THE REVIEW OF SYMBOLIC LOGIC
Volume 9, Number 3, September 2016

MORITA EQUIVALENCE

THOMAS WILLIAM BARRETT and HANS HALVORSON
Department of Philosophy, Princeton University

Abstract. Logicians and philosophers of science have proposed various formal criteria for theoretical equivalence. In this paper, we examine two such proposals: definitional equivalence and categorical equivalence. In order to show precisely how these two well-known criteria are related to one another, we investigate an intermediate criterion called Morita equivalence.

§1. Introduction. Many theories admit different formulations, and these formulations often bear interesting relationships to one another. One relationship that has received significant attention from logicians and philosophers of science is theoretical equivalence.[1] In this paper we will examine two formal criteria for theoretical equivalence. The first criterion, called *definitional equivalence*, has been known to logicians since the middle of the twentieth century.[2] It was introduced into philosophy of science by Glymour (1971, 1977, 1980). The second criterion is called *categorical equivalence*. It was first described by Eilenberg & Mac Lane (1942, 1945), but was only recently introduced into philosophy of science by Halvorson (2012, 2015) and Weatherall (2015a).

In order to illustrate the relationship between these two criteria, we will consider a third criterion for theoretical equivalence called *Morita equivalence*. We will show that these three criteria form the following hierarchy, where the arrows in the figure mean "implies."

$$\begin{tikzcd}\begin{array}{c}\text{Definitional}\\ \text{equivalence}\end{array} \arrow[r, shift left] & \begin{array}{c}\text{Morita}\\ \text{equivalence}\end{array} \arrow[l, shift left, "/" marking] \arrow[r, shift left] & \begin{array}{c}\text{Categorical}\\ \text{equivalence}\end{array} \arrow[l, shift left, "/" marking]\end{tikzcd}$$

Our discussion will allow us to evaluate definitional equivalence against categorical equivalence. Indeed, it will demonstrate a precise sense in which definitional equivalence is too strict a criterion for theoretical equivalence, while categorical equivalence is too liberal. There are theories that are not definitionally equivalent that one nonetheless has good reason to consider equivalent. And on the other hand, there are theories that are categorically equivalent that one has good reason to consider *in*equivalent.

Received: June 15, 2015.

[1] See Quine (1975), Sklar (1982), Halvorson (2012, 2013, 2015), Glymour (2013), Van Fraassen (2014), and Coffey (2014) for discussion of theoretical equivalence in philosophy of science.

[2] Artigue, Isambert, Perrin, & Zalc (1978) and de Bouvére (1965) attribute the concept of definitional equivalence to Montague (1957). Definitional equivalence was certainly familiar to logicians by the late 1960s, as is evident from the work of de Bouvére (1965), Shoenfield (1967), and Kanger (1968).

doi:10.1017/S1755020316000186

§2. Many-sorted logic. All of these criteria for theoretical equivalence are most naturally understood in the framework of first-order many-sorted logic. We begin with some preliminaries about this framework.[3]

2.1. Syntax. A *signature* Σ is a set of sort symbols, predicate symbols, function symbols, and constant symbols. Σ must have at least one sort symbol. Each predicate symbol $p \in \Sigma$ has an *arity* $\sigma_1 \times \cdots \times \sigma_n$, where $\sigma_1, \ldots, \sigma_n \in \Sigma$ are (not necessarily distinct) sort symbols.[4] Likewise, each function symbol $f \in \Sigma$ has an *arity* $\sigma_1 \times \cdots \times \sigma_n \rightarrow \sigma$, where $\sigma_1, \ldots, \sigma_n, \sigma \in \Sigma$ are again (not necessarily distinct) sort symbols. Lastly, each constant symbol $c \in \Sigma$ is assigned a sort $\sigma \in \Sigma$. In addition to the elements of Σ we also have a stock of variables. We use the letters x, y, and z to denote these variables, adding subscripts when necessary. Each variable has a sort $\sigma \in \Sigma$.

A $\boldsymbol{\Sigma}$*-term* can be thought of as a "naming expression" in the signature Σ. Each Σ-term has a sort $\sigma \in \Sigma$. The Σ-terms of sort σ are recursively defined as follows. Every variable of sort σ is a Σ-term of sort σ, and every constant symbol $c \in \Sigma$ of sort σ is also a Σ-term of sort σ. Furthermore, if $f \in \Sigma$ is a function symbol with arity $\sigma_1 \times \cdots \times \sigma_n \rightarrow \sigma$ and $t_1, \ldots, t_n$ are Σ-terms of sorts $\sigma_1, \ldots, \sigma_n$, then $f(t_1, \ldots, t_n)$ is a Σ-term of sort σ. We will use the notation $t(x_1, \ldots, x_n)$ to denote a Σ-term in which all of the variables that appear in t are in the sequence $x_1, \ldots, x_n$, but we leave open the possibility that some of the x_i do not appear in the term t.

A $\boldsymbol{\Sigma}$*-atom* is an expression either of the form $s(x_1, \ldots, x_n) = t(x_1, \ldots, x_n)$, where s and t are Σ-terms of the same sort $\sigma \in \Sigma$, or of the form $p(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are Σ-terms of sorts $\sigma_1, \ldots, \sigma_n$ and $p \in \Sigma$ is a predicate of arity $\sigma_1 \times \cdots \times \sigma_n$. The $\boldsymbol{\Sigma}$*-formulas* are then defined recursively as follows.

- Every Σ-atom is a Σ-formula.
- If ϕ is a Σ-formula, then $\neg\phi$ is a Σ-formula.
- If ϕ and ψ are Σ-formulas, then $\phi \rightarrow \psi$, $\phi \wedge \psi$, $\phi \vee \psi$ and $\phi \leftrightarrow \psi$ are Σ-formulas.
- If ϕ is a Σ-formula and x is a variable of sort $\sigma \in \Sigma$, then $\forall_\sigma x\phi$ and $\exists_\sigma x\phi$ are Σ-formulas.

In addition to the above formulas, we will use the notation $\exists_{\sigma=1} y\phi(x_1, \ldots, x_n, y)$ to abbreviate the formula $\exists_\sigma y(\phi(x_1, \ldots, x_n, y) \wedge \forall_\sigma z(\phi(x_1, \ldots, x_n, z) \rightarrow y = z))$. As above, the notation $\phi(x_1, \ldots, x_n)$ will denote a Σ-formula ϕ in which all of the free variables appearing in ϕ are in the sequence $x_1, \ldots, x_n$, but we again leave open the possibility that some of the x_i do not appear as free variables in ϕ. A $\boldsymbol{\Sigma}$*-sentence* is a Σ-formula that has no free variables.

2.2. Semantics. A $\boldsymbol{\Sigma}$*-structure* A is an "interpretation" of the symbols in Σ. In particular, A satisfies the following conditions.

- Every sort symbol $\sigma \in \Sigma$ is assigned a nonempty set A_σ. The sets A_σ are required to be pairwise disjoint.
- Every predicate symbol $p \in \Sigma$ of arity $\sigma_1 \times \cdots \times \sigma_n$ is interpreted as a subset $p^A \subset A_{\sigma_1} \times \cdots \times A_{\sigma_n}$.

[3] Our notation follows Hodges (2008). We present the more general case of many-sorted logic, however, while Hodges only presents single-sorted logic.

[4] The symbol $\sigma_1 \times \cdots \times \sigma_n$ has no intrinsic meaning. To say that, "p has arity $\sigma_1 \times \cdots \times \sigma_n$" is simply an abbreviated way of saying that p can be combined with n terms, whose sorts must respectively be $\sigma_1, \ldots, \sigma_n$.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

- Every function symbol $f \in \Sigma$ of arity $\sigma_1 \times \cdots \times \sigma_n \to \sigma$ is interpreted as a function $f^A : A_{\sigma_1} \times \cdots \times A_{\sigma_n} \to A_\sigma$.
- Every constant symbol $c \in \Sigma$ of sort $\sigma \in \Sigma$ is interpreted as an element $c^A \in A_\sigma$.

Given a Σ-structure A, we will often refer to an element $a \in A_\sigma$ as "an element of sort σ."

Let A be a Σ-structure with $a_1, \dots, a_n \in A$ elements of sorts $\sigma_1, \dots, \sigma_n$. We let $t(x_1, \dots, x_n)$ be a Σ-term of sort σ, with $x_1, \dots, x_n$ variables of sorts $\sigma_1, \dots, \sigma_n$, and we recursively define the element $t^A[a_1, \dots, a_n] \in A_\sigma$. If t is the variable x_i, then $t^A[a_1, \dots, a_n] = a_i$, and if t is the constant symbol $c \in \Sigma$, then $t^A[a_1, \dots, a_n] = c^A$. Furthermore, if t is of the form $f(t_1, \dots, t_m)$ where each t_i is a Σ-term of sort $\tau_i \in \Sigma$ and $f \in \Sigma$ is a function symbol of arity $\tau_1 \times \cdots \times \tau_m \to \sigma$, then

$$t^A[a_1, \dots, a_n] = f^A\big(t_1^A[a_1, \dots, a_n], \dots, t_m^A[a_1, \dots, a_n]\big).$$

One can think of the element $t^A[a_1, \dots, a_n] \in A_\sigma$ as the element of the Σ-structure A that is denoted by the Σ-term $t(x_1, \dots, x_n)$ when $a_1, \dots, a_n$ are substituted for the variables $x_1, \dots, x_n$.

Our next aim is to define when a sequence of elements $a_1, \dots, a_n \in A$ *satisfy* a Σ-formula $\phi(x_1, \dots, x_n)$ in the Σ-structure A. When this is the case we write $A \vDash \phi[a_1, \dots, a_n]$. We begin by considering Σ-atoms. Let $\phi(x_1, \dots, x_n)$ be a Σ-atom with $x_1, \dots, x_n$ variables of sorts $\sigma_1, \dots, \sigma_n$ and let $a_1, \dots, a_n \in A$ be elements of sorts $\sigma_1, \dots, \sigma_n$. There are two cases to consider. First, if $\phi(x_1, \dots, x_n)$ is the formula $s(x_1, \dots, x_n) = t(x_1, \dots, x_n)$, where s and t are Σ-terms of sort σ, then $A \vDash \phi[a_1, \dots, a_n]$ if and only if

$$s^A[a_1, \dots, a_n] = t^A[a_1, \dots, a_n].$$

Second, if $\phi(x_1, \dots, x_n)$ is the formula $p(t_1, \dots, t_m)$, where each t_i is a Σ-term of sort τ_i and $p \in \Sigma$ is a predicate symbol of arity $\tau_1 \times \cdots \times \tau_m$, then $A \vDash \phi[a_1, \dots, a_n]$ if and only if

$$\big(t_1^A[a_1, \dots, a_n], \dots, t_m^A[a_1, \dots, a_n]\big) \in p^A.$$

This definition is extended to all Σ-formulas in the following standard way.

- $A \vDash \neg\phi[a_1, \dots, a_n]$ if and only if it is not the case that $A \vDash \phi[a_1, \dots, a_n]$.
- $A \vDash \phi \wedge \psi[a_1, \dots, a_n]$ if and only if $A \vDash \phi[a_1, \dots, a_n]$ and $A \vDash \psi[a_1, \dots, a_n]$. The cases of $\vee$, $\to$, and $\leftrightarrow$ are defined analogously.
- Suppose that $\phi(x_1, \dots, x_n)$ is $\forall_\sigma y \psi(x_1, \dots, x_n, y)$, where $\sigma \in \Sigma$ is a sort symbol. Then $A \vDash \phi[a_1, \dots, a_n]$ if and only if $A \vDash \psi[a_1, \dots, a_n, b]$ for every element $b \in A_\sigma$. The case of $\exists_\sigma$ is defined analogously.

If ϕ is a Σ-sentence, then $A \vDash \phi$ just in case $A \vDash \phi[]$, i.e., the empty sequence satisfies ϕ in A.

2.3. Relationships between structures. There are different relationships that Σ-structures can bear to one another. An *isomorphism* $h : A \to B$ between Σ-structures A and B is a family of bijections $h_\sigma : A_\sigma \to B_\sigma$ for each sort symbol $\sigma \in \Sigma$ that satisfies the following conditions.

- For every predicate symbol $p \in \Sigma$ of arity $\sigma_1 \times \cdots \times \sigma_n$ and all elements $a_1, \dots, a_n \in A$ of sorts $\sigma_1, \dots, \sigma_n$, $(a_1, \dots, a_n) \in p^A$ if and only if $(h_{\sigma_1}(a_1), \dots, h_{\sigma_n}(a_n)) \in p^B$.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

- For every function symbol $f \in \Sigma$ of arity $\sigma_1 \times \cdots \times \sigma_n \to \sigma$ and all elements $a_1, \dots, a_n \in A$ of sorts $\sigma_1, \dots, \sigma_n$,

$$h_\sigma\big(f^A(a_1, \dots, a_n)\big) = f^B\big(h_{\sigma_1}(a_1), \dots, h_{\sigma_n}(a_n)\big).$$

- For every constant symbol $c \in \Sigma$ of sort σ, $h_\sigma(c^A) = c^B$.

When there is an isomorphism $h : A \to B$ one says that A and B are *isomorphic* and writes $A \cong B$.

There is another important relationship that Σ-structures can bear to one another. An *elementary embedding* $h : A \to B$ between Σ-structures A and B is a family of injective maps $h_\sigma : A_\sigma \to B_\sigma$ for each sort symbol $\sigma \in \Sigma$ that satisfies

$$A \vDash \phi[a_1, \dots, a_n] \text{ if and only if } B \vDash \phi[h_{\sigma_1}(a_1), \dots, h_{\sigma_n}(a_n)]$$

for all Σ-formulas $\phi(x_1, \dots, x_n)$ and elements $a_1, \dots, a_n \in A$ of sorts $\sigma_1, \dots, \sigma_n$. Given an isomorphism or elementary embedding $h : A \to B$, we will often use the notation $h(a_1, \dots, a_n)$ to denote the sequence of elements $h_{\sigma_1}(a_1), \dots, h_{\sigma_n}(a_n)$. Every isomorphism is an elementary embedding, but in general the converse does not hold.

There is an important relationship that can hold between structures of different signatures. Let $\Sigma \subset \Sigma^+$ be signatures and suppose that A is a Σ^+-structure. One obtains a Σ-structure $A|_\Sigma$ by "forgetting" the interpretations of symbols in $\Sigma^+ - \Sigma$. We call $A|_\Sigma$ the *reduct* of A to the signature Σ, and we call A an *expansion* of $A|_\Sigma$ to the signature Σ^+. Note that in general a Σ-structure will have more than one expansion to the signature Σ^+.

We can now discuss first-order theories in many-sorted logic. A $\boldsymbol{\Sigma}$*-theory* T is a set of Σ-sentences. The sentences $\phi \in T$ are called the axioms of T. A Σ-structure M is a *model* of a Σ-theory T if $M \vDash \phi$ for all $\phi \in T$. A theory T *entails* a sentence ϕ, written $T \vDash \phi$, if $M \vDash \phi$ for every model M of T.

We begin our discussion of theoretical equivalence with the following preliminary criterion.

DEFINITION. *Theories T_1 and T_2 are* logically equivalent *if they have the same class of models.*

One can easily verify that T_1 and T_2 are logically equivalent if and only if $\{\phi : T_1 \vDash \phi\} = \{\psi : T_2 \vDash \psi\}$.

§3. Definitional equivalence. Logical equivalence is a particularly strict criterion for theoretical equivalence. Indeed, theories can only be logically equivalent if they are formulated in the same signature. There are many cases, however, of theories in different signatures that are nonetheless intuitively equivalent. For example, the theory of groups can be formulated in a signature with a binary operation $\cdot$ and a constant symbol e, or it can be formulated in a signature with a binary operation $\cdot$ and a unary function -1 (Barrett & Halvorson, 2015a). Similarly, the theory of linear orders can be formulated in a signature with the binary relation $<$, or it can be formulated in a signature with the binary relation $\leq$. Since logical equivalence does not capture any sense in which these theories are equivalent, logicians and philosophers of science have proposed more general criteria for theoretical equivalence.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

One such criterion is *definitional equivalence*. This criterion is well known among logicians, and many results about it have been proven.[5] The basic idea behind definitional equivalence is simple. Theories T_1 and T_2 are definitionally equivalent if T_1 can define all of the symbols that T_2 uses, and in a compatible way, T_2 can define all of the symbols that T_1 uses. In order to state this criterion precisely, we need to do some work.

3.1. Definitional extensions. We first need to formalize the concept of a definition. Let $\Sigma \subset \Sigma^+$ be signatures and let $p \in \Sigma^+ - \Sigma$ be a predicate symbol of arity $\sigma_1 \times \cdots \times \sigma_n$. An *explicit definition of p in terms of* $\boldsymbol{\Sigma}$ is a Σ^+-sentence of the form

$$\forall_{\sigma_1} x_1 \ldots \forall_{\sigma_n} x_n \big(p(x_1, \ldots, x_n) \leftrightarrow \phi(x_1, \ldots, x_n)\big),$$

where $\phi(x_1, \ldots, x_n)$ is a Σ-formula, and the variables $x_1, \ldots, x_n$ are pairwise disjoint. Note that an explicit definition of p in terms of Σ can only exist if $\sigma_1, \ldots, \sigma_n \in \Sigma$. An explicit definition of a function symbol $f \in \Sigma^+ - \Sigma$ of arity $\sigma_1 \times \cdots \times \sigma_n \to \sigma$ is a Σ^+-sentence of the form

$$\forall_{\sigma_1} x_1 \ldots \forall_{\sigma_n} x_n \forall_\sigma y \big(f(x_1, \ldots, x_n) = y \leftrightarrow \phi(x_1, \ldots, x_n, y)\big) \tag{1}$$

and an explicit definition of a constant symbol $c \in \Sigma^+ - \Sigma$ of sort σ is a Σ^+-sentence of the form

$$\forall_\sigma y \big(y = c \leftrightarrow \psi(y)\big), \tag{2}$$

where $\phi(x_1, \ldots, x_n, y)$ and $\psi(y)$ are both Σ-formulas. Note again that these explicit definitions of f and c can only exist if $\sigma_1, \ldots, \sigma_n, \sigma \in \Sigma$.

Although they are Σ^+-sentences, (1) and (2) have consequences in the signature Σ. In particular, (1) and (2) imply the following sentences, respectively:

$$\forall_{\sigma_1} x_1 \ldots \forall_{\sigma_n} x_n \exists_{\sigma=1} y \phi(x_1, \ldots, x_n, y)$$
$$\exists_{\sigma=1} y \psi(y).$$

These two sentences are called the *admissibility conditions* for the explicit definitions (1) and (2).

A *definitional extension* of a Σ-theory T to the signature Σ^+ is a theory

$$T^+ = T \cup \{\delta_s : s \in \Sigma^+ - \Sigma\}$$

that satisfies the following two conditions. First, for each symbol $s \in \Sigma^+ - \Sigma$ the sentence δ_s is an explicit definition of s in terms of Σ, and second, if s is a constant symbol or a function symbol and α_s is the admissibility condition for δ_s, then $T \vDash \alpha_s$.

3.2. Three results. A definitional extension of a theory “says no more” than the original theory. There are a number of ways to make this idea precise. Of particular interest to us will be the following three. The reader is encouraged to consult Hodges (2008, pp. 58–62) for proofs of these results.

The first result captures a sense in which the models of a definitional extension T^+ are “determined” by the models of the original theory T. In order to specify a model of T^+, one needs to interpret all of the symbols in Σ^+. The interpretation of the symbols in $\Sigma^+ - \Sigma$, however, “comes for free” given an interpretation of the symbols in Σ.

[5] For example, see de Bouvére (1965), Kanger (1968), Pinter (1978), Pelletier & Urquhart (2003), Andréka, Madarász, & Németi (2005), Friedman & Visser (2014), and Barrett & Halvorson (2015a).

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

THEOREM 3.1. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a definitional extension of T to Σ^+, then every model M of T has a unique expansion M^+ that is a model of T^+.*

Theorem 3.1 provides a semantic sense in which a definitional extension T^+ "says no more" than the original theory T. The models of T^+ are completely determined by the models of T.

In order to state the second result, we need to introduce some terminology. Let $\Sigma \subset \Sigma^+$ be signatures. A Σ^+-theory T^+ is an *extension* of a Σ-theory T if $T \vDash \phi$ implies that $T^+ \vDash \phi$ for every Σ-sentence ϕ. A Σ^+-theory T^+ is a *conservative extension* of a Σ-theory T if $T \vDash \phi$ if and only if $T^+ \vDash \phi$ for every Σ-sentence ϕ. All conservative extensions are extensions, but in general the converse does not hold. We have the following simple result about definitional extensions.

THEOREM 3.2. *If T^+ is a definitional extension of T, then T^+ is a conservative extension of T.*

If T^+ is a conservative extension of T, then T^+ entails precisely the same Σ-sentences as T. Theorem 3.2 therefore shows that a definitional extension T^+ "says no more" in the signature Σ than the original theory T does.

The third result shows something stronger. If T^+ is a definitional extension of T, then every Σ^+-formula $\phi(x_1, \dots, x_n)$ can be "translated" into an equivalent Σ-formula ϕ^* $(x_1, \dots, x_n)$. The theory T^+ might use some new language that T did not use, but everything that T^+ says with this new language can be "translated" back into the old language of T. This result captures another robust sense in which the theory T^+ "says no more" than the theory T.

THEOREM 3.3. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a definitional extension of T to Σ^+ then for every Σ^+-formula $\phi(x_1, \dots, x_n)$ there is a Σ-formula $\phi^*(x_1, \dots, x_n)$ such that $T^+ \vDash \forall_{\sigma_1} x_1 \dots \forall_{\sigma_n} x_n (\phi(x_1, \dots, x_n) \leftrightarrow \phi^*(x_1, \dots, x_n))$.*

These results capture three different senses in which a definitional extension has the same expressive power as the original theory. With this in mind, we have the resources necessary to state definitional equivalence.

DEFINITION. *Let T_1 be a Σ_1-theory and T_2 be a Σ_2-theory. T_1 and T_2 are* definitionally equivalent *if there are theories T_1^+ and T_2^+ that satisfy the following three conditions:*

- *T_1^+ is a definitional extension of T_1,*
- *T_2^+ is a definitional extension of T_2,*
- *T_1^+ and T_2^+ are logically equivalent $\Sigma_1 \cup \Sigma_2$-theories.*

One often says that T_1 and T_2 are definitionally equivalent if they have a "common definitional extension." Theorems 3.1, 3.2, and 3.3 demonstrate a robust sense in which theories with a common definitional extension "say the same thing," even though they might be formulated in different signatures.

One trivially sees that if two theories are logically equivalent, then they are definitionally equivalent. But there are many examples of theories that are definitionally equivalent and not logically equivalent. The theory of groups formulated in the signature $\{\cdot, e\}$ is definitionally equivalent to the theory of groups formulated in the signature $\{\cdot, -1\}$. And likewise, the theory of linear orders formulated in the signature $\{<\}$ is definitionally equivalent to the theory of linear orders formulated in the signature $\{\leq\}$. Definitional equivalence

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

is therefore a weaker criterion for theoretical equivalence than logical equivalence. It is capable of capturing a sense in which theories formulated in different signatures might nonetheless be equivalent.

§4. Morita equivalence. Glymour (1971) claims that definitional equivalence is a necessary condition on the equivalence of scientific theories. However, there are several reasons to believe that this criterion is too strict.

First, it is frequently argued that many-sorted logic is reducible to single-sorted logic (see Schmidt, 1951; Manzano, 1996). What is actually shown in these arguments is that for any many-sorted theory T, a corresponding single-sorted theory T' can be constructed. But what is the relation between T and T'? Obviously, the two theories T and T' cannot be definitionally equivalent, since that criterion applies only to single-sorted theories. Therefore, to make sense of the claim that many-sorted logic can be reduced to single-sorted logic, we need a generalization of definitional equivalence (see Barrett & Halvorson, 2015b).

Second, there are well-known examples of theories that could naturally be formulated either within a single-sorted framework, or within a many-sorted framework—and we need a generalization of definitional equivalence to explain in what sense these two formulations are equivalent. For example, category theory can be formulated as a many-sorted theory, using both a sort of "objects" and a sort of "arrows" (Eilenberg & Mac Lane, 1942, 1945); and category theory can also be formulated as a single-sorted theory using only "arrows" (Mac Lane, 1948).[6] These two formulations of category theory are in some sense equivalent, and we would like an account of this more general notion of equivalence.

Third, definitional equivalence is too restrictive even for single-sorted theories. For example, affine geometry can be formalized in a way that quantifies over points; or it can be formalized in a way that quantifies over lines (see Schwabhäuser, Szmielew, & Tarski, 1983). But saying that the point theory (T_p) and the line theory (T_ℓ) both are formulations of the same theory indicates again that T_p and T_ℓ are in some sense equivalent—although T_p and T_ℓ are *not* definitionally equivalent. Indeed, the smallest model of T_p has five elements, which we can think of as the four corners of a square, and its center point. On the other hand, the smallest model of T_ℓ has six elements. But if T_p and T_ℓ were definitionally equivalent, then every model M of T_ℓ would be the reduct of an expansion of a model M' of T_p (de Bouvére, 1965). In particular, we would have $|M| = |M'|$, which entails that T_ℓ has a model of cardinality five—a contradiction. Therefore, T_p and T_ℓ are not definitionally equivalent.

Finally, even if we ignore the complications mentioned above, and even if we assume that each many-sorted theory T can be replaced by a single-sorted variant T' (by the standard procedure of unifying sorts), definitional equivalence is still inadequate. For example, let T_1 be the objects-and-arrows formulation of category theory, and let T_2 be the arrows-only formulation of category theory. Intuitively, T_1 and T_2 are equivalent theories; but, as we will now show, their single-sorted versions T_1' and T_2' are not definitionally equivalent. Indeed, $T_2' = T_2$, since T_2 is single-sorted. However, T_1' has a single sort that includes both objects and arrows. Thus, while T_2' has a model with one element (i.e., the category with a single arrow), T_1' has no models with one element (since every model of T_1' has at least one object and at least one arrow). Therefore, T_1' and T_2' are not definitionally equivalent.

These examples all show that definitional equivalence does not capture the sense in which some theories are equivalent. If one wants to capture this sense, one needs a more

[6] Freyd (1964, p. 5) and Mac Lane (1971, p. 9) also describe this alternative formulation.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

general criterion for theoretical equivalence than definitional equivalence. Our aim here is to introduce one such criterion. We will call it *Morita equivalence*.[7] This criterion is a natural generalization of definitional equivalence. In fact, Morita equivalence is essentially the same as definitional equivalence, except that it allows one to define new sort symbols in addition to new predicate symbols, function symbols, and constant symbols. In order to state the criterion precisely, we again need to do some work. We begin by defining the concept of a Morita extension. We then make precise the sense in which Morita equivalence is a natural generalization of definitional equivalence by proving analogues of Theorems 3.1, 3.2, and 3.3.

4.1. Morita extensions. As we did for predicates, functions, and constants, we need to say how to define new sorts. Let $\Sigma \subset \Sigma^+$ be signatures and consider a sort symbol $\sigma \in \Sigma^+ - \Sigma$. One can define the sort σ as a product sort, a coproduct sort, a subsort, or a quotient sort. In each case, one defines σ using old sorts in Σ and new function symbols in $\Sigma^+ - \Sigma$. These new function symbols specify how the new sort σ is related to the old sorts in Σ. We describe these four cases in detail.

In order to define σ as a product sort, one needs two function symbols $\pi_1, \pi_2 \in \Sigma^+ - \Sigma$ with π_1 of arity $\sigma \to \sigma_1$, π_2 of arity $\sigma \to \sigma_2$, and $\sigma_1, \sigma_2 \in \Sigma$. The function symbols π_1 and π_2 serve as the "canonical projections" associated with the product sort σ. A sort definition of the symbols σ, π_1, and π_2 as a *product sort* in terms of Σ is a Σ^+-sentence of the form

$$\forall_{\sigma_1} x \forall_{\sigma_2} y \exists_{\sigma=1} z (\pi_1(z) = x \wedge \pi_2(z) = y).$$

One should think of a product sort σ as the sort whose elements are ordered pairs, where the first element of each pair is of sort σ_1 and the second is of sort σ_2.

One can also define σ as a coproduct sort. One again needs two function symbols $\rho_1, \rho_2 \in \Sigma^+ - \Sigma$ with ρ_1 of arity $\sigma_1 \to \sigma$, ρ_2 of arity $\sigma_2 \to \sigma$, and $\sigma_1, \sigma_2 \in \Sigma$. The function symbols ρ_1 and ρ_2 are the "canonical injections" associated with the coproduct sort σ. A sort definition of the symbols σ, ρ_1, and ρ_2 as a *coproduct sort* in terms of Σ is a Σ^+-sentence of the form

$$\forall_\sigma z \big(\exists_{\sigma_1=1} x (\rho_1(x) = z) \vee \exists_{\sigma_2=1} y (\rho_2(y) = z)\big) \wedge \forall_{\sigma_1} x \forall_{\sigma_2} y \neg\big(\rho_1(x) = \rho_2(y)\big).$$

One should think of a coproduct sort σ as the disjoint union of the elements of sorts σ_1 and σ_2.

When defining a new sort σ as a product sort or a coproduct sort, one uses two sort symbols in Σ and two function symbols in $\Sigma^+ - \Sigma$. The next two ways of defining a new sort σ only require one sort symbol in Σ and one function symbol in $\Sigma^+ - \Sigma$.

In order to define σ as a subsort, one needs a function symbol $i \in \Sigma^+ - \Sigma$ of arity $\sigma \to \sigma_1$ with $\sigma_1 \in \Sigma$. The function symbol i is the "canonical inclusion" associated with

[7] This criterion is already familiar in certain circles of logicians. See Andréka, Madarász, & Németi (2008) and Mere & Veloso (1992). The name "Morita equivalence" descends from Kiiti Morita's work on rings with equivalent categories of modules. Two rings R and S are called *Morita equivalent* just in case there is an equivalence $\text{Mod}(R) \cong \text{Mod}(S)$ between their categories of modules. The notion was generalized from rings to algebraic theories by Dukarm (1988). See also Adámek, Sobral, & Sousa (2006). More recently, topos theorists have defined theories to be Morita equivalent just in case their classifying toposes are equivalent (Johnstone, 2003). See Tsementzis (2015) for a comparison of the topos-theoretic notion of Morita equivalence with ours.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

the subsort σ. A sort definition of the symbols σ and i as a *subsort* in terms of Σ is a Σ^+-sentence of the form

$$\forall_{\sigma_1} x\big(\phi(x) \leftrightarrow \exists_\sigma z(i(z) = x)\big) \wedge \forall_\sigma z_1 \forall_\sigma z_2\big(i(z_1) = i(z_2) \rightarrow z_1 = z_2\big), \tag{3}$$

where $\phi(x)$ is a Σ-formula. One can think of the subsort σ as consisting of "the elements of sort σ_1 that are ϕ." The sentence (3) entails the Σ-sentence $\exists_{\sigma_1} x\phi(x)$. As before, we will call this Σ-sentence the *admissibility condition* for the definition (3).

Lastly, in order to define σ as a quotient sort one needs a function symbol $\epsilon \in \Sigma^+ - \Sigma$ of arity $\sigma_1 \rightarrow \sigma$ with $\sigma_1 \in \Sigma$. A sort definition of the symbols σ and ϵ as a *quotient sort* in terms of Σ is a Σ^+-sentence of the form

$$\forall_{\sigma_1} x_1 \forall_{\sigma_1} x_2\big(\epsilon(x_1) = \epsilon(x_2) \leftrightarrow \phi(x_1, x_2)\big) \wedge \forall_\sigma z \exists_{\sigma_1} x(\epsilon(x) = z), \tag{4}$$

where $\phi(x_1, x_2)$ is a Σ-formula. This sentence defines σ as a quotient sort that is obtained by "quotienting out" the sort σ_1 with respect to the formula $\phi(x_1, x_2)$. The sort σ should be thought of as the set of "equivalence classes of elements of σ_1 with respect to the relation $\phi(x_1, x_2)$." The function symbol ϵ is the "canonical projection" that maps an element to its equivalence class. One can verify that the sentence (4) implies that $\phi(x_1, x_2)$ is an equivalence relation. In particular, it entails the following Σ-sentences:

$$\forall_{\sigma_1} x(\phi(x, x))$$
$$\forall_{\sigma_1} x_1 \forall_{\sigma_1} x_2(\phi(x_1, x_2) \rightarrow \phi(x_2, x_1))$$
$$\forall_{\sigma_1} x_1 \forall_{\sigma_1} x_2 \forall_{\sigma_1} x_3\big((\phi(x_1, x_2) \wedge \phi(x_2, x_3)) \rightarrow \phi(x_1, x_3)\big).$$

These Σ-sentences are the *admissibility conditions* for the definition (4).

Now that we have presented the four ways of defining new sort symbols, we can define the concept of a Morita extension. A Morita extension is a natural generalization of a definitional extension. The only difference is that now one is allowed to define new sort symbols. Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. A *Morita extension* of T to the signature Σ^+ is a Σ^+-theory

$$T^+ = T \cup \{\delta_s : s \in \Sigma^+ - \Sigma\}$$

that satisfies the following conditions. First, for each symbol $s \in \Sigma^+ - \Sigma$ the sentence δ_s is an explicit definition of s in terms of Σ. Second, if $\sigma \in \Sigma^+ - \Sigma$ is a sort symbol and $f \in \Sigma^+ - \Sigma$ is a function symbol that is used in the sort definition of σ, then $\delta_f = \delta_\sigma$. (For example, if σ is defined as a product sort with projections π_1 and π_2, then $\delta_\sigma = \delta_{\pi_1} = \delta_{\pi_2}$.) And third, if α_s is an admissibility condition for a definition δ_s, then $T \vDash \alpha_s$.

Note that unlike a definitional extension of a theory, a Morita extension can have more sort symbols than the original theory.[8] The following is a particularly simple example of a Morita extension.

[8] Also note that if T^+ is a Morita extension of T to Σ^+, then there are restrictions on the arities of predicates, functions, and constants in $\Sigma^+ - \Sigma$. If $p \in \Sigma^+ - \Sigma$ is a predicate symbol of arity $\sigma_1 \times \cdots \times \sigma_n$, we immediately see that $\sigma_1, \ldots, \sigma_n \in \Sigma$. Taking a single Morita extension does not allow one to define predicate symbols that apply to sorts that are not in Σ. One must take multiple Morita extensions to do this. Likewise, any constant symbol $c \in \Sigma^+ - \Sigma$ must be of sort $\sigma \in \Sigma$. And a function symbol $f \in \Sigma^+ - \Sigma$ must either have arity $\sigma_1 \times \cdots \times \sigma_n \rightarrow \sigma$ with $\sigma_1, \ldots, \sigma_n, \sigma \in \Sigma$, or f must be one of the function symbols that appears in the definition of a new sort symbol $\sigma \in \Sigma^+ - \Sigma$.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

EXAMPLE 4.1. *Let $\Sigma = \{\sigma, p\}$ and $\Sigma^+ = \{\sigma, \sigma^+, p, i\}$ be a signatures with σ and σ^+ sort symbols, p a predicate symbol of arity σ, and i a function symbol of arity $\sigma^+ \to \sigma$. Consider the Σ-theory $T = \{\exists_\sigma x p(x)\}$. The following Σ^+-sentence defines the sort symbol σ^+ as the subsort consisting of "the elements that are p."*

$$\forall_\sigma x\big(p(x) \leftrightarrow \exists_{\sigma^+} z(i(z) = x)\big) \wedge \forall_{\sigma^+} z_1 \forall_{\sigma^+} z_2\big(i(z_1) = i(z_2) \to z_1 = z_2\big) \tag{δ_{σ^+}}$$

The Σ^+-theory $T^+ = T \cup \{\delta_{\sigma^+}\}$ is a Morita extension of T to the signature Σ^+. The theory T^+ adds to the theory T the ability to quantify over the set of "things that are p."

4.2. Three results. As with a definitional extension, a Morita extension "says no more" than the original theory. We will make this idea precise by proving analogues of Theorems 3.1, 3.2, and 3.3. These three results also demonstrate how closely related the concept of a Morita extension is to that of a definitional extension.

Theorem 3.1 generalizes in a perfectly natural way. When T^+ is a Morita extension of T, the models of T^+ are "determined" by the models of T.

THEOREM 4.2. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a Morita extension of T to Σ^+, then every model M of T has a unique expansion (up to isomorphism) M^+ that is a model of T^+.*

Before proving Theorem 4.2, we introduce some notation and prove a lemma. Suppose that a Σ^+-theory T^+ is a Morita extension of a Σ-theory T. Let M and N be models of T^+ with $h : M|_\Sigma \to N|_\Sigma$ an elementary embedding between the Σ-structures $M|_\Sigma$ and $N|_\Sigma$. The elementary embedding h naturally induces a map $h^+ : M \to N$ between the Σ^+-structures M and N.

We know that h is a family of maps $h_\sigma : M_\sigma \to N_\sigma$ for each sort $\sigma \in \Sigma$. In order to describe h^+, we need to describe the map $h^+_\sigma : M_\sigma \to N_\sigma$ for each sort $\sigma \in \Sigma^+$. If $\sigma \in \Sigma$, we simply let $h^+_\sigma = h_\sigma$. On the other hand, when $\sigma \in \Sigma^+ - \Sigma$, there are four cases to consider. We describe h^+_σ in the cases where the theory T^+ defines σ as a product sort or a subsort. The coproduct and quotient sort cases are described analogously.

First, suppose that T^+ defines σ as a product sort. Let $\pi_1, \pi_2 \in \Sigma^+$ be the projections of arity $\sigma \to \sigma_1$ and $\sigma \to \sigma_2$ with $\sigma_1, \sigma_2 \in \Sigma$. The definition of the function h^+_σ is suggested by the following diagram.

$$\begin{tikzcd}
 & M_{\sigma_2} \arrow[r, "h^+_{\sigma_2}"] & N_{\sigma_2} & \\
M_\sigma \arrow[ur, "\pi_2^M"] \arrow[dr, "\pi_1^M"'] \arrow[rrr, dotted, "h^+_\sigma"] & & & N_\sigma \arrow[ul, "\pi_2^N"'] \arrow[dl, "\pi_1^N"] \\
 & M_{\sigma_1} \arrow[r, "h^+_{\sigma_1}"] & N_{\sigma_1} &
\end{tikzcd}$$

Let $m \in M_\sigma$. We define $h^+_\sigma(m)$ to be the unique $n \in N_\sigma$ that satisfies both $\pi_1^N(n) = h^+_{\sigma_1} \circ \pi_1^M(m)$ and $\pi_2^N(n) = h^+_{\sigma_2} \circ \pi_2^M(m)$. We know that such an n exists and is unique because N is a model of T^+ and T^+ defines the symbols σ, π_1, and π_2 to be a product sort. One can verify that this definition of h^+_σ makes the above diagram commute.

Suppose, on the other hand, that T^+ defines σ as the subsort of "elements of sort σ_1 that are ϕ." Let $i \in \Sigma^+$ be the inclusion map of arity $\sigma \to \sigma_1$ with $\sigma_1 \in \Sigma$. As above, the definition of h^+_σ is suggested by the following diagram.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

$$\begin{tikzcd}
M_\sigma \arrow[rrr, dotted, "h^+_\sigma"] \arrow[dr, "i^M"'] & & & N_\sigma \arrow[dl, "i^N"] \\
& M_{\sigma_1} \arrow[r, "h^+_{\sigma_1}"] & N_{\sigma_1} &
\end{tikzcd}$$

Let $m \in M_\sigma$. We see that following implications hold:

$$\begin{aligned} M \vDash \phi[i^M(m)] &\Rightarrow M|_\Sigma \vDash \phi[i^M(m)] \\ &\Rightarrow N|_\Sigma \vDash \phi[h^+_{\sigma_1}(i^M(m))] \Rightarrow N \vDash \phi[h^+_{\sigma_1}(i^M(m))]. \end{aligned}$$

The first and third implications hold since $\phi(x)$ is a Σ-formula, and the second holds because $h_{\sigma_1} = h^+_{\sigma_1}$ and h is an elementary embedding. T^+ defines the symbols i and σ as a subsort and M is a model of T^+, so it must be that $M \vDash \phi[i^M(m)]$. By the above implications, we see that $N \vDash \phi[h^+_{\sigma_1}(i^M(m))]$. Since N is also a model of T^+, there is a unique $n \in N_\sigma$ that satisfies $i^N(n) = h^+_{\sigma_1}(i^M(m))$. We define $h^+_\sigma(m) = n$. This definition of h^+_σ again makes the above diagram commute.

When T^+ defines σ as a coproduct sort or a quotient sort one describes the map h^+_σ analogously. For the purposes of proving Theorem 4.2, we need the following simple lemma about this map h^+.

LEMMA 4.3. *If $h : M|_\Sigma \to N|_\Sigma$ is an isomorphism, then $h^+ : M \to N$ is an isomorphism.*

Proof. We know that $h_\sigma : M_\sigma \to N_\sigma$ is a bijection for each $\sigma \in \Sigma$. Using this fact and the definition of h^+, one can verify that $h^+_\sigma : M_\sigma \to N_\sigma$ is a bijection for each sort $\sigma \in \Sigma^+$. So h^+ is a family of bijections. And furthermore, the commutativity of the above diagrams implies that h^+ preserves any function symbols that are used to define new sorts.

It only remains to check that h^+ preserves predicates, functions, and constants that have arities and sorts in Σ. Since $h : M|_\Sigma \to N|_\Sigma$ is a isomorphism, we know that h^+ preserves the symbols in Σ. So let $p \in \Sigma^+ - \Sigma$ be a predicate symbol of arity $\sigma_1 \times \cdots \times \sigma_n$ with $\sigma_1, \ldots, \sigma_n \in \Sigma$. There must be a Σ-formula $\phi(x_1, \ldots, x_n)$ such that $T^+ \vDash \forall_{\sigma_1} x_1 \ldots \forall_{\sigma_n} x_n (p(x_1, \ldots, x_n) \leftrightarrow \phi(x_1, \ldots, x_n))$. We know that $h : M|_\Sigma \to N|_\Sigma$ is an elementary embedding, so in particular it preserves the formula $\phi(x_1, \ldots, x_n)$. This implies that $(m_1, \ldots, m_n) \in p^M$ if and only if $(h_{\sigma_1}(m_1), \ldots, h_{\sigma_n}(m_n)) \in p^N$. Since $h^+_{\sigma_i} = h_{\sigma_i}$ for each $i = 1, \ldots, n$, it must be that h^+ also preserves the predicate p. An analogous argument demonstrates that h^+ preserves functions and constants. □

We now turn to the proof of Theorem 4.2.

Proof of Theorem 4.2. Let M be a model of T. First note that if M^+ exists, then it is unique up to isomorphism. For if N is a model of T^+ with $N|_\Sigma = M$, then by letting h be the identity map (which is an isomorphism) Lemma 4.3 implies that $M^+ \cong N$. We need only define the Σ^+-structure M^+. To guarantee that M^+ is an expansion of M we interpret every symbol in Σ the same way that M does. We need to say how the symbols in $\Sigma^+ - \Sigma$ are interpreted. There are a number of cases to consider.

Suppose that $p \in \Sigma^+ - \Sigma$ is a predicate symbol of arity $\sigma_1 \times \cdots \times \sigma_n$ with $\sigma_1, \ldots, \sigma_n \in \Sigma$. There must be a Σ-formula $\phi(x_1, \ldots, x_n)$ such that $T^+ \vDash \forall_{\sigma_1} x_1 \ldots \forall_{\sigma_n} x_n (p(x_1, \ldots, x_n) \leftrightarrow \phi(x_1, \ldots, x_n))$. We define the interpretation of the symbol p in M^+ by letting $(a_1, \ldots, a_n) \in p^{M^+}$ if and only if $M \vDash \phi[a_1, \ldots, a_n]$. It is easy to see that this definition of p^A implies that $M^+ \vDash \delta_p$. The cases of function and constant symbols are handled similarly.

Let $\sigma \in \Sigma^+ - \Sigma$ be a sort symbol. We describe the cases where T^+ defines σ as a product sort or a subsort. The coproduct and quotient sort cases follow analogously. Suppose first that σ is defined as a product sort with π_1 and π_2 the projections of arity $\sigma \to \sigma_1$ and $\sigma \to \sigma_2$, respectively. We define $M^+_\sigma = M^+_{\sigma_1} \times M^+_{\sigma_2}$ with $\pi_1^{M^+} : M^+_\sigma \to M^+_{\sigma_1}$ and $\pi_2^{M^+} : M^+_\sigma \to M^+_{\sigma_2}$ the canonical projections. One can easily verify that $M^+ \vDash \delta_\sigma$. On the other hand, suppose that σ is defined as a subsort with defining Σ-formula $\phi(x)$ and inclusion i of arity $\sigma \to \sigma_1$. We define $M^+_\sigma = \{a \in M_{\sigma_1} : M \vDash \phi[a]\}$ with $i^{M^+} : M^+_\sigma \to M^+_{\sigma_1}$ the inclusion map. One can again verify that $M^+ \vDash \delta_\sigma$. □

We have shown that the exact analogue of Theorem 3.1 holds for Morita extensions. Theorem 3.2 also generalizes in a perfectly natural way. Indeed, the generalization follows as a simple corollary to Theorem 4.2.

THEOREM 4.4. *If T^+ is a Morita extension of T, then T^+ is a conservative extension of T.*

Proof. Suppose that T^+ is not a conservative extension of T. One can easily see that $T \vDash \phi$ implies that $T^+ \vDash \phi$ for every Σ-sentence ϕ. So there must be some Σ-sentence ϕ such that $T^+ \vDash \phi$, but $T \nvDash \phi$. This implies that there is a model M of T such that $M \vDash \neg\phi$. This model M has no expansion that is a model of T^+ since $T^+ \vDash \phi$, contradicting Theorem 4.2. □

Theorems 3.1 and 3.2 therefore generalize naturally from definitional extensions to Morita extensions. In order to generalize Theorem 3.3, however, we need to do some work. Theorem 3.3 said that if T^+ is a definitional extension of T to Σ^+, then for every Σ^+-formula $\phi(x_1, \dots, x_n)$ there is a corresponding Σ-formula $\phi^*(x_1, \dots, x_n)$ that is equivalent to $\phi(x_1, \dots, x_n)$ according to the theory T^+. The following example demonstrates that this result does not generalize to the case of Morita extensions in a perfectly straightforward manner.

EXAMPLE 4.5. *Recall the theories T and T^+ from Example 4.1 and consider the Σ^+-formula $\phi(x, z)$ defined by $i(z) = x$. One can easily see that there is no Σ-formula $\phi^*(x, z)$ that is equivalent to $\phi(x, z)$ according to the theory T^+. Indeed, the variable z cannot appear in any Σ-formula since it is of sort $\sigma^+ \in \Sigma^+ - \Sigma$. A Σ-formula simply cannot say how variables with sorts in Σ relate to variables with sorts in Σ^+.*

In order to generalize Theorem 3.3, therefore, we need a way of specifying how variables with sorts in $\Sigma^+ - \Sigma$ relate to variables with sorts in Σ. We do this by defining the concept of a "code."[9] Let $\Sigma \subset \Sigma^+$ be signatures with T a Σ-theory and T^+ a Morita extension of T to Σ^+. A *code* for the variables $x_1, \dots, x_n$ of sorts $\sigma_1, \dots, \sigma_n \in \Sigma^+ - \Sigma$ is a Σ^+-formula

$$\xi_1(x_1, y_{11}, y_{12}) \wedge \cdots \wedge \xi_n(x_n, y_{n1}, y_{n2}),$$

where the conjuncts ξ_i are defined as follows. Suppose that T^+ defines σ_i as a product sort with π_1 and π_2 the projections of arity $\sigma_i \to \sigma_{i1}$ and $\sigma_i \to \sigma_{i2}$. The conjunct $\xi_i(x_i, y_{i1}, y_{i2})$ is then the Σ^+-formula $\pi_1(x_i) = y_{i1} \wedge \pi_2(x_i) = y_{i2}$, where y_{i1} and y_{i2} are variables of sorts $\sigma_{i1}, \sigma_{i2} \in \Sigma$. On the other hand, suppose that T^+ defines σ_i as a coproduct sort with injections ρ_1 and ρ_2 of arity $\sigma_{i1} \to \sigma_i$ and $\sigma_{i2} \to \sigma_i$. Then the

[9] One can compare this concept with the one employed by Szczerba (1977).

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

conjunct ξ_i is either the Σ^+-formula $\rho_1(y_{i1}) = x_i$ or the Σ^+-formula $\rho_2(y_{i2}) = x_i$, where y_{i1} and y_{i2} are again variables of sorts $\sigma_{i1}, \sigma_{i2} \in \Sigma$.

The subsort and quotient sort cases are handled analogously. Suppose that T^+ defines σ_i as a subsort with i the inclusion map of arity $\sigma_i \to \sigma_{i1}$. Then the conjunct ξ_i is the Σ^+-formula $i(x_i) = y_{i1}$, where y_{i1} is a variable of sort $\sigma_{i1} \in \Sigma$. And finally, suppose that T^+ defines σ_i as a quotient sort with ϵ the projection of arity $\sigma_{i1} \to \sigma_i$. The conjunct ξ_i is then the Σ^+-formula $\epsilon(y_{i1}) = x_i$, where y_{i1} is again a variable of sort $\sigma_{i1} \in \Sigma$. Given the empty sequence of variables, we let the *empty code* be the tautology $\exists_\sigma x(x = x)$, where $\sigma \in \Sigma$ is a sort symbol.

Given the conjuncts $\xi_1, \dots, \xi_n$, we will use the notation $\xi(x_1, \dots, y_{n2})$ to denote the code $\xi_1(x_1, y_{11}, y_{12}) \wedge \cdots \wedge \xi_n(x_n, y_{n1}, y_{n2})$ for the variables $x_1, \dots, x_n$. Note that the variables y_{i1} and y_{i2} have sorts in Σ for each $i = 1, \dots, n$. One should think of a code $\xi(x_1, \dots, y_{n2})$ for $x_1, \dots, x_n$ as encoding one way that the variables $x_1, \dots, x_n$ with sorts in $\Sigma^+ - \Sigma$ might be related to variables $y_{11}, \dots, y_{n2}$ that have sorts in Σ. One additional piece of notation will be useful in what follows. Given a Σ^+-formula ϕ, we will write $\phi(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ to indicate that the variables $x_1, \dots, x_n$ have sorts $\sigma_1, \dots, \sigma_n \in \Sigma^+ - \Sigma$ and that the variables $\overline{x}_1, \dots, \overline{x}_m$ have sorts $\overline{\sigma}_1, \dots, \overline{\sigma}_m \in \Sigma$.

We can now state our generalization of Theorem 3.3. One proves this result by induction on the complexity of $\phi(x_1, \dots, x_n)$. The proof has been placed in an appendix.

THEOREM 4.6. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. Suppose that T^+ is a Morita extension of T to Σ^+ and that $\phi(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ is a Σ^+-formula. Then for every code $\xi(x_1, \dots, y_{n2})$ for the variables $x_1, \dots, x_n$ there is a Σ-formula $\phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})$ such that*

$$T^+ \vDash \forall_{\sigma_1} x_1 \dots \forall_{\sigma_n} x_n \forall_{\overline{\sigma}_1} \overline{x}_1 \dots \forall_{\overline{\sigma}_m} \overline{x}_m \forall_{\sigma_{11}} y_{11} \dots \forall_{\sigma_{n2}} y_{n2} \big(\xi(x_1, \dots, y_{n2}) \to (\phi(x_1, \dots x_n, \overline{x}_1, \dots, \overline{x}_m) \leftrightarrow \phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}))\big).$$

The idea behind Theorem 4.6 is simple. Although one might not initially be able to translate a Σ^+-formula ϕ into an equivalent Σ-formula ϕ^*, such a translation is possible after one specifies how the variables in ϕ with sorts in $\Sigma^+ - \Sigma$ are related to variables with sorts in Σ. Theorem 4.6 has the following immediate corollary.

COROLLARY 4.7. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a Morita extension of T to Σ^+, then for every Σ^+-sentence ϕ there is a Σ-sentence ϕ^* such that $T^+ \vDash \phi \leftrightarrow \phi^*$.*

Proof. Let ϕ be a Σ^+-sentence and consider the empty code ξ. Theorem 4.6 implies that there is a Σ-sentence ϕ^* such that $T^+ \vDash \xi \to (\phi \leftrightarrow \phi^*)$. Since ξ is a tautology we trivially have that $T^+ \vDash \phi \leftrightarrow \phi^*$. □

Theorems 4.2, 4.4, and 4.5 capture different senses in which a Morita extension of a theory “says no more” than the original theory. The definition of Morita equivalence is analogous to definitional equivalence.

DEFINITION. *Let T_1 be a Σ_1-theory and T_2 a Σ_2-theory. T_1 and T_2 are* Morita equivalent *if there are theories $T_1^1, \dots, T_1^n$ and $T_2^1, \dots, T_2^m$ that satisfy the following three conditions:*

- *Each theory T_1^{i+1} is a Morita extension of T_1^i,*
- *Each theory T_2^{i+1} is a Morita extension of T_2^i,*
- *T_1^n and T_2^m are logically equivalent Σ-theories with $\Sigma_1 \cup \Sigma_2 \subset \Sigma$.*

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

Two theories are Morita equivalent if they have a “common Morita extension.” The situation can be pictured as follows, where each arrow in the figure indicates a Morita extension.

$$\begin{tikzcd}
 & & T_1^n \cong T_2^m & & \\
 & \cdots \arrow[ur] & & \cdots \arrow[ul] & \\
T_1^1 \arrow[ur] & & & & T_2^1 \arrow[ul] \\
T_1 \arrow[u] & & & & T_2 \arrow[u]
\end{tikzcd}$$

At first glance, Morita equivalence might strike one as different from definitional equivalence in an important way. To show that theories are Morita equivalent, one is allowed to take any finite number of Morita extensions of the theories. On the other hand, to show that two theories are definitionally equivalent, it appears that one is only allowed to take *one* definitional extension of each theory. One might worry that Morita equivalence is therefore not perfectly analogous to definitional equivalence.

Fortunately, this is not the case. Theorem 3.3 implies that if theories $T_1, \ldots, T_n$ are such that each T_{i+1} is a definitional extension of T_i, then T_n is in fact a definitional extension of T_1. (One can easily verify that this is not true of Morita extensions.) To show that two theories are definitionally equivalent, therefore, one actually *is* allowed to take any finite number of definitional extensions of each theory.

If two theories are definitionally equivalent, then they are trivially Morita equivalent. Unlike definitional equivalence, however, Morita equivalence is capable of capturing a sense in which theories with different sort symbols are equivalent. The following example demonstrates that Morita equivalence is a more liberal criterion for theoretical equivalence.

EXAMPLE 4.8. *Let $\Sigma_1 = \{\sigma_1, p, q\}$ and $\Sigma_2 = \{\sigma_2, \sigma_3\}$ be signatures with σ_i sort symbols, and p and q predicate symbols of arity σ_1. Let T_1 be the Σ_1-theory that says: p and q are nonempty, mutually exclusive, and exhaustive. Let T_2 be the empty theory in Σ_2. Since the signatures Σ_1 and Σ_2 have different sort symbols, T_1 and T_2 can’t possibly be definitionally equivalent. Nonetheless, it’s easy to see that T_1 and T_2 are Morita equivalent. Let $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \{i_2, i_3\}$ be a signature with i_2 and i_3 function symbols of arity $\sigma_2 \to \sigma_1$ and $\sigma_3 \to \sigma_1$. Consider the following Σ-sentences.*

$$\begin{aligned}&\forall_{\sigma_1} x\big(p(x) \leftrightarrow \exists_{\sigma_2} y(i_2(y) = x)\big)\\ &\qquad \wedge \forall_{\sigma_2} y_1 \forall_{\sigma_2} y_2\big(i_2(y_1) = i_2(y_2) \to y_1 = y_2\big)\end{aligned} \tag{δ_{σ_2}}$$

$$\begin{aligned}&\forall_{\sigma_1} x\big(q(x) \leftrightarrow \exists_{\sigma_3} z(i_3(z) = x)\big)\\ &\qquad \wedge \forall_{\sigma_3} z_1 \forall_{\sigma_3} z_2\big(i_3(z_1) = i_3(z_2) \to z_1 = z_2\big)\end{aligned} \tag{δ_{σ_3}}$$

$$\begin{aligned}&\forall_{\sigma_1} x\big(\exists_{\sigma_2 = 1} y(i_2(y) = x) \vee \exists_{\sigma_3 = 1} z(i_3(z) = x)\big)\\ &\qquad \wedge \forall_{\sigma_2} y \forall_{\sigma_3} z \neg\big(i_2(y) = i_3(z)\big)\end{aligned} \tag{δ_{σ_1}}$$

$$\forall_{\sigma_1} x\big(p(x) \leftrightarrow \exists_{\sigma_2} y(i_2(y) = x)\big) \tag{δ_p}$$

$$\forall_{\sigma_1} x\big(q(x) \leftrightarrow \exists_{\sigma_3} z(i_3(z) = x)\big) \tag{δ_q}$$

The Σ-theory $T_1^1 = T_1 \cup \{\delta_{\sigma_2}, \delta_{\sigma_3}\}$ is a Morita extension of T_1 to the signature Σ. It defines σ_2 to be the subsort of “elements that are p” and σ_3 to be the subsort of “elements that are q.” The theory $T_2^1 = T_2 \cup \{\delta_{\sigma_1}\}$ is a Morita extension of T_2 to the signature $\Sigma_2 \cup \{\sigma_1, i_2, i_3\}$.

It defines σ_1 to be the coproduct sort of σ_2 and σ_3. Lastly, the Σ-theory $T_2^2 = T_2^1 \cup \{\delta_p, \delta_q\}$ is a Morita extension of T_2^1 to the signature Σ. It defines the predicates p and q to apply to elements in the "images" of i_2 and i_3, respectively. One can verify that T_1^1 and T_2^2 are logically equivalent, so T_1 and T_2 are Morita equivalent.

§5. Categorical equivalence. Morita equivalence captures a clear and robust sense in which theories might be equivalent, but it is a difficult criterion to apply outside of the framework of first-order logic. Indeed, without a formal language one does not have the resources to say what an explicit definition is. Questions of equivalence and inequivalence of theories, however, still come up outside of this framework. It is well known, for example, that there are different ways of formulating the theory of smooth manifolds (Nestruev, 2002). There are also different formulations of the theory of topological spaces (Kuratowski, 1966). None of these formulations are first-order theories. Physical theories too are rarely formulated in first-order logic, and there are many pairs of physical theories that are often considered equivalent.[10]

Morita equivalence is incapable of capturing any sense in which these theories are equivalent. We need a criterion for theoretical equivalence that is applicable outside the framework of first-order logic. *Categorical equivalence* is one such criterion.[11] It was first described by Eilenberg & Mac Lane (1942, 1945), but was only recently introduced into philosophy of science by Halvorson (2012, 2015) and Weatherall (2015a). In this section, we describe categorical equivalence and then show how it is related to Morita equivalence.

Categorical equivalence is motivated by the following simple observation: First-order theories have categories of models. A *category* C is a collection of objects with arrows between the objects that satisfy two basic properties. First, there is an associative composition operation $\circ$ defined on the arrows of C, and second, every object c in C has an identity arrow $1_c : c \to c$. If T is a Σ-theory, we will use the notation $\text{Mod}(T)$ to denote the *category of models* of T. An object in $\text{Mod}(T)$ is a model M of T. For the arrows of $\text{Mod}(T)$, we have a couple of salient choices. On the one hand, we could choose arrows to be homomorphisms, i.e., $f : M \to N$ is a function (or family of functions) that preserves the extensions of the terms in the signature Σ. On the other hand, we could choose arrows to be elementary embeddings, i.e., $f : M \to N$ is an injective function (or family of functions) that preserves the extensions of all Σ formulas.

Let $\text{Mod}(T)$ denote the category with elementary embeddings as arrows, and let $\text{Mod}_h(T)$ denote the category with homomorphisms as arrows. But which of these two categories, $\text{Mod}(T)$ or $\text{Mod}_h(T)$ should we think of as representing the theory T? We will choose the category $\text{Mod}(T)$, with elementary embeddings as arrows, for the following reasons.

First, the image of a model of T under a homomorphism f is not necessarily a model of T. For example, let T be the theory (in a single-sorted signature) that says there are exactly two things. Then a model M of T is a set with two elements. However, the mapping

[10] For example, see Glymour (1977), Knox (2014), and Weatherall (2015a) for discussion of whether or not Newtonian gravitation and geometrized Newtonian gravitation are equivalent. See North (2009), Halvorson (2011), Swanson & Halvorson (2012), Curiel (2014), and Barrett (2015) for discussion of whether or not Hamiltonian and Lagrangian mechanics are equivalent. See Rosenstock, Barrett, & Weatherall (2015) for a discussion of general relativity and the theory of Einstein algebras and Weatherall (2015b) for a summary of many of these results.

[11] The reader is encouraged to consult Mac Lane (1971), Borceux (1994), or Awodey (2010) for preliminaries.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

$f : M \to M$ that takes both elements to a single element is a homomorphism, and its image $f(M)$ is not a model of T. Such a situation is not necessarily a disaster, but it shows that homomorphisms do not mesh well with full first-order logic.

Second, $\mathrm{Mod}_h(\cdot)$ does not even preserve definitional equivalence, i.e., there are definitionally equivalent theories T_1 and T_2 such that $\mathrm{Mod}_h(T_1)$ is not categorically equivalent to $\mathrm{Mod}_h(T_2)$.

EXAMPLE 5.1. *Let $\Sigma_1 = \{\sigma\}$, where σ is a sort symbol, and let T_1 be the theory in Σ_1 that says there are exactly two things. Let $\Sigma_2 = \{\sigma, \theta\}$ where θ is a relation of arity $\sigma \times \sigma$, and let T_2 be the theory in Σ_2 that says there are exactly two things, and $T_2 \vDash \theta(x, y) \leftrightarrow (x \neq y)$. Obviously T_2 is a definitional extension of T_1. Now, every arrow of* $\mathrm{Mod}_h(T_2)$ *is an injection, since it preserves θ and hence $\neq$. But arrows of* $\mathrm{Mod}_h(T_1)$ *need not be injections. Therefore,* $\mathrm{Mod}_h(T_1)$ *and* $\mathrm{Mod}_h(T_2)$ *are not categorically equivalent.*

Because of these issues with homomorphisms, we will henceforth associate a theory T with the category $\mathrm{Mod}(T)$ of its models and elementary embeddings.

Before describing categorical equivalence, we need some additional terminology. Let C and D be categories. A *functor* $F : C \to D$ is a map from objects and arrows of C to objects and arrows of D that satisfies

$$F(f : a \to b) = Ff : Fa \to Fb \qquad F(1_c) = 1_{Fc} \qquad F(g \circ h) = Fg \circ Fh$$

for every arrow $f : a \to b$ in C, every object c in C, and every composable pair of arrows g and h in C. Functors are the "structure-preserving maps" between categories; they preserve domains, codomains, identity arrows, and the composition operation. A functor $F : C \to D$ is *full* if for all objects c_1, c_2 in C and arrows $g : Fc_1 \to Fc_2$ in D there exists an arrow $f : c_1 \to c_2$ in C with $Ff = g$. F is *faithful* if $Ff = Fg$ implies that $f = g$ for all arrows $f : c_1 \to c_2$ and $g : c_1 \to c_2$ in C. F is *essentially surjective* if for every object d in D there exists an object c in C such that $Fc \cong d$. A functor $F : C \to D$ that is full, faithful, and essentially surjective is called an *equivalence of categories*. The categories C and D are *equivalent* if there exists an equivalence between them.[12]

A first-order theory T has a category of models $\mathrm{Mod}(T)$. This categorical structure, however, is not particular to first-order theories. Indeed, one can easily define categories of models for the different formulations of the theory of smooth manifolds and for the different formulations of the theory of topological spaces. The arrows in these categories are simply the structure-preserving maps between the objects in the categories. One can also define categories of models for physical theories.[13] This means that the following criterion for theoretical equivalence is applicable in a more general setting than definitional equivalence and Morita equivalence. In particular, it can be applied outside of the framework of first-order logic.

DEFINITION. *Theories T_1 and T_2 are* categorically equivalent *if their categories of models Mod(T_1) and Mod(T_2) are equivalent.*

[12] The concept of a "natural transformation" is often used to define when two categories are equivalent. C and D are equivalent if there are functors $F : C \to D$ and $G : D \to C$ such that FG is naturally isomorphic to the identity functor 1_D and GF is naturally isomorphic to 1_C. See Mac Lane (1971) for the definition of a natural transformation and for proof that these two characterizations of equivalence are the same.

[13] See the examples in (Weatherall, 2015a,b,c) and Rosenstock, Barrett, & Weatherall (2015).

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

Categorical equivalence captures a sense in which theories have “isomorphic semantic structure.” If T_1 and T_2 are categorically equivalent, then the relationships that models of T_1 bear to one another are “isomorphic” to the relationships that models of T_2 bear to one another.

In order to show how categorical equivalence relates to Morita equivalence, we focus on first-order theories. We will show that categorical equivalence is a strictly weaker criterion for theoretical equivalence than Morita equivalence is. We first need some preliminaries about the category of models $\text{Mod}(T)$ for a first-order theory T. Suppose that $\Sigma \subset \Sigma^+$ are signatures and that the Σ^+-theory T^+ is an extension of the Σ-theory T. There is a natural “projection” functor $\Pi : \text{Mod}(T^+) \to \text{Mod}(T)$ from the category of models of T^+ to the category of models of T. The functor Π is defined as follows.

- $\Pi(M) = M|_\Sigma$ for every object M in $\text{Mod}(T^+)$.
- $\Pi(h) = h|_\Sigma$ for every arrow $h : M \to N$ in $\text{Mod}(T^+)$, where the family of maps $h|_\Sigma$ is defined to be $h|_\Sigma = \{h_\sigma : M_\sigma \to N_\sigma \text{ such that } \sigma \in \Sigma\}$.

Since T^+ is an extension of T, the Σ-structure $\Pi(M)$ is guaranteed to be a model of T. Likewise, the map $\Pi(h) : M|_\Sigma \to N|_\Sigma$ is guaranteed to be an elementary embedding. One can easily verify that $\Pi : \text{Mod}(T^+) \to \text{Mod}(T)$ is a functor.

The following three propositions will together establish the relationship between $\text{Mod}(T^+)$ and $\text{Mod}(T)$ when T^+ is a Morita extension of T. They imply that when T^+ is a Morita extension of T, the functor $\Pi : \text{Mod}(T^+) \to \text{Mod}(T)$ is full, faithful, and essentially surjective. The categories $\text{Mod}(T^+)$ and $\text{Mod}(T)$ are therefore equivalent.

PROPOSITION 5.2. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a Morita extension of T to Σ^+, then Π is essentially surjective.*

Proof. If M is a model of T, then Theorem 4.2 implies that there is a model M^+ of T^+ that is an expansion of M. Since $\Pi(M^+) = M^+|_\Sigma = M$ the functor Π is essentially surjective. □

PROPOSITION 5.3. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a Morita extension of T to Σ^+, then Π is faithful.*

Proof. Let $h : M \to N$ and $g : M \to N$ be arrows in $\text{Mod}(T^+)$ and suppose that $\Pi(h) = \Pi(g)$. We show that $h = g$. By assumption $h_\sigma = g_\sigma$ for every sort symbol $\sigma \in \Sigma$. We show that $h_\sigma = g_\sigma$ also for $\sigma \in \Sigma^+ - \Sigma$. We consider the cases where T^+ defines σ as a product sort or a subsort. The coproduct and quotient sort cases follow analogously.

Suppose that T^+ defines σ as a product sort with projections π_1 and π_2 of arity $\sigma \to \sigma_1$ and $\sigma \to \sigma_2$. Then the following equalities hold:

$$\pi_1^N \circ h_\sigma = h_{\sigma_1} \circ \pi_1^M = g_{\sigma_1} \circ \pi_1^M = \pi_1^N \circ g_\sigma.$$

The first and third equalities hold since h and g are elementary embeddings and the second since $h_{\sigma_1} = g_{\sigma_1}$. One can verify in the same manner that $\pi_2^N \circ h_\sigma = \pi_2^N \circ g_\sigma$. Since N is a model of T^+ and T^+ defines σ as a product sort, we know that $N \vDash \forall_{\sigma_1} x \forall_{\sigma_2} y \exists_{\sigma=1} z(\pi_1(z) = x \wedge \pi_2(z) = y)$. This implies that $h_\sigma = g_\sigma$.

On the other hand, if T^+ defines σ as a subsort with injection i of arity $\sigma \to \sigma_1$, then the following equalities hold:

$$i^N \circ h_\sigma = h_{\sigma_1} \circ i^M = g_{\sigma_1} \circ i^M = i^N \circ g_\sigma.$$

These equalities follow in the same manner as above. Since i^N is an injection it must be that $h_\sigma = g_\sigma$. □

Before proving that Π is full, we need the following simple lemma.

LEMMA 5.4. *Let M be a model of T^+ with $a_1, \dots, a_n$ elements of M of sorts $\sigma_1, \dots, \sigma_n \in \Sigma^+ - \Sigma$. If $x_1, \dots, x_n$ are variables sorts $\sigma_1, \dots, \sigma_n$, then there is a code $\xi(x_1, \dots, x_n, y_{11}, \dots, y_{n2})$ and elements $b_{11}, \dots, b_{n2}$ of M such that $M \models \xi[a_1, \dots, a_n, b_{11}, \dots, b_{n2}]$.*

Proof. We define the code $\xi(x_1, \dots, y_{n2})$. If T^+ defines σ_i as a product sort, quotient sort, or subsort then we have no choice about what the conjunct $\xi_i(x_i, y_{i1}, y_{i2})$ is. If T^+ defines σ_i as a coproduct sort, then we know that either there is an element b_{i1} of M such that $\rho_1(b_{i1}) = a_i$ or there is an element b_{i2} of M such that $\rho_2(b_{i2}) = a_i$. If the former, we let ξ_i be $\rho_1(y_{i1}) = x_i$ and if the latter, we let ξ_i be $\rho_2(y_{i2}) = x_i$. One defines the elements $b_{11}, \dots, b_{n2}$ in the obvious way. For example, if σ_i is a product sort, then we let $b_{i1} = \pi_1^M(a_i)$ and $b_{i2} = \pi_2^M(a_i)$. By construction, we have that $M \models \xi[a_1, \dots, a_n, b_{11}, \dots, b_{n2}]$. □

We now use this lemma to show that Π is full.

PROPOSITION 5.5. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. If T^+ is a Morita extension of T to Σ^+, then Π is full.*

Proof. Let M and N be models of T^+ with $h : \Pi(M) \to \Pi(N)$ an arrow in Mod(T). This means that $h : M|_\Sigma \to N|_\Sigma$ is an elementary embedding. We show that the map $h^+ : M \to N$ is an elementary embedding and therefore an arrow in Mod(T^+). Since $\Pi(h^+) = h$ this will imply that Π is full.

Let $\phi(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ be a Σ^+-formula and let $a_1, \dots, a_n, \overline{a}_1, \dots, \overline{a}_m$ be elements of M of the same sorts as the variables $x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m$. Lemma 5.4 implies that there is a code $\xi(x_1, \dots, x_n, y_{11}, \dots, y_{n2})$ and elements $b_{11}, \dots, b_{n2}$ of M such that $M \models \xi[a_1, \dots, a_n, b_{11}, \dots, b_{n2}]$. The definition of the map h^+ implies that $N \models \xi[h^+(a_1, \dots, a_n, b_{11}, \dots, b_{n2})]$. We now show that $M \models \phi[a_1, \dots, a_n, \overline{a}_1, \dots, \overline{a}_m]$ if and only if $N \models \phi[h^+(a_1, \dots, a_n, \overline{a}_1, \dots, \overline{a}_m)]$. By Theorem 4.6 there is a Σ-formula $\phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})$ such that

$$\begin{aligned} T^+ \models \forall_{\sigma_1} x_1 \dots \forall_{\sigma_n} x_n \forall_{\overline{\sigma}_1} \overline{x}_1 \dots \forall_{\overline{\sigma}_m} \overline{x}_m \forall_{\sigma_{11}} y_{11} \dots \forall_{\sigma_{n2}} y_{n2} \big(\xi(x_1, \dots, y_{n2}) \rightarrow \\ \big(\phi(x_1, \dots x_n, \overline{x}_1, \dots, \overline{x}_m) \leftrightarrow \phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})\big)\big). \end{aligned} \tag{5}$$

We then see that the following string of equivalences holds.

$$\begin{aligned} M \models \phi[a_1, \dots, a_n, \overline{a}_1, \dots, \overline{a}_m] &\Longleftrightarrow M \models \phi^*[\overline{a}_1, \dots, \overline{a}_m, b_{11}, \dots, b_{n2}] \\ &\Longleftrightarrow M|_\Sigma \models \phi^*[\overline{a}_1, \dots, \overline{a}_m, b_{11}, \dots, b_{n2}] \\ &\Longleftrightarrow N|_\Sigma \models \phi^*[h(\overline{a}_1, \dots, \overline{a}_m, b_{11}, \dots, b_{n2})] \\ &\Longleftrightarrow N \models \phi^*[h(\overline{a}_1, \dots, \overline{a}_m, b_{11}, \dots, b_{n2})] \\ &\Longleftrightarrow N \models \phi^*[h^+(\overline{a}_1, \dots, \overline{a}_m, b_{11}, \dots, b_{n2})] \\ &\Longleftrightarrow N \models \phi[h^+(a_1, \dots, a_n, \overline{a}_1, \dots, \overline{a}_m)] \end{aligned}$$

The first and sixth equivalences hold by (5) and the fact that M and N are models of T^+, the second and fourth hold since ϕ^* is a Σ-formula, the third since $h : M|_\Sigma \to N|_\Sigma$ is an elementary embedding, and the fifth by the definition of h^+ and the fact that the elements $\overline{a}_1, \dots, \overline{a}_m, b_{11}, \dots, b_{n2}$ have sorts in Σ. □

These three propositions provide us with the resources to show how categorical equivalence is related to Morita equivalence. Our first result follows as an immediate corollary.

THEOREM 5.6. *Morita equivalence entails categorical equivalence.*

Proof. Suppose that T_1 and T_2 are Morita equivalent. Then there are theories $T_1^1, \ldots, T_1^n$ and $T_2^1, \ldots, T_2^m$ that satisfy the three conditions in the definition of Morita equivalence. Propositions 5.2, 5.3, and 5.5 imply that the Π functors between these theories, represented by the arrows in the following figure, are all equivalences.

$$\begin{tikzcd}
& & \mathrm{Mod}(T_1^n)=\mathrm{Mod}(T_2^m) \arrow[dl] \arrow[dr] & & \\
& \iddots \arrow[dl] & & \ddots \arrow[dr] & \\
\mathrm{Mod}(T_1^1) \arrow[d] & & & & \mathrm{Mod}(T_2^1) \arrow[d] \\
\mathrm{Mod}(T_1) & & & & \mathrm{Mod}(T_2)
\end{tikzcd}$$

This implies that $\mathrm{Mod}(T_1)$ is equivalent to $\mathrm{Mod}(T_2)$, and so T_1 and T_2 are categorically equivalent. □

The converse to Theorem 5.6, however, does not hold. There are theories that are categorically equivalent but not Morita equivalent.[14] In order to show this, we need one piece of terminology. A category C is *discrete* if it is equivalent to a category whose only arrows are identity arrows.

THEOREM 5.7. *Categorical equivalence does not entail Morita equivalence.*

Proof. Let $\Sigma_1 = \{\sigma_1, p_0, p_1, p_2, \ldots\}$ be a signature with a single sort symbol σ_1 and a countable infinity of predicate symbols p_i of arity σ_1. Let $\Sigma_2 = \{\sigma_2, q_0, q_1, q_2, \ldots\}$ be a signature with a single sort symbol σ_2 and a countable infinity of predicate symbols q_i of arity σ_2. Define the Σ_1-theory T_1 and Σ_2-theory T_2 as follows:

$$\begin{aligned}
T_1 &= \{\exists_{\sigma_1=1} x(x = x)\} \\
T_2 &= \{\exists_{\sigma_2=1} y(y = y), \forall_{\sigma_2} y(q_0(y) \to q_1(y)), \forall_{\sigma_2} y(q_0(y) \to q_2(y)), \ldots\}.
\end{aligned}$$

The theory T_2 has the sentence $\forall_{\sigma_2} y(q_0(y) \to q_i(y))$ as an axiom for each $i \in \mathbb{N}$.

We first show that T_1 and T_2 are categorically equivalent. It is easy to see that $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ both have $2^{\aleph_0}$ (nonisomorphic) objects. Furthermore, $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ are both discrete categories. We show here that $\mathrm{Mod}(T_1)$ is discrete. Suppose that there is an elementary embedding $f : M \to N$ between models M and N of T_1. It must be that f maps the unique element $m \in M$ to the unique element $n \in N$. Furthermore, since f is an elementary embedding, $M \vDash p_i[m]$ if and only if $N \vDash p_i[n]$ for every predicate $p_i \in \Sigma_1$. This implies that $f : M \to N$ is actually an isomorphism. Every arrow $f : M \to N$ in $\mathrm{Mod}(T_1)$ is therefore an isomorphism, and there is at most one arrow between any two objects of $\mathrm{Mod}(T_1)$. This immediately implies that $\mathrm{Mod}(T_1)$ is discrete. An analogous argument demonstrates that $\mathrm{Mod}(T_2)$ is discrete. Any bijection between the objects of $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ is therefore an equivalence of categories.

[14] Halvorson (2012) mentions the following example to illustrate a different point.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

But T_1 and T_2 are not Morita equivalent. Suppose for contradiction that T is a "common Morita extension" of T_1 and T_2. Corollary 4.7 implies that there is a Σ_1-sentence ϕ such that $T \vDash \forall_{\sigma_2} y q_0(y) \leftrightarrow \phi$. One can verify using Theorem 4.4 and Corollary 4.7 that the sentence ϕ has the following property: If ψ is a Σ_1-sentence and $T_1 \vDash \psi \rightarrow \phi$, then either (i) $T_1 \vDash \neg\psi$ or (ii) $T_1 \vDash \phi \rightarrow \psi$. But ϕ cannot have this property. Consider the Σ_1-sentence

$$\psi := \phi \wedge \forall_{\sigma_1} x p_i(x),$$

where p_i is a predicate symbol that does not occur in ϕ. We trivially see that $T_1 \vDash \psi \rightarrow \phi$, but neither (i) nor (ii) hold of ψ. This implies that T_1 and T_2 are not Morita equivalent. □

§6. Conclusions and open questions.

We have discussed three formal criteria for theoretical equivalence, and we have shown that they form the following hierarchy.

$$\begin{tikzcd}
\text{Definitional equivalence} \arrow[r, shift left] & \text{Morita equivalence} \arrow[l, shift left, "/" marking] \arrow[r, shift left] & \text{Categorical equivalence} \arrow[l, shift left, "/" marking]
\end{tikzcd}$$

This hierarchy suggests that definitional equivalence is too strict a criterion for theoretical equivalence. One often has good reason to consider two theories equivalent even if they make use of different quantifiers—as in the case of point-based versus line-based formulations of geometry. But definitional equivalence does not allow this flexibility. Morita equivalence, on the other hand, explains the precise sense in which such theories are equivalent.

The hierarchy also suggests that categorical equivalence is too liberal a criterion for theoretical equivalence. Indeed, the example from Theorem 5.7 captures the general fact that any two theories with discrete categories of models are categorically equivalent, as long as they have the same number of models. But one often has good reason to consider two such theories inequivalent. For example, there is a sense in which the two theories from Theorem 5.7 do not "say the same thing." According to the theory T_2, there is a special predicate q_0: in any model of T_2, if q_0 holds of the unique element, then so do all the other q_i. The theory T_1, however, singles out no such predicate. If one takes categorical equivalence as the standard for theoretical equivalence, then one is forced to consider T_1 and T_2 equivalent. Morita equivalence, on the other hand, allows one to consider them inequivalent.

It would be interesting to know whether there are special cases where the implications in the above diagram are reversible. To this end, we conclude with a few conjectures and open questions.

First, the two categorically equivalent theories from Theorem 5.7 have infinite signatures. We conjecture that among theories with finite signatures, categorical equivalence implies Morita equivalence.

QUESTION 6.1. *For theories with finite signatures, does categorical equivalence imply Morita equivalence?*

It is obvious that at least for propositional theories, this question has an affirmative answer. (Indeed, finite propositional theories are definitionally equivalent iff they have the same number of models.) But even if this question turns out to have a negative answer in general, it would nonetheless be interesting to describe some subclasses of theories for which categorical equivalence implies Morita equivalence.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

Second, considering the examples we gave of Morita equivalent theories that are not definitionally equivalent, we see that they are all relatively weak theories.[15] Indeed, it seems that stronger first-order theories—such as Peano arithmetic or ZF set theory—are in a sense "Morita complete," since they can explicitly define new sorts. For example, since Peano Arithmetic (PA) allows sequences of integers to be coded as single integers, product sorts in PA can be identified with subsorts of the original sort. Furthermore, since PA has uniform elimination of imaginaries, quotient sorts can be identified with subsorts of the original sort. Thus, we raise the following question:

QUESTION 6.2. *Suppose T is a single-sorted theory that is Morita equivalent to PA (respectively ZF). Then is T definitionally equivalent to PA (respectively ZF)?*

And more generally:

QUESTION 6.3. *For sufficiently strong single-sorted theories, does Morita equivalence imply definitional equivalence?*

Of course, the really interesting question here is what "sufficiently strong" could mean in this context. Could a theory's ability to explicitly define product and quotient sorts provide an interesting new measure of that theory's strength?

Third, it would be interesting to know whether categories of models naturally carry additional structure such that categorical equivalence *plus* preservation of this additional structure entails Morita equivalence. Such a possibility is suggested by the Stone duality theorem for propositional logic: in that case, the set $\mathrm{Mod}(T)$ of models of T is naturally equipped with the Stone topology, and $\mathrm{Mod}(T_1)$ is homeomorphic to $\mathrm{Mod}(T_2)$ iff T_1 and T_2 are definitionally equivalent. A partial generalization of Stone duality for first-order logic has already been obtained by Makkai (1991), and in a different manner by Awodey & Forssell (2010). Makkai shows that if the *ultracategories* $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ are equivalent, then T_1 and T_2 are Morita equivalent. Awodey and Forssell show that if the *topological groupoids* $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ are equivalent, then T_1 and T_2 are Morita equivalent. But there is still more work to be done before we completely understand the relationship between Morita equivalence and categorical equivalence.

Finally, our proof of Theorem 4.6 shows that if T^+ is a Morita extension of T, then there is a sort of interpretation $J : T^+ \to T$. We conjecture that this map $J : T^+ \to T$ is, in fact, a *generalized interpretation* in the sense of van Benthem & Pearce (1984). What's more, there is a natural interpretation $I : T \to T^+$, and we conjecture that I and J are inverses of each other—which would strengthen the case that T and T^+ should be considered to be equivalent theories.[*]

Appendix. This appendix contains a proof of Theorem 4.6, which we restate here for convenience.

THEOREM 4.6. *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. Suppose that T^+ is a Morita extension of T to Σ^+ and that $\phi(x_1, \ldots, x_n, \overline{x}_1, \ldots, \overline{x}_m)$ is a Σ^+-formula. Then for every*

[15] By "weak" here, we are thinking in terms of relative interpretability. So, for example, affine geometry is weaker than ZF set theory since the former can be interpreted in the latter, but not vice versa.

[*] This material is based upon work supported by the National Science Foundation under Grant No. DGE 1148900.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

code $\xi(x_1, \dots, y_{n2})$ for the variables $x_1, \dots, x_n$ there is a Σ-formula $\phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})$ such that

$$T^+ \vDash \forall_{\sigma_1} x_1 \dots \forall_{\sigma_n} x_n \forall_{\overline{\sigma}_1} \overline{x}_1 \dots \forall_{\overline{\sigma}_m} \overline{x}_m \forall_{\sigma_{11}} y_{11} \dots \forall_{\sigma_{n2}} y_{n2} \big(\xi(x_1, \dots, y_{n2}) \rightarrow \big(\phi(x_1, \dots x_n, \overline{x}_1, \dots, \overline{x}_m) \leftrightarrow \phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})\big)\big).$$

We first prove the following lemma. Given a Σ^+-term t, we will again write $t(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ to indicate that the variables $x_1, \dots, x_n$ have sorts $\sigma_1, \dots, \sigma_n \in \Sigma^+ - \Sigma$ and that the variables $\overline{x}_1, \dots, \overline{x}_m$ have sorts $\overline{\sigma}_1, \dots, \overline{\sigma}_m \in \Sigma$.

LEMMA A.1. *Let $t(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ be a Σ^+-term of sort σ and x a variable of sort σ. Let $\xi(x, x_1, \dots, x_n, y_1, y_2, y_{11}, \dots, y_{n2})$ be a code for the variables $x, x_1, \dots, x_n$, where the variables y_1 and y_2 are used for coding the variable x. Then there is a Σ-formula $\phi_t(x, \overline{x}_1, \dots, \overline{x}_m, y_{01}, \dots, y_{n2})$ such that*

$$T^+ \vDash \forall_\sigma x \forall_{\sigma_1} x_1 \dots \forall_{\sigma_n} x_n \forall_{\overline{\sigma}_1} \overline{x}_1 \dots \forall_{\overline{\sigma}_m} \overline{x}_m \forall_{\sigma_{11}} y_{11} \dots \forall_{\sigma_{n2}} y_{n2} \big(\xi(x, \dots, y_{n2}) \rightarrow \big(t(x_1, \dots, \overline{x}_m) = x \leftrightarrow \phi_t(x, \overline{x}_1, \dots, \overline{x}_m, y_1, \dots, y_{n2})\big)\big).$$

If $\sigma \in \Sigma$, then x will not appear in the code ξ. If $\sigma \in \Sigma^+ - \Sigma$, then x will not appear in the Σ-formula ϕ_t.

Proof. We induct on the complexity of t. First, suppose that t is a variable x_i of sort σ. If $\sigma \in \Sigma$, then there are no variables in t with sorts in $\Sigma^+ - \Sigma$. So ξ must be the empty code. Let $\phi_t(x, x_i)$ be the Σ-formula $x = x_i$. This choice of ϕ_t trivially satisfies the desired property. If $\sigma \in \Sigma^+ - \Sigma$, then there are four cases to consider. We consider the cases where σ is a product sort and a subsort. The coproduct and quotient cases follow analogously. Suppose that T^+ defines σ as a product sort with projections π_1 and π_2 of arity $\sigma \rightarrow \sigma_1$ and $\sigma \rightarrow \sigma_2$. A code ξ for the variables x and x_i must therefore be the formula

$$\pi_1(x) = y_1 \wedge \pi_2(x) = y_2 \wedge \pi_1(x_i) = y_{i1} \wedge \pi_2(x_i) = y_{i2}.$$

One defines the Σ-formula ϕ_t to be $y_1 = y_{i1} \wedge y_2 = y_{i2}$ and verifies that it satisfies the desired property. On the other hand, suppose that T^+ defines σ as a subsort with injection i of arity $\sigma \rightarrow \sigma_1$. A code ξ for the variables x and x_i is therefore the formula

$$i(x) = y \wedge i(x_i) = y_{i1}.$$

Let ϕ_t be the Σ-formula $y = y_{i1}$. The desired property again holds.

Second, suppose that t is the constant symbol c. Note that it must be the case that c is of sort $\sigma \in \Sigma$. If $c \in \Sigma$, then letting ϕ_t be the Σ-formula $x = c$ trivially yields the result. If $c \in \Sigma^+ - \Sigma$, then there is some Σ-formula $\psi(x)$ that T^+ uses to explicitly define c. Letting $\phi_t = \psi$ yields the desired result.

For the third (and final) step of the induction, we suppose that t is a term of the form

$$f\big(t_1(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m), \dots, t_k(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)\big),$$

where $f \in \Sigma^+$ is a function symbol. We show that the result holds for t if it holds for all of the terms $t_1, \dots, t_k$. There are three cases to consider. First, if $f \in \Sigma$, then it must be that f has arity $\sigma_1 \times \dots \times \sigma_k \rightarrow \sigma$, where $\sigma, \sigma_1, \dots, \sigma_k \in \Sigma$. Let ξ be a code for $x_1, \dots, x_n$. We define ϕ_t to be the Σ-formula

$$\exists_{\sigma_1} z_1 \dots \exists_{\sigma_k} z_k \big(\phi_{t_1}(z_1, \overline{x}_1, \dots, y_{n2}) \wedge \dots \wedge \phi_{t_k}(z_k, \overline{x}_1, \dots y_{n2}) \wedge f(z_1, \dots, z_k) = x\big),$$

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

where each of the ϕ_{t_i} exists by our inductive hypothesis. One can verify that ϕ_t satisfies the desired property. Second, if $f \in \Sigma^+ - \Sigma$ is defined by a Σ-formula $\psi(z_1, \dots, z_k, x)$ then one defines ϕ_t in an analogous manner to above. (Note that in this case the arity of f is again $\sigma_1 \times \cdots \times \sigma_k \to \sigma$ with $\sigma_1, \dots, \sigma_k, \sigma \in \Sigma$.)

Third, we need to verify that the result holds if f is a function symbol that is used in the definition of a new sort. We discuss the cases where f is π_1 and where f is ϵ. Suppose that f is π_1 with arity $\sigma \to \sigma_1$. Then it must be that the term t_1 is a variable x_i of sort σ since there are no other Σ^+-terms of sort σ. So the term t is $\pi_1(x_i)$. Let $\xi(x_i, y_{i1}, y_{i2})$ be a code for x_i. It must be that ξ is the formula

$$\pi_1(x_i) = y_{i1} \wedge \pi_2(x_i) = y_{i2}.$$

Letting ϕ_t be the formula $y_{i1} = x$ yields the desired result. On the other hand, suppose that f is the function symbol ϵ of arity $\sigma_1 \to \sigma$, where σ is a quotient sort defined by the Σ-formula $\psi(z_1, z_2)$. The term t in this case is $\epsilon(t_1(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m))$ and we assume that the result holds for the Σ^+-term t_1 of sort $\sigma_1 \in \Sigma$. Let ξ be a code for the variables $x, x_1, \dots, x_n$. This code determines a code $\overline{\xi}$ for the variables $x_1, \dots, x_n$ by "forgetting" the conjunct $\epsilon(y) = x$ that involves the variable x. We use the code $\overline{\xi}$ and the inductive hypothesis to obtain the formula ϕ_{t_1}. Then we define ϕ_t to be the Σ-formula

$$\exists_{\sigma_1} z\big(\phi_{t_1}(z, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}) \wedge \psi(y, z)\big).$$

Considering the original code ξ, one verifies that the result holds for ϕ_{t_1}. □

We now turn to the proof of the main result.

Proof of Theorem 4.6. We induct on the complexity of ϕ. Suppose that ϕ is the formula $t(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m) = s(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ where t and s are Σ^+-terms of sort σ. Let $\xi(x_1, \dots, y_{n2})$ be a code for $x_1, \dots, x_n$ and let x be a variable of sort σ. By Lemma A.1, there are corresponding Σ-formulas $\phi_t(x, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})$ and $\phi_s(x, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})$. The Σ-formula ϕ^* is then defined to be

$$\exists_\sigma x\big(\phi_t(x, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}) \wedge \phi_s(x, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})\big).$$

One can verify that this definition of ϕ^* satisfies the desired result.

If t and s are of sort $\sigma \in \Sigma^+ - \Sigma$, then there are four cases to consider. We show that the result holds when T^+ defines σ as a product sort or a quotient sort. The coproduct and subsort cases follow analogously. If T^+ defines σ as a product sort with projections π_1 and π_2 of arity $\sigma \to \sigma_1$ and $\sigma \to \sigma_2$, then we define a code $\overline{\xi}(x, x_1, \dots, y_{n2}, v_1, v_2)$ for the variables $x, x_1, \dots, x_n$ by

$$\xi(x_1, \dots, y_{n2}) \wedge \pi_1(x) = v_1 \wedge \pi_2(x) = v_2.$$

Lemma A.1 and the code $\overline{\xi}$ for the variables $x, x_1, \dots, x_n$ generate the Σ-formulas $\phi_t(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_1, v_2)$ and $\phi_s(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_1, v_2)$. We then define the Σ-formula ϕ^* to be

$$\exists_{\sigma_1} v_1 \exists_{\sigma_2} v_2\big(\phi_t(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_1, v_2) \\ \wedge \phi_s(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_1, v_2)\big).$$

One can verify that ϕ^* again satisfies the desired result.

If T^+ defines σ as a quotient sort with projection ϵ of arity $\sigma_1 \to \sigma$, then we again define a new code $\overline{\xi}(x, x_1, \dots, y_{n2}, v)$ for the variables $x, x_1, \dots, x_n$ by

$$\xi(x_1, \dots, y_{n2}) \wedge \epsilon(v) = x.$$

Lemma A.1 and the code $\overline{\xi}$ for the variables $x, x_1, \dots, x_n$ again generate the Σ-formulas $\phi_t(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v)$ and $\phi_s(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v)$. We define the Σ-formula ϕ^* to be

$$\exists_{\sigma_1} v\big(\phi_t(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v) \wedge \phi_s(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v)\big).$$

One again verifies that this ϕ^* satisfies the desired property. So the result holds when ϕ is of the form $t = s$ for Σ^+-terms t and s.

Now suppose that $\phi(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ is a Σ^+-formula of the form

$$p(t_1(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m), \dots, t_k(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)),$$

where p has arity $\sigma_1 \times \cdots \times \sigma_k$. Note that it must be that $\sigma_1, \dots, \sigma_k \in \Sigma$. Either $p \in \Sigma$ or $p \in \Sigma^+ - \Sigma$. We consider the second case. (The first is analogous.) Let $\psi(z_1, \dots, z_k)$ be the Σ-formula that T^+ uses to explicitly define p and let $\xi(x_1, \dots, y_{n2})$ be a code for $x_1, \dots, x_n$. Lemma A.1 and ξ generate the Σ-formulas $\phi_{t_i}(z_i, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2})$ for each $i = 1, \dots, k$. We define ϕ^* to be the Σ-formula

$$\exists_{\sigma_1} z_1 \dots \exists_{\sigma_k} z_k \big(\phi_{t_1}(z_1, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}) \wedge \cdots \\ \wedge \phi_{t_k}(z_k, \overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}) \wedge \psi(z_1, \dots, z_k)\big).$$

One can again verify that the result holds for this choice of ϕ^*.

We have covered the "base cases" for our induction. We now turn to the inductive step. We consider the cases of $\neg$, $\wedge$, and $\forall$. Suppose that the result holds for Σ^+-formulas ϕ_1 and ϕ_2. Then it trivially holds for $\neg\phi_1$ by letting $(\neg\phi)^*$ be $\neg(\phi^*)$. It also trivially holds for $\phi_1 \wedge \phi_2$ by letting $(\phi_1 \wedge \phi_2)^*$ be $\phi_1^* \wedge \phi_2^*$.

The $\forall_{\sigma_i}$ case requires more work. If x_i is a variable of sort $\sigma_i \in \Sigma$, we let $(\forall_{\sigma_i} x_i \phi_1)^*$ be $\forall_{\sigma_i} x_i(\phi_1^*)$. The only nontrivial part of the inductive step is when one quantifies over variables with sorts in $\Sigma^+ - \Sigma$. Suppose that $\phi(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$ is a Σ^+-formula and that the result holds for it. We let x_i be a variable of sort $\sigma_i \in \Sigma^+ - \Sigma$ and we show that the result also holds for the Σ-formula $\forall_{\sigma_i} x_i \phi(x_1, \dots, x_n, \overline{x}_1, \dots, \overline{x}_m)$. There are again four cases. We show that the result holds when σ_i is a product sort and a coproduct sort. The cases of subsorts and quotient sorts follow analogously.

Suppose that T^+ defines σ_i as a product sort with projections π_1 and π_2 of arity $\sigma_i \to \sigma_{i1}$ and $\sigma_i \to \sigma_{i2}$. Quantifying over a variable x_i of product sort σ_i can be thought of as "quantifying over pairs of elements of sorts σ_{i1} and σ_{i2}." Indeed, let $\xi(x_1, \dots, y_{n2})$ be a code for the variables $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ (these are all of the free variables in $\forall_{\sigma_i} x_i \phi$ with sorts in $\Sigma^+ - \Sigma$). We define a code $\overline{\xi}$ for the variables $x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n$ by

$$\xi(x_1, \dots, y_{n2}) \wedge \pi_1(x_i) = v_1 \wedge \pi_2(x_i) = v_2.$$

One uses the code $\overline{\xi}$ and the inductive hypothesis to generate the Σ-formula $\phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_1, v_2)$. We then define the Σ-formula $(\forall_{\sigma_i} x_i \phi)^*$ to be

$$\forall_{\sigma_{i1}} v_1 \forall_{\sigma_{i2}} v_2 \phi^*(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_1, v_2).$$

And one verifies that the desired result holds for this choice of $(\forall_{\sigma_i} x_i \phi)^*$.

Suppose that T^+ defines σ_i as a coproduct sort with injections ρ_1 and ρ_2 of arity $\sigma_{i1} \to \sigma_i$ and $\sigma_{i2} \to \sigma_i$. Quantifying over a variable x_i of coproduct sort σ_i can be thought of as "quantifying over *both* elements of sort σ_{i1} and elements of sort σ_{i2}." Indeed, let $\xi(x_1, \dots, y_{n2})$ be a code for the variables $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ (these are again all of the free variables in $\forall_{\sigma_i} x_i \phi$ with sorts in $\Sigma^+ - \Sigma$). We define two different codes $\overline{\xi}$ for the variables $x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n$ by

$$\xi(x_1, \dots, y_{n2}) \wedge \rho_1(v_1) = x_i$$
$$\xi(x_1, \dots, y_{n2}) \wedge \rho_2(v_2) = x_i.$$

We will call the first code $\xi'(x_1, \dots, y_{n2}, v_1)$ and the second $\xi''(x_1, \dots, y_{n2}, v_2)$. We use these two codes and the inductive hypothesis to generate Σ-formulas $\phi^{*'}$ and $\phi^{*''}$. We then define the Σ-formula $(\forall_{\sigma_i} x_i \phi)^*$ to be

$$\forall_{\sigma_{i1}} v_1 \forall_{\sigma_{i2}} v_2 \big(\phi^{*'}(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_2) \\ \wedge \phi^{*''}(\overline{x}_1, \dots, \overline{x}_m, y_{11}, \dots, y_{n2}, v_2)\big).$$

One can verify that the desired result holds again for this definition of $(\forall_{\sigma_i} x_i \phi)^*$. □

§5. Acknowledgments. We would like to thank Dimitris Tsementzis, Jim Weatherall, JB Manchak, and two anonymous referees for helpful comments and discussion.

BIBLIOGRAPHY

Adámek, J., Sobral, M., & Sousa, L. (2006). Morita equivalence of many-sorted algebraic theories. *Journal of Algebra*, **297**(2), 361–371.

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.

Andréka, H., Madaraśz, J. X., & Németi, I. (2008). Defining new universes in many-sorted logic. *Mathematical Institute of the Hungarian Academy of Sciences, Budapest*, 93.

Artigue, M., Isambert, E., Perrin, M., & Zalc, A. (1978). Some remarks on bicommutability. *Fundamenta Mathematicae*, **101**(3), 207–226.

Awodey, S. (2010). *Category Theory*. Oxford: Oxford University Press.

Awodey, S. & Forssell, H. (2010). First-order logical duality. Manuscript.

Barrett, T. W. (2015). On the structure of classical mechanics. *The British Journal for the Philosophy of Science*, **66**(4), 801–828.

Barrett, T. W. & Halvorson, H. (2015a). Glymour and Quine on theoretical equivalence. *the Journal of Philosophical Logic*. Forthcoming.

Barrett, T. W. & Halvorson, H. (2015b). Quine's conjecture on many-sorted logic. Manuscript. URL: http://philsci-archive.pitt.edu/11660/.

Borceux, F. (1994). *Handbook of Categorical Algebra*, Vol. 1. Cambridge: Cambridge University Press.

Coffey, K. (2014). Theoretical equivalence as interpretative equivalence. *The British Journal for the Philosophy of Science*, **65**(4), 821–844.

Curiel, E. (2014). Classical mechanics is Lagrangian; it is not Hamiltonian. *The British Journal for the Philosophy of Science*, **65**(2), 269–321.

de Bouvére, K. L. (1965). Synonymous theories. *Symposium on the Theory of Models*. Amsterdam: North-Holland Publishing Company, pp. 402–406.

Dukarm, J. J. (1988). Morita equivalence of algebraic theories. *Colloquium Mathematicae*, **55**(1), 11–17.

Eilenberg, S. & Mac Lane, S. (1942). Group extensions and homology. *Annals of Mathematics*, 757–831.

Eilenberg, S. & Mac Lane, S. (1945). General theory of natural equivalences. *Transactions of the American Mathematical Society*, **58**(2), 231–294.

Freyd, P. (1964). *Abelian categories*. New York: Harper and Row.

Friedman, H. M. & Visser, A. (2014). When bi-interpretability implies synonymy. *Logic Group Preprint Series*, **320**, 1–19.

Glymour, C. (1971). Theoretical realism and theoretical equivalence. *PSA 1970*. Springer, pp. 275–288.

Glymour, C. (1977). The epistemology of geometry. *Noûs*, **11**, 227–251.

Glymour, C. (1980). *Theory and Evidence*. Princeton, NJ: Princeton University Press.

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

Halvorson, H. (2011). Natural structures on state space. Manuscript.

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

Halvorson, H. (2013). The semantic view, if plausible, is syntactic. *Philosophy of Science*, **80**(3), 475–478.

Halvorson, H. (2015). Scientific theories. *Oxford Handbooks Online*. Oxford, UK: Oxford University Press, Forthcoming

Hodges, W. (2008). *Model Theory*. Cambridge University Press.

Johnstone, P. T. (2003). *Sketches of an Elephant*. Oxford: Oxford University Press.

Kanger, S. (1968). Equivalent theories. *Theoria*, **34**(1), 1–6.

Knox, E. (2014). Newtonian spacetime structure in light of the equivalence principle. *The British Journal for the Philosophy of Science*, **65**(4), 863–880.

Kuratowski, K. (1966). *Topology*, Vol. 1. Warsaw and New York: Academic Press.

Mac Lane, S. (1948). Groups, categories, and duality. *Proceedings of the National Academy of Sciences*, **34**(6), 263–267.

Mac Lane, S. (1971). *Categories for the working mathematician*. New York: Springer.

Makkai, M. (1991). *Duality and definability in first order logic*, Vol. 503. Providence, RI: American Mathematical Society.

Manzano, M. (1996). *Extensions of First Order Logic*. Cambridge University Press.

Mere, M. C. & Veloso, P. (1992). On extensions by sorts. *Monografias em Ciências da Computaçao, DI, PUC-Rio*, **38**, 92.

Montague, R. (1957). *Contributions to the Axiomatic Foundations of Set Theory*. PhD thesis, University of California, Berkeley.

Nestruev, J. (2002). *Smooth Manifolds and Observables*. New York: Springer.

North, J. (2009). The ‘structure’ of physics: A case study. *The Journal of Philosophy*, **106**, 57–88.

Pelletier, F. J. & Urquhart, A. (2003). Synonymous logics. *Journal of Philosophical Logic*, **32**(3), 259–285.

Pinter, C. C. (1978). Properties preserved under definitional equivalence and interpretations. *Mathematical Logic Quarterly*, **24**(31–36), 481–488.

Quine, W. V. O. (1975). On empirically equivalent systems of the world. *Erkenntnis*, **9**(3), 313–328.

Rosenstock, S., Barrett, T. W., & Weatherall, J. O. (2015). On Einstein algebras and relativistic spacetimes. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **52**, 309–316.

Schmidt, A. (1951). Die Zulässigkeit der Behandlung mehrsortiger Theorien mittels der üblichen einsortigen Prädikatenlogik. *Mathematische Annalen*, **123**(1), 187–200.

Schwabhäuser, W., Szmielew, W., & Tarski, A. (1983). *Metamathematische Methoden in der Geometrie*. New York: Springer.

Shoenfield, J. (1967). *Mathematical Logic*. Addison-Wesley.

Sklar, L. (1982). Saving the noumena. *Philosophical Topics*, **13**(1), 89–110.

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press

Swanson, N. & Halvorson, H. (2012). On North's 'The structure of physics'. Manuscript.

Szczerba, L. (1977). Interpretability of elementary theories. *Logic, Foundations of Mathematics, and Computability Theory*. Amsterdam: Springer.

Tsementzis, D. (2015). A syntactic characterization of Morita equivalence. Manuscript.

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

van Fraassen, B. C. (2014). One or two gentle remarks about Hans Halvorson's critique of the semantic view. *Philosophy of Science*, **81**(2), 276–283.

Weatherall, J. O. (2015a). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis*. Forthcoming.

Weatherall, J. O. (2015b). Categories and the foundations of classical field theories. In Landry, E., editor, *Categories for the Working Philosopher*. Oxford University Press. Forthcoming.

Weatherall, J. O. (2015c). Understanding gauge. *Philosophy of Science*. Forthcoming.

DEPARTMENT OF PHILOSOPHY
PRINCETON UNIVERSITY
PRINCETON, NJ, USA
E-mail: thomaswb@princeton.edu
E-mail: hhalvors@princeton.edu

https://doi.org/10.1017/S1755020316000186 Published online by Cambridge University Press