we will exhibit a concrete group* $\mathfrak{G}$ *that occurs as the automorphism group for finitely many models altogether, but more models of* T_1 *than of* T_2 *have* $\mathfrak{G}$ *as their automorphism group. Let the universe of* $\mathfrak{G}$ *consist of one member, the identity map on* $H = \{0, 1\}$*. There are* 12 *binary relations on* H *altogether whose automorphism group consists only of the identity on* H*,* 2 *of these are asymmetric. Thus there are* 24 *models in* $\mathsf{Mod}(\mathsf{T}_1)$ *with automorphism group* $\mathfrak{G}$*, because in each such model of* T_1 *either* S *is empty and* R *is one of the* 12 *binary relations or the other way round. However, only* 14 *models in* $\mathsf{Mod}(\mathsf{T}_2)$ *have* $\mathfrak{G}$ *as automorphism group because either* R *is empty and* S *is one of the* 12 *above, or* S *is empty and* R *is one of the two antisymmetric relations. This shows that there is no bijection between the models of* T_1 *and* T_2 *which is such that corresponding models have the same automorphism group. Therefore, they are not model mergeable and so not definitionally equivalent.*

It may be interesting to have only infinite models for our theories. An easy modification of T_1 *and* T_2 *will do. Namely, we add both to* T_1 *and to* T_2 *the infinitely many sentences that together state that their models are infinite. We then have to modify* $\mathfrak{G}$*. The universe of the new* $\mathfrak{G}$ *consists of all permutations on* $H = \{0, 1, 2, \ldots\}$*, the set of non-negative integers, that leave* 0 *fixed.*

The previous example suggests that multiplicity of concrete automorphism groups has to be taken into account when testing definitional equivalence. We define the *spectrum of concrete automorphism groups* of a theory T as a function that to each permutation group associates the number of non-isomorphic models of T that have this group as concrete automorphism group, i.e.,

$$\mathsf{AutSpec}(\mathsf{T}) := \{\langle \mathfrak{G}, \nu(\mathfrak{G}, \mathsf{T})\rangle : \mathfrak{G} \text{ is a permutation group}\},$$

where

$$\nu(\mathfrak{G}, \mathsf{T}) := |\{\mathfrak{M} \in \mathsf{Mod}(\mathsf{T}) : \mathsf{Aut}(\mathfrak{M}) = \mathfrak{G}\}/\cong|.$$

Note that if two models have the same concrete automorphism group, then they must have the same universe.

Definitionally equivalent theories have the same spectrum of concrete automorphism groups. Therefore, for two theories to be definitional equivalent, it is necessary that they have the same spectrum of concrete automorphism groups. The most natural way of ensuring this is to require a bijection between their classes of models which preserves concrete automorphism groups as well as isomorphisms. This leads to the notion of a category of models formed from the models of a theory.

The most common way of forming a *category from the models* of a first-order logic theory is to take the models of the theory as the objects of the category and take

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

the elementary embeddings[2] between these models as morphisms of the category. Let $\mathcal{M}od(\mathsf{T})$ denote this category of models of T. Often, it is useful to investigate a category of models with fewer morphisms taken into account. The *model-iso-category* $\mathcal{M}od^{iso}(\mathsf{T})$ of a theory is defined by having $\mathsf{Mod}(\mathsf{T})$ as its class of objects and having as morphisms only the isomorphisms between models. The arguments in [29] point in the direction to deal with the category of models when only isomorphisms are taken as arrows, and not all elementary embeddings. The idea is that in many realistic cases, just as ones dealt with in [29], the scientific theory is not defined by a first-order logic theory, yet one has a clear sense of what models and isomorphisms between these models can be.

Model categories come with a natural *forgetful functor* to the category Set of all sets. These functors assign the universe M to a model $\mathfrak{M}$ and they assign the "function content" to a morphism between two models. These are so natural in model theory that they are called *the* forgetful functor. For definitions, see [1, definition 5.1(1)]. A functor F between model categories is called a *concrete functor* iff it commutes with these natural forgetful functors. Thus a functor F between model categories is a concrete one iff the universes of connected models are the same and if connected morphisms are the same as functions between the universes of models. Two model categories are called *concretely isomorphic* iff there is a concrete isomorphism between them.

Existence of concrete isomorphism between model-iso-categories is a natural generalization of having the same spectrum of concrete automorphism groups. The next theorem says that, in fact, it is not a generalization.

THEOREM 1. *Two theories have the same spectrum of concrete automorphism groups if and only if their model-iso-categories are concretely isomorphic.*

Proof. Let T_1 and T_2 be first-order theories and assume that $\mathsf{AutSpec}(\mathsf{T}_1) = \mathsf{AutSpec}(\mathsf{T}_2)$. We are going to define a concrete isomorphism b between their model-iso-categories.

The identity element of a permutation group is always of the form $\{(a, a) : a \in A\}$ for some A; let us call this A the base of the permutation group. Let $\mathfrak{G}, \mathfrak{H}$ be permutation groups, let $h : A \to B$ be a bijection between the bases of $\mathfrak{G}$ and $\mathfrak{H}$, and define $\overline{h}(g) = h \circ g \circ h^{-1}$ for all $g \in G$. Then it is easy to see that $\overline{h}$ is an isomorphism between $\mathfrak{G}$ and $\mathfrak{H}$; we say that it is the base-isomorphism induced by h. A *base-isomorphism* between two permutation groups $\mathfrak{G}, \mathfrak{H}$ is an isomorphism between them that is induced by some h. We will also use the fact that if $h : \mathfrak{M} \to \mathfrak{N}$ is an isomorphism between the structures $\mathfrak{M}, \mathfrak{N}$, then $\overline{h}$ is a base-isomorphism between their automorphism groups.

Let $\mathcal{G}$ be a class of representatives for the base-isomorphism classes of permutation groups. That is, each permutation group has a base-isomorphic copy in $\mathcal{G}$ and the elements of $\mathcal{G}$ are pairwise non-base-isomorphic. For any permutation group $\mathfrak{G} \in \mathcal{G}$ choose $\nu(\mathfrak{G}, \mathsf{T}_1)$-many non-isomorphic models $\mathfrak{M}(\mathfrak{G}, i)$ of T_1, for $i < \nu(\mathfrak{G}, \mathsf{T}_1)$, and similarly choose $\nu(\mathfrak{G}, \mathsf{T}_2) = \nu(\mathfrak{G}, \mathsf{T}_1)$ non-isomorphic models $\mathfrak{M}'(\mathfrak{G}, i)$ of T_2, with concrete automorphism group $\mathfrak{G}$. Then the models $\mathfrak{M}(\mathfrak{G}, i)$ for $\mathfrak{G} \in \mathcal{G}$ are pairwise non-isomorphic, i.e., $\mathfrak{M}(\mathfrak{G}, i) \cong \mathfrak{M}(\mathfrak{H}, j)$ for some $\mathfrak{G}, \mathfrak{H}, i, j$ implies $\mathfrak{G} = \mathfrak{H}$ and $i = j$. Similarly, the models $\mathfrak{M}'(\mathfrak{G}, i)$ are pairwise non-isomorphic.

[2] For the definition of elementary embedding, see [11, p. 84].

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

Let $\mathfrak{M} \in \mathsf{Mod}(\mathsf{T}_1)$. There is a unique $\mathfrak{M}(\mathfrak{G}, i)$ isomorphic to $\mathfrak{M}$, as follows. Let $\mathfrak{H}$ be the concrete automorphism group of $\mathfrak{M}$ and let $\mathfrak{G} \in \mathcal{G}$ be base-isomorphic to $\mathfrak{H}$ via the base-isomorphism $\overline{h} : \mathfrak{H} \to \mathfrak{G}$. Then the automorphism group of $h(\mathfrak{M})$ is $\mathfrak{G} \in \mathcal{G}$; thus, $h(\mathfrak{M})$ is isomorphic to $\mathfrak{M}(\mathfrak{G}, i)$ for some i, by our construction. Choose any isomorphism f mapping $\mathfrak{M}(\mathfrak{G}, i)$ to $\mathfrak{M}$ and let us define

$$b(\mathfrak{M}) := f(\mathfrak{M}'(\mathfrak{G}, i)).$$

We show that $b(\mathfrak{M})$ is well-defined, i.e., it does not depend on which isomorphism f we choose. Let g be any other isomorphism between $\mathfrak{M}(\mathfrak{G}, i)$ and $\mathfrak{M}$; we show that $f(\mathfrak{M}'(\mathfrak{G}, i)) = g(\mathfrak{M}'(\mathfrak{G}, i))$. Indeed, $g = f \circ \alpha$ for $\alpha = f^{-1} \circ g \in \mathsf{Aut}(\mathfrak{M}(\mathfrak{G}, i)) = \mathfrak{G}$. But $\alpha \in \mathfrak{G} = \mathsf{Aut}(\mathfrak{M}'(\mathfrak{G}, i))$, so $g(\mathfrak{M}'(\mathfrak{G}, i)) = f(\alpha(\mathfrak{M}'(\mathfrak{G}, i))) = f(\mathfrak{M}'(\mathfrak{G}, i))$.

We define b on the morphisms. Let $h : \mathfrak{M} \to \mathfrak{N}$ be an isomorphism between $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}(\mathsf{T}_1)$. We have seen that $f : \mathfrak{M}(\mathfrak{G}, i) \to \mathfrak{M}$ for some $f, \mathfrak{G}, i$ and so $g : \mathfrak{M}(\mathfrak{G}, i) \to \mathfrak{N}$ for $g = h \circ f$. Thus, by definition, $b(\mathfrak{M}) = f(\mathfrak{M}'(\mathfrak{G}, i))$ and $b(\mathfrak{N}) = g(\mathfrak{M}'(\mathfrak{G}, i))$. Hence, $h : b(\mathfrak{M}) \to b(\mathfrak{N})$ is an isomorphism by $g \circ f^{-1} = h \circ f \circ f^{-1}$. We define

$$b(h) := h.$$

We now show that b is an isomorphism between the model-iso-categories of T_1 and T_2. First we show that the function $b : \mathsf{Mod}(\mathsf{T}_1) \to \mathsf{Mod}(\mathsf{T}_2)$ defined this way is a bijection between $\mathsf{Mod}(\mathsf{T}_1)$ and $\mathsf{Mod}(\mathsf{T}_2)$. Indeed, let $\mathfrak{M}' \in \mathsf{Mod}(\mathsf{T}_2)$ be any model. There is a unique $\mathfrak{M}'(\mathfrak{G}, i)$ isomorphic to it, say via $f : \mathfrak{M}'(\mathfrak{G}, i) \to \mathfrak{M}'$. Let $\mathfrak{M} = f(\mathfrak{M}(\mathfrak{G}, i))$, then $\mathfrak{M}' = b(\mathfrak{M})$, by the definition of b. Thus, the range of b is $\mathsf{Mod}(\mathsf{T}_2)$. To see that b is one-to-one, let $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}(\mathsf{T}_1)$. Assume that $b(\mathfrak{M}) = b(\mathfrak{N})$. By the definition of b, there are $\mathfrak{M}(\mathfrak{G}, i), \mathfrak{M}(\mathfrak{H}, j)$ and isomorphisms $f : \mathfrak{M}(\mathfrak{G}, i) \to \mathfrak{M}$, $g : \mathfrak{M}(\mathfrak{H}, j) \to \mathfrak{N}$ such that $b(\mathfrak{M}) = f(\mathfrak{M}'(\mathfrak{G}, i))$ and $b(\mathfrak{N}) = g(\mathfrak{M}'(\mathfrak{H}, j))$. By $b(\mathfrak{M}) = b(\mathfrak{N})$ then $\mathfrak{M}'(\mathfrak{G}, i)$ is isomorphic to $\mathfrak{M}'(\mathfrak{H}, j)$; therefore, $(\mathfrak{G}, i) = (\mathfrak{H}, j)$ and $f(\mathfrak{M}'(\mathfrak{G}, i)) = g(\mathfrak{M}'(\mathfrak{G}, i))$. Thus, $f^{-1} \circ g \in \mathsf{Aut}(\mathfrak{M}'(\mathfrak{G}, i)) = \mathfrak{G}$. So, $\mathfrak{M} = f(\mathfrak{M}(\mathfrak{G}, i)) = f((f^{-1} \circ g)(\mathfrak{M}(\mathfrak{G}, i))) = g(\mathfrak{M}(\mathfrak{G}, i)) = g(\mathfrak{M}(\mathfrak{H}, j)) = \mathfrak{N}$.

We turn to the proof for b being a bijection between the set of isomorphisms from $\mathfrak{M}$ to $\mathfrak{N}$ and the set of isomorphisms from $b(\mathfrak{M})$ to $b(\mathfrak{N})$, for any $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}(\mathsf{T}_1)$. To show surjectivity, let $h : b(\mathfrak{M}) \to b(\mathfrak{N})$. By the definition of $b(\mathfrak{M})$, we have that $\mathfrak{M} = f(\mathfrak{M}(\mathfrak{G}, i))$ and $b(\mathfrak{M}) = f(\mathfrak{M}'(\mathfrak{G}, i))$, for some $f, \mathfrak{G}, i$. Thus, $f : \mathfrak{M}'(\mathfrak{G}, i) \to b(\mathfrak{M})$, and so $h \circ f : \mathfrak{M}'(\mathfrak{G}, i) \to b(\mathfrak{N})$, by $h : b(\mathfrak{M}) \to b(\mathfrak{N})$. Let $\mathfrak{N}' = f(h(\mathfrak{M}(\mathfrak{G}, i))$. By the definition of $b(\mathfrak{N}')$ then $b(\mathfrak{N}') = (h \circ f)(\mathfrak{M}'(\mathfrak{G}, i)) = b(\mathfrak{N})$. Thus, $\mathfrak{N}' = \mathfrak{N}$ because b is one-to-one on $\mathsf{Mod}(\mathsf{T}_1)$, i.e., $\mathfrak{N} = (h \circ f)(\mathfrak{M}(\mathfrak{G}, i)) = h(f(\mathfrak{M}(\mathfrak{G}, i))) = h(\mathfrak{M})$. Thus, $h : \mathfrak{M} \to \mathfrak{N}$ is an isomorphism and $b(h) = h$. By definition, it is clear that b is one-to-one on the morphisms, and also that it preserves composition of morphisms in both directions. This finishes the proof for b being a category theoretical isomorphism between the model-iso-categories of T_1 and T_2. It is concrete, by its definition.

In the other direction, assume that b is a concrete isomorphism between $\mathcal{M}od^{iso}(\mathsf{T}_1)$ and $\mathcal{M}od^{iso}(\mathsf{T}_2)$. Then $\mathsf{Aut}(\mathfrak{M}) = \mathsf{Aut}(b(\mathfrak{M}))$, and $\mathfrak{M} \cong \mathfrak{N}$ iff $b(\mathfrak{M}) \cong b(\mathfrak{N})$, for all $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}(\mathsf{T}_1)$. Therefore, $\mathsf{AutSpec}(\mathsf{T}_1) = \mathsf{AutSpec}(\mathsf{T}_2)$. □

The next example shows that having the same spectrum of automorphism groups still does not entail definitional equivalence. It is more refined than the previous one. We will see that it shows, in a sense, a limit till we still can get failure of definitional equivalence (compare Lemma 2 with Theorem 2). It also serves as a counterexample to Barrett and Halvorson's conjecture that, among first-order logic theories with finite signatures,

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

categorical equivalence implies many-sorted (Morita) definitional equivalence. With this, we answer in the negative [8, question 6.1] as well as [6, question 1, p. 77].

EXAMPLE 2 (Uncountably many theories with the same model category). *We present continuum many complete theories on a finite similarity type with the same automorphism spectrum such that no two of them are definitionally equivalent. Moreover, their model categories are isomorphic via concrete functors which preserve ultraproducts up to isomorphism, and further, no two of the theories are even many-sorted definitionally (Morita) equivalent. (The latter notion will be introduced later, below Lemma 2.)*

We work in the similarity type which contains one constant symbol 0, *one unary function symbol* suc, *and one unary relation symbol* R. *Let* n *be a natural number, then* $\mathsf{suc}^n(x)$ *denotes the term where* suc *is* n*-times applied to* x, *i.e.,* $\mathsf{suc}^0(x) = x$ *and* $\mathsf{suc}^{(n+1)}(x) = \mathsf{suc}(\mathsf{suc}^n(x))$. *For each subset S of the natural numbers* ω, *let*

$$\mathsf{T}(S) := \{\mathsf{R}(\mathsf{suc}^n(0)) : n \in S\} \cup \{\neg\mathsf{R}(\mathsf{suc}^n(0)) : n \notin S\} \cup \mathsf{Th}(\langle\omega, 0, \mathsf{suc}\rangle),$$

where $\langle\omega, 0, \mathsf{suc}\rangle$ *denotes natural numbers* ω *with zero as* 0 *and the successor function as* suc.

A set S of natural numbers is called irregular *if all finite patterns occur in it. In more detail, let* $n > 0$ *be a positive number and let* $P \subseteq \{0, 1, ..., n-1\}$. *We say that the P, n-pattern occurs at* x *in S if* $\{m < n : suc^m(x) \in S\} = P$. *For example,* $S = \{0, 2, 4, 6, ...\}$ *is not irregular, because the pattern* $\{0, 1\}, 2$ *does not occur in it (i.e.,* $x, \mathsf{suc}(x) \in S$ *does not hold for any* $x \in \omega$).

There are continuum many irregular subsets of ω. *This can be seen as follows. Construct an infinite sequence of* $0, 1, x$ *by first laying the two* 0, 1*-sequences of length* 1 *after each other in alphabetical order, then mark the next number by an* x, *then lay the four* 0, 1*-sequences of length* 2 *after each other in alphabetical order and mark the next number by an* x, *etc. This sequence will begin like* $\langle 0, 1, x, 0, 0, 0, 1, 1, 0, 1, 1, x, 0, 0, 0, ...\rangle$. *There are infinitely many* x*s in this sequence and so there are continuum many ways of replacing the* x*s with* 0 *or* 1. *Each of the continuum many* 0, 1*-sequences that are obtained this way is a characteristic function of an irregular set. This proves that there are at least continuum many irregular sets. There can be at most continuum many irregular subsets of* ω *since there are continuum many subsets of* ω.

We are going to show that the model categories $\mathcal{M}od(\mathsf{T}(S))$ *for irregular sets S are isomorphic to each other in a strong constructive way (see Lemma 2).*

We say that $\mathfrak{N}$ is an *induced subalgebra* of $\mathfrak{M}$ when the R-free part of $\mathfrak{N}$ is a subalgebra of the R-free part of $\mathfrak{M}$ and the R-relation of $\mathfrak{N}$ is that of $\mathfrak{M}$ restricted to the universe of $\mathfrak{N}$. For the definition of elementary submodel, see [11, p. 84].

LEMMA 1. *Let* $S \subseteq \omega$ *be irregular. Then (i) and (ii) below hold.*

(i) *The elementary submodels of a model of* $\mathsf{T}(S)$ *are exactly its induced subalgebras.*
(ii) $\mathsf{T}(S)$ *is a complete theory.*

Proof. Let $\mathbb{N}$ denote the set of natural numbers with 0 as constant 0 and the successor function as unary distinguished function suc, and let $\mathbb{Z}$ denote the set of integers with the successor function as unary distinguished function suc. Note that $\mathbb{Z}$ does not have 0 in its language. Any model of $\mathsf{Th}(\langle\omega, 0, \mathsf{suc}\rangle)$ is a disjoint union of one copy of $\mathbb{N}$ together with some copies of $\mathbb{Z}$. When k is negative, $\mathsf{suc}^k(x) = y$ means $\mathsf{suc}^{-k}(y) = x$, and we say that $\mathsf{suc}^k(x)$ exists when such a y exists. In models of $\mathsf{Th}(\langle\omega, 0, \mathsf{suc}\rangle)$ such

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

a y is unique when it exists. When $\mathfrak{N}$ is a model of $\mathsf{Th}(\langle \omega, 0, \mathsf{suc}\rangle)$, by a $\mathbb{Z}$-*part* of $\mathfrak{N}$ we mean a subset of N of the form $\{\mathsf{suc}^n(a) : n \in \omega\} \cup \{\mathsf{suc}^{-n}(a) : n \in \omega\}$ for some $a \in N$. By a $\mathbb{Z}$-model we mean $\mathbb{Z}$ together with a unary relation R and by $\langle \mathbb{N}, S\rangle$ we mean $\mathbb{N}$ expanded with S as the unary relation R. We are going to prove the following statement $(*)$.

In $(*)$ as well as later on, we will use ultraproducts [11, chap. 4]. As in [11], when U is an ultrafilter on the set I and $\langle \mathfrak{M}_i : i \in I\rangle$ is an I-sequence of similar models, $\prod_U \langle \mathfrak{M}_i : i \in I\rangle$, or sloppily just $\prod_U \mathfrak{M}_i$, denotes the U-ultraproduct of the models $\mathfrak{M}_i$ and y_U denotes the equivalence-class of y in $\Pi_U \mathfrak{M}_i$, for $y \in \Pi_{i\in I} M_i$. When each $\mathfrak{M}_i = \mathfrak{A}$ for some $\mathfrak{A}$, we call $\prod_U \mathfrak{M}_i$ an ultrapower of $\mathfrak{A}$ and we denote it by $\Pi_U \mathfrak{A}$.

$(*)$ Assume that $\mathfrak{M}$ is a countable model of $\mathsf{T}(S)$ and U is a nonprincipal ultrafilter on a countable set I. Then $\Pi_U \mathfrak{M}$ is isomorphic to a disjoint union of a copy of $\langle \mathbb{N}, S\rangle$ with continuum many copies of each possible $\mathbb{Z}$-model.

Indeed, $(*)$ is true because each $\mathbb{Z}$-model can be put together in the ultrapower from its finite parts which are patterns occurring in $\mathfrak{M}$, and in fact, each such pattern occurs infinitely many times in $\mathfrak{M}$. In more detail: Let $\langle \mathbb{Z}, R\rangle$ be any $\mathbb{Z}$-model; we show that continuum many disjoint copies of it occurs in the ultrapower of $\mathfrak{M}$. We may assume that $I = \omega$ because I is countable. For each $n > 0$ let $R_n := \{m \leq 2n : m - n \in R\}$. The pattern $R_n, 2n + 1$ occurs in S because S is irregular. In fact, each pattern occurs in an irregular set infinitely many times because each finite pattern has infinitely many different extensions to other finite patterns and each of these patterns occurs in the irregular set. Let X_n be the set of xs where $R_n, 2n + 1$ occurs in S and let $Y_n := \{x + n : x \in X_n\}$. First we show that in $\Pi_U \mathfrak{M}$ each element of $\Pi_U Y_n$ lies on a copy of $\langle \mathbb{Z}, R\rangle$. Indeed, let $x_n \in X_n$ and $y_n := x_n + n$ for all $n \in \omega$. Let $y := \langle y_n : n \in \omega\rangle$, and let $k \in \mathbb{Z}$ be arbitrary. We will show that $\mathsf{suc}^k(y_U)$ exists and $k \in R$ iff $R(\mathsf{suc}^k(y_U))$ in $\Pi_U \mathfrak{M}$. By our definitions, for all n such that $2n \geq k$ we have that $k \in R$ iff $k + n \in R_n$ iff $x_n + k + n \in S$ iff $y_n + k \in S$ iff $R(\mathsf{suc}^k(y_n))$ in $\mathfrak{M}$. Since U is nonprincipal on $I = \omega$, this means that $R(\mathsf{suc}^k(y_U))$ in $\Pi_U \mathfrak{M}$. We have seen that y_U is in a copy of $\langle \mathbb{Z}, R\rangle$ for all $y \in \Pi_{n\in\omega} Y_n$. Since each Y_n is countably infinite, the cardinality of $\Pi_U Y_n$ is continuum (see [11, proposition 4.3.9]). Since each copy of $\langle \mathbb{Z}, R\rangle$ is countable, this means that $\Pi_U \mathfrak{M}$ contains continuum many disjoint copies of $\langle \mathbb{Z}, R\rangle$, and we are done with proving $(*)$.

Proof of (i): An elementary submodel of $\mathfrak{M}$ has to be an induced subalgebra. Conversely, assume that $\mathfrak{N}$ is an induced subalgebra of $\mathfrak{M}$; we show that it is an elementary submodel. We will use the testing method in [11, proposition 3.1.2]. Thus, assume that $\varphi(\bar{x}, y)$ is a first-order logic formula in the language of $\mathfrak{M}$, assume that $\bar{a}$ is an appropriate sequence of elements of $\mathfrak{N}$, and $\mathfrak{M} \models \exists y \varphi(\bar{a}, y)$. We have to show the existence of $a' \in \mathfrak{N}$ such that $\mathfrak{M} \models \varphi(\bar{a}, a')$. We have $\Pi_U \mathfrak{M} \models \exists y \varphi(d(\bar{a}), y)$ since the diagonal (or natural) embedding d of a model into its ultrapower is an elementary one [11, corollary 4.1.13]. Let $b \in \Pi_U \mathfrak{M}$ be such that $\Pi_U \mathfrak{M} \models \varphi(d(\bar{a}), b)$. Now, $\Pi_U \mathfrak{N}$ is an induced subalgebra of $\Pi_U \mathfrak{M}$, by $\mathfrak{N}$ being an induced subalgebra of $\mathfrak{M}$. There are infinitely many $\mathbb{Z}$-parts in $\Pi_U \mathfrak{N}$ that do not contain any element of $d(\bar{a})$ and that are isomorphic to the $\mathbb{Z}$-part of $\Pi_U \mathfrak{M}$ containing b, by $(*)$. Take an automorphism of $\Pi_U \mathfrak{M}$ that interchanges the $\mathbb{Z}$-part of b with any of such a $\mathbb{Z}$-part of $\Pi_U \mathfrak{N}$ and leaves anything else fixed. There is such an automorphism by the choice of the $\mathbb{Z}$-part of $\Pi_U \mathfrak{N}$ and since $\mathfrak{M} \in \mathsf{Mod}\mathsf{T}(S)$. Let c be the image of b under such an automorphism, then $\Pi_U \mathfrak{M} \models \varphi(d(\bar{a}), c)$, since the automorphism leaves the elements of $d(\bar{a})$ fixed.

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

Then $\mathfrak{M} \models \varphi(\bar{a}, a')$ for some $a' \in \mathfrak{N}$ by the fundamental theorem of ultraproducts [11, theorem 4.1.9(ii)] since $c \in \Pi_U \mathfrak{N}$. We have shown that $\mathfrak{N}$ is an elementary submodel of $\mathfrak{M}$.

Proof of (ii): Assume that $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}\mathsf{T}(S)$; we have to show that $\mathfrak{N}$ is elementarily equivalent to $\mathfrak{M}$. We may assume that $\mathfrak{M}$ and $\mathfrak{N}$ are countable, by the downward Löwenheim–Skolem–Tarski theorem [11, corollary 2.1.4]. Now, $\mathfrak{M}$ and $\mathfrak{N}$ are elementarily equivalent by $(*)$, since they have isomorphic ultrapowers. The proof of Lemma 1 is complete. □

By using Lemma 1, we now specify a functor F between the model categories of $\mathsf{T}(S)$ and $\mathsf{T}(Z)$, for any irregular sets S and Z. Let $\mathfrak{M} = \langle M, 0, \mathsf{suc}, R\rangle \in \mathsf{Mod}(\mathsf{T}(S))$. We define

$$F(\mathfrak{M}) := \langle M, 0, \mathsf{suc}, (R \setminus \{\mathsf{suc}^k(0) : k \in S\}) \cup \{\mathsf{suc}^k(0) : k \in Z\}\rangle.$$

That is, $F(\mathfrak{M})$ is defined to be $\mathfrak{M}$ except that R on the $\mathbb{N}$-part of $\mathfrak{M}$ is changed to be the R of the $\mathbb{N}$-part of a $\mathsf{T}(Z)$ model. For an elementary embedding $f : \mathfrak{M} \to \mathfrak{N}$ between $\mathfrak{M}, \mathfrak{N} \in \mathsf{Mod}(\mathsf{T}(S))$, let us define

$$F(f) := f.$$

Lemma 2. *Let S and Z be irregular sets and let F be the function defined above.*

(i) *F is a concrete isomorphism between $\mathcal{M}od(\mathsf{T}(S))$ and $\mathcal{M}od(\mathsf{T}(Z))$.*
(ii) *F preserves ultraproducts of models up to isomorphism, i.e., F takes an ultraproduct of models of $\mathsf{T}(S)$ to a model isomorphic to the corresponding ultraproduct of the F-images of the models.*

Proof. F is a functor, since (f is an elementary embedding of $\mathfrak{M}$ into $\mathfrak{N}$ if and only if it is an elementary embedding of $F(\mathfrak{M})$ into $F(\mathfrak{N})$), by Lemma 1 and the construction of F. Thus F is a concrete isomorphism by its construction.

To show that F preserves ultraproducts up to isomorphism, let U be an ultrafilter on a set I and let $\mathfrak{M}_i \in \mathsf{Mod}(\mathsf{T}(S))$ for all $i \in I$. We will define an isomorphism j between $F(\Pi_U \mathfrak{M}_i)$ and $\Pi_U F(\mathfrak{M}_i)$. Let $\mathfrak{N}_i$ denote the $\mathbb{N}$-part of $\mathfrak{M}_i$, for each $i \in I$. Then each $\mathfrak{N}_i$ is isomorphic to $\langle \mathbb{N}, \{\mathsf{suc}^k(0) : k \in S\}\rangle$ by $\mathfrak{M}_i \in \mathsf{Mod}\mathsf{T}(S)$. Let $y := \langle y_i : i \in I\rangle \in \Pi_{i\in I} M_i$. We define

$$j(y_U) := y_U \quad \text{if} \quad \{i \in I : y_i \notin N_i\} \in U.$$

To define j on the rest, assume first that U is not ω^+-complete. Then by a straightforward modification of the proof of $(*)$ we get that both $\Pi_U \mathfrak{N}_i$ and $\Pi_U F(\mathfrak{N}_i)$ consist of one $\mathbb{N}$-model together with continuum many copies of all possible $\mathbb{Z}$-models. If U is ω^+-complete, then both $\Pi_U \mathfrak{N}_i$ and $\Pi_U F(\mathfrak{N}_i)$ consist of one $\mathbb{N}$-model only by [11, proposition 4.2.4]. In both cases, there is an isomorphism between $F(\Pi_U \mathfrak{N}_i)$ and $\Pi_U F(\mathfrak{N}_i)$. We define

$$j \text{ be any isomorphism between } F(\Pi_U \mathfrak{N}_i) \text{ and } \Pi_U F(\mathfrak{N}_i)$$

and be identity on the rest. It is not difficult to check that $j : F(\Pi_U \mathfrak{M}_i) \to \Pi_U F(\mathfrak{M}_i)$ is an isomorphism. This finishes the proof of Lemma 2. □

We have seen that, for any two irregular sets S and Z, the model categories of $\mathsf{T}(S)$ and $\mathsf{T}(Z)$ are rather close to each other in a constructive way. We now turn to definability issues between $\mathsf{T}(S)$ and $\mathsf{T}(Z)$. In logic, there are two weaker versions

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

of definitional equivalence between theories in use. One is called many-dimensional [16, 28] or many-sorted [4, 24] definitional equivalence, and it is also called Morita equivalence of theories [8, 15]. The other is called bi-interpretability between theories [18, 28]. Both notions are weaker than definitional equivalence between first-order logic theories in the sense that when T_1 and T_2 are definitionally equivalent, then they are also many-dimensionally equivalent and bi-interpretable. For a comparison of these notions, see [8]. We will rely on the definitions in the mentioned references; we do not recall them.

Corollary 1.

(i) *All the theories* $\mathsf{T}(S)$ *with* S *irregular have the same spectrum of automorphism groups.*

(ii) *There is an uncountable set* $\mathcal{S}$ *of irregular sets such that no* $\mathsf{T}(S)$ *and* $\mathsf{T}(Z)$ *for distinct* $S, Z \in \mathcal{S}$ *are definitionally equivalent, many-sorted definitionally equivalent or bi-interpretable.*

Proof. (i) follows from Lemmas 1 and 2. Each of definitional equivalence, many-sorted definitional equivalence and bi-interpretability of two theories can be specified by the use of finitely many formulas on the language of the theories (see the references given for their definitions). Therefore, a concrete theory can be definitionally equivalent to at most countably many theories on a given other similarity type. This implies that of the continuum many theories $\mathsf{T}(S)$ on the same language; there are continuum many pairwise non-equivalent theories (neither many-sorted equivalent nor bi-interpretable). This finishes the proof of Corollary 1. With this, the presentation of Example 2 is finished. □

The essence of Example 2 above is that the model categories of $\mathsf{T}(S)$ for irregular sets S are almost the same because the R on the $\mathbb{N}$-parts do not play a role in this category. However, the R on the $\mathbb{N}$-part can code more "information" than available (syntactical) translations between theories and therefore many such theories have to be definitionally inequivalent.

Remark 3 (F does not preserve ultraproducts). *The functor* F *constructed above Lemma 2 does not preserve ultraproducts; it preserves ultraproducts only up to isomorphism. This follows from Theorem 2 in the next section and Corollary 1(ii). We now want to provide a concrete example that shows that* F *does not preserve ultraproducts. Recall the continuum many irregular sets constructed above Lemma 1. Let* S_0 *and* S_1 *be the irregular sets we obtain by filling all the* x*s with* 0 *and by filling all the* x*s with* 1*, respectively. Then* $S_0 \subseteq S_1$ *and* $S_1 \setminus S_0$ *is infinite. Let* $\mathfrak{N}_i := \langle \omega, 0, \mathsf{suc}, R_i \rangle$ *where* $R_i = \{\mathsf{suc}^k(0) : k \in S_i\}$ *for* $i = 0, 1$*. Consider the functor* F *between* $\mathsf{T}(S_0)$ *and* $\mathsf{T}(S_1)$*. Then* $F(\mathfrak{N}_0) = \mathfrak{N}_1$ *by definition of* F*. Let* $X \subseteq \omega$ *be an infinite set which is disjoint from* S_0 *but is contained in* S_1*, let* U *be a nonprincipal ultrafilter on* $I = \omega$ *such that* $X \in U$ *and let* $y = \langle \mathsf{suc}^k(0) : k \in \omega \rangle$*. Then* $R(y_U)$ *does not hold in* $F(\Pi_U \mathfrak{N}_0)$*, while* $R(y_U)$ *holds in* $\Pi_U F(\mathfrak{N}_0)$ *showing that the two structures are not the same (though, isomorphic). We will see in the next section that in fact* $\mathsf{T}(S_0)$ *is not definitionally equivalent to* $\mathsf{T}(S_1)$ *because there is no concrete isomorphism between their model categories that would preserve ultraproducts (see Theorem 3).*

Remark 4 (More striking example). *We can modify the above example to give a more striking counterexample to the conjecture in* [8] *which at the same time is analogous to*

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

the example in the proof of [8, *theorem* 5.7]. *The similarity type of* T_1 *and* T_2 *will be as in Example* 2. *The first theory,* T_1 *states only that* 0 *is not in relation* R:

$$\mathsf{T}_1 = \{\neg \mathsf{R}(0)\}.$$

For defining T_2, *take any irregular set* S *such that* $0 \in S$, *and then* T_2 *is*

$$\mathsf{T}_2 = \{\mathsf{R}(0) \to \varphi \ : \ \varphi \in \mathsf{T}(S)\}.$$

That is, the models of T_2 *are those of* T_1 *together with all the models of* $\mathsf{T}(S)$. *Now,* T_1 *is finitely axiomatized, while it is easy to see that* T_2 *cannot be axiomatized finitely* (*e.g., by showing that the complement of* $\mathsf{Mod}(\mathsf{T}_2)$ *is not closed under ultraproducts*). *Since intertranslatability is an essence of definitional equivalence both for the classical and the many-sorted versions, as, e.g., Halvorson* [15] *argues, being finitely axiomatized is preserved, for theories of finite similarity types, by the weaker many-sorted* (*Morita*) *definitional equivalence also. So,* T_1 *and* T_2 *are not Morita definitionally equivalent. However, their model categories are equivalent, in fact isomorphic, as in* [8, *theorem* 5.7]: *a model category consists of isolated islands of* $\mathcal{M}od(\mathsf{Th}(\mathfrak{M}))$ *for the models* $\mathfrak{M}$ *of the theory* (*because if there is a morphism between* $\mathfrak{M}$ *and* $\mathfrak{N}$ *then* $\mathfrak{M}$ *and* $\mathfrak{N}$ *are elementarily equivalent since this morphism is an elementary embedding of* $\mathfrak{M}$ *into* $\mathfrak{N}$). *Now, by Lemma* 1, *the extra island of* $\mathcal{M}od(\mathsf{T}_2)$ *is isomorphic to any one of the continuum many islands* $\mathcal{M}od(\mathsf{T}(Z))$ *of* $\mathcal{M}od(\mathsf{T}_1)$ *where* Z *is an irregular set with* $0 \notin Z$.

§3. Testing with automorphism groups and ultraproducts. We are ready to turn to the positive results of this paper. Lemma 2 suggests that, besides automorphism groups, ultraproducts have to be taken into account in testing definitional equivalence. Indeed, Theorem 2 gives such a characterization making our search for a complete testing method successful.

The following theorem is a semantic characterization of definitional equivalence. It is a slight modification of the theorem in [27] which is a semantic characterization of restricted interpretations between theories. For a closely related theorem, see also [21, theorem 12.1].

THEOREM 2. *Two theories* T_1 *and* T_2 *are definitionally equivalent if and only if there is a bijection* b *between their model classes that satisfies the following two conditions.*

(i) *An isomorphism between different models of* T_1 *is an isomorphism between their* b*-images and vice versa. In particular, the universes of* $\mathfrak{M}$ *and* $b(\mathfrak{M})$ *are the same.*
(ii) *Ultraproducts are preserved by* b *in the sense that* $b(\prod_U \mathfrak{M}_i) = \prod_U b(\mathfrak{M}_i)$ *for all ultrafilters* U *and models* $\mathfrak{M}_i$ *in* $\mathsf{Mod}(\mathsf{T}_1)$.

Proof. The proof follows that of [27, theorem]. Let assume first that the languages of T_1 and T_2 are disjoint. Assume that we have a bijection b satisfying (i) and (ii). We define a class K of models in the similarity type as the union of the similarity types of T_1 and T_2 and we will show that the first-order logic theory of K is a joint definitional extension for both T_1 and T_2. For a model $\mathfrak{M} \in \mathsf{Mod}(\mathsf{T}_1)$, let

$$\overline{\mathfrak{M}} = \langle \mathfrak{M}, b(\mathfrak{M}) \rangle$$

denote the model whose universe is the joint universe of $\mathfrak{M}$ and $b(\mathfrak{M})$, the relation and function symbols of the language of T_1 are interpreted as in $\mathfrak{M}$, and the relation and

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

function symbols of the language of T_2 are interpreted as in $b(\mathfrak{M})$. Let

$$\mathsf{K} = \{\langle \mathfrak{M}, b(\mathfrak{M})\rangle : \mathfrak{M} \in \mathsf{Mod}(\mathsf{T}_1)\}.$$

We will show that K is axiomatizable, i.e., $\mathsf{K} = \mathsf{ModThK}$. We use [11, corollary 6.1.16(i)], which states that a class is elementary if and only if it is closed under taking ultraproducts and isomorphic images, and the complement is closed under ultrapowers. Now, K is closed under ultraproducts and isomorphisms by conditions (ii) and (i), since $\mathsf{Mod}(\mathsf{T}_1)$ is elementary. Assume that $\mathfrak{A} = \langle \mathfrak{M}, \mathfrak{N}\rangle$ is such that an ultrapower $\Pi_U\mathfrak{A}$ is in K. We have to show that $\mathfrak{A} \in \mathsf{K}$. Now, $\Pi_U\mathfrak{A} = \langle \Pi_U\mathfrak{M}, \Pi_U\mathfrak{N}\rangle$, and then $\Pi_U\mathfrak{A} \in \mathsf{K}$ means that $\Pi_U\mathfrak{N} = b(\Pi_U\mathfrak{M})$. By condition (ii) we have $b(\Pi_U\mathfrak{M}) = \Pi_U b(\mathfrak{M})$. Thus we have $\Pi_U\mathfrak{N} = \Pi_U b(\mathfrak{M})$. This implies $\mathfrak{N} = b(\mathfrak{M})$ since any structure $\mathfrak{B}$ can be recovered from $\Pi_U\mathfrak{B}$. We have seen that K is an elementary class; let

$$\mathsf{T} = \mathsf{Th}(\mathsf{K}).$$

Now, we show that T is a definitional extension of T_1. When the language of T_2 has only one non-logical symbol, this follows immediately from Beth's definability theorem (see [11, theorem 2.2.22]), since for each $\mathfrak{M} \in \mathsf{Mod}(\mathsf{T}_1)$ there is at most one relation satisfying T, namely that of $b(\mathfrak{M})$. However, a generalized version of Beth's theorem is well-known as folklore: if the $\mathcal{R}$-free reduct of each model of T can be extended to at most one model of T, then T explicitly defines each member of $\mathcal{R}$ by a formula on the language of the $\mathcal{R}$-free reducts.[3] The proof that T is a definitional extension of T_2 is completely analogous. Thus, T_1 and T_2 are definitionally equivalent theories.

Assume now that the languages of T_1 and T_2 are not disjoint. Rename the symbols in the language of T_2 so that the new symbols be distinct from any one used in T_1 and T_2, call the new theory T_2'. Now, there is a natural bijection $b_1 : \mathsf{Mod}(\mathsf{T}_2) \to \mathsf{Mod}(\mathsf{T}_2')$ satisfying conditions (i) and (ii), and $b_1 \circ b : \mathsf{Mod}(\mathsf{T}_1) \to \mathsf{Mod}(\mathsf{T}_2')$ also satisfies (i) and (ii). These bijections are between models of theories of disjoint languages. Apply the previous case to b_1 and $b_1 \circ b$, and use that definitional equivalence is a transitive relation by [22]. □

REMARK 5 (Relationship of Theorem 2 with the van Benthem and Pearce result). *The theorem in* [27], *call it BP-theorem for van Benthem and Pearce theorem, seems to be neither stronger nor weaker than Theorem 2. It is not weaker because the kind of interpretation it deals with is restricted interpretation which is in between classical and Morita-interpretation. It is not stronger because it deals with interpretation and not with equivalence. In more detail, assume that there is a bijection between* $\mathsf{Mod}(\mathsf{T}_1)$ *and* $\mathsf{Mod}(\mathsf{T}_2)$ *satisfying* (*i*) *and* (*ii*) *of Theorem 2. By applying the BP-theorem, we get that there are two restricted interpretations, one from* T_1 *to* T_2 *and the other from* T_2 *to* T_1. *However, we know that mutual interpretability even with strong properties does not imply definitional equivalence* (*see, e.g.,* [3]). *Although the BP-theorem does not seem to imply Theorem 2, the proof of Theorem 2 here is just a slight modification of the proof of the BP-theorem in* [27, *p.* 296].

REMARK 6 (Automorphism groups and elementary embeddings in Theorem 2). *The word "different" can be omitted from condition* (*i*) *of Theorem 2 and the theorem remains true. This is true because condition* (*i*) *implies that the automorphism groups are preserved by* b *in the sense that* $\mathsf{Aut}(\mathfrak{M}) = \mathsf{Aut}(b(\mathfrak{M}))$ *for all* $\mathfrak{M} \in \mathsf{Mod}(\mathsf{T}_1)$. *Indeed, if*

[3] We give a short proof of this in the Appendix.

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

$\alpha \in \mathrm{Aut}(\mathfrak{M})$, then let $f : \mathfrak{M} \to \mathfrak{M}'$ be any isomorphism where $\mathfrak{M}'$ is different from $\mathfrak{M}$; there is always such an f. *Then both* f *and $\alpha \circ f$ are isomorphisms between $b(\mathfrak{M})$ and $b(\mathfrak{M}')$ by condition (ii); thus, $\alpha = \alpha \circ f \circ f^{-1}$ is an automorphism of $b(\mathfrak{M})$. In a sense, this corollary about the automorphism groups is the essential part of condition (i).*

Also, Theorem 2 remains true if in (i) we require to preserve all elementary embeddings in place of all isomorphisms. The reason is that elementary embeddings are preserved by definitional equivalence.

The proof of the following theorem intends to illustrate the use of Theorem 2 for proving definitional inequivalence. Recall the definitions of S_0 and S_1 from Remark 3.

THEOREM 3. *$\mathsf{T}(S_0)$ and $\mathsf{T}(S_1)$ are not definitionally equivalent.*

Proof. Let $b : \mathsf{Mod}(S_0) \to \mathsf{Mod}(S_1)$ be any bijection that preserves isomorphisms between distinct models. (We note that there is such a function b [see Lemma 2].) It preserves automorphism groups also (see Remark 3). We will show that b cannot preserve all ultrapowers. By Theorem 2, this will prove that $\mathsf{T}(S_0)$ and $\mathsf{T}(S_1)$ are not definitionally equivalent.

Let $\mathfrak{M} = \langle \omega, 0, \mathsf{suc}, S_0 \rangle \in \mathsf{Mod}(\mathsf{T}(S_0))$. Let $b(\mathfrak{M}) = \langle \omega, o, F, R \rangle$ and let U be any nonprincipal ultrafilter on ω. First we show that $b(\Pi_U \mathfrak{M}) \neq \Pi_U b(\mathfrak{M})$ if $b(\mathfrak{M})$ contains any copy of a $\mathbb{Z}$-model. Indeed, assume that $F^k(n)$ exists for all $k \in \mathbb{Z}$ for some $n \in \omega$. Let $\langle \mathbb{Z}, P \rangle$ be the $\mathbb{Z}$-model that is isomorphic to the induced subalgebra of $b(\mathfrak{M})$ with universe $\{F^k(n) : k \in \mathbb{Z}\}$. By $(*)$ in the proof of Lemma 1, $\Pi_U b(\mathfrak{M})$ contains infinitely many copies of this $\mathbb{Z}$-model. Therefore, the image of the $\mathbb{Z}$-model in $b(\mathfrak{M})$ under the diagonal embedding can be interchanged with a distinct copy of this $\mathbb{Z}$-model in $\Pi_U b(\mathfrak{M})$. On the other hand, all automorphisms of $\Pi_U \mathfrak{M}$ leave the diagonal embedding of $\mathfrak{M}$ unchanged. Thus $b(\Pi_U \mathfrak{M})$ cannot be $\Pi_U b(\mathfrak{M})$ since the two have different automorphism groups. Therefore, we assume in the rest

(1) $\omega = \{F^n(o) : n \in \omega\}$ and thus $R = \{F^n(o) : n \in S_1\}$.

Next we show that $b(\Pi_U \mathfrak{M}) \neq \Pi_U b(\mathfrak{M})$ if $F(y) \notin \{\mathsf{suc}^k(y) : k \in \mathbb{Z}\}$ for some $y \in \Pi_U \omega$. Indeed, assume the latter. Choose an automorphism of $\Pi_U \mathfrak{M}$ that interchanges the copy of the $\mathbb{Z}$-model containing y with another copy that does not contain either y or $F(y)$ and is identity on the rest. There is such an automorphism by $(*)$ in the proof of Lemma 1. Now, this is not an automorphism of $\Pi_U b(\mathfrak{M})$ since F is one-to-one in $b(\mathfrak{M})$ (by $b(\mathfrak{M}) \in \mathsf{Mod}(\mathsf{T}(S_1))$). Therefore, we assume in the rest

(2) $F(y) \in \{\mathsf{suc}^k(y) : k \in \mathbb{Z}\}$ for all $y \in \Pi_U \omega$.

Now, (2) implies that there is a bound on "how far F can jump," i.e., there is $N_0 \in \omega$ such that for all $n \in \omega$ we have

(2a) $F(n) = n + k$ implies $|k| < N_0$.

Indeed, let $J := \{k \in \mathbb{Z} : F(n) = n + k \text{ for some } n \in \omega\}$ and assume that J is infinite. Let $f : \omega \to J$ be a bijection; there is such a bijection because J is countably infinite. For all $j \in J$ let $n_j \in \omega$ be such that $F(n_j) = n_j + j$ and let $y_i := n_{f(i)}$ for all $i \in \omega$. Let $y := \langle y_i : i \in \omega \rangle$. Then $F(y_U) \notin \{\mathsf{suc}^k(y_U) : k \in \mathbb{Z}\}$ because U is nonprincipal. This contradicts (2), and thus J is finite, which implies the existence of the bound N_0.

Next we show that $b(\Pi_U \mathfrak{M}) \neq \Pi_U b(\mathfrak{M})$ if F does not agree with suc on copies of $\mathbb{Z}$-models in $\Pi_U \mathfrak{M}$ all elements of which are in R or no elements of which are

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

in R. Indeed, assume $\mathsf{R}(\mathsf{suc}^m(y))$ in $\Pi_U\mathfrak{M}$ for all $m \in \mathbb{Z}$. There is $k \in \mathbb{Z}$ such that $F(y) = \mathsf{suc}^k(y)$, by (2). There is an automorphism α in $\Pi_U\mathfrak{M}$ that "shifts with 1 step in $Y := \{\mathsf{suc}^m(y) : m \in \mathbb{Z}\}$," i.e., $\alpha(z) = \mathsf{suc}(z)$ for all $z \in Y$, because $\mathsf{R}(z)$ for all $z \in Y$. Now, if $F(z) \neq \mathsf{suc}^k(z)$ for some $z \in Y$, then α is not an automorphism in $\Pi_U b(\mathfrak{M})$. So, assume that $F(z) = \mathsf{suc}^k(z)$ for all $z \in Y$. Now, if $k \notin \{1, -1\}$, then $Y \neq \{F^m(y) : m \in \mathbb{Z}\} = \{\mathsf{suc}^{mk}(y) : m \in \mathbb{Z}\}$. However, there is an automorphism β of $\Pi_U b(\mathfrak{M})$ that "shifts $\{F^m(y) : m \in \mathbb{Z}\}$ with one step" and leaves all the other elements fixed. This β is not an automorphism of $\Pi_U\mathfrak{M}$. We show now that $F(z) = \mathsf{suc}^{-1}(z)$ for all $z \in Y$ cannot happen. Indeed, assume that $F(z) = \mathsf{suc}^{-1}(z)$ for all $z \in Y$. Then there is an "N_0-long descending F-chain in $b(\mathfrak{M})$," i.e., there is $n \in \omega$ such that $F(k) = k - 1$ for all $n - N_0 \leq k \leq n$ in $b(\mathfrak{M})$. Then F has to stay below n since then on, by (2a) and F being one-to-one, i.e., $F^k(n) \leq n$ for all $k \in \omega$. This again contradicts F being one-to-one. The same argument works if $\neg\mathsf{R}(\mathsf{suc}^m(y))$ in $\Pi_U\mathfrak{M}$ for all $m \in \mathbb{Z}$. By the above, we assume in the rest

(3) $F(z) = \mathsf{suc}(z)$ for all $z \in Y := \{\mathsf{suc}^k(y) : k \in \mathbb{Z}\}$ if $y \in \Pi_U\omega$ is such that either $\mathsf{R}(z)$ in $\Pi_U\mathfrak{M}$ for all $z \in Y$ or $\neg\mathsf{R}(z)$ in $\Pi_U\mathfrak{M}$ for all $z \in Y$.

Now, (3) has implications on the behavior of F on long R-chains or $\neg$R-chains in $\mathfrak{M}$, as follows. Let us say that $\langle y + k : k < n\rangle = \langle y, y + 1, y + 2, ..., y + n - 1\rangle$ is an n-long R-chain in $\mathfrak{M}$ beginning with y if $\mathsf{R}(y + k)$ in $\mathfrak{M}$ for all $k < n$. The definition of a $\neg$R-chain is analogous. First we show the existence of a bound N such that for all R-chains longer than $2N$, F agrees with suc on the chain, except for N-long chains at the beginning and at the end of the chain, and the same holds for $\neg$R-chains.

(3a) There is $N > N_0$ such that for all R-chains longer than $2N$ and beginning with y we have $F(\mathsf{suc}^k(y)) = \mathsf{suc}^{k+1}(y)$ for all $y + N \leq k \leq y + n - N$ and the same holds for $\neg$R-chains, too.

Indeed, assume that there is no such bound. Then n is not such a bound for any $n \in \omega$, i.e., there is an m-long R-chain with beginning y such that $m \geq 2n$ and $F(\mathsf{suc}^k(y)) \neq \mathsf{suc}^{k+1}(y)$ for some $y + n \leq k \leq y + m - n$. For each $n \in \omega$, let $y_n := \mathsf{suc}^k(y)$ for such a chain and let $z = \langle y_n : n \in \omega\rangle$. Then in the ultrapower $\Pi_U\mathfrak{M}$ we have $F(z) \neq \mathsf{suc}(z)$, while $\mathsf{R}(\mathsf{suc}^k(z))$ for all $k \in \mathbb{Z}$. This contradicts (3). The proof for the $\neg$R-chains is analogous. This completes the proof of (3a).

From now on we assume that N is as in (3a). Next we prove that if there is an $n \geq 3N$-long R-chain ending with $y - 1$ and there is an $n \geq 3N$-long $\neg$R-chain starting with $y + 1$, then the behavior of F is rather close to that of suc in these chains. Namely, $F^k(o) = k$ in the interval $[y - n + N, y + n - N]$ except in $[y - N, y + N]$, and F enumerates the elements of $[y - N, y + N]$.

(3b) Assume that $n > 3N$ and there is an n-long R-chain in $\mathfrak{M}$ ending with $y - 1$ and there is an n-long $\neg$R-chain starting with $y + 1$. Then $F^k(o) = k$ for all $y - n + N \leq k \leq y - N$ and $y + N \leq k \leq y + n - N$. Further, $\{F^k(o) : y - N \leq k \leq y + N\} = \{k : y - N \leq k \leq y + N\}$.

Indeed, assume that n and y are as in (3b). There is an $n \geq 2N$-long R-chain beginning with $y - n$, so by (3a) we have $F(y - n + k) = y - n + k + 1$ for all $y - n + N \leq k \leq y - N$. Let $v := y - n + N$. Then

$$F(v + k) = v + k + 1 \text{ for all } k \leq n - 2N. \tag{a}$$

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

Then $F(w) \notin \{k : v < k \leq v + n - 2N\}$ for all $w < v$ since F is one-to-one by $b(\mathfrak{M}) \models \mathsf{T}(S_1)$. By $n - 2N \geq N > N_0$ and (2a) then $F(w) \leq v$ for all $w < v$ and hence F enumerates $[0, v]$, i.e.,

$$\{F^k(o) : k \leq v\} = \{k : k \leq v\}. \tag{b}$$

There is $m \in \omega$ such that $v = F^m(o)$, by (1). As before, by (2a) and (a) we have that $m \leq v$ and then $m = v$ by (b). Thus, $F^v(o) = v$ and by (a) we have $F^k(o) = k$ for all $v \leq k \leq y - N$. The rest of (3a) can be obtained similarly.

We are ready to show $b(\Pi_U \mathfrak{M}) \neq \Pi_U b(\mathfrak{M})$, finishing the proof of Theorem 3. Let X be the infinite set where S_0 and S_1 differ. Then X is disjoint from S_0 and $X \subseteq S_1$, by definition. Let x_n denote the nth member of X according to the natural ordering of ω. Then $\neg\mathsf{R}(x_n)$ in $\mathfrak{M}$ by $x_n \notin S_0$ and the definition of R in $\mathfrak{M}$. Also, $\mathsf{R}(x_n - k - 1)$ and $\neg\mathsf{R}(x_n + k)$ for all $k < n$, because the 0, 1-sequences between two xs are laid by alphabetical order; thus, before the nth $x \in X$ there are n many 1s and after it there are $n + 1$ many 0s. Let $x := \langle x_n : n \in \omega \rangle$. Then x_U is contained in $\Pi_U \mathfrak{M}$ in a copy of the $\mathbb{Z}$-model $\langle \mathbb{Z}, \{k : k < 0\} \rangle$, i.e., all members of the $\mathbb{Z}$-model below x_U are in R, and no member after x_U, including x_U is in R.

How does the set $C := \{\mathsf{suc}^k(x_U) : k \in \mathbb{Z}\}$ look like in $\Pi_U b(\mathfrak{M})$? Note that we cannot assume $F = \mathsf{suc}$ and $o = 0$ in $b(\mathfrak{M})$. Thus, for example, we cannot infer $\mathsf{R}(x_n)$ in $b(\mathfrak{M})$ from $x_n \in S_1$. However, we can use our assumptions (1)–(3) and their implications. Especially, we can use (3b). Let $n \geq 3N$, where N is the bound in (3b). We have seen in the previous paragraph that, in $\mathfrak{M}$, the assumptions hold for $y = x_n$. By (1), the definition of S_1, and (3b) then $\mathsf{R}(F^k(o))$ for $x_n - n + N \leq k \leq x_n - N$ and $\neg\mathsf{R}(F^k(o))$ for $x_n + N \leq k \leq x_n + n - N$, in $b(\mathfrak{M})$. Also, by (3b) we get that F agrees with suc "below" $\mathsf{suc}^{-N}(x_U)$ and "above" $\mathsf{suc}^N(x_U)$, in $\Pi_U b(\mathfrak{M})$. Further, F enumerates the interval $I := [\mathsf{suc}^{-N}(x_U), \mathsf{suc}^N(x_U)]$. However, there is a difference between $\Pi_U \mathfrak{M}$ and $\Pi_U b(\mathfrak{M})$ concerning I. Namely, in $\Pi_U \mathfrak{M}$ exactly N elements of I are in R because $\neg\mathsf{R}(x_n)$ in $\mathfrak{M}$. At the same time, due to the definition of S_1, by (1) we get $\mathsf{R}(F^w(o))$ for all $w \in X$. Hence, exactly $N + 1$ elements are in R in the corresponding intervals in $b(\mathfrak{M})$, so exactly $N + 1$ elements of I are in R, in $\Pi_U b(\mathfrak{M})$.

For all $n \in \omega$ let $y_n \in \omega$ be similar to x_n in that $\neg\mathsf{R}(y_n)$, there is an n-long R-chain ending with $y_n - 1$, there is an n-long $\neg$R-chain starting with $y_n + 1$, and such that neither y_n nor any element of these chains belong to X. There are such y_ns by the construction of S_0, S_1. Let $y := \langle y_n : n \in \omega \rangle$. Then there is an automorphism in $\Pi_U \mathfrak{M}$ that interchanges x_U with y_U. We will show that there is no automorphism in $\Pi_U b(\mathfrak{M})$ that interchanges $\mathsf{suc}^{-N}(x_U)$ and $\mathsf{suc}^{-N}(y_U)$. Indeed, such an automorphism has to be a bijection between the intervals I and J because it can be seen that F enumerates $J := [\mathsf{suc}^{-N}(y_U), \mathsf{suc}^N(y_U)]$ in $\Pi_U b(\mathfrak{M})$ and F agrees with suc outside J. We have seen that there are $N + 1$ elements of I that are in R in $\Pi_U b(\mathfrak{M})$. It can be seen just the same way that there are only N elements of J because $\neg\mathsf{R}(y_U)$ in $\Pi_U b(\mathfrak{M})$. Therefore, no bijection between I and J can preserve R. The proof of Theorem 3 is complete. □

We close the paper with some implications of the results for questions raised in the wider literature.

Glymour [13] raises an interesting question about definitional equivalence. The common understanding is that definitionally equivalent theories have essentially the same content and we would think that all important properties are shared by them. Theorem 2 implies that a property of a theory is preserved by definitional equivalence

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

when it can be expressed in terms of universes, isomorphisms and ultraproducts of its models. Therefore, having a one-element model, having only finite models, being categorical in a power or being complete are preserved by classical definitional equivalence (since two models are elementarily equivalent if and only if they have isomorphic ultrapowers). Glymour [13, p. 296] conjectures that also the model class being closed under substructures, the model class being closed under unions of chains, and having an equational axiomatization are preserved. We now show that neither one of these three properties is preserved by definitional equivalence.

Indeed, let T_1 be the empty theory on the language with one constant symbol c. Let T_2 be the definitional extension of T_1 with $\forall x(R(x) \leftrightarrow [\exists yz(y \neq z) \wedge x = c]$. Then $\mathsf{Mod}(\mathsf{T}_1)$ is closed under taking substructures but $\mathsf{Mod}(\mathsf{T}_2)$ is not. The counterexample to preservation of unions of chains is similar in spirit. Let T_1 be the empty theory on the language with a binary relation symbol $\leq$. Let T_2 be the definitional extension of T_1 with defining R to be the set of $\leq$-minimal elements when there is a $\leq$-maximal element and R is the empty set when there is no $\leq$-maximal element (i.e., $\forall x[R(x) \leftrightarrow (\exists y \forall z(z \leq y) \wedge \forall z(z \leq x))]$). Clearly, T_1 is closed under taking unions of chains. However, T_2 is not closed under taking unions of chains, as the following models show. For each natural number n let $\mathfrak{M}_n$ have the set of natural numbers smaller than n as universe, let $\leq$ be the "smaller" relation and let only 0 be in relation R. Then each $\mathfrak{M}_n$ is a model of T_2 but their union is not a model of T_2 since it does not have a maximal element yet R is nonempty in it. For showing that having an equational axiomatization is not preserved by definitional equivalence, one could take groups as counterexamples; this is mentioned in [17, p. 56]. Indeed, let T_1 be the class of semigroups in which inverses exist and let T_2 be its extension with the inverse operation and the zero element as constant. Then T_1 does not have a universal axiomatization because its model class is not closed under subalgebras, while T_2 is an equational class.

It is known that definitionally equivalent theories have isomorphic Lindenbaum–Tarski formula-algebras; they only differ from each other in what definable properties they take to be as basic ones. The proofs above show that this latter choice can influence the existence of axiom systems of given forms. For example, being substructure is not preserved by definitional expansion because in this notion the basic relations are treated differently from the rest, namely being a substructure is formulated in terms of basic relations only. Similarly for homomorphism, union, etc. However, being an elementary substructure is preserved by definitional expansion because in the definition of elementary substructure all definable relations are treated alike (and indeed, this notion can be characterized by means of isomorphisms and ultraproducts as follows: $\mathfrak{N}$ is an elementary substructure of $\mathfrak{M}$ if and only if $N \subseteq M$ and there is an ultrafilter U such that $\Pi_U \mathfrak{N}$ is isomorphic to $\Pi_U \mathfrak{M}$ via an isomorphism that is identity on the diagonal image of N in $\Pi_U \mathfrak{N}$).

The following corollary of Theorem 2 states that an associated structure to be defined below, namely the concrete ultracategory of a theory, is an invariant characteristic to definitional equivalence of first-order logic theories.

By a *concrete ultracategory*, we mean a triple (C, F, p) where (C, F) is a concrete category,[4] and the additional structure p is a system of infinitary functions $\langle p_U : U \text{ an ultrafilter} \rangle$ on $Ob(C)$ such that if U is an ultrafilter on the set I then $F(p_U(m_i)_{i \in I}) = \Pi_U F(m_i)$ for all $m : I \to Ob(C)$. A functor between two

[4] For the notions of a concrete category and a concrete functor, see [1, chap. 5].

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

ultracategories (C, F, p) and (C', F', p') is a concrete functor between (C, F) and (C', F') that preserves all the functions p_U. Two concrete ultracategories are isomorphic if there is a functor between them that is a category theoretical isomorphism.

Let T be a theory. Its *concrete ultracategory* is (C, F, p) where (C, F) is $\mathcal{M}od^{iso}(\mathsf{T})$ with the natural forgetful functor, and for all ultrafilters U on I and all systems $(\mathfrak{M}_i)_{i\in I}$ we have $p_U((\mathfrak{M}_i)_{i\in I}) = \Pi_U \mathfrak{M}_i$. Notice that an isomorphism between the ultracategories of two theories preserves only the universes of the models (through the forgetful functors) and the behaviour of isomorphisms and ultraproducts as functions on $\mathcal{M}od^{iso}(\mathsf{T})$.

THEOREM 4. *Two first-order logic theories are definitionally equivalent if and only if their concrete ultracategories are isomorphic.*

Proof. This is just a reformulation of Theorem 2. □

We note that one can define the concrete ultracategory of a theory to contain all elementary embeddings in place of all isomorphisms only, as is usual. Theorem 4 is true with this modified definition, too. The reason is that elementary embeddings are preserved by definitional equivalence.

REMARK 7 (Connection with Stone duality). *Halvorson* [14, *sec.* 7] *proposes the programme to investigate what structure a model class naturally has. This program involves to endow the model class of a theory in such a way that from this structure on the model class, the theory can be recovered up to definitional equivalence. Theorem* 4 *offers an answer, namely concrete ultracategory of a theory. In category theoretical logic, Makkai* [26, *theorem* 4.1] *offers the notion of* (*abstract*) *ultracategory and Awodey and Forssell* [5] *offer the notion of topological groupoid in place of our concrete ultracategory. These three structures are quite similar to each other, so there seems to be a convergence here in finding a natural structure on the model classes. Unlike our concrete ultracategory, Makkai's ultracategory and Awodey and Forssell's topological groupoids characterize first-order theories only up to many-sorted definitional equivalence, which is weaker than classical definitional equivalence. Halvorson* [14] *points out the connection of his programme with generalizing Stone duality from propositional logic to predicate logic. We believe that a full-fledged Stone duality can be based on Theorem* 4. *See also* [16, 25, 26] *and* [8, *p.* 576].

Definability theory is used quite extensively in recent philosophy of science papers to investigate what symmetries tell about theories and how to compare "structure" (see, for example, [7, 10, 15, 19]). When one theory is an expansion of the other, there is a natural functor between their model categories. This is the "reduct-formation" functor denoted by Π in [7, above example 9]. It is shown in [7] that the question investigated in the present paper gets rather nice answers in this special case. We now show how one of the attractive theorems in [7] follows from Theorem 2. In fact, Theorem 2 in the present paper is a generalization of [7, corollary 2] to the general case concerning two arbitrary theories.

COROLLARY 2 (Corollary 2 in [7]). *Let* T^+ *be an expansion of* T. *Then* T^+ *is definitionally equivalent to* T *if and only if the reduct-formation functor* Π *is an equivalence between their model iso-categories.*

Proof. The reduct-formation functor Π is a concrete functor and it always preserves isomorphisms and ultraproducts "forwards," i.e., from T^+ to T. It is a bijection up to

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

isomorphism if and only if it is a bijection because the range of Π is always closed under isomorphisms. Thus, if Π is a category theoretical equivalence, then each model of T has a unique expansion in $\mathrm{Mod}(\mathsf{T}^+)$; therefore, Π preserves isomorphisms and ultraproducts also backwards. Thus, if Π is a category theoretical equivalence, then it satisfies (i) and (ii) in Theorem 2; hence, T and T^+ are definitionally equivalent. The other direction is easy. □

Categorical equivalence of theories is investigated in [8] as a weaker form of definitional equivalence. Two theories are defined to be *categorically equivalent* iff there is a categorical equivalence between their model categories. It is shown in [8] that categorical equivalence, many-dimensional (Morita) equivalence and definitional equivalence are strictly stronger in this order. The question naturally arises about how "large" the gaps between them are and under what additional properties these are the same.

According to Corollary 2, the reduct-formation functor Π bridges the gap between definitional equivalence and categorical equivalence between a theory and its expansion. It is asked in [7, below corollary 2] what special property $\mathfrak{P}$ of Π allows it to fill the gap between categorical and definitional equivalence of theories. Theorem 2 gives an answer to this question. The answer it offers is that this special property $\mathfrak{P}$ of Π is that it is a concrete functor which preserves ultraproducts in both directions when it is an equivalence.

Question 2 in [6] asks for an additional property $\mathfrak{P}$ of functors such that two theories are definitionally equivalent iff there is a category theoretical equivalence between their model categories which has property $\mathfrak{P}$. This question is also mentioned in [29, note 23], where it is written: "It is not known how much weaker categorical isomorphism is than definitional equivalence, or Morita equivalence, which is a weakening of definitional equivalence that allows one to define new sorts." Now, Corollary 3 below says, roughly, that categorical equivalence is just as much weaker than definitional equivalence as it misses how ultraproducts behave and what the universes of models as well as the set theoretical contents of morphisms are. In other words, two theories are definitionally equivalent if and only if there is an equivalence between their model categories which is a concrete isomorphism and preserves ultraproducts. We note that [19, theorem 3] gives an answer to the above questions that is different in spirit from our Corollary 3.

COROLLARY 3. *Two theories* T_1 *and* T_2 *are definitionally equivalent if and only if there is a concrete ultraproduct-preserving functor* F *that is an equivalence between* $\mathcal{M}od(\mathsf{T}_1)$ *and* $\mathcal{M}od(\mathsf{T}_2)$.

Ultraproducts are intimately connected to first-order logic. It would be interesting to see whether analogous theorems hold for other languages where ultraproducts can be omitted or replaced with some other additional structure. Hudetz [19, 20] contain interesting generalizations and results in the direction of broadening definability theory in order to be more applicable in philosophy of science. These results may be used perhaps to get an analogue of Theorem 2 in which ultraproducts do not occur.

§A. Appendix. The following generalized version of Beth's theorem is well-known as folklore. Both [26, 27] use this generalized version of Beth's theorem without proof. Since Theorem 2 relies heavily on this folklore theorem, here we give a short proof for it. For simplicity, we assume that we have only relation symbols.

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press

THEOREM 5. *Assume that* T *is a theory on the language* $\Sigma \cup \mathcal{R}$ *and the* Σ*-reduct of each model of* T *has at most one extension to a model of* T. *Then each element of* $\mathcal{R}$ *is explicitly definable in* T *by a* Σ*-formula.*

Proof. Let T' denote the theory T where each relation symbol $R \in \mathcal{R}$ is replaced by a new relation symbol R' not occurring in the language of T (and having the same arity). Then $\mathsf{T} \cup \mathsf{T}' \models \forall \bar{x}[R(\bar{x}) \leftrightarrow R'(\bar{x})]$ for all $\mathsf{R} \in \mathcal{R}$, since the $\mathcal{R}$-free reduct of each model of T has at most one expansion to a model of T. Let $R \in \mathcal{R}$ be arbitrary. By the compactness theorem, there is a finite subset T_0 of T such that $\mathsf{T}_0 \cup \mathsf{T}'_0 \models \forall \bar{x}[R(\bar{x}) \leftrightarrow R'(\bar{x})]$. Therefore, R has to occur in T_0, since otherwise both the empty set and the biggest relation of the same rank as R can be chosen in a model to satisfy T_0. Since T_0 is finite, it contains only finitely many elements from $\mathcal{R}$, let the set of these elements be $\mathcal{R}_0 := \{R_1, \dots, R_n\}$, and we may assume R_1 is R. By the usual Beth's theorem, there is a formula φ_R on the language $\Sigma \cup \{R_2, \dots, R_n\}$ which defines R in T_0. Now, let T_1 be the theory we obtain from T_0 by replacing R in it everywhere with φ_R. Then T_1 follows from T_0, only $R_2, \dots, R_n$ occur in T_1 and $\mathsf{T}_1 \cup \mathsf{T}'_1 \models \forall \bar{x}[R_2(\bar{x}) \leftrightarrow R'_2(\bar{x})]$. By the usual Beth's theorem, there is a formula φ_{R1} on the language $\Sigma \cup \{R_3, \dots, R_n\}$ which defines R_1 in T_1. And so on. At the end we get T_{n-1} on the language $\Sigma \cup \{R_n\}$ and a formula φ_{Rn} on the language Σ which defines R_n in T_{n-1}. Let ψ_n be φ_{Rn}, let ψ_{n-1} be the formula we get from φ_{Rn-1} by replacing R_n in it by ψ_n, etc. Then ψ_1 is in the language Σ which defines R in $\mathsf{T}_0 \subseteq \mathsf{T}$. □

Acknowledgements. We are indepted to the two anonymous referees for their very useful feedbacks.

Funding. This research is supported by the Hungarian National Research, Development and Innovation Office (NKFIH), grant no. FK-134732.

BIBLIOGRAPHY

[1] Adámek, J., Herrlich, H., & Strecker, G. E. (2004). *Abstract and Concrete Categories* (online edition). The Joy of Cats. http://katmat.math.uni-bremen.de/acc/acc.pdf

[2] Andréka, H., Madarász, J. X., & Németi, I. (2002). *On the Logical Structure of Relativity Theories*. Research report. Budapest: Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, 1312 pp., with contributions from A. Andai, G. Sági, I. Sain and Cs. Tőke. Available from: http://www.math-inst.hu/pub/algebraiclogic/Contents.html.

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

[4] Andréka, H., & Németi, I. (2014). Comparing theories: The dynamics of changing vocabulary. In Baltag, A. and Smets, S., editors. *Johan Van Benthem on Logic and Information Dynamics*. Springer Series Outstanding Contributions to Logic, Vol. 5. Dordrecht: Springer, pp. 143–172.

[5] Awodey, S., & Forssell, H. (2013). First-order logical duality. *Annals of Pure and Applied Logic*, **164**(3), 319–348.

[6] Barrett, T. W. (2017). On the Structure and Equivalence of Theories. Ph.D. Thesis, Princeton University.

[7] ———. (2018). What do symmetries tell us about structure? *Philosophy of Science*, **85**(4), 617–639.

[8] Barrett, T. W., & Halvorson, H. (2016). Morita equivalence. *The Review of Symbolic Logic*, **9**(3), 556–582.

[9] ———. (2022). Mutual translatability, equivalence, and the structure of theories. *Synthese*, **200**(3), 1–36.

[10] Barrett, T. W., Manchak, J. B., & Weatherall, J. (2023). On automorphism criteria for comparing amounts of mathematical structure. *Synthese*, **201**, Article no. 19.

[11] Chang, C. C., & Keisler, H. J. (1990). *Model Theory* (third edition). Studies in Logic and the Foundations of Mathematics, Vol. 73. Amsterdam: North-Holland.

[12] de Bouvère, K. (1965). Synonymous theories. In Addison, J. W., Henkin, L., and Tarski, A., editors. *The Theory of Models*. Amsterdam: North-Holland, pp. 402–406.

[13] Glymour, C. (2013). Theoretical equivalence and the semantic view of theories. *Philosophy of Science*, **80**(2), 286–297.

[14] Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science*, **79**(2), 183–206.

[15] ———. (2019). *The Logic in Philosophy of Science*. Cambridge: Cambridge University Press.

[16] Harnik, V. (2011). Model theory vs. categorical logic: Two approaches to pretopos completion (a.k.a. T^{eq}). In *Models, Logics, and Higher-Dimensional Categories*. CRM Proceedings and Lecture Notes, Vol. 53. Providence: American Mathematical Society, pp. 79–106.

[17] Henkin, L., Monk, J. D., & Tarski, A. (1971 and 1985). *Cylindric Algebras. Parts I–II*. Amsterdam: North-Holland.

[18] Hodges, W. (2008). *Model Theory*. Cambridge: Cambridge University Press.

[19] Hudetz, L. (2019). Definable categorical equivalence. *Philosophy of Science*, **86**, 47–75.

[20] ———. (2019). The semantic view of theories and higher-order languages. *Synthese*, **196**, 1131–1149.

[21] Kochen, S. (1961). Ultraproducts in the theory of models. *Annals of Mathematics*, **74**(2), 221–261.

[22] Lefever, K., & Székely, G. (2019). On generalization of definitional equivalence to non-disjoint languages. *Journal of Philosophical Logic*, **48**(4), 709–729.

[23] Lutz, S. (2017). What was the syntax–semantics debate in the philosophy of science about? *Philosophy and Phenomenological Research*, **95**(2), 319–352.

[24] Madarász, J. X. (2002). Logic and Relativity (in the Light of Definability Theory). Ph.D. Thesis, Eötvös Loránd University. Available from: http://www.math-inst.hu/pub/algebraic-logic/diszi.pdf.

[25] Makkai, M. (1985). Ultraproducts and categorical logic. In Prisco, C. A., editor. *Methods in Mathematical Logic*. Lecture Notes in Mathematics, Vol. 1130. Berlin–Heidelberg: Springer, pp. 222–309.

[26] ———. (1987). Stone duality for first order logic. *Advances in Mathematics*, **65**, 97–170.

[27] van Benthem, J., & Pearce, D. (1984). A mathematical characterization of interpretation between theories. *Studia Logica*, **43**(3), 295–303.

[28] Visser, A. (2006). Categories of theories and interpretations. In Enayat, A., Kalantari, I., and Moniri, M., editors. *Logic in Tehran. Proceedings of the Workshop and Conference on Logic, Algebra and Arithmetic, Held October 18–22, 2003*. Lecture Notes in Logic, Vol. 26. Wellesley: ASL and A. K. Peters, pp. 284–341.

[29] Weatherall, J. O. (2016). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis*, **81**, 1073–1091.

ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS
REÁLTANODA STREET 13–15
H-1053 BUDAPEST, HUNGARY
E-mail: andreka.hajnal@renyi.hu
E-mail: madarasz.judit@renyi.hu
E-mail: nemeti.istvan@renyi.hu

UNIVERSITY OF PUBLIC SERVICE
2 LUDOVIKA SQUARE
H-1053 BUDAPEST, HUNGARY
E-mail: szekely.gergely@renyi.hu

https://doi.org/10.1017/S1755020323000242 Published online by Cambridge University Press